Lecture Notes on Data Engineering
and Communications Technologies 14

Arun Kumar Sangaiah
Arunkumar Thangavelu
Venkatesan Meenakshi Sundaram *Editors*

Cognitive Computing for Big Data Systems Over IoT

Frameworks, Tools and Applications

Springer

Lecture Notes on Data Engineering and Communications Technologies

Volume 14

Series editor

Fatos Xhafa, Technical University of Catalonia, Barcelona, Spain
e-mail: fatos@cs.upc.edu

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It publishes latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.
The series has a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

More information about this series at http://www.springer.com/series/15362

Arun Kumar Sangaiah · Arunkumar Thangavelu
Venkatesan Meenakshi Sundaram
Editors

Cognitive Computing for Big Data Systems Over IoT

Frameworks, Tools and Applications

Springer

Editors
Arun Kumar Sangaiah
School of Computing Science
and Engineering
VIT University
Vellore, Tamil Nadu
India

Arunkumar Thangavelu
School of Computing Science
and Engineering
VIT University
Vellore, Tamil Nadu
India

Venkatesan Meenakshi Sundaram
Department of Computer Science
and Engineering
National Institute of Technology, Surathkal
Mangalore, Karnataka
India

ISSN 2367-4512 ISSN 2367-4520 (electronic)
Lecture Notes on Data Engineering and Communications Technologies
ISBN 978-3-319-70687-0 ISBN 978-3-319-70688-7 (eBook)
https://doi.org/10.1007/978-3-319-70688-7

Library of Congress Control Number: 2017959915

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Cognitive systems attract major attention in the new era of computing. In addition, cognitive computing delivers an extended guidance towards building a new class of systems with convergence of big data and Internet of Things (IoT). The human life is driven by the smart electronic devices called IoT. Moreover, IoT devices generate and exchange more amounts of data. Extracting the valid truth from this data becomes a hectic task. Consequently, machine learning techniques have been proposed to analyse large amounts of data and enhance decision-making. The development of new techniques helps to provide relevant information to users with high efficiency. Today's world is driven by the era of digital data. People not only look into conceiving information from the data but also perform exploratory data analysis, thus, the area of big data analytics has emerged. Analysis of data sets can find new correlations to spot business trends, prevent diseases, and combat crime, etc.

In cognitive computing, new hardware or software devices mimic human brain and take a decision appropriate to the situation. Moreover, cognitive computing is used in numerous artificial intelligence (AI) applications, including expert systems, natural language programming, neural networks, robotics and virtual reality. Further, cognitive computing has lots of applications in every area of our lives, from travel, sports and entertainment, to fitness, health and wellness, etc. In the business domain, entrepreneurs from different industries have already created products and services based on cognitive technologies. Cognitive computing helps them by giving intelligent recommendation through data analysis. Cognitive computing is not helping only humans, it is also helping veterinarians take better care of the animals that come into their practices. In future, cognitive systems provide expert assistance to a problem without the intervention of human beings. Self-learning capability of human beings is adapted to the system by applying artificial intelligence to it. Thus, the combination of big data analysis and cognitive computing methodologies over IoT devices can change the world with new colour of Intelligence.

This book attempts to explicate the state-of-the-art research in cognitive computing for big data systems and provide a comprehensive and in-depth coverage of the key subjects in the field of IoT. The book is invaluable, topical and timely and can serve nicely as a reference book for courses at both undergraduate and postgraduate levels. It can also serve as a key source of knowledge for scientists, professionals, researchers and academicians, who are interested in new challenges, theories, practice and advanced applications of cognitive computing.

I am happy to inform the readers that this book titled "Cognitive Computing for Big Data Systems over IoT" addresses important research directions in cognitive computing and the development of innovative big data models for analysing the data generated by IOT devices. It marks an important step in the maturation of this field and will serve to unify, advance and challenge the scientific community in many ways. I commend the editors and the authors on their accomplishment, and hope that the readers will find the book useful and a source of inspiration for their research and professional activity.

September 2017
Rochester, MI, USA

Vijayan Sugumaran, Ph.D.
Professor of Management Information Systems
Chair, Department of Decision
and Information Sciences
School of Business Administration
Oakland University

Preface

The paradigm shift in towards Internet of Things (IoT) is becoming the vital component of Internet. Low-cost sensing and actuation are available to the whole world, which enable seamless information exchange and networked interactions of physical and digital objects. This interconnectivity together with large-scale data processing, advanced machine learning, robotics and new fabrication techniques are steadily bringing in innovation and business models of digital space into the physical world. Further, IoT is expected to improve the intelligence, promote the interaction between the human and the environment, as well to enhance reliability, resilience, operational efficiency, energy efficiency and resource consumption. Subsequently, many of the IoT systems and technologies are relatively novel, there are still many untapped applications areas, numerous challenges and issues that need to be improved.

Cognitive science has broad horizons, which cover different characteristics of cognition. The field is highly transdisciplinary in nature, combining ideas, principles and methods of psychology, computer science, linguistics, philosophy, neuroscience, etc. In addition, cognitive computing is the creation of self-learning systems that use data mining, pattern recognition and natural language processing (NLP) to solve complicated problems without constant human oversight.

Cognitive computing will bring a high level of fluidity to analytics. The chapters included in this book aim on addressing recent trends, innovative ideas, challenges and cognitive computing solutions in big data and IoT. Moreover, these chapters specify novel in-depth fundamental research contributions from a methodological/application in data science accomplishing sustainable solution for the future perspective. Further, this book provides a comprehensive overview of constituent paradigms underlying cognitive computing methods, which are illustrating more attention to big data over IoT problems as they evolve. Hence, the main objective of the book is to facilitate a forum to a large variety of researchers, where decision-making approaches under cognitive computing paradigms are adapted to demonstrate how the proposed procedures as well as big data and IoT problems can be handled in practice.

Need for a Book on the Proposed Topics

Data science techniques have been adopted to improve the IoT in terms of data throughput, optimization and management. Moreover, the data science techniques carry major impact on the future of IoT, allowing researchers to reproduce scenarios and optimize the acquisition, analysis and visualization of the data acquired by IoT devices. One of the most ambitious and exciting challenges in data science is to model and replicate how people think and learn. This book explores domain knowledge and reasoning of data science and cognitive methods over the IoT. The idea of embodying this concept would be to extend existing data science approaches by incorporating knowledge from experts as well as a notion of artificial intelligence, and performing inference on the knowledge. The main focus is design of best cognitive embedded data science technologies to process and analyse the large amount of data collected through IoT and help for good decision-making. Consequently, the cognitive data science research facilitates a platform to scientific community to work for the best solution of challenges related with cognitive methods and data science model issues to support IoT solutions towards smart infrastructure and meet the requirement of modern world. This book addresses a wide spectrum of cognitive computing paradigms, making decisions of an industry or organization happened at all the levels of data science challenges. In addition, this book aims to provide relevant theoretical frameworks and the latest empirical research findings in the area. Solutions for big data over IoT problems have been effectively handled through wide range of algorithmic and cognitive computing frameworks, such as optimization, machine learning, decision support systems and meta-heuristics. In addition, the main contributions to this volume address big data over IoT problems in computing and information processing environments and technologies, and at various levels of the cognitive computing paradigms.

Organization of the Book

The volume is organized into 15 chapters. A brief description of each chapter is given as follows:

Chapter "Beyond Automation: The Cognitive IoT. Artificial Intelligence Brings Sense to the Internet of Things" gives an overview of Intelligent IoT is named as Cognitive IoT (CIoT) describes the convergence of IoT with artificial intelligence techniques. In addition, this chapter introduces a preliminary idea of cognitive computing that discusses several aspects of CIoT through different forms such as Cognitive Network, Cognitive Devices and Cognitive Analytics.

Chapter "Cybercrimes Investigation and Intrusion Detection in Internet of Things Based on Data Science Methods" introduces the principles of Digital Forensics, Intrusion Detection and Internet of Things as well as exploring the data science concepts and methods. The chapter authors highlight the need for

employing data science, data mining and big data analytics methods in cybercrime investigation area because they have many advantages to support the digital investigation.

Chapter "Modelling and Analysis of Multi-Objective Service Selection Scheme in IoT-Cloud Environment" presents computational intelligent paradigm based on fuzzy multi-criteria decision-making approaches (AHP and TOPSIS) to select an optimal cloud for accessing different services of cloud. The offloading data is evaluating the weights of important criteria's and by calculating the final ranking of alternative clouds. In addition, this chapter addresses the significance of the proposed approach in better understanding of the whole evaluation process and their efficiency of decision-making process in cloud path selection for offloading in Mobile Cloud Computing (MCC) environment.

Chapter "Cognitive Data Science Automatic Fraud Detection Solution, Based on Benford's Law, fuzzy Logic with Elements of Machine Learning" presents computational intelligence based heuristic approach for maximizing energy efficiency in the Internet of Things (IoT). The authors present the Modified Multi-Objective Particle Swarm Optimization (MMOPSO) algorithm based on the concept of dominance to solve the mobile cloud task scheduling problem. Overall, this chapter explores IoT and cloud computing as well as their symbiosis based on the common environment of distributed processing.

Chapter "Reliable Cross Layer Design for E-health Applications—IoT Perspective" illustrates cognitive data science automatic fraud detection solution, based on Benford' s law, fuzzy logic with elements of machine learning. Moreover, proposed methodology gives solution for automatic seeking patterns within data with focus on fraud detection.

Chapter "Erasure Codes for Reliable Communication in Internet of Things (IoT) embedded with Wireless Sensors" provides a comprehensive overview of existing erasure codes in the wireless sensor networks which are integral part of Internet of Things communication. This chapter presents the construction methods of extensively used Reed–Solomon codes and Fountain codes that are provided in addition to decentralized erasure codes. Further, basic communication paradigm of information transmission namely end-to-end transmission and hop-by-hop transmission are discussed in detail with and without emphasis on erasure codes.

Chapter "Review: Security and Privacy Issues of Fog Computing for the Internet of Things (IoT)" highlights the security and privacy issues of fog computing through a comprehensive review of fog computing and suggests solutions for identified problems. The chapter authors highlighted the areas that need attention in fog computing research.

Chapter "A Review on Security and Privacy Challenges of Big Data" emphasizes on certain gaps in the literature to evaluate possible solutions to a rising problem in various privacy and security issues in different areas of big data. The chapter authors have addressed gaps in the literature by highlighting security and privacy issues that big companies face with recent technological advancements in corporate societies.

Chapter "Recent Trends in Deep Learning with Applications" presents the overview review of understanding the deep learning methods and their recent advances in Internet of things. The deep learning methods are divided into four classifications such as Convolutional Neural Networks, Restricted Boltzmann Machines, Auto-encoder and Sparse Coding. The applications with respect to Internet of things such as image caption, object detection and visual tracking are also discussed in this chapter.

Chapter "High-Level Knowledge Representation and Reasoning in a Cognitive IoT/WoT Context" presents an overview of the Generalized World Entities (GWEs) paradigm, used to add a semantic/conceptual dimension to the ordinary IoT/WoT procedures. This chapter have focused on development in effective SWOT (Semantic Web of Things) applications via high-level Cognitive Science/Artificial Intelligence techniques. It is necessary to overcome the shortcomings of the present cognitive/conceptual IoT/WoT approaches.

Chapter "Applications of IoT in Healthcare" focuses on how Internet of things (IoT) capabilities can be leveraged in providing better healthcare. This chapter also discusses the key enabling technologies of the IoT (e.g., sensors and Wireless Sensor Networks (WSN)), their characteristics and challenges.

Chapter "Security Stipulations on IoT Networks" discusses on various attacks which are possible in the IoT connected network. This chapter will provide readers with an understanding about the security policies and mechanisms in complex IoT systems. Moreover, the chapter authors have illustrated the various security aspects and its countermeasures were analysed and discussed.

Chapter "A Hyper Heuristic Localization Based Cloned Node Detection Technique using GSA Based Simulated Annealing in Sensor Networks" presents a Residual Energy and GSA based Simulated Annealing (RE-GSASA) for detecting and isolating the cloned attack node in WSN. The chapter authors have proposed a novel Residual Energy and GSA based Simulated Annealing (RE-GSASA) method is introduced to reduce the energy consumption during data aggregation and improve the packet delivery ratio.

Chapter "Review on Analysis of the Application Areas and Algorithms used in Data Wrangling in Big Data" presents an extended review on the analysis of the application areas and algorithms used in data wrangling in Big Data. This chapter results show that data wrangling and clustering algorithm can solve medical data storage issues.

Chapter "An Innovation Model for Smart Traffic Management System Using Internet of Things (IoT)" discusses about an architecture which integrates Internet of things and other moving components like data management techniques to create a model for traffic management and monitoring. The model comprises of a single platform where this platform will communicate with the large number of decentralized heterogeneous components.

Audience

The intended audience of this book includes scientists, professionals, researchers and academicians, who deal with the new challenges and advances in the specific areas mentioned above. Designers and developers of applications in these fields can learn from other experts and colleagues through studying this book. Many universities have started to offer courses on cognitive computing, big data analytics on the graduate/postgraduate level in information technology and management disciplines. This book starts with an introduction to cognitive computing and data science approaches, hence suitable for university level courses as well as research scholars. Their insightful discussions and knowledge, based on references and research work, will lead to an excellent book and a great knowledge source.

Vellore, India — Arun Kumar Sangaiah
Vellore, India — Arunkumar Thangavelu
Mangalore, India — Venkatesan Meenakshi Sundaram

Acknowledgement

The editors would like to acknowledge the help of all the people involved in this project and, more specifically, to the authors and reviewers who took part in the review process. Without their support, this book would not have become a reality.

First, the editors would like to thank each one of the authors for their contributions. Our sincere gratitude goes to the chapters' authors who contributed their time and expertise to this book.

Second, the editors wish to acknowledge the valuable contributions of the reviewers regarding the improvement of quality, coherence and content presentation of chapters. We deeply appreciate the comments of the reviewers who helped us to refine the context of this book. Most of the authors also served as referees; we highly appreciate their double task.

Finally, our gratitude goes to all of our friends and colleagues, who were so generous with their encouragement, advice and support.

Arun Kumar Sangaiah
Arunkumar Thangavelu
Venkatesan Meenakshi Sundaram

Contents

Beyond Automation: The Cognitive IoT. Artificial Intelligence Brings Sense to the Internet of Things

Pijush Kanti Dutta Pramanik, Saurabh Pal and Prasenjit Choudhury

> "*The ability to interact with a computer presence like you would a human assistant is becoming increasingly feasible*."
>
> —Vint Cerf.

Abstract The Internet of Things (IoT) has already been infiltrated to our everyday life in the forms of smartphones, smart TVs, fitness trackers, health monitoring systems, smart watches, vending machines, smart meters, city traffic, building security systems and much more like this. It has steered the automation to a new high. But IoT alone has limited capability. To reap the actual benefit of IoT, it has to be intelligent. In this chapter, we have reasoned why IoT needs Artificial Intelligence (AI). The intelligent IoT that we discuss here, is termed as the Cognitive IoT (CIoT) . Cognitive IoT uses a new computing paradigm called Cognitive Computing, often popularly dubbed as the third era of computing. Cognitive Computing will make IoT more sophisticated, more intelligent, and more interactive. This chapter essentially focuses on the convergence of IoT with Cognitive AI. In addition to introducing a preliminary idea of Cognitive Computing, the chapter discusses several aspects of CIoT. It explains how Cognitive Computing has been used in IoT through different forms such as Cognitive Network, Cognitive Devices, and Cognitive Analytics. The desirable properties of the CIoT are mentioned. A special discussion is presented on how CIoT has taken automation to a new level. Realizing CIoT is not straightforward for the reason that it is a complex system and is quite different from other computing systems we have been familiar with. The implementation challenges along with the societal and ethical apprehensions of CIoT have been identified and discussed in detail. The business world is showing great interest in IoT and subsequently in CIoT. Considering that, we have highlighted the business values of CIoT and some prospective application areas along

Vint Cerf, a "father of the internet," in "Your Life: Vinton Cerf" interview by David Frank in AARP Bulletin (December 2016, Vol. 57, No. 10, p. 30.).

P. K. D. Pramanik (✉) · P. Choudhury
National Institute of Technology, Durgapur, India
e-mail: pijushjld@yahoo.co.in

S. Pal
Bengal Institute of Technology, Kolkata, India

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_1

with a few prominent use cases. But, will excessive adoption of AI and tendering human-like intellect to inorganic devices be a threat to the human race? Do we need to be warned? A rational discussion on that has been presented before winding up.

Keywords Cognitive IoT · Internet of things · Cognitive computing
Artificial intelligence · Machine learning · Automation · Cognitive automation
Business analytics · Cognitive analytics

1 Introduction

We wondered when in 1997, Deep Blue, a supercomputer from IBM, defeated Garry Kasparov by 3,5: 2,5 in a most inciting showdown in the chess history. We were awed how a machine can outperform a human brain! Since then, we have witnessed the further revolution in computing as the machines have become more intelligent day-by-day. Today we are on the verge of the era of the smart world where everything is supposed to be done automated and proactively without or minimal human intervention. And the technology that is supposed to take us to that fairy world is known as Internet of Things (IoT) . IoT refers to a connected system where anything that is addressable in the digital universe, can be connected via the Internet. This omni-connectivity allows devices to share data effortlessly and sharing data makes systems smart. For example, combining the traffic sensor data can give an overall picture of the traffic status of a city. Similarly, the activities going on in a smart home can be perceived by interpreting the data collected from different sensors planted at the home. The pervasive presence of IoT devices has made our life smart. For example, while returning home from office we can instruct our micro oven, through our smartphone, to warm up the pizza; or a wearable gizmo that has an embedded sensor for checking sugar level can remind us to take the insulin dose when the sugar level is high. But we want more. For instance, the room light will be off when we say good night; or on the way to our office, our car radio should brief the morning headlines on its own. Moreover, if the morning flight is delayed, the alarm clock will adjust itself accordingly, allowing us to slumber a bit longer. Well, these are achievable, but not by IoT alone. We have to make IoT more intelligent. We have to impulse cognition to the IoT as such that it becomes the extended version of human. For that, we require infusing IoT with Artificial Intelligence (AI). These two technologies can blur the line between sci-fi and a new high-tech reality. In fact, without AI we cannot attain the full potential and vision of IoT; we won't be able to pluck the entire bunch of fruits from the IoT-tree. AI can add more advanced control features and autonomous behavior to the smart world. The AI which goes beyond normal machine intelligence to cognitive reasoning and rationalizing problem solutions like human intelligence is cognitive AI. Cognitive AI, a special arm of AI, in which a machine is loaded with the features of human thought process, plays a vital role in bridging the gap between human and

machine. It turns the unintelligent machine to one with humanoid intelligence, enabling machines to learn, reason and communicate with humans (and vice versa) in natural language. Cognitive Computing that refers to the practical enactment of Cognitive AI through computing model, adds a new layer of functionality to the existing IoT architecture. This is known as Cognitive IoT (CIoT) that learns, applies sense, and takes a decision and convey that to the humans. CIoT will allow realizing the full potential of the IoT and will take the relationship between the human and the physical things to an unforeseen echelon by making inorganic things more intelligent, intuitive and interactive [1]. In short, CIoT will augment the current IoT with the added cognitive ability very much similar to human cognition.

The rest of the chapter is organized as follows. The next section discusses how IoT is related to automation and AI, in general, and the need for AI in IoT. Section 3 discusses the basics of cognitive systems and Cognitive Computing with a special discussion on how Cognitive Computing is different from automation. Section 4 explores different aspects of CIoT including the properties and implementation challenges. Section 5 highlights the business values of CIoT. Sections 6 and 7 look at the prospective application areas and some use cases of CIoT respectively. Section 8 brings up the hypothetical apprehensions about CIoT and excessive adoption of AI. And at the end, Sect. 9 concludes the chapter.

2 IoT and AI

2.1 Internet of Things

The Internet is the interconnection of several computers in a global range. IoT has extended the vision of Internet further. Here, not only computers are interconnected, rather all the entities on the earth should be connected. The motto is that "Anything that can be connected will be connected" [2]. The term 'Internet of Things' was first coined by Kevin Ashton, a British technology pioneer, in 1999. The basic goal was to capture and share data in an automated and pervasive manner. The 'things' in IoT can include anything from a smart-watch to a cruise control system. But the common factor is that every object is consisting of some type of sensors (e.g. temperature, light, motion, etc.) or actuators (e.g. displays, sound, motors, etc.). These sensors and actuators are often part of a larger 'things' where a number of them are embedded. The other components of the IoT ecosystem include the computing resources (to process the sensed data locally or remotely) and the communication medium (Bluetooth, ZigBee, RFID, etc., for short range and the Internet for long range). The IoT devices sense or read their surroundings or the environment where they are embedded. The sensed data are processed and analyzed to acquire information and knowledge on the basis of which some definitive action may be taken. Actually, the essence of IoT is to make any information available to anyone and anytime across all the barriers [3]. IoT devices can do their work

autonomously i.e. without any explicit external command. They independently collect information and exchange them proactively with other IoT devices within the network. The gathered data are analyzed in runtime or later either manually or by the machine to infer information. The applications of IoT have been seen in the wide range of domains such as manufacturing sector, logistic, transportation, agriculture, medical and healthcare, home and building automation, smart grid, service sector, etc. The advancements in technologies that have led to cheap sensors, cheap processing, and cheap bandwidth with ubiquitous wireless coverage and smartphones, have provided an effective ground for IoT to harvest information from the environment and interact with the physical world [4]. This has abetted in minimizing the gap between the physical objects and the cyber world, which has made easier to control these devices. In short, by means of the ubiquitously and universally connected devices, IoT has taken us to the new possibilities of a smart world for a smart living.

2.2 IoT and Automation

Earlier, we stated that the motto of IoT is to connect everything wherever possible. But why do we need to connect all the things? The answer is very clear and simple —to automate. As we pointed out in the last section that improved relations between the physical objects and the cyber world with intensified connectivity have allowed us to control these devices better and more effortlessly. Indeed, a sense of automation came into the picture.

The advantageous aspects of IoT have led industries/businesses to take IoT as a new step towards automation empowering centralized controlling and management [5]. The IoT-based automation can reduce the operational cost, as compared to the manual procedure, through automated control and management of isolated and independent devices by connecting and making them communicate with each other [6].

In the manufacturing plants, IoT has immense potential for automation for more responsive and effective operations. Smooth communication among devices like sensors, actuators, analyzers, and robots enhances the manufacturing performance and flexibility [3]. The data obtained from these devices can be used for governing equipment status, energy management, condition monitoring, load balancing, tracking and tracing systems, etc. [3]. Integrating these data with the ERP system can take the automation in the organization to a ‘never achieved before’ level through real-time observation of the overall operation and production process. If required, machines communicate to each other to adapt to changed conditions even without any centralized controlling [3].

IoT integration and tracking goods for delivery using GPS and RFID technologies automate the supply chain management. Logistic service can be automated and controlled centrally through continuous monitoring of the location, expected

time to arrive, and the status of the product and environment (e.g. temperature, humidity, etc.) while in transit ensuring the quality [7].

Connecting an equipment to the IoT will enable the manufacturer or the dealer of that equipment to provide better preventative maintenance as they will be well aware of the condition of the equipment through the stream of data that the equipment sends with the help of IoT [8].

The retail sector also has huge advantages in inventory management by incorporating IoT. For instance, detecting low stock by smart shelves, notifying customers about discount offers as they enter the store, tracking goods for smoother supply-chain management, etc. [5].

Besides business and manufacturing, IoT-based automation has been adopted in other segments such as automated home and building, cities, shopping mall, transport etc. Automated homes and buildings will enhance the comfort and standard of living to a great extent. For example, temperature and luminosity of lights will be automatically adjusted depending on the situation and peoples' presence in the room [5]. Likewise, building security systems can automatically detect any unusual activities and irregularities, etc. IoT applications in transport can send the information about a traffic jam and potential delay to the traveller's smartphone by assessing the real-time traffic condition automatically.

2.3 IoT Is Not AI

Going through all the wonderful stories about 'smart'ness and automation, if you get confused and think that IoT is nothing but AI only donning a latest fancy technical jargon, you are surely not to be panned, particularly if you are a newbie in the world of IoT. Yes, the promises IoT has offered very much sound like AI. In fact, they are closely related, though not exactly same. To make things clearer, let us refer to the standard definitions of the both.

AI is defined in Merriam-Webster as:

> An area of computer science that deals with giving machines the ability to seem like they have human intelligence.

Whereas, IoT is being defined by Gartner as:

> Network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment.

The primary object of IoT is to connect the objects which are connectable. Here the 'intelligence' (to be technically correct, we should state it as knowledge) come from the sharing of data and streamlining these data to the proper channel. But the purpose of AI is to program the machines such a way so that they can take decisions independently without any human intervention as such that they have their own intelligence to think and reason necessary for taking decisions. So, the emphasis in IoT is the 'connection' whereas, in AI, it is the 'intelligence'.

In the perspective of Computer Science studies, AI is a stream of study having many sub-branches such as Machine Learning (ML) , Neural Networks, Fuzzy Systems, Genetic Algorithm, Natural Language Processing (NLP) and much more. Every sub-branch of AI contributes to putting intelligence into the machines in some way or other. Whereas IoT is a descendant technology which incorporates the properties from many streams of Computer and Electronic Sciences such as Computer Networks, Communication Technology, Wireless Technology, Sensor Technology, Mobile Technology, Distributed Computing, Data Science and lots more. To sum up, many fields of study diverge from AI whereas IoT is the outcome of converging multiple fields of study.

2.4 Need for AI in IoT

From the previous section's discussion, we understood that IoT and AI are not synonymous but far-off distinct. So why is it necessary for IoT to get married to AI? The answer is very obvious—to get the best out of IoT. Initial IoT applications did not possess any decision-making skill, which caused failure to achieve the desired performance level. The word 'intelligence' was missing from the IoT. AI makes IoT intelligent. AI has transformed IoT to an intelligent entity, capable of behaving decisively on the basis of past data and events. Furthermore, it can automatically train, learn and troubleshoot future issues up to some good extent. Automating IoT incorporates AI so as the decision can be taken autonomously for the management and controlling of IoT. IoT comprises a complex network that includes billions devices and sensors. Managing this kind of system is very costly and gruesome task. This demands automatic monitoring, management and self-* (self-configuring, self-protecting, self-organizing, self-reliant, self-healing, self-aware, self-learning, and self-adjusting) properties of IoT to minimize the human intervention and thereby reducing operational cost [9]. For that, it is essential to imbue AI into IoT. The increasing connected devices, demands for real-time adaptability and smartness to manage up the intended service operations. AI incorporates reasoning and decision making skills to IoT, leading to smart functioning. AI adds a new layer of functionality to the existing IoT architecture, leading devices to decision making, reasoning, and learning. Thereby transforming IoT into an intelligent entity capable of behaving decisively over past data and events. Furthermore, it may automatically train, learn, and troubleshoot future issues up to some good extent. AI enabled IoT sense other things it needs when connected to the Internet or local network. It builds up intelligence from the pattern it observes on various data it acquires from the surrounding and the network and makes intelligent decisions. ML can be used to improve efficiency. Using AI in IoT applications has manifold advantages:

- Enhance user experience.
- Learn pattern and behavior of an applied system automatically.
- Identifies anomalies and conflicting situation.

Only IoT devices themselves are not worthwhile, much like our body without a brain. It is the data that IoT devices generate is valuable. The data that would help doctors to have real-time information about a patient condition, enabling decision making across the industry, managing home appliances, communication and navigation in transportation, and other numerous applications. The rapid expansion of devices, sensors, and connectivity produce sheer volume of data. But again, just to capture the data is not sufficient. To utilize IoT to its full potential, the vast amount of data generated out of millions of IoT devices are to be utilized judiciously. We need to take out the actionable knowledge from the data. These data contain valuable insights on what is working and what not. The big challenge is to analyze the data to crave right information at right time. Maximum of the IoT data has actually never been put to use. The traditional methods for data analysis are incapable to handle the sheer amount of data IoT generate. AI can boost the accuracy, quality and speed of data analysis. By applying intelligent algorithms, human-like intelligence can be extracted from the IoT data [10]. Various AI, ML, and other related algorithms like evolutionary algorithm, genetic algorithm, etc., have been found to be effective in achieving automated decision making skill. Incorporation of intelligent algorithms/software in IoT data analysis combined with data from other business sources breed statistical information and summarization, reinforcing a future prediction scenario. This alleviates in new system plans, resulting in increased productivity and operational efficiency, which was otherwise unattainable before [6].

Besides sensing data for a situation, IoT must be quite aware of different anomalies like faulty sensors, incorrect data acquisition, missing data, and data ambiguity to keep the system integrated and sustained. Use of AI plays a crucial role in monitoring sensors, network, and other devices using intelligent security algorithms.

Unquestionably, AI leads to autonomous decision making and self-management, but subjected to the question—to what extent intelligence should be applied to IoT? Because of the scattered network architecture of IoT with varying heterogeneous devices having a different energy level, it is quite reasonable to have reservations about the degree of intelligence conferred to each IoT nodes. For the nodes are enabled with limited AI, resulting in, autonomous decision making on what, when and how much to process. This will restrict devices to over-process and be over-active thus saving a significant amount of energy which is very crucial for IoT devices that are typically energy scanty. Again, too much autonomy can make the whole system unpredictable which may defy the system design goal. There is also need to understand the trade-off between centralized and distributed intelligence. Centralized intelligence is sophisticated and less complex while distributing intelligence at the edges and nodes of network make IoT more autonomous and scalable but at the cost of complexity in terms of implementation.

Nevertheless, AI is already being used in IoT for the better upshot and the future offers incredible opportunities for us to look beyond horizons. We just need to make sure to grab the right opportunity that transforms our lives.

Various areas of IoT and AI enabled IoTs are evolving and require further research. Various automation related opportunities in the field of IoTs are still in their infancy and AI to resolve the challenges and issues. The current research area is quite emphasizing to train things to perform as expected and furthermore on how well they communicate. The future research may focus on how to make things better and smarter.

3 Cognitive AI and Cognitive Computing

3.1 Cognition, AI, and Cognitive AI

The term 'cognition' has been derived from the Latin word *cognito* which means 'to think'. The Cambridge dictionary defines 'cognition' (noun) as "the use of a conscious mental process" and 'cognitive' (verb) as "connected with thinking or conscious mental processes". Cognition emphasizes on the process how one learns, remembers, and reasons rather than any particular fact that one has learned. Cognition had been originated from philosophy and gradually transcended to psychology. Cognition is a general term used in diverse disciplines. For example, the psychologists use the term 'cognitive psychology' where they study how people think and their behavioral process whereas the term 'cognitive therapy' is used by the psychotherapists to refer the treatment of the people suffering from the mental illness that tries to change the way they think.

'Intelligence' has come from the Latin word *intelligere* which means 'to pick up'. Fundamentally, AI refers to the capacity of a machine to learning or understanding and out of some given options, which action to be taken and when. AI can perform only those tasks which are defined by clear-cut rules and are hard-coded by the human. This means that the success of AI systems depends largely on the perception and anticipation of the probable complex scenarios by human and their ability to code the solution logic and assimilate it to the system. To get rid of this restraint, the AI people have espoused cognition and propelled beyond the study of human thinking and behavioral process. They termed this amalgamation as Cognitive AI regarded as the natural extension of the present AI. The purpose was to emulate the cognition process into the non-living objects, to be specific, the computers and machines, to make them intelligent, really brainy which otherwise cannot be possible through other conventional AI techniques. Cognitive AI plays a vital role in bridging the gap between human and machine.

3.2 Cognitive Computing

Over the last hundred years, computing technologies have evolved dramatically. Advancement in AI has lead machines to learn and empowers to reason,

transforming machines to a 'thinking device', to what we refer as cognitive device/system. It is the evolution of software approach which mimics human cognition into machines leading to the dawn of new era of computing a 'cognitive era' [11].

Cognitive systems are fundamentally different in comparison to other forms of computing available or practiced in past. The cognitive system continuously learns from its interaction with data (structured or unstructured), people, and situation and thus eventually improves its learning and reasoning capabilities [12]. As the learning of a human brain gradually evolves from birth to adulthood, a cognitive system also learns over time through the experience and information it gains every time it interacts with other systems [13]. The auto-learning features ensure that cognitive system may never go outmoded but only get smarter with time. It reasons with purpose, infers meaning and interacts with human, naturally. This paradigm shift in computing approach from rule-based to learning and reasoning will change the future computing [14].

Cognitive Computing profoundly applies agent-based technology. In situations, intelligent agents comprehend high-level objectives and learn autonomously on how to accomplish the objectives. The system interacts iteratively with outside world to learn, purposefully reason, and consequently, update till the objectives are not achieved [11]. By the processes of learning, Cognitive Computing gets better over time and builds its own world of knowledge. With time, the error margin reduces and the quality of analysis and prediction improves. As with time, Cognitive Computing develops deep domain expertise, thus dependency in comparison to the expert systems could lead to a better decision whether in healthcare, finance or customer service [1].

Cognitive systems are not programmed in prior rather they are designed to augment themselves by learning through training, interactions, previous experience, and past reference datasets [13]. Thus, in contrast to conventional programmable (von Neumann) computers, Cognitive Computing does not limit itself within the deterministic boundaries. It brings a dynamic essence to the systems by continuously sensing and learning from the surroundings and enhancing its decision-making capabilities [15]. Cognitive technologies extend the capability of computers in executing the tasks usually performed by humans such as handwriting recognition, face recognition, and other tasks which require human cognitive skills such as planning, reasoning, and learning from complete or incomplete data [16]. The fact that the applications that incorporate Cognitive Computing do not follow any pre-programmed logic with specific rules raises the obvious question on the performance, complexity, and effectiveness of these applications. The performance of a cognitive system for a database with n records and m fields, essentially by brute-force method it could be $O(nm^2)$. But for the system trained well, it could perform close to the best case which asymptotes to O(n). The longer the system run, the faster it performs [17].

Real world scenarios are random, chaotic, uncertain and ambiguous which leads to problems that are dynamic, information-rich, changing and conflicting in nature. Cognitive Computing apprehends the situation by taking into account the information source, influencing factors, and contextual insights. Context includes

features which describe what, when, where, and how an entity is engaged in its environment and the specific process to which it is involved with. Context provides serendipity to find suitable solution pattern in the massive and diverse collection of information, which may be a suitable response to the need of the moment.

Cognitive Computing is the blend of different capabilities as ML, NLP, cognitive vision, reasoning and learning, etc. Leveraging these capabilities, Cognitive Computing unlocks the information from massive data to develop deep and predictive insights [18]. The output of Cognitive Computing may be prescriptive, suggestive or instructive [19].

By exploiting past errors and success, cognitive systems help machines to learn and teach humans new concepts and/or behaviors [20]. The unique combination of analytics, problem-solving and communication with a human in the natural form [21] redefines the relationship between human and machines, whereby machines augment human in decision support and reasoning.

3.3 Beyond Automation

Is there really any difference between automated systems and the cognitive systems? Or is it just another technological gimmick? Well, practically speaking, they are different, fundamentally, by far. For instance, a sensor device can take some action according to the data it senses. This is automation but not cognitive. Let us elaborate. When we enter to our smart room the light is automatically turned on. This is automation. If the light changes its incandescence and probably the color according to time, weather and my mood, it is cognitive. When we step out of our room and shut the door, the lights are off. This is automation. But when we go to bed and wish good night to the room lights, they turn off. This is cognitive. Traditional automated systems can not perceive or express humanly emotion. In contrast, cognitive systems incorporate emotional behavior in their interaction [1]. The automated systems are pre-configured and hard-coded by the human. They follow rigid definitions and rules. Cognitive systems are soft coded and able to define their own rule. They make themselves complete through continuous learning by experience. The automated systems are good in operating on structured data, grasp it and organize into meaningful and well-directed data whereas cognitive systems, in addition to these basic chores, can handle unstructured data, learn and infer extra knowledge through each process and reuse the learning [1]. They are able to filter out and focus on the meaningful events only. Almost a century ago, when Ford introduced the assembly line, the elementary form of early industrial automation, the purpose was to save working hours from repetitive manual tasks. Since then, automation has been adopted in several industries so that the lower level workers can concentrate on more complex and creative tasks that machines can't handle. It helped organizations to reduce labor cost and time-to-market, eliminate inefficiency and augment overall production process. The application of cognitive technologies is extending automation to new areas that have never been thought of before. By adding cognitive

abilities to the existing automation, not only the basic processes will be executed faster, but human aptitude and judgment can be emulated to the automated systems quickly and at large scale [22]. Not only machine automation, it offers automation of knowledge work that can help companies become more efficient and agile in their businesses. It will take industries beyond mere automating processes at the lower level to provide significant decision-making knowledge to all level of employees and allow them to acquire high-end skills to take up and solve problems they previously did not have the source and time to work on [23]. The original purpose of automation was pure business oriented and that was to save operational cost. Though CIoT is receiving significant attention from the business community, the primary purpose of CIoT is to make people's life smarter and chilled out by taking us way beyond the unadorned and artless automation.

4 Internet of Things and Cognitive Computing

4.1 The Cognitive IoT

The purpose of the IoT is to eliminate the boundary between human and the physical world by personally connecting to our surrounding objects and sharing information about them with us, as naturally as possible. But owing to the complexity and the scale of IoT, that cannot be realized by the basic form of IoT [24]. To reap the absolute benefit from IoT, we need to employ Cognitive Computing as an add-on what we call as Cognitive IoT (CIoT). CIoT is aimed at improving performance and to achieve intelligence of IoT through cooperative mechanisms with Cognitive Computing [25]. Today's IoT generally focuses on sensing its surroundings and act accordingly. The decisions taken by the devices connected to IoT are generally based on pre-programmed models. They can infer based on the sensed data available. But they are not full-fledged autonomous systems which can take their decisions very much depending on the immediate context. By infusing sense into IoT, Cognitive Computing enables IoT to interact dynamically with other connected objects, as well as adapt to the present context through continuous learning from the environment. They will be able to observe, filter, and recognize, very similar to humans, and also assimilate that information to excerpt actionable knowledge and meaningful patterns [1]. They will understand the context based on the domain where IoT is applied and act accordingly [13]. Cognitive Computing can be exercised, principally, in three different facets of IoT as described below.

1. **Networking aspect**: The networking aspect of CIoT is basically an extended concept of cognitive radio [26] and cognitive networks [27, 28]. Cognitive networks try to achieve the optimum performance by adapting to the present condition of the network. Using cognitive networks, the CIoT can make intelligent decisions through comprehending the current network condition and analyzing the perceived knowledge [25]. Thus, CIoT can take up necessary

adaptive measures to maximize network performance and minimize latency which is crucial especially to the time-constrained applications. It also considers the route that requires less energy for data communication because the sensor devices normally have limited battery life. It helps saving energy also by dynamically establishing demand-based connection and disconnection among IoT devices. The cognitive network can increase network capacity through intelligent multi-domain cooperation that will be a big boon as the volume of IoT data is growing incredibly [25].

2. **Behavioral aspect**: Like humans, the IoT devices are also intended to sense the inputs, in the ideal case, in the form of vision, sound, taste, smell, and touch. The human mind is naturally capable of negotiating and collaborating with these ecological inputs, process them, and reason the output [1]. Embedding cognitive technologies to the IoT devices makes them intelligent and sensible much like a human. In addition to their normal properties i.e. sensing the surroundings and sharing this information, by mimicking human cognition capability, they will be capable of learning, thinking, and comprehending, by themselves, both the physical and social worlds [29]. The cognitive things can take actions on their own and interact directly with humans by understanding and use their (human's) language [15]. To recognize and to differentiate among users are the preconditions of learning. Cognitive devices should be able to recognize different users through a variety identification attributes such as the face, fingerprints/touch, voice, and usage pattern [1]. According to the learning from previous interaction experience, they will be able to adjust their future interactions. CIoT not only can sense but also learn to anticipate, along with time, different emotions and respond accordingly [1]. They are able to adjust their response according to the mood and the changing emotions of other systems they are interacting with.
3. **Data analytics aspect**: The extensive application of IoT is producing enormous data and in the coming years, it will become the largest source of digital data in the world. Processing and channelizing this huge amount of data into the right direction and utilizing them for purposeful usage will become a challenge. In most of the IoT applications, the maximum chunk of these data is discarded just because of lack of capacity to handle the data of this size. Even if in some cases, the Big Data platforms such as Hadoop are used to store all these data, they are not being fully utilized due to the technical limitations of these platforms and the computers [30]. The traditional approach of programming that is based on series of conditional statements are not sufficient to handle this enormous amount of data [24]. Something special is required to facilitate all the blended transactions for the smooth functioning of these smart devices. Something special is required to mine and understand the multifarious and unstructured data generated out of these devices. Cognitive Computing is that special thing. The probabilistic nature of cognitive systems enables us to square with the voluminous, complex, and unpredictable IoT data [24]. IoT data can be converted into intelligent data through cognitive techniques which should aid organizations in automating tasks, designing better products, and innovate new customer-centric services

[31]. CIoT will recognize the organizational goals, accumulate, integrate, and analyze relevant data to help businesses achieve those goals [24].

4.2 Properties of Cognitive IoT

CIoT is inherently distributed as well as pervasive in nature. Hence most of the aspects of these two computing paradigms are expected to be engrossed by CIoT as well. Apart from those, some other standard (and desirable) properties of CIoT are mentioned below [19].

Self-learning: CIoT should require minimum explicit programming. It will continuously learn from the environment, the dealing with other entities, and from the events and continuously improve itself. Ideally, the learning should be unsupervised rather than supervised i.e. they should learn by themselves without any pre-set parameters. Cognitive Computing works based on continual hypotheses formation. Every time CIoT learns something new it seeks approval from the existing hypotheses [17]. Depending on the outcome it updates its memory.

Probabilistic: CIoT is not structured either deterministic i.e. it cannot be defined by formal language. It interprets the input and the context probabilistically. Deterministic systems are designed to follow the algorithms that are expected to adhere to the pre-set pattern and support specific data sets. Whereas CIoT does not follow any pre-set pattern. It sets out with a hypothesis, learns as it goes on, and exercises statistical methods to find a correlation between the new findings and its hypothesis. Based on the outcome it may change the previous hypotheses and possibly will find a new course of operation.

Adaptive: It must adapt itself to changed conditions (both physical and logical) [32]. It should also learn to adjust itself as information changes as well as when new objectives and requirements develop and update the acquired knowledge in the case of any modification [19].

Flexible: The self-learning and adaptive capabilities give CIoT a lot of flexibility in ingestion and processing of input data. If, for example, while processing, an unwanted data or variable comes up, CIoT will adjust its processing model to incorporate that deviance, unlike the traditional IoT where the program may have to be rewritten [17]. This becomes very expedient in the unpredictable environments, especially where the content itself keeps evolving.

Dynamic: CIoT must be able to handle real-time (both soft and hard) data. It should possess the dynamic capacity to process and analyze data on-the-fly. Moreover, it must resolve ambiguity dynamically and also deal with unpredictability.

Interactive: CIoT is designed to be interactive with people (users), machines, and other automated services. CIoT is especially remarkable for its ability to interact with people in a fully human way. Interacting capability using natural language makes it very powerful. It can take input from natural language or

unstructured text and give output in the same form [17] and mainly due to this, CIoT has gained widespread applicability be it personal or business application.

Iterative: Self-learning is always an iterative process. CIoT also follows the iterative pattern: make a hypothesis—learn further—update the hypothesis.

Stateful: CIoT has memory. It remembers the states of every transaction and is capable to trek backwards and forward to locate something if needed [17]. Because it is a self-learning, adaptive, and iterative system, it must retain the previous interactions and acquired knowledge at on every occasion so that they can reminisce whenever required.

Unstructured data friendly: CIoT can handle most of the unstructured data types [11]. This has allowed it to be integrated with the majority of the data sources in the world thus, increasing the scalability and the scope of knowledge acquiring.

Highly integrated: Though each individual device in CIoT works (sense and learn) independently and most often autonomously, they all team up focusing to contribute to a central learning system. In that sense, every device is closely bonded to each other through continuous interaction, sharing information, and updating own knowledge accordingly [33].

Scalable: Working range of an IoT can be confined within a small room or expanded over a whole city. The pervasive and ubiquitous application of IoT makes it essential to support appending (or removing) of mobile devices to (or from) the system dynamically. Therefore, CIoT should support highly real-time scalability. In this aspect, one thing goes in favor of CIoT that it makes fewer redundant calculations, which is very crucial for scalable systems [17].

Context- and situation-aware: The validity of IoT data and inferred knowledge depends on the particular contextual and situational information such as time, location, application and administrative domain, regulations, user's profile, process, task and goal [19]. CIoT is able to identify, read, and extract this information and apply correspondingly.

Self-management: The extraordinary scale of CIoT makes it impossible to manage manually or even using automated management tools. It is ought to be self-managed. There are different elements of management of CIoT. Of them, some important ones are:

- *Diagnosis, troubleshooting, and maintenance.* Ideally, CIoT should diagnose and troubleshoot itself besides regular maintenance, without any human intervention. It should be able to observe the behavior of its components, assess the state, and predict likely trouble spots [1]. Through this proactive tricks, the unnecessary interruptions can be avoided. For scheduled maintenance, which includes battery replacement, network inspection, other hardware looking over, etc., CIoT will automatically find a suitable slot for this so that the maintenance downtime is minimum and least hurting for the business. The system will intuitively decide when to repair parts and to replace in case of worn out.
- *Fault tolerance.* Since IoT devices are delicate and error-prone and generally works over wireless networks, there is every possibility of frequent fault occurrence. CIoT should be able to recover, if possible, or mask these faults.

Most of the IoT applications gives output as events based on some triggered action. So, it may be confusing to differentiate between an error condition and a trigger event. CIoT should mark this distinction correctly by applying its intelligence.

- *Performance management.* CIoT will not only be able to point out the reason for performance degradation, it will try to resolve the snag either by reconfiguring, if possible, the deficient component or replacing it or bypassing it [1].
- *Configuration management.* As discussed above, CIoT is supposed to be scalable to accommodate new devices and applications. When new devices are added (or exits) or application program is modified, it should spontaneously adjust to the new configuration according to the modified settings. Any problem caused due to change in network configuration, especially in the case of cognitive networks, should be smoothly resolved on its own.
- *Security management.* It is no way deniable that IoT is seriously vulnerable to security threats. CIoT should anticipate these threats and shield itself accordingly.

4.3 The Pillars of Cognitive IoT

Endowing cognition to non-living objects is not easy. We require to provide them with the interfaces through which they will interact with the outer world like eyes and ears in humans. They have to be empowered to be able to comprehend the inputs they get through these interactions, interpret them in the context of the event and deduce to some logical and rational response. They also have to be trained to understand and respond in the language used by humans for natural communication. In fact, to build successful cognitive systems, every facet of human cognition should be meticulously simulated on those objects. For that, different fields of study in AI (Fig. 1) are to be consulted. For example, ML will mimic the brain to enable CIoT to learn themselves without any supervision. Machine vision will be used to give CIoT the eye. Speech recognition will give CIoT the ear as along with NLP they will be able to understand human language and also respond in the same. The ontology will give a uniform knowledge scheme irrespective of any differentiation in vocabularic terms and names. Out of those, the three most significant fields are discussed briefly in this section.

Machine learning. ML is one of the major wings of AI in which we keep track of the past events and activities of what a machine has performed and create a future expectation from this learning. It refers to a learning technique whereby machines/computers (more precisely the algorithms) learn from a set of past data (input and outputs), thereby detecting patterns and inferring information by using a bunch of mechanisms based on statistics and mathematical models. The methods iteratively learn from the data and find the hidden insight without being notified where to look in the data set. Based on specific training over datasets from a number of test-cases,

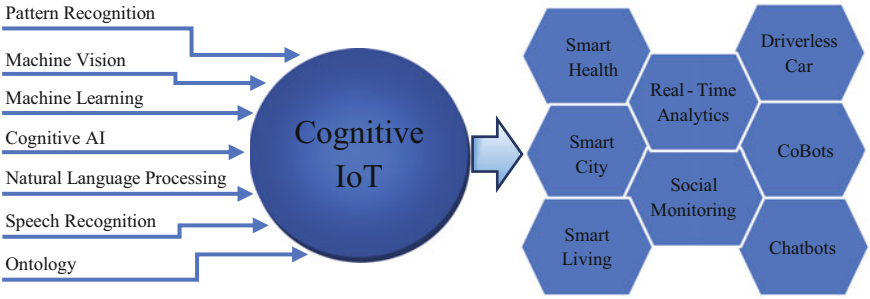


Fig. 1 Constituents and applications of cognitive IoT

it can make an opinion about data and learns intrinsic data pattern. Upon exposition to some unforeseen event/data, based on its learning experience ML can classify or annotate the data and can take necessary action.

ML involves various techniques and learning algorithms from Statistics, Mathematics, Computer Science and Neuroscience. These algorithms and techniques are categorized as supervised and unsupervised learning algorithms. In supervised learning, machine learns from set labeled test case whereas in unsupervised learning the machine learns from unlabeled test data. In the absence of labeling, machine identifies and correlates the data itself. Bayesian Statistics, Hidden Markov, Neural Network, Support Vector Machine, Decision Tree, Principle Component Analysis, k-mean algorithm, etc., are the popular examples of the different algorithms and techniques in ML. Further, with the involvement of cognitive science, this ML becomes more human-prone behaved.

ML applied to business and industry, produces accurate real-time information and predicts future events [34]. ML has a significant potential in IoT appliance in real life and business application domain. ML in IoT follows an iterative process where the necessary action taken by the 'things' is taken back as feedback to learn more and thus improving the learning model a bit more. ML is employed to analyze the IoT data and build an analytical model for future predictions. Real-time data sensed by the 'things' are analyzed over the model to provide real-time information. The data is transformed into machine knowledge which gives a clear insight into the IoT world [35]. Examples of ML in the IoT-based applications are: detecting anomalies and defective parts in the avionic industry, finding an anomaly in logistic operation and also their cause and effect, monitoring running status of elevators and resource consumptions by a building etc. ML is required for different purposes in IoT applications such as speech recognition, computer vision, bio-surveillance and robot control [36], smart meters, analyzing traffic congestion, pollution measurement, weather prediction, traffic control, city management, health monitoring etc. [35].

Computer vision. Computer vision involves processing visual data (video or image), to extract the meaningful information out of it. One of the persistent challenges that have attracted researchers' attention is identifying objects or things

in video and their relations. Techniques like ML algorithm and geometrical operations have been found commonly applied in various research to process video data in the frame by frame manner. Processes like feature extraction, feature matching, keypoint detection and description, key point matching and object class recognition are frequently used to recognize the contents of the image and their spatial and temporal relationships. In general, computer vision process involves feature extraction of the image in question and matching that to stored images. This enables to recognize the class of the object and further drawing the relationship among them over the time. In this direction, over the last decade computer vision has really moved towards cognitive vision with the capability of semantically identifying the objects in the captured image and video and further reasoning and correlating them logically.

Cognitive computer vision definition is presented in [37] as:

> A cognitive vision system can achieve the four levels of generic computer vision functionality of detection, localization, recognition, and understanding. It can engage in purposive goal-directed behavior, adapting to unforeseen changes of the visual environment, and it can anticipate the occurrence of objects or events. It achieves these capabilities through learning semantic knowledge (i.e. contextualized understanding of form, function, and behavior); through the retention of knowledge about the environment, about itself, and about its relationship with the environment; and through deliberation about objects and events in the environment (including itself).

Cognitive vision presents a strong ground to achieve robust, resilient and adaptable computer vision having ability to learn, cognition to adapt to the new strategy for interpretation and analysis. It requires attention from two separate fields: computer vision and cognition involving an array of disciplines like Computer Science, Psychology, Neuroscience, AI, Mathematics, and Statistics etc. [37]. In [38] neural network technology has been used to realize how human visualize. in the form of symbols. Correlation between the symbols in terms of spatial dimension and their relationship in the physical world is modeled. Semantic interpretation is inferred between the object and the physical world.

IoT interacts with the environment in various ways. To interact meaningfully with physical objects, information about them are required to be known first. Computer vision has advanced by a big leap in identifying real life objects, which serve better visibility in identifying what entity the system is interacting with. In recent years, with technological development, mobile devices are getting smarter with almost all the mobiles have camera and internet connection. A huge population across the globe do possess smart mobile devices which has enabled to realize IoT in large-scale. The potential of these devices can be harnessed for reaching IoT to each and every corner in the physical world. IoT associated with computer vision can endeavor endless possibilities, to reason the events that are taking place in the physical world. Visually identification is more natural in comparison to object identification using marker technique (RFID, QR code, Barcode) in terms of scalability and anonymous object identification. The advances in computer vision have led IoT in the voyage to discover the endless limit to recognize an object, its pattern, and situation which enables the machine to have a more humane sense of

real world [39]. Examples of IoT with computer vision are like traffic management, security and vigilance, patient health monitoring, and automatic navigation. Swimming pools equipped with computer vision are able to monitor and raise alert for a person struggling or drowning. In agriculture and farming, computer vision can monitor for cattle grazing in a field and sick animals in a farm.

In spite of recent advancements, a true cognitive computer vision has not been reached yet, it is a long road ahead before the cognitive computer vision reach its maturity [38].

NLP. The effort to make computers understand natural (human) language is not new. Over the years, NLP technology has evolved with development of computing algorithm and computing performance. NLP involves processing the human natural language to extract its meaning and the intention. NLP begins with understanding and processing of letter, word, and phrases. Sentences are analyzed for grammar (syntactic analysis) and meaning (semantic analysis) using different techniques like tokenization, lexical analysis, word sense disambiguation, morphological analysis etc. For understanding the deep structural meaning and semantics, grammar models are constructed. Textual sentences tend to be abstract and have hidden meaning within. Appliance of Hidden Markov Model is useful to understand the hidden meaning within textual sentence. NLP associated with IoT could be quite beneficial in exchanging dialogue and query in various search and automation process [40]. One of the key requirements of Cognitive Computing is to make systems capable of taking input in natural unstructured text form. The unstructured text data are analyzed to find the valuable insight and which helps to reason and predict user over a long time. NLP bridges the communication barrier between human and machines. In our daily life, we come across many NLP application like Droid, Siri, Cortana, etc. which gives interface to interact naturally in textual form. But these applications lack cognition to comprehend and reason user on the basis of text input. Processing natural language without human intervention is one of the persistent problems, though many research had taken place but leading solution is far reach [17].

4.4 Challenges of Cognitive IoT

Theoretically, the idea of CIoT may sound brilliant but practical realization of it is surely not going to be straightforward. To make the vision of CIoT true, a number of technical challenges should be resolved. Below some of these challenges are summarized.

Limited battery: For any wireless device, power is a serious issue. CIoT devices are no different. They require constant power; in fact, more power than traditional IoT because besides usual sensing and data transferring, CIoT devices functions added cognitive activities. Recharging these devices and replacing batteries require continuous human involvement. Which means the amenity that is earned through CIoT may be outdone by this hassle.

Diverse data types: To find the relevancy, the IoT data is required to be analyzed appropriately which is a challenging job because the amount of these data will be enormous and most of them are unstructured. The data are drawn from different sensors which include visual, audio, gesture, text or other form of input. The data inputs are completely heterogeneous in type, most representing different meaning and source of information; quite apparently, the data includes both structured and unstructured form of information in it [19].

Privacy and security: Privacy and security are always among the biggest challenges in a networked system. The threat increases as more devices joins the network, as in the case of IoT. Every device becomes the probable entryway for the hackers. The application of cognition over data may reveal valuable insights, but become vulnerable to exposition raising privacy and security issues. Privacy to personal and business data and information is one of the critical challenges. Different encryption algorithms allow securing the data. These algorithms are complex, time-consuming and high computing resource constraint. For fitting suitable encryption algorithms, the challenge lies in developing an algorithm which supports distributed key mechanism, fast and is energy efficient. The issues which need to be addressed are:

- Decentralized authentication and security model for IoT application.
- Data protection algorithms and technologies which are fast and resource and energy efficient.

General privacy concerns are associated with IoT. For example, will it be safe to share user's personal information (e.g. location) with other IoT devices or applications? Severe privacy concerns are there for business aspects also. IoT system seeks customer behavior pattern (preference, buying, and usage) to generate market analysis figures. Different business planning and campaigning are staged based on these analyses. Connected devices with intelligence have deepened the security and privacy issues further. Cognitive IoT enable to foresee the personal preferences, needs, and suitability, these data are vulnerable to accidental leaks and are lucrative to fraudsters [1].

Training: The efficacy of the CIoT depends on the quality training it receives. In CIoT, the cognition of things get better with continuous iterative training. This put question forward as how the things can be trained iteratively when the environment is changing, furthermore how long the things need to be trained before it can be applied to perform with all its senses [1]. Cognition gives things the ability to remember to what they have learned. This bridges the learning gap among things, but can they learn new things autonomously and furthermore can they apply their 'common sense' creatively to assess effects if the situation changes dramatically. While training for specific subject domain with task specialization it may need additional background knowledge to deal with exceptions of specific types [1]. Finding background material appropriate for training is one challenging issue. Cognitive things learn from past interactions. Based on the success and failures, it meticulously calculates the best option for new problem situation and improves the

future interactions. However, things are not creative in their cognition, this emphasis that the training must explicitly deal with how to respond appropriately to unknown situations [1].

Device organization and data flow management: IoT devices generate data in continuous time series and in most of the cases the data must be processed and analyzed on the go. The outcome of one device may be fed to other or may be fed back to itself for actuation. Hence, it is important that information should be flowing rightly from one device to another. It is a challenge to manage and routing data flow among different device [34]. A crowdy network like IoT with multiple devices connected to it for information exchange has challenges in searching device and devising protocol to exchange different data types. Right flow of information helps in steady information generation and synchronous consumption that will lead to streamlining the business process (in case of organizational IoT). Hence it is crucial to keep the devices, in the network, in harmony. But what should be the ideal device organization for optimum data flow? It has to be determined whether the devices will have full autonomy in job execution or the organization will have the central controlling with common standardization for the entire organization. Decentralized organization gives better operation in a big and distributed organization which has its own complexity, pace, and separate policies to run business. While centralized organization of things offers better control and management. However, the arrangement of the IoT devices should be in accord to the organizational policies.

Network and communication challenges: IoT is a complex network with varying devices and sensors in wired and wireless mode, fixed and mobile modes. This had raised challenges in managing, interoperating and troubleshooting the network. Realizing IoT for a practical application has many communication challenges, especially when the sensor network is a wireless one. Since wireless network sensors consume high energy, it demands the development of energy efficient communication protocols. Further, the heterogeneity of wireless sensors demands multi-frequency spectrum to overcome the frequency spectrum conflicts. Besides these, the ad hoc networks made of collaborative mobile things introduces other issues such as scalability, sustainability, adaptability etc.

Being autonomous, CIoT network is able to dynamically change, evolve, and discover self and other networks. The network should have the adaptability to add new ‘things’ to its existing topology. The automated discovery of things and mapping capability is a big challenge but required for accurate and efficient management of network in terms of scalability and operation.

Software and algorithms: The software is integrated into ‘things’ at various network levels to communicate and assess the network data. Software works in different environments having heterogeneous types of ‘things’ communicating using different protocols. One big challenge is to develop CIoT applications which would integrate different software module, operating at different environment coherently. The big issue is how to fabricate the distributed software into one knit, addressing service oriented issue and supporting machine to machine and things to

things interaction over the network. The demands which are needed to be addressed including:

- Distributed self-adaptive software capable of self-management, self-optimizing, self-configuring, and self-healing.
- Open and energy efficient software platform capable of integrating the distributed software modules as coherent application abstracting the network resources and communication.

It is essential to identify and minimize the biasness in the programs. The bias is introduced into a CIoT program either by the training data or the way the algorithm is designed. The presence of a specific type of information (e.g. demographic) in the training data makes the algorithm inherently bend to resolve the specific type of problems while proving inefficient for other problem cases. The system could be biased by the way algorithm process that information. The bias may also be introduced by the developer into the code to solve any particular type of problem either intentionally or unintentionally. To maintain the neutrality and objectivity of an algorithm, bias management is really challenging. It is important for an AI system to state how and why the system arrived at a particular conclusion so that a human can evaluate the system's rationale [41].

Societal apprehensions and ethics: Social concerns must also be taken care of while training the machine. Researchers have shared their views in terms of societal concerns of AI in IoTs, having similar opinion that too much dependency on AI may lead to a cyber war. Technological developments should be intended to augment human intelligence to enhance their capability, expertise, and potential but not to replace it. The leading players in AI like IBM, Google, Facebook, Apple, Microsoft, Amazon etc. are taking extra caution in developing and implementing AI so that it should only benefit the society [42]. It is not enough to foresee how the CIoT should be used, but it should be judiciously anticipated how it might be abused. It is equally important to study and consider the "misuse" cases as the use cases of CIoT [43]. In future, our life will be highly influenced by Cognitive AI and CIoT. These technologies, in a way or other, would change the social paradigm. We will be responsible for architecting the social changes, and this is a profound and daunting responsibility [42].

5 Business Value of Cognitive IoT

IoT has already created a huge hype among the businesses. Not only big players, SMEs are also sensing lucrative potential in adopting IoT. It promises to bring value to all types of businesses by reinventing the business processes and operations that will eventually enhance level and quality of products and services as well the customer experience [44]. As IoT has provided them the most important element of business—the data acquisition cog, organizations are emancipated to

collect any sort of data (e.g. contextual, locational, etc.) either related to business process or customer. IoT can contribute in improving business operations in several directions [45–47]:

Improved business process: The massive connected data from IoT makes business processes smarter. Analyzing the data collected from every division of the business will give new insights and knowledge.

Increase business opportunities: IoT opens the door for new and innovative business opportunities and creates further revenue inlets. Exploiting acquired knowledge through IoT, corporates will be able to develop advanced and new business models, locate new markets to extend services and diversify their product line.

Uplifting business moment: Businesses can earn competitive velocity and agility by capitalizing and toning with the influx of dynamic and crucial business data generated through IoT devices across the domains.

Increase productivity: IoT helps to identify the need and lack of workforce expertise and also enables organizations to train the employees just-in-time. This improves workers' efficiency and reduces mismatch of skills which in turn increases organizational productivity.

Improved operational efficiencies: The real-time sensor data from IoT devices enables organizations to monitor business operations observantly, minimizing human intervention. If IoT data collected from logistics network, factory floor, and supply chain are utilized judiciously, inventory management can be optimized, and time to market as well as downtime due to maintenance can be curtailed significantly.

Enhanced asset utilization: Industrial IoT enables tracking of the production equipment, machinery, and tools. Examining the real-time status, better asset utilization can be achieved.

Faster decision-making: The real-time business process and operational knowledge will help organizations to make faster and smarter business decisions. The connected nature of IoT facilitates dispensing the intelligence and hence decision makers are able to prioritize all business decisions.

Cost saving: All the above-mentioned gains of adopting IoT will eventually lead to saving the business expenditures. IoT supported organizations have the opportunity to well harmonize their physical assets which minimize the energy costs also.

The business potential of IoT lies on the effective utilizing of the massive volumes of data it generates as precisely reflected by the statement—"Data is money", stated by Nick Jones, research vice president and distinguished analyst at Gartner. However, only large quantifiable data means nothing; the honey is the relevant data [44]. But measurable business and financial benefits may not be achieved even through relevant IoT data alone. To transmute the IoT data into money, they are required to be analyzed decorously to extract decisive knowledge so that the CEOs and the CIOs can take pivotal and purposeful decisions. Employing analytics tools on data accumulated from multiple diverse data points (thanks to the pervasive use of IoT) has enabled businesses to understand the customer demands and spot trends better. But these insights are general in nature. It would be even better if a business

tries to explore these data with a specific purpose and to find solutions to specific problems as possible [44]. As Nick further asserted, "Computers can make sophisticated decisions based on data and knowledge, and they can communicate those decisions in our native language. To succeed at the pace of a digital world, you'll have to allow them to do so." Here, Cognitive Computing will play the big role. These technologies can be leveraged to find patterns and regulate the flow of information through all the interconnected devices in the organization to streamline and alleviating the business processes [34]. By infusing Cognitive Computing with IoT, businesses can discover patterns, opportunities, and actionable business propositions that would never be accomplished through the IoT without intelligence [24]. CIoT has allowed companies to gather, observe and share an unprecedented amount of various data about customers, personnel, products, and business processes and operations throughout the organization [1]. Running different business operations becomes seamless by treasuring specific and accurate real-time information generated by the IoT devices [34]. This will result in the significant enhancement in workforce's performance which keeps organization's stride on a speedy and ascendant trajectory [24]. Decision making also becomes more straightforward and meticulous as the prediction of the future events can be done more accurately with the help of intelligent analysis. CIoT, with the use of visual analytics and data visualization techniques, can visually portray the analytical outcomes that equip humans with a better perception and set them in a better position to take a decision [1]. Since IoT can provide personalized and focused data, endowing it by cognition capabilities will help companies to analyze trends and find unexpected patterns for better decision making by narrowing down to the small but precise sets of highly imperative data [44]. Businesses will be galvanized to find new business insights and achieve improved productivity and efficiency [18]. Cognitive technologies can impact organization's workforce either by augmenting their proficiency or by replacing them. In both the cases, organizations will gain financially. Cognitive systems are able to master the professional linguistic of different professions (e.g. manufacturing, medicine etc.) and also can communicate with the users in natural language [24]. Hence CIoT may benefit organizations by eliminating the need of investing in the employees to become experts.

CIoT can bring significant developments to the service sector by offering highly individualized services. Though the service sector already has experienced remarkable innovations, in accord to the new approaches and developments in IT, consumer behavior, and consumption passages new service models need to be invented as well. With the advent of new digital platforms, dealers are finding new avenues to directly connect with consumers. To be more responsive to the increasing customer engagement sellers need to understand their customers much better which leads to better customer satisfaction. They require to collect and analyze consumer data. IoT has enabled them to collect contextual data of the users and the products they use. This will allow cognitive things to become the ideal shopping assistants by providing real-time product recommendations [1]. Differentiated and personalized shopping experience will be delivered pervasively through targeted recommendations and individualized communication. At the same

time, these cognitive things also collect the shopping manners/pattern of the customer that allows the companies to remold their marketing strategy and pricing. The cognitive analysis of IoT data will assist retailers in recognizing changing behavior and expectations of the customers and respond accordingly [48]. They can foresee consumer behaviors and desires in order to anticipate requirements before they are needed. CIoT will allow companies to design innovative personalized products and services that complement the user's choice. The users will be benefitted by having served individualized and optimized services that are capable of adapting to local factors automatically [44]. Through CIoT, new kind of personalized data can be captured that probably was not possible earlier. These data include sentiment and emotional state, speaking tone, and the strength and nature of a person's relationships [24]. Businesses can use these data to engage with customers more deeply and recommend and deliver services that are more emotionally relevant. The CIoT will enable companies to take proactive measures to avoid customer dissatisfaction by pinpointing the actual cause for a performance degradation [1]. Service providers can assess their service-providing infrastructure so that they can tune themselves in order to provide the highest level of quality service to the clients [49].

To summarize, businesses now can explore unforeseen possibilities that were previously either indiscernible or inaccessible [24]. In fact, business organizations are already employing cognitive technologies for quite some time now to boost their products, business process and business analytics [16]. According to IDC, a premier global market intelligence firm, by 2020, as much as half of the business analytics software in the market will be based on Cognitive Computing [50].

6 Applications

Machines are slowly evolving with the appliance of AI. AI has made machines more intelligent, adaptive, durable, aware and efficient. In the course of time, perhaps in very near future, AI will be transformed fully into Cognitive AI, thereby changing the social and technological paradigm. We will be advancing to a future where machine and human will synchronize in a harmonious way whereby machine will augment humans at every step of life. Perhaps this will reduce the chaos and randomness of the realistic world. Even though work has been done on IoT, AI, and Cognitive AI individually, the CIoT is merely in a proposition state. Though a lot of research are coming up, the realization of CIoT to its fullest extent is way ahead. Innumerable applications of AI and IoT will come up in the next five years of span [27]. In general, cognitive systems will be the next major technology which will significantly impact business and economy, healthcare, society and living, get recommendation and make purchases, etc. [50]. AI will elevate as well alleviate the existing technologies. Gradually, all the systems will be updated with the support of AI and Cognitive AI. This will give rise to the various opportunities to the early adopters of this technology. There is a lot of scope for existing IoTs where AI and Cognitive AI may boost the smartness and efficiency of the system.

Further, Cognitive AI supported IoT will create a scenario where there will be very less intervention of the human in handling these devices. A few applications of AI and Cognitive AI based IoT are mentioned below.

6.1 Smart Living

Smart home and smart environment are no longer a mere fantasy, various products which we come across in our daily life are smart. They can enhance our comfort of living by supporting in several ways in our daily lives. For instance, a smart home senses the presence of a person in the home and depending on the context of the person, suitable services are activated. Furthermore, future requirements are predicted on the present action of the person. For example, for a person's presence found in the room, the light gets switched on or off depending upon the environment luminosity. Similarly, if it is very hot and humid, the person's presence in the room starts the air conditioner. At morning, the alarm clock starts the coffee machine, and refrigerator communicates to the salesperson at the shopping mall with the grocery list to be purchased. At late night, the television sound automatically adjusts itself, and if the person found falling asleep, the television automatically gets switched off. Smart living simplifies our lives and reduces dependencies on others. Smart living has an important role in easing the life of ageing and elderly persons. But a lot has to be done in this regard and it needs more attention from the researchers in the field of IoT and Cognitive AI.

6.2 Smart Health

One of the biggest impacts of CIoT will be in the health care. Medical care devices embedded with intelligence could monitor the health status of sick people [51]. Any detritions found could be analyzed to avert dangerous medical condition, saving the life in time. Medical aid attached to a person could monitor heart beat rate, blood pressure, the oxygen level in blood, blood sugar, tiredness or fatigue-ness, seizures etc. Information could be aggregated to suggest the person what activity, food or medicine is suitable for him to avert any serious illness. In case of medical emergency, the health monitoring device attached to the person may alert the medical services. IoT helps to monitor elder sick people living remotely/alone at home. An early alert about an emergency medical condition may prevent unfortunate incidents [51].

6.3 Household Appliances

Cognitive IoT has enabled the home appliances to be smart. The application connected to the Internet enables to associate cognition which allows the device to

learn user requirements, activity pattern, and their daily schedule. The auto-learning makes home appliances smart, whether it's a television, washing machine, dryer, coffee machine, refrigerator or security system. Cognitive devices must recognize the user by voice, face, touch or fingerprint and thereby providing service based on user past interaction [1]. In this direction, commercial companies are updating their devices to make them smarter. For example, Whirlpool is manufacturing smart washing machine which could be controlled by mobile device [52]. So, that user can program the washing and drying based on their fabric need. Moreover, these machines can detect the detergent use and defects in the machine. The valuable data generated by the device can be used as feedback to engineering and development for building more intelligent machines that can assist human more smartly. The future washing machines will be more intelligent in scheduling workload, identifying the wash pattern for clothes, and minimizing the water uses [1].

In the near future, we will be surrounded by the smart home appliances, which will have their own cognition and would assist the user by detecting the user's intention and schedule/uses pattern. People will be less intervened to interact with job process; jobs will be performed in an automated fashion.

6.4 Smart Cities

It is expected that by 70% of world population will live in urban area by 2050 [53]. This will put enormous strain on city resources and infrastructure resulting in resource scarcity. The appliance of CIoT in day-to-day management and activities in city can help in long term development and planning. Information gathering like water usage, electricity and other source of energy usage, public transportation, people rush, parking space occupancy, traffic condition, weather condition, environmental condition (noise and air pollution), water contamination, etc., can lead to better decisions on city infrastructure and resources management [54]. The appliance of CIoT in city planning and management will make cost-effective municipal services. To mention another application, CIoT-captured traffic data would help in efficient transportation planning and operational activities. This would make public transportation more reliable by proper traffic rerouting and signaling, and further addressing the issues like commuters rush, road blockages, congestion, etc. In future, Internet connected parking meters would communicate with driverless cars for available parking space [1]. Furthermore, CIoT will augment garbage management by automated communication with and among separated garbage bins for better garbage collection etc. The appliance of CIoT along with loaded information will provide a robust platform for information processing and communication, thereby delivering quality services and information like current happenings in the city to citizens.

6.5 Wiki Cities

The concept of WikiCity has been derived from the Wikipedia, the online crowdsourced knowledge repository. A WikiCity is the knowledge repository of IoT data for a specific location of a city. It gathers various information about a city such as temperature, humidity, weather, garbage level and disposition, pollution, green area irrigation system, traffic light and traffic movement, etc. [55]. This information is collected through numerous physically connected sensor that are scattered throughout the city. These sensors sense their surrounding events and send the information to the knowledge repositories which are updated regularly. Just like Wikipedia, people can edit and access this information through simple web pages. For example, pollution particle level in the atmosphere at any time of the day may continually be updated in a database that records the daily pollution level and can be viewed on the page say, "Daily Pollution". Similarly, weather related information such as temperature, humidity, sunlight, snowfall, rain, etc., and city traffic related information such as the traffic status at any instant of time over any part of the city can be checked on the same website [55]. The combined information taken out of the database can be used by the administration to get the overall picture of the city on different parameters and in the case of any crisis, the possible solution also might be found depending on available data. For example, checking out the pollution levels in different part of the city, the traffic could be regulated accordingly. Similarly, traffic may be controlled in certain areas based on the information about rain and the drainage system of that area to avoid traffic jam due to water logging. This concept of WikiCity is already has been deployed at some of the areas. For example, Smart Santander (Spain), Amsterdam Smart City (Netherlands), and Songdo IBD (South Korea) [55].

6.6 Driverless Cars

Driverless cars which appear to be a fiction story is no more an object of the imagination but a highly sophisticated reality. The blend of different technologies like IoT, AI, and cognition has made it a reality. Driving is a cognitive activity. Driver while driving looks for road obstacles, road turning, stop signals etc. and apply cognitive activities like braking, turning the steering, changing gears, and speeding or slowing. If the driver, for any reason, cannot apply his cognitive abilities, it may cause a fatal accident. The driverless cars imitate human driving skills. They include a large number of sensors or 'things' which collect dynamic information of the car, the road, and the surroundings [1]. But how IoT/sensors could replace driver? The collection of sensors around the car and inside continually sense the engine status, gear state, road condition, the vehicle ahead, congestion on road etc. By processing the sensor data intelligently, different cognitive actions like gear change, cruise control, automated lane change and parking can be

performed. The various jobs which a driver can do could be replaced by CIoT with sensors mimicking driver eyes and ears, furthermore, cognitive AI mimicking the driver intelligence. A car with cognition communicates with others to negotiate the passage, thereby eliminating signaling through sound and light. The cognitive AI in the car automatically calls for service assistance for if the sensors detect any failure. Often necessary action like stopping the car and raising alert is enforced for any serious anomaly is found in machine functioning. Besides these, the cognitive AI can automatically select the journey route by taking the car data (journey source and destination, in between hops, etc.) and the traffic data into consideration. Moreover, the observed past data of car and the traffic data could be used for traffic management, signaling, traffic rerouting etc. [1].

6.7 Social Monitoring

Another important application which centers on IoT is the social monitoring. IoT monitors events or situations and initiates the appropriate preventive measure for the safeguard of people. For example, a car monitoring the emotional and behavioral state of the driver, consequently take appropriate preventive measures to avert any mishap such as alerting concerning authority or another vehicle [56]. Automatic climate and light control in office, based on the mood or stress of people. People accidentally leave their belongings while in transit or often items are left unattended. Automated monitoring could help to locate the right person or safeguarding people by alerting authority for any suspicious item left unattended. New sensors enable IoT to "hear" sonic information from people grouping [44]. The other social impact of CIoT is assisting humans to interact/communicate with each other effectively increasing community dynamics [57].

6.8 CoBots

Cobot (collaborative robot) is a robot which is capable of physically interacting with the human. It was invented and patented in 1996 and 1999 respectively by J. Edward Colgate and Michael Peshkin, professors of mechanical engineering at Northwestern University. The purpose was to make robots work together with the humans and assist them proactively as far as possible. IFA (Institut für Arbeitsmedizin der) has defined Cobot as [58]:

> Collaborative robots are complex machines which work hand in hand with human beings. In a shared work process, they support and relieve the human operator.

Typically, Cobots operate autonomously, though in cases they may need a little guidance as well. Cobots are empowered with the heavily simulated characteristics of human behaviors. The Cobots employed in indoor office environments perform

service tasks much like a helping hand. In executing their duties if they find themselves in a position with limited perceptual or physical (as they do not have arms) or reasoning abilities, they can also proactively solicit help from human workers [59]. They may also be employed in the production floors as the friendly robots. In the context of industrial employment, the significant difference between Cobots and the traditional industrial robots is that Cobots have cast out the requirement of safety guards in most of the floor operations as it was very much necessary in case of robots to avoid any mishaps.

6.9 Chatbots

Chatbots is a conversational entity that is programmed to carry out conversation, over the Internet, with the humans in natural language by mimicking human conversational behavior and pattern. The conversation may either be textual or aural. Chatbots use AI and applied NLP to process chatter's textual data and to synthesize its own chat text. Chatbots have been successful in text-based messaging. People raise a query in textual form and Chatbots with its fullest AI and NLP capability, response meaningfully by exchanging textual dialogues. Chatbot finds its extensive applications in e-commerce, customer services, call centers and Internet gaming. Facebook Messenger, Snapchat, and Kik have applied chatbots for entertainment, while company's like Pizza Hut and Disney have used chatbot to engage customers and promote products and services. However, you may raise a legitimate question —what IoT has to do at all with an automated chatting program? Well, the Chatbot certainly is not an IoT. But it may be a part of IoT; a significantly important one. Chatbots may become the natural interface of CIoT. We can talk to the CIoT using usual language. It will free us from remembering any command structure and the interface sequence [60]. For example, Chatbots will allow the driver of a smart car to communicate with several devices embedded in the car without taking hands off the steering. Apple Siri and Android Auto are the good examples of this [61]. Intelligent Chatbots can refine IoT user's commands based on the context, for subsequent interactions and operations [60]. Not only human, Chatbots will allow IoT to communicate with other devices outside the local network. So, chatbots will make CIoT cognitive in the true sense by empowering us to talk to IoT anytime and most importantly through our own language.

6.10 Weather Forecasting

Today's IoT device can sense and measure environment condition. These sensed data could suitably be used to infer atmospheric information. Data collected from sensors attached to wearable devices, cars, buildings, smartphones even from social media post produces enough information to accurately forecast weather for any city

or locality. For example, at The Weather Company, atmospheric pressure data, sensed by plenteous cell phones, are collected and processed by cloud technology to forecast weather accurately and instantly [62]. The weather forecast predicted in advance could be significant information. It might help in suggesting daily commutators for early departures or take the umbrella if it is to be raining. Children suffering from cold or asthma are intimidated to take precautionary measures if the temperature is about to fall. This type of applications need sophisticated atmospheric models which can be achieved by assimilating Cognitive Computing and IoT with atmospheric science. These models can predict the weather rightly, which will improve the decision-making that is correct and timely and also personalized to individual need/demand to cope up with the weather crisis [62].

6.11 Real-Time Analytics

Connected IoT devices often make a large complex network. These devices continuously generate data which need to be analyzed in real time for taking appropriate action in right time and at the right place. The appliance of AI offers real-time analytics capabilities into IoT system [34]. Real-time analytics is all about data analysis in actual time, so that system could react in the very same time window frame in which information or action request is made. Often no or low human intervention is required for starting the process. The smart sensors will automatically help in the sophisticated analytical job. Real-time analytics is a time critical process which in terms depends on factors like network latency, data processing speed, pattern recognition, information inference from past data, storage and retrieval of stream data etc. Using CIoT, business divisions can infer insights from the sensed data and communicate to another business divisions for real-time and context-specific decision-making [1].

7 Use Cases

CIoT is still in its neonatal stage. Hence, not many actual implementations of CIoT can be found in practice. But people have started noticing it and realizing the potential. Though in small-scale, CIoT has been implemented successfully in different applications. A few most talked about the commercial application of CIoT have been mentioned below.

Nest labs' smart thermostat: Nest lab's smart thermostat, utilizing IoT, checks and controls the room temperature (either heating or cooling) on a personalized basis, thus saving huge energy. It learns the habit and pattern of the occupant's comfort level for room temperature and prepares a personalized cooling/heating schedule for the room. It detects people's presence and set the temperature appropriately suitable for that particular person and when he leaves the house, it

switches off the air conditioning system. Furthermore, it checks the air flow to determine a defect in the cooling system and thus raises alerts timely. The system is connected to a smartphone through Wi-Fi or the Internet allows to control and monitor the room temperature remotely [63]. Leveraging AI, Nest's devices learn and adapt energy uses thus saving energy, a novel application with pragmatic benefits.

Tesla Motors' self-driving vehicles: At Tesla, IoT technology has been significantly applied to make the cars smart. Different technologies like sensors, AI, cognition, and cognitive vision has been applied to make fully automatic cars. The different sensors around and inside the cars are linked, the data gathered are very quickly processed to take a run-time decision. The computer vision augmented with ultra-sonic sensors and radars has provided vision to the car to detect living objects, track/road, road markings and signals. It nicely calculates the distance from other cars on road, objects, and turnings. The different images captured by the camera are being processed by the neural network to recognize the object and the path. The intelligence embedded in the car controls the motor, increasing and decreasing speed, turning etc. The whole system is connected to the Internet, which provides the road maps, alternative routes between source and destination. The car is linked to a person by the mobile device, supports personalization like the place to go in a day and rationalize the traveling need. All the cars sold by Tesla motors act as IoT devices/things. These cars similar to IoT devices makes a network among themselves. With one car learns something new about driving, the entire fleet will learn the same through collaborative learning [34]. Tesla cars leverage intelligence to share and learn, putting automotive technology ahead of the curve. Tesla has demonstrated the power AI enabled IoT, which lead to future where machine intelligently augment humans [63].

IBM Watson: Watson and Bluemix are the pioneering technologies conceived by IBM with a perspective of providing cognition in terms of business data analysis, customer based services, and IoT-based services. Design and implementation of application specific AI enabled IoT is very challenging in terms of data aggregation and applying machine intelligence over it. The application requirements of IoT varies from application to application. Watson is programmed to have intelligence like NLP and cognitive vision. Watson provides generalized machine intelligence to analyze and learn textual and image data applicable in a different context [64]. Along to this, it provides cognitive services like conversations, discovery, and intelligent virtual agent. It has different capabilities such as, understanding a different language, natural language classification, text synthesis, voice to text and vice versa conversion, personality insight, tone analyzing etc. Besides, it is capable of discovering insight/pattern in data (textual or visual). This help designers and manufacturers to design new and better solutions and services putting a high impression in our life—making it better. Since each people are different to the way they communicate, their needs and constraints thus demand personalized attention. Watson virtual agents comprehend people need and communicate back with specific service suitably fitting to the person's need. The cognitive technology used in the Watson allows it interacting with things and associated people using natural

language and voice commands. This has essentially dramatically improved the adaptability of the system to people. The entire set of services could be accessed by IoT devices or other over Bluemix platform. Bluemix is a cloud-based solution which ensures computing across devices pervasively. Watson Cognitive Computing boost the potentiality of IoT influencing a strong integration of social lives with other physical and digital world, making our homes, office, cars, elevator, electronic and other appliance more smart, safer, better, intuitive and more interactive.

Ambient assisted living: Ambient Assisted Living (AAL) is a joint program initiated by European technology and innovation. The aim of the project is to provide people a smart living environment at home. It enhances the comfort and the quality of living of the occupants, especially elderly and disabled ones [65]. Using smart technologies and remote care services, the project aims at increasing the quality of life, independence, social involvement and reducing the cost of health and care. It especially focuses on to simplify the daily lives of the elderly and disabled people by abating their dependency on others. Using CIoT, AAL has been able to compensate some of the disabilities by the means of the smart devices. Ambient intelligence allows things to use all the functionality of environment by themselves, thus reinforcing their independence [66]. Application of ambient intelligence has turned our surrounding object intelligent, leading to less human activity. People are augmented with automatic assistance at right time and right place based on personal requirement. It is beneficial especially to the elderly people who are using smart technologies and remote care service can stay longer at home comfortably.

8 Concerns

Stephen Hawkins and Elon Musk recently have warned of too much AI that can subdue human race. The big question is—do we really have to worry? Can really intelligent systems outdo human beings? What will happen if AI takes over our life through IoT on a large scale? Do we need to be concerned of the evil of Cognitive AI?

Frankly speaking, nobody is sure about the far reaching consequences of AI in human race but it is certain that intelligent systems can become disastrous if they turn out to be stupidly intelligent or quasi-intelligent. Cognitive AI is all about to train the devices like a child. Just as the child grows, his knowledge and decision making increases, same as is the role of Cognitive AI in the intelligent systems with cognitive ability. The systems grow smarter as they learn from their surroundings and past experience. In that process, their cognitive power also increases. But what happens if the child grows as a spoiled brat? What if it gets a terrible environment to grow and learn terrible things? What if it becomes really bothering and uncontrollable? This is the exact point people are concern about the overabundance of AI especially the Cognitive AI. If the training goes wrong, the cognitive systems will become really a matter of worry. The philosophy is ‘garbage in garbage out’.

The faulty training will be reflected in the behavior of the system. So, to avoid fatal consequences, special care must be taken to devise cognitive systems. The developers have to be wise and smart enough to create the interaction and learning model and design the algorithms that define these systems.

On an optimistic note, individual flawed and unruly intelligent systems are not going to be a major threat. For instance, a smart robot suddenly gone mad or a smart car all at once started acting weird—the impact of these types of glitches are negligible. This is like our body is being inconvenient in a small way. But it will not be tense-free if a vital organ fails. For example, consider the scenarios—(a) all the robots on a large production floor gone mad together and start playing havoc, (b) all the smart and connected cars in the city get out of control simultaneously, and (c) in the era of smart grid the whole grid system of a country/continent collapses bringing the doomsday! Yes, that will be a crisis. Indeed, a big crisis for which we certainly have to be worried. An ominous CIoT can cause catastrophe considering the scale of its application. In fact, it is indeed uncertain about the extent of the adverse consequences when a system that interconnects almost every device on the planet and has cognitive abilities to make decisions and act on its own, goes wrong. It is very much like the unexpected behavior of our nervous system as the omnipresence of IoT and the span of its connectivity actually can be compared to the nervous system. However, these are all hypothetical concerns and hope they end up as hypothetical.

But there is a real concern which cannot be ignored—the intentional or unintentional misuse of CIoT by the human. IoT is still in its adolescent age and has an entirely new genre of devices for which there are no architectural and security standards yet [67]. This exposes several open patches for potential new security vulnerabilities. To deal with these unforeseen security threats, people are seeking the help of AI to develop smart, autonomous security systems. And considering the enormous magnitude of IoT, taking help of AI is the best course for simulating threats and find and fix bugs instinctively. But as AI is used to strengthen the security, a Frankenstein can also be made out of it by infusing the reverse logic into the system. Now with its devil intelligence, it can push bugs in more bizarre ways that are not anticipated at all and hence cannot be defended and quelled. Our own devices may be metamorphosed against us. So, in either way, over dependence on AI may not be an intelligent option.

CIoT undoubtedly is the most powerful technology of the recent years. And power accompanies responsibility. Ill-handling and wrong implementation of intelligent ‘things’ can take down the utility of the whole idea. Therefore, utmost care must be taken to develop and deploy CIoT. Just as it is our necessity to apply CIoT to make our life easy and smart, it is our obligation to develop it in a way that engenders trust and safeguards humanity [41]. Nonetheless, let us be positive about CIoT, leave all apprehensions and relish only the silver lines. Instead of worrying about what may go wrong we should focus on how to make it right and cherish its enormous potential.

9 Conclusion

Cloning human behavior to the machines is not new and has been tried and tested since long. The recent addition to this is the Cognitive IoT (CIoT) that intends to infuse cognition to the IoT. IoT has certainly been instrumental in taking automation to a new level. But the basic IoT lacks intelligence. Adding intelligence in the form of human-like cognition, the full potential of IoT can be brought out as CIoT will make the devices smart enough to learn dynamically from the environment and interact with the human through natural language. They can take contextual decisions autonomously. Cognition in IoT can be perceived in three essences—(1) Networking, where IoT tries to adapts itself according to the network condition to maximise the communication performance and (2) Behavioral, where IoT aims to learn, think and cognize on its own and (3) Data analytics, where IoT data are processed and analyzed to obtain knowledge that can be used for augmenting business. CIoT is typically a self-learned and self-managed system. It also poses other properties such as probabilistic, adaptive, flexible, dynamic, interactive, integrated, iterative, stateful etc. Smart systems are not easy to develop. Likewise, to realize CIoT, various challenges, for example, limited battery, diverse data types, train the system accurately, social and ethical concerns etc. are to be sorted out. Cognitive AI and CIoT have the potential to go beyond basic automation to deliver business benefits such as better business analysis and decisions, augmented business operations, more customer satisfaction, and increased revenues. Ideally, the future CIoT will be able to create a problem statement based on its learning from an existing problem that it is experiencing and will come out with a best possible solution by applying its AIQ (artificial intelligence quotient) earned through artificial cognition. While CIoT offers a lot of promises, there is always a high probability of ill-consequences when a system that interconnects almost every device on the planet earns cognitive abilities and do things they are not meant to. The ill-handling and wrong implementation of intelligent ‘things’ may negate the utility of the CIoT. We should be really careful in designing, implementing and using CIoT to successfully realization of this revolutionary vision.

References

1. Sathi, A.: Cognitive (Internet of) Things: Collaboration to Optimize Action. Palgrave Macmillan (2016)
2. Morgan, J.: A simple explanation of the internet of things. 13 May 2014. https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#6a3922fb1d09. Accessed 2 May 2017
3. Lydon, B.: Internet of things: industrial automation industry exploring and implementing IoT. Mar–Apr 2014. https://www.isa.org/standards-and-publications/isa-publications/intech-magazine/2014/mar-apr/cover-story-internet-of-things/. Accessed 2 May 2017
4. Banafa, A.: Internet of things: opportunities and challenges. https://www.bbvaopenmind.com/en/internet-of-things-opportunities-and-challenges/. Accessed 14 Dec 2015

5. Kothari, M.: Experience the magic of internet of things automation (IoT). Feb 2017. https://thenextweb.com/contributors/internet-of-things-automation-iot/. Accessed 2 May 2017
6. Pinto, J.: The automation internet of things. 25 Sep 2012. https://www.automationworld.com/sensors-discrete/automation-internet-things. Accessed 2 May 2017]
7. Shankar, U.: How the internet of things impacts supply chains. http://www.inboundlogistics.com/cms/article/how-the-internet-of-things-impacts-supply-chains/. Accessed 3 May 2017
8. Michel, R.: 4 Ways the internet of things will reshape manufacturing. 1 July 2014. http://www.mmh.com/article/4_ways_the_internet_of_things_will_reshape_manufacturing. Accessed 3 May 2017
9. Tarkoma, S., Katasonov, A. (eds.): Internet of things strategic research agenda (IoT–SRA). In: Finnish Strategic Centre for Science, Technology, and Innovation: For Information and Communications (ICT) Services, Businesses, and Technologies, Finland, Sept 2011
10. Cole, A.: AI and the IoT: are we truly prepared for what's coming? 5 Sep 2016. http://www.itbusinessedge.com/blogs/infrastructure/ai-and-the-iot-are-we-truly-prepared-for-whats-coming.html. Accessed 9 November 2016
11. Srivastava, B., Marecki, J., Tesauro, G.: In: 2nd Workshop on Cognitive Computing and Applications for Augmented Human Intelligence, in conjunction with International Joint Conference on Artifi cial Intelligence (IJCAI), Buenos Aires, Argentina. July 2015
12. Kelly, J.E.: Computing, cognition, and the future of knowing. IBM Research. Oct 2015
13. Balani, N.: Cognitive IoT (2015)
14. Kelley, J.: Smart Machines: IBM's Watson and the Era of Cognitive Computing. Columbia Business School Publishing. Sep 2013
15. Garrett, M.: Big data analytics and cognitive computing—future opportunities for astronomical research (2014)
16. Schatsky, D., Muraskin, C., Gurumurthy, R.: Cognitive technologies: the real opportunities for business. Deloitte review, no. 16, pp. 115–129 (2015)
17. How is cognitive computing different from big data and NLP? www.coseer.com
18. Pearson, N.: How to boost competitive advantage with cognitive computing. 21 Oct 2016. https://www.ibm.com/blogs/think/2016/10/boost-advantage-with-cognitive/. Accessed 21 Feb 2017
19. Cognitive computing defined, cognitive computing consortium. https://cognitivecomputingconsortium.com/resources/cognitive-computing-defined/#1467829079735-c0934399–599a. Accessed 18 February 2017]
20. Coccoli, M., Maresca, P., Stanganelli, L.: Cognitive computing in education. J. e-Learn. Knowl. Soc. **12**(2), 55–69 (2016)
21. Reynolds, H.: Cognitive computing: big data and cognitive computing-Part 1. 30 Dec 2015. www.kmworld.com/Articles/News/News-Analysis/Cognitive-computing-Big-data-and-cognitive-computing–Part-1–108248.aspx. Accessed 21 Feb 2017
22. Streamlining knowledge processes through cognitive automation: beyond the human brain. Deloitte. https://www2.deloitte.com/us/en/pages/deloitte-analytics/articles/cognitive-automation.html
23. Justice, C.: Cognitive technology and the automation of everything. 15 Sep 2015. http://www.cio.com/article/2977565/robotics/cognitive-technology-and-the-automation-of-everything.html. Accessed 28 April 2017
24. Green, H.: Five ways cognitive computing will power the internet of things. 15 Dec 2015. http://www.forbes.com/sites/ibm/2015/12/15/five-ways-cognitive-computing-will-power-the-internet-of-things/#2336705466c7. Accessed 20 Feb 2017
25. Zhang, M., Zhao, H., Zheng, R., Wu, Q., Wei, W.: Cognitive internet of things: concepts and application example. Int. J. Comput. Sci. Issues (IJCSI) **9**(6), 151–158 (2012)
26. Mitola, J., Maguire, G.Q: Cognitive radio: making software radios more personal. IEEE Personal Commun. **6**(4), 13–18 (1999)
27. Thomas, R.W., Friend, D.H., DaSilva, L.A.: Cognitive networks: adaptation and learning to achieve end-to-end performance objectives. IEEE Commun. Mag. **44**(12), 51–57 (2006)

28. Fortuna, C., Mohorcic, M.: Trends in the development of communication networks Cognitive networks. Comput. Netw. **53**(9), 1354–1376 (2009)
29. Wu, Q., Ding, G., Xu, Y., Feng, S., Du, Z., Wang, J., Long, K.: Cognitive internet of things: a new paradigm beyond connection. IEEE J. Int. Things (2014)
30. Sathi, A.: Big Data Analytics: Disruptive Technologies for Changing the Game. MC Press (2012)
31. Millman, R.: Artificial intelligence needed to make sense of IoT data. 6 Sep 2016. https://internetofbusiness.com/artificial-intelligence-needed-make-sense-iot-data/. Accessed 9 Nov 2016
32. Li, W., Taheri, J., Zomaya, A.Y., Seredyński, F., Landfeldt, B.G.: Nature-inspired computing for autonomic wireless sensor networks. In: Sarbazi-Azad, H., Zomaya, A.Y. (eds.) Large Scale Network-Centric Distributed Systems, Wiley-IEEE Computer Society Press, pp. 219–253 (2013)
33. Noor, A.K.: Potential of cognitive computing and cognitive. Open Eng. **5**, 75–88 (2015)
34. Bhardwaj, A.: AI could be the catalyst to unleash the power of IoT. 20 Oct 2016. http://www.oodlestechnologies.com/blogs/AI-Could-Be-The-Catalyst-To-Unleash-The-Power-of-IoT. Accessed 9 Nov 2016
35. Xu, Y.: Recent machine learning applications to internet of things (IoT). http://www.cse.wustl.edu/~jain/cse570-15/ftp/iot_ml/
36. Garcia, J.: Machine learning and cognitive systems: the next evolution of enterprise intelligence (Part I). https://www.wired.com/insights/2014/07/machine-learning-cognitive-systems-next-evolution-enterprise-intelligence-part/
37. Vernon, D.: Cognitive vision-the development of a discipline. In: Proceedings of IST 2004 Event Participate in Your Future
38. Benjamin, D.P., Lyons, D.: Toward A Cognitive Computer Vision System
39. Quack, T., Bay, H., Van Gool, L.: Object recognition for the internet of things. In: The Internet of Things, pp. 230–246. Springer, Berlin (2008)
40. Hurwitz, J.S., Kaufman, M., Bowles, A.: Cognitive Computing And Big Data Analysis, 10475 Crosspoint Boulevard, Indianapolis, IN 46256, Wiley Inc. (2015)
41. Banavar, G.: What it will take for us to trust AI. 6 Dec 2016. https://www.ibm.com/blogs/think/2016/12/trusting-ai/. Accessed 21 Feb 2017
42. Banavar, G.: The science of AI and the art of social responsibility. 20 Feb 2017. https://www.ibm.com/blogs/think/2017/02/ai-and-the-art-of-social-responsibility/. Accessed 20 Feb 2017
43. Blog, I.T.: Transparency and trust in the cognitive era. 17 Jan 2017. https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/. Accessed 20 Feb 2017
44. Gupta, M.: Artificial intelligence IoT: why AI-flying IoT is a big idea. 26 April 2016. http://www.digitalservicecloud.com/insights/ai-iot.html. Accessed 9 Nov 2016
45. Tejaswi, S.: 6 Ways businesses can take advantage of IoT. https://vmokshagroup.com/blog/6-ways-businesses-can-take-advantage-of-iot/. Accessed 15 April 2017
46. Angeles, S.: Internet of things has big startup potential. 12 Nov 2013. http://www.businessnewsdaily.com/5450-internet-of-things-business-opportunities.html. Accessed 15 April 2017
47. Prolim: The internet of things—bringing greater business benefits. 10 Sep 2015. https://prolim.com/the-internet-of-things-bringing-greater-business-benefits/. Accessed 15 April 2017
48. Zimmerman, M.: The future of retail is consumer-connected and cognitive. 18 Jan 2016. https://www.ibm.com/blogs/think/2016/01/the-future-of-retail-in-2025/. Accessed 21 Feb 2017
49. Jarratt, B.: Cognitive solutions: the next chapter in streaming video on demand. 27 Oct 2016. https://www.ibm.com/blogs/think/2016/10/cognitive-solutions-the-next-chapter-in-streaming-video-on-demand/. Accessed 21 Feb 2017
50. Vesset, D., Nadkarni, A., McDonough, B., Bond, S., Li, Q., Olofson, C.W., Zaidi, A., Schubmehl, D., Kusachi, S., Carnelley, P.: IDC FutureScape: Worldwide Big Data and Analytics 2016 Predictions. IDC (2015)

51. Hamm, S.: Embedding intelligence in the internet of things. 23 Feb 2016. https://www.ibm.com/blogs/think/2016/02/embedding-intelligence-in-the-internet-of-things/. Accessed 21 Feb 2017
52. Whirlpool Smart Top Load Connected Laundry Pair. Whirlpool® USA, 23 Oct 2015. https://youtu.be/2BGcGdNcSOk
53. IoT Applications for Smart Cities. IoT Innovation (2016). http://internet-of-things-innovation.com/insights/the-blog/iot-applications-smart-cities/#.WVfrMoTfrDd. Accessed 1 July 2017
54. Talari, S., Shafie-khah, M., Siano, P., Loia, V., Tommasetti, A., Catalão, J.P.S.: A review of smart cities based on the internet of things concept. Energies **10**(4) (2017)
55. Vázquez, J.I.: The internet of things: outlook and challenges (2013). https://www.bbvaopenmind.com/en/article/the-internet-of-things-outlook-and-challenges/?fullscreen=true. Accessed 7 March 2017
56. Banafa, A.: What is affective computing? 6 June 2016. https://www.bbvaopenmind.com/en/what-is-affective-computing/. Accessed 7 March 2017
57. Banavar, G.: The cognitive era presents opportunities for enhanced collaboration. 14 Dec 2015. http://www.forbes.com/sites/ibm/2015/12/14/the-cognitive-era-presents-opportunities-for-enhanced-collaboration/#6eece8a667d5. Accessed 20 Feb 2017
58. Collaborative robots (COBOTS): Safe co-operation between human beings and robots. Institut für Arbeitsmedizin der (IFA). http://www.dguv.de/ifa/fachinfos/kollaborierende-roboter/index-2.jsp. Accessed 1 May 2017
59. CORAL Group: CoBot Robots. Carnegie Mellon University. www.cs.cmu.edu/~coral/projects/cobot/. Accessed 1 May 2017
60. Tolcher, R.: Using smart chatbots as an IoT interface. 23 June 2016. https://iot.telefonica.com/blog/using-smart-chatbots-as-an-iot-interface. Accessed 2 May 2017
61. Haque, S.: Why chatbots can be used as internet of things (IoT) interface. 9 Aug 2016. http://www.ameyoemerge.in/blog/why-chatbots-can-be-used-as-internet-of-things-iot-interface. Accessed 2 May 2017
62. Zimmerman, M.: Weathering hurricane season with cognitive, IoT. 24 May 2016. https://www.ibm.com/blogs/think/2016/05/weathering-hurricane-season-with-cognitive-iot/. Accessed 21 Feb 2017
63. Faggella, D.: Artificial intelligence plus the internet of things (IoT)—3 Examples worth learning from. 8 Feb 2016. http://techemergence.com/artificial-intelligence-plus-the-internet-of-things-iot-3-examples-worth-learning-from/. Accessed 9 Nov 2016
64. Orii, Y., Horibe, A., Matsumoto, K., Aoki, T., Sueoka, K., Kohara, S., Okamoto, K., Yamamichi, S., Hosokawa, K., Mori, H.: Advanced interconnect technologies in the era of cognitive computing. In: Pan Pacific Microelectronics Symposium (Pan Pacific) (2016)
65. Chan, M., Estève, D., Escriba, C., Campo, E.: A review of smart homes-present state and future challenges. Comput. Methods Programs Biomed. **91**(1), 55–81 (2008)
66. Nussbaum, G.: Smart environments: introduction to the special thematic session. In: ICCHP (2008)
67. Meek, A.: Connecting artificial intelligence with the internet of things. 28 July 2015. https://www.theguardian.com/technology/2015/jul/24/artificial-intelligence-internet-of-things. Accessed 9 Nov 2016

Cybercrimes Investigation and Intrusion Detection in Internet of Things Based on Data Science Methods

Ezz El-Din Hemdan and D. H. Manjaiah

Abstract In recent times, Internet of Things (IoT) has paying attention from different organization ranging from academia to industry. The IoT is an internetworking of connecting and integrating several types of devices and technologies that comprising sensors, Radio Frequency Identification (RFID), cloud computing, the Internet, smart grids, and vehicle networks, and many other devices and new technologies. The IoT becomes a subject for illegal and criminals activities. Cyber criminals and terrorists are highly qualified persons in the computer, network, digital systems and new technologies. An enormous amount of data is gathered about criminals and their behavior from different data sources over the Internet can be processed using data science methods to monitor and trace them in real-time and online. The massive amount of data needs new fast and efficient processing tools and techniques for data extracting and analyzing in less period of time. Data science methods can be used for this purpose to investigate and detect a different type of severe attacks and intrusions. This chapter introduce principles of Digital Forensics, Intrusion Detection and Internet of Things as well as exploring data science concepts and methods that can help the digital investigators and security professionals to develop and propose new data science techniques and methods that can be adapted to the unique context of Internet of Things environment for performing intrusion detection and digital investigation process in forensically sound and timely fashion manner.

Keywords Cybercrimes · Digital forensics · Intrusion detection
Internet of things · Data science methods

E. E.-D. Hemdan (✉) · D. H. Manjaiah
Department of Computer Science, Mangalore University, Mangalore, India
e-mail: ezzvip@yahoo.com

D. H. Manjaiah
e-mail: manju@mangaloreuniversity.ac.in

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_2

1 Introduction

Data science is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured, which is a continuation of some of the data analysis fields such as statistics, machine learning, data mining and knowledge discovery, and predictive analytics. There are several advantages and benefits of applying data science methods in cybercrime investigation and intrusion detection over Internet of Things (IoT). This is to look for crucial information that can be used in the digital investigation and help refute or support a claim or put together a missing piece, this has seen a rapid increase in the field of digital investigation, and intrusion detection and prevention. Internet of Things has paying attention from different organization ranging from academia to industry. A Huge amount of data is congregated about criminals and their behavior from different data sources in the IoT environment through using data science methods to observing and tracing them. New fast and efficient processing tools and techniques are required for data extracting and analyzing in less period of time. Data science methods can be used for this purpose to investigate and detect a different type of severe attacks and intrusions. Digital Investigators, examiners and system administrators can use numerous innovative statistics, machine learning, data mining, and predictive analytics to recognize data patterns from the gigantic collected large data from IoT devices to discover any digital evidence about hackers or detect and trace them through their criminal activities by identifying suspicious behavior patterns to identify threats that are likely to happen.

Presently, existing anomaly detection is often associated with high false alarm with moderate accuracy and detection rates when it's unable to detect all types of attacks properly as well as digital investigation of cybercrimes is become two important area in information security. Thus, this chapter focus on applying both of them over the Internet of Things based on using data science methods and approaches that can help to improve detecting and investigating crimes in efficient and effective manner. This chapter will explore and identify challenges and opportunities of digital forensic and intrusion detection and how can apply data science approaches and cognitive methods to fight and investigate serve attacks and crimes over the Internet of Things environment in forensically sound and timely way.

This chapter introduce principles of Digital Forensics, Intrusion Detection and Internet of Things as well as exploring data science concepts and methods that can help the digital investigators and security professionals in developing and proposing new techniques that can be adapted to the unique context of Internet of Things environment which can help in performing intrusion detection and digital investigation process in forensically sound and timely fashion manner.

The remainder of this chapter is structured as follows: Sect. 2, presents an overview about digital forensics, intrusion detection, and the internet of things as well as exploring data science concepts and methods while cybercrimes investigation in the internet of things is presented in Sect. 3. Section 4 provides intrusion

detection in the internet of things while applying data science methods for the cybercrimes investigation and intrusion detection in the internet of things is introduced in Sect. 5. Finally, the chapter conclusions and future directions in this innovative subject are presented in Sect. 6.

2 Background

This section provides basics of digital forensics, intrusion detection, and Internet of Things (IoT) along with identifying data science concepts and methods.

2.1 Digital Forensics

Digital forensics is a branch of forensics science that concern with finding and collecting digital evidence then analysis and examine them to find any traces related to crimes against digital systems. Digital forensics has many directions such as Computer Forensics, Mobile Forensics, Network Forensics and Cloud Forensics. This section discusses digital forensics definition as well as digital forensics investigation process which the digital investigators follow it during the investigation of crimes to reconstruct the crime events that occurred.

2.1.1 Digital Forensics Definition

The process of collecting, identifying, preserving and examining digital evidence is known as 'Digital Forensics'. One of the popular definition for the digital forensics is introduced by first Digital Forensic Research Workshop (DFRWS). The DFRWS defined the digital forensics as: "*The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations*" [1].

2.1.2 Digital Forensic Investigation Process

Criminals and attackers after committing their cybercrimes some trails that remain behind them. Collecting, extracting and preserving digital evidence from the crime scene need careful strategies to handle and manage them to become ready for presenting in the court of law. In digital forensics, there are four crucial steps for performing the digital forensic process as shown in Fig. 1 as follows [2]:

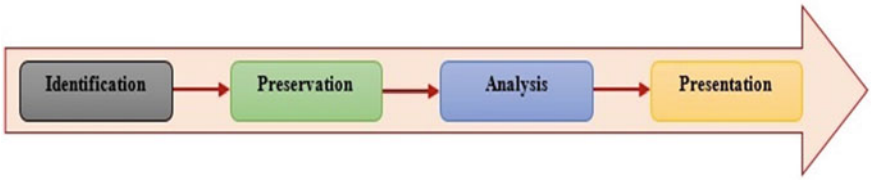


Fig. 1 Digital forensic investigation process

- *Identification*: It is the identification of sources of digital evidence, which will be required to prove the committed crime.
- *Preservation*: In the preservation process, the digital investigator preserves the collected digital evidence such as Hard Disks, Laptops, Tablets and Mobile Phones.
- *Analysis*: In the analysis process, the examiners and digital investigators interpret and correlate the evidential data to come to a summary and conclusion, which can prove or disprove civil, or criminal actions.
- *Presentation*: In this process, the digital investigators make a forensic report to summarise their findings from the analysis process. This report should be suitable to present to the court of law.

2.2 *Intrusion Detection System*

Intrusion Detection System (IDS) is can be a software or hardware system that observers the system or actions of the network for policy criminal and malicious activities and produces reports to the central administration system. The main focus of intrusion detection systems is to recognize the possible incidents, logging information about them and in report attempts. Furthermore, organizations use the intrusion detection systems for other objectives, such as detecting problems with security policies, deterring individuals and documenting present threats from infringing security policies. The intrusion detection systems have become an important addition to the security infrastructure of nearly each organization. Several procedures can be used to identify criminals and intrusions but each one is specific to a particular method. The main aim of intrusion detection system is to detect and identify the attacks professionally. In addition, it is equally imperative to detect attacks at an early stage in order to reduce their impacts.

2.2.1 Categorization of Intrusion Detection Systems

Intrusion Detection System is used to analyze packet traffic to match it with any anomalies found in comparison with the normal traffic. For anomalous traffic, the

IDS tries to identify the pattern of common threats and alerts the system administrator. Ever since an IoT-based platform is frequently a high-speed network, it essential be protected using an entirely automated intrusion detection system. The intrusion detection system briefly classified into two basic categories:

- *Network Intrusion Detection System*: This type tries to secure all machine systems in the network.
- *Host-based Intrusion Detection System*: This type tries to secure a single host. A highly scalable intrusion detection system is able to provide support for proficient utilization of recent high-performance architectures.

Detection models are divided into statistical or signature based models [3]. The statistical model maintains profiles regarding applications, hosts, users, and connections. It then matches current activity with the attributes of the profile for any anomalies. In the other side, the signature-based model compares the traffic against a collection of existing signatures. Also, there are three detection approaches that are used by host-based or network intrusion detection systems which are used to analyze events and discover attacks as follows [4, 5]:

- *Signature based System*: This model is also known as misuse detection system which still the utmost method and focuses on the identification of known bad patterns and searching for similar activities such as vulnerabilities or acknowledged intrusion signatures. As with each system that uses a blacklist approach, it is vulnerable to attacks for which the signature is unknown, such as zero-day exploits or use of encoding, obfuscation or packing methods.
- *Anomaly based System*: This system work on searching unusual behavior on network traffic as well as observing system behavior to fix whether an observed activity is anomalous or normal, according to a heuristic analysis, can be used to identify unknown attacks. Anomaly detection based IDS model has the ability to detect attack indications without specifying attack models, but these models are very sensitive to false alarms.
- *Specification based IDS*: This type of IDS is like anomaly detection system. In this system, the normal behavior of the network is defined by manually, so it gives less incorrect positives rate. This system attempts to excerpt best between signature-based and anomaly based detection methods by trying to clarify deviations from normal behavioral patterns that are produced neither by the training data nor by the machine learning techniques. The development of attack specification is done manually so it takes more time.

2.3 Internet of Things

Recently, Internet of Things (IoT) has an urgent economic and societal impact for the future construction of communication and network systems to exchange

information between things and people. The novel planning of future will be eventually, "everything will be connected and intelligently controlled". The idea of IoT is becoming more relevant to the real world due to the development of mobile devices, cloud computing, embedded and ubiquitous communication technologies, data science and data analytics. The IoT made up of devices connected to the Internet to collect information about the environment using sensors connected to devices (i.e. things). These devices communicate and interact together to acquire, process and storage information in smart and intelligent manner.

With the IoT, millions of devices are connected to each other which need to exchange information through the network (i.e. Internet) with the need to massive capabilities such as processing, storage, and high bandwidth. These capabilities can be delivered through using cloud computing technology. Researchers and scientist who are working in the IoT can use the cloud computing services to design and develop applications that can create of smart environments like Smart Cities. The devices that are used in the IoT system produces enormous data (i.e. big data) which often need to leverage the technology of cloud computing to scale cost effectively. Big data analytic is an important direction nowadays to help business to predict about future and so make correct decisions in business marketing.

2.3.1 Internet of Things Operations

In Internet of Things, there are various operation phases include collection phase, transmission phase, and processing, management and utilization phase [6] as follows:

- *Collection Phase*: The principal aim is to collect data about the physical environment. Sensing devices and technologies for short range communication are combined to reach this objective. Devices of the collection phase are usually small and resource-constrained. Communication technologies and protocols for this phase are designed to operate at limited data rates and short distances, with constrained low energy consumption and memory capacity. Due to these characteristics, collection phase networks often are referred to as Low power and Lossy Networks (LLN).
- *Transmission Phase*: The goal of this phase is to transmit the data collected during the collection phase to applications and, therefore, to users. Here, technologies such as Ethernet, WiFi, Hybrid Fiber Coaxial (HFC) and Digital Subscriber Line (DSL) are united with TCP/IP protocols to construct a network that interconnects objects and users across longer distances. Gateways are necessary to integrate LLN protocols of the collection phase with traditional Internet protocols employed in the transmission phase.
- *Processing, Management and Utilization Phase*: Applications process gather data to obtain useful data about the physical environment. These applications

may take decisions based on this data, controlling the physical objects to act on the physical environment. This phase also contains a middleware, which is responsible for facilitating the integration and communication between different physical objects and multi-platform applications.

2.3.2 Internet of Things Categorization

Internet of Things can be categorized into four categories such as Internet of Nano Things, Internet of WiFi-enabled Things, Internet of Things for Smart Society, and Global-scaled Internet of Things [7–13] as follows:

1. *Internet of Nano Things (IoNT)*: The IoNT consist of nano-devices that are communicating with each other over a nanonetwork. In the IoNT, it becomes possible to add a new dimension to the IoT by embedding nano-sensors to the numerous things and devices that surround us. Also, it can be used in different areas such as biology.
2. *Internet of WiFi-enabled Things*: Currently, the WiFi is a significant category of wireless networks for connecting several devices to the Internet. When the WiFi enabled devices are connected together over the Internet which it offered new kind of the IoT named Internet of WiFi-enabled Things.
3. *Internet of Things for Smart Society*: The idea of smart city become an attractive research topic for numerous scientists and researchers to introduce novel methods for connecting things or devices in the society in a smart manner to make a smart society through embedding sensors in all surrounding devices and things to allow them to interact and communicate together in an intelligent manner. This gets a new category of IoT known as the internet of things for smart society.
4. *Global-scaled Internet of Things*: It is utilizing in a global-scaled area such as unmanned aerial vehicle and satellite system. Using remote connections, these systems communicated and interacted with several devices which are connected to sensors to sense data in an effective way. The Tsunami Detection System is a real world example about the global-scaled internet of things [13].

2.3.3 Internet of Things Applications

Internet of Things is an imperative paradigm for providing smart applications which can improve and enhance the quality of our lives to the better level of life. Recently, the IoT has several applications in various areas such as; industrial control system smart society, smart manufacturing, smart agriculture, healthcare, military, and trade and logistics [7–13] as follows:

1. *Smart City:* The smart city is the idea of making smart cities which make people life more comfortable and easy. The innovative development of smart technologies assists the IoT in changing people life style. The IoT can be used in smart cities to provide many services as; intelligent highways with warning messages for unexpected actions such as accidents. Also, for monitoring of vehicles and pedestrian levels to optimize driving and walking routes, monitoring of parking spaces availability inside the city.
2. *Tracking Animals Movement*: recently, a large sensor network can be deployed to study the effect of micro climate issues in habitat choice of sea birds. Researchers located their sensors in burrows and used heat to detect the presence of nesting birds, providing invaluable data to biological researchers. The deployment is heterogeneous in that it employed burrow nodes and weather nodes.
3. *Industrial and Manufacturing Systems*: The IoT can be used in industrial control systems to enhance their performance through making them smarter with taking in consideration necessary factors like safety and availability to guarantee continues in business and save people life. The industrial control system can use the IoT for several purposes such as; auto-diagnosis of machines in control system, observing of toxic gas and oxygen levels inside chemical plants and monitoring of ozone levels during the drying meat process in factories of food engineering.
4. *Smart Agriculture*: Agriculture is a significant domain that provides people and society with food so that there are serious desires to improve the agriculture system through using smart technologies that are presented by the IoT. The IoT will improve operational efficiency and productivity in agriculture system. The benefits of using IoT in agriculture field are; monitoring soil moisture and trunk diameter in vineyards to control the amount of sugar in grapes and grapevine health, and control micro-climate conditions to maximize the production of fruits and vegetables and its quality.
5. *Healthcare*: The IoT can provide several benefits in healthcare domain. These benefits such as remote monitoring of patients, tracking of drugs, identification, and authentication of people. It also can use for the assistance of elderly or disabled people living independently and monitoring of conditions of patients inside hospitals.
6. *Trade and Logistics*: Innovative development in the IoT systems support to develop and manage products shopping from online websites. In addition to, using IoT in trade and logistics through embedding sensors and tags in roads and products for monitoring and tracking them. In Trade, IoT can be used for product tracking, monitoring of storage conditions and payment processing based on location. In logistics, the IoT can be used for observing of vibrations, strokes, container openings for insurance purposes, a search of individual items in big surfaces such warehouse.

2.4 Data Science

Data science, or more especially, big data analytics, become a hot and popular topic that has attracted attention among researchers in computer science and statistics. It concerns with a wide variety of data processing jobs, such as data analysis, data collection, data management, data visualization, and real-world applications. Today, the volume of data is increasing very quickly, the existing data processing tasks exceed the computing ability of classical computational models to store, validate, analyze, visualize, and extract knowledge. To analyze immense data, there are numerous complications, such as dynamical changes of data, a large volume of data, and data noise so that there is a serious need to develop novel and efficient methods to handle complex data analytics problems.

2.4.1 Data Science Definition

Data science is the interdisciplinary domain of computer science and statistics about scientific methods, processes and systems to extract knowledge from data in various forms, either structured or semi-structured and unstructured. The combination of computer science and statistics to take advantages of them to handle the massive amount of data in an efficient manner. The statistics are the science that concerns with collection, analysis, and organization of data. From the perspective of statistics, there are various objectives in data analyses such as predict the response/output of future input variables and deduce the association among response variables and input variables. While From the perspective of computer science, the data science is a process of data mining for converting raw data into useful knowledge and attempts to discover valuable patterns in large data storage.

2.4.2 Data Science Mission

The task of performing data science methods is to store, validate, analyze, visualize, and extract knowledge from the massive amount of data using computer science and statistical algorithms. Briefly, the data science area comprises of many sub areas, such as classification, clustering, and association analysis. The clustering and classification are two different types of basic problems, important methods in data mining research. Clustering is the process of grouping similar objects together. The data clustering analysis is a technique that divides data into several groups (i.e. clusters).The aim of clustering is to categorize objects being similar to one another in the similar cluster and place objects being distant from each other in dissimilar clusters. Data classification is a problem that finds the correct category(s) for data objects when a set of categories and a group of data set are given.

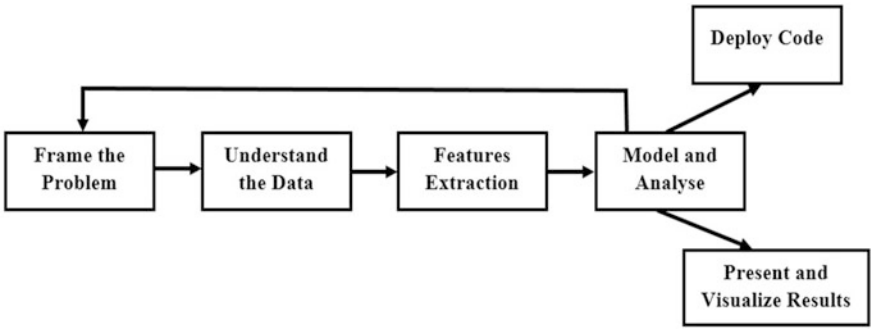


Fig. 2 Data science road map (DSRM)

2.4.3 Data Science Road Map

This part provides the idea of Data Science Road Map (DSRM) that refers to how to perform solving process for a data science problem as shown in Fig. 2. This DSRM consists of different stages as follows [14]:

1. *Frame the Problem*: It is a very important stage to understand the type of problem that will be solved using data science methods.
2. *Understand the Data*: Study and understand the data and the real-world things that it describes, which are related to the problem can help in choosing best methods to handle and manage this data in a given frame time.
3. *Feature Extraction*: It is the process of extracting features and hidden patterns from given data that will feed into the data science model for solving a certain problem.
4. *Model and Analyse*: In this stage, the data scientists are building a model which is suitable to a given problem as well as analyze the data sets related to the problem.
5. *Deploy Code*: Here, the data scientists write code, and they use several of the same tools as software engineers.
6. *Results Presentation and Visualization*: It is the final stage of designing and implementation of data science model.

2.4.4 Programming Languages for Data Science

There is various programming language which can be useful for data scientists. Here, will provide some of the most popular ones [14].

- *Python*: Python is a high-level scripting language, with functionality similar to Ruby and Perl and with an unusually clean and self-consistent syntax. Outside of the core language, Python has many open-source technical computing

libraries that make it a powerful tool for analytics. Python is considered as one of the best programming language available for general-purpose use. It is also a very popular choice among data scientists, who feel like it balances the flexibility of a conventional scripting language with the numerical muscles of a good mathematics package.

- *R*: R is probably the most popular programming language among data scientists. Python is a scripting language designed for computer programmers, which has been augmented with libraries for technical computing. In contrast, R was designed by and for statisticians, and it is natively integrated with graphics capabilities and extensive statistical functions. One of the key reason to use R is just that there are so many special libraries that have been written for it over the years, and Python has not covered all the little special use cases yet.
- *MATLAB*: The data science community skews strongly toward open-source software, so good proprietary programs such as MATLAB often get less credit than they deserve. Developed and sold by the MathWorks Corporation, MATLAB is an excellent package for numerical computing. It has a more consistent syntax compared to R and more numerical muscle compared to Python. A lot of people coming from physics or mechanical/electrical engineering backgrounds are well-versed in MATLAB. It is not as well-suited to large software frameworks or string-based data munging, but it is best-in-class for numerical computing.

3 Cybercrimes Investigation in Internet of Things

Recently, The Internet of Things (IoT) has become an attractive research topic for academia and industry. It has several application domains such as medical, industry and military. The Internet of Things represents a network of connected devices or machines include mobile handsets, wireless sensors, refrigerators, cars, Radio Frequency Identification (RFID), fitness trackers, watches, eBooks, vending machines, and parking meters, and other types of devices are likely to grow exponentially over the next years. These devices are already generating, gathering and communicating enormous volumes of data about themselves, which is collated, curated, and harvested by a growing number of smart applications.

The IoT devices are considered as sources for massive volumes of data. The variety of these sources provides complex challenges to digital forensics community especially digital investigators who will be required to interact with this new technology to investigate IoT-based crimes. In the IoT environment, a lot of devices or machines are interconnecting together. This refers to the possibility of interconnecting various different threats and attacks such a malware can easily propagate through the IoT at an unprecedented rate. In the following design aspects of the IoT system, there may be various threats and attacks as follows [15]:

- *Data Perception and Collection*: In this part, typical attacks involve sovereignty and control, data leakage and authentication.
- *Data Storage*: Here many these attacks may happen such as denial-of-service attacks, data integrity, impersonation, and modification and tampering of sensitive data.
- *Data Processing*: In this stage, it may be computational attacks that have the objective of producing wrong data processing outcomes and results.
- *Data Transmission*: During the transmission process may occur severe type of attacks like session hijacks, routing attacks, flooding, and channel attacks. So, effective defense procedures and strategies are of the extreme significance to guarantee the security of the IoT infrastructure.

3.1 Digital Forensics in IoT Systems

In the last years, some researchers provide work related to the IoT Forensics area. Some of them explained the concept of the IoT Forensics while the others provided new methods for performing the digital investigation process in the IoT environment. Perumal et al. [16], proposed an integrated model which is planned based on triage model and 1-2-3 zone model for volatile based data preservation. This model started with the following authorization, planning and obtaining a warrant as fundamental steps in the digital forensic investigation process. Then starts to investigate the IoT infrastructure and finally after seizing the IoT device from the selected area or zone, the investigator completes the digital forensic method which includes a chain of custody, lab analysis, result and proof, and archive and storage.

Zawoad et al. [17], proposed a Forensics-Aware IoT (FAIoT) model for supporting digital forensics investigations in the IoT environment in a reliable manner. The FAIoT model provides secure evidence preservation module and secure province module as well as access to evidence using Application Programming Interface (API) that will reduce the challenge in performing investigation process. To facilitate the digital investigators a centralized trusted evidence repository in the FAIoT is used to ease the process of evidence collection and analysis. The IoT devices need to register this secure evidence repository service. The FAIoT architecture consists of three main parts as follows:

- *Secure Evidence Preservation Module*: This module can be used to monitor all the registered IoT devices and store evidence securely in the evidence repository. Also, segregating of the data according to the IoT devices and its owner will do in this module. Hadoop Distributed File System (HDFS) can be used to handle a large volume of data.
- *Secure Provenance Module*: This module guarantees the proper chain of custody of digital evidence by preserving the access history of the evidence.

- *Access to Evidence through API*: In this model, a secure read-only APIs to law enforcement agencies is proposed. Only digital investigators and the court member will have access to these APIs. Through these APIs, they can gather the preserved digital evidence and the provenance information.

Oriwoh et al. [18], they proposed two methods for digital investigation in IoT environment which are 1-2-3 Zones Digital Forensics and Next-Best-Thing Triage as follows:

1. ***1-2-3 Zones Digital Forensics***: This approach divides the IoT infrastructure into three areas or zones to help in performing digital investigation process. These zones are zone 1, zone 2 and zone 3 as follows:
   - *Zone 1*: This zone is called the internal zone that includes all IoT smart devices like a smart refrigerator and TV that can contain valuable data about committed crime in IoT infrastructure.
   - *Zone 2*: This zone includes all intermediate components between resides between the internal and external networks to support the communication process. These devices may be protection devices such as Intrusion Detection and Prevention Systems and Firewalls. The digital investigators can find evidential data that help them to extract facts about committed crime related to IoT.
   - *Zone 3*: This zone includes hardware and software components that reside in the external part of IoT infrastructures such as cloud services and other service providers that used to IoT devices and users. These components with hardware devices and software in zone 1 and zone 2 will help digital practitioners to perform their investigation mission in a timely fashion manner.

   This approach reduces the challenges that will be encountered in IoT environments and ensures that investigators can focus on clearly identified areas and objects in preparation for investigations.
2. ***Next-Best-Thing Triage***: The Next-Best-Thing Triage (NBT) can use in conjunction with the 1-2-3 Zones approach. This approach discusses to find an alternative source in the crime scene if it unavailable after a crime occurred in IoT environment. The NBT approach can be used to determine what devices were connected to the Objects of Forensic Interest (OOFI) and find anything which left behind the devices after they removed from the network. Direct access to the OOFI may not always be possible. Therefore, in such circumstances, the option of recognizing and considering the next best source of related evidence may have to be taken. The design of a technique of systematically deciding what this next best thing might be in different situations and scenarios can be the subject of further research.

4 Intrusion Detection in Internet of Things

Intrusion Detection System (IDS) is used to monitor network traffic, check for suspicious activities and notifies the network administrator or the system. In some instances, the IDS might also react to malicious or anomalous traffic and will take action such as barring the user or perhaps the IP address source from accessing the system. Detection and prevention malicious activities in Internet of Things environment becomes very important topic in the coming years.

A typical IDS is consist of sensors, an analysis engine, and a reporting system. Sensors are deployed at diverse network places or hosts [6, 19]. Their mission is to gather network or host data such as packet headers, traffic statistics, service requests, operating system calls, and file-system changes. The sensors send the gathered data to the analysis engine, which is responsible to investigate the gathered data and detect ongoing intrusions. When the analysis engine detects an intrusion, the reporting system generates an alert to the network administrator.

4.1 Attacks in Internet of Things

IoT infrastructure is exposed to various types of severe attacks both from internal and external so these attacks are mainly categorized by two types inside and outside attacks. In an inside attack, the attack can be originated by compromised or malicious nodes that are part of the infrastructure while in an outside attack, the attacker is not a part of the infrastructure. There are several types of attacks against IoT applications as follows [20]:

- *Sinkhole Attack*: In this attack, The criminal creates an attack by introducing false node inside IoT network where the malicious node attracts network traffic towards it. To launch these types of attack, a criminal node attracts all neighboring nodes to forward their packets through the malicious node by showing its routing cost minimum.
- *Wormhole Attack*: In this attack, the enemy node creates a virtual tunnel between two ends. An enemy node works as a forwarding node between two nodes. The two criminal nodes usually claim that they are one hop away from the base station. The wormhole attack can also be used to convince two different nodes that they are the neighbors by relaying packets between two of them.
- *Selective Forwarding Attack*: In this attack, criminal node works as a normal node but it selectively drops some packets. One of the simplest forms of selective forwarding attack is black hole attack where in it all packets are dropped by the criminal node.
- *Sybil Attack*: In this attack, the node has many identities. The routing protocol, detection algorithm, and cooperation processes can be attacked by a criminal node.

- *Hello Flood Attack*: In a network, the routing protocol broadcast hello message to announce its presence to its neighbors. A node which receives the hello message may assume that the source node is within its communication range and add this source node to its neighbor list.
- *Denial of Service (DOS) Attack*: This attack can damage the availability of resources to legitimate users. Such type of attacks, when launched by various criminal nodes is called Distributed Denial of Service (DDoS). This attack may affect the network resources, such as bandwidth and CPU time.

4.2 Categorization of IDS in Internet of Things

In [6], they classified Intrusion Detection in IoT regarding the following attributes: IDS placement strategy, detection method, security threat and validation strategy as shown in Fig. 3.

1. **IDS Placement Approaches**

In IoT infrastructure, the IDS can be located in the border router, in one or more dedicated hosts, or in every physical object. The advantage of placing the IDS in the edge router is the intrusion detection from the Internet against the devices in the physical domain. However, an IDS in the edge router might produce communication overhead between the LLN nodes and the edge router because of the IDS regular querying of the network state. There are three possible placement approaches for IDSs as follows:

1. *Distributed IDS Placement*: In this placement strategy, IDSs are employed in every single physical object of the LLN. The IDS deployed in each node must be optimized since these nodes are resource-constrained. In the distributed

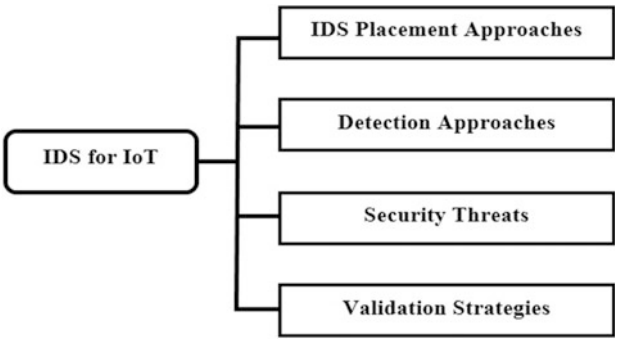


Fig. 3 Intrusion detection in internet of things

placement, the nodes may also be responsible for monitoring their neighbors. Nodes that audit their neighbors are called as watchdogs.

2. *Centralized IDS Placement*: In the centralized IDS placement, the IDS is located in a centralized component, for example, in the border router or a dedicated host. All the data that the LLN nodes collect and transmit to the Internet cross the border router along with the requests that Internet clients send to the LLN nodes. Consequently, the IDS placed in a border router can analyze all the traffic exchanged between the LLN and the Internet. However, analyzing the traffic that traverses the border router is not enough to detect attacks that involve only nodes within the LLN. Then, scientists must propose IDSs that can observe the traffic switched between LLN nodes, without ignoring the impact that this monitoring activity may have on low capacity nodes operation. Also, the centralized IDS may have trouble in auditing the nodes through an attack that compromises part of the network.
3. *Hybrid IDS Placement*: Hybrid IDS placement combines ideas of centralized and distributed placement to take benefit of their strong points and avoid their weaknesses. The first approach for hybrid placement organizes the network into clusters or regions, and only the main node in each host of cluster an IDS instance. Then, this node becomes responsible for auditing the other nodes of its cluster. In the second approach for hybrid placement, IDS modules are placed both in the edge router and in the remaining network nodes. The key difference of this approach to the first one is the presence of a central element. The IDS modules in the edge router are responsible for tasks that demand more resource capacity, while the IDS modules in regular nodes are usually lightweight.

2. **Detection Approaches**

Intrusion detection approaches in IoT are categorized into four types depending upon the detection mechanism which is anomaly-based, signature-based, specification-based and hybrid.

1. *Signature-based Approaches*: In signature-based approaches, IDSs detect attacks when the system or network behavior matches an attack signature warehoused in the IDS in-house databases. If any system or network activity matches with stored patterns/signatures, then an alert will be triggered. Signature-based IDSs are accurate at detecting known threats, and their technique is easy to understand. However, this approach is ineffective to detect new attacks and variants of known attacks, since a matching signature for these attacks is still unidentified.
2. *Anomaly-based Approaches*: Anomaly-based IDS compare the activities of a system at an instant against a normal behavior profile and produces the alarm whenever a deviation from normal behavior goes beyond a threshold. This

approach is effective to detect new attacks, in particular, those attacks related to misuse of resources. Nevertheless, whatever that does not match to a normal behavior is considered an intrusion and learning the entire scope of the normal behavior is not a simple task. Thereby, this method usually has high false positive rates. To construct the normal behavior profile, scientists usually employ statistical methods or machine learning algorithms that may be too heavy for low capacity nodes of IoT networks. Consequently, anomaly-based methods for IoT networks should take this particularity into account.

3. *Specification-based Approaches*: A specification is a group of rules and thresholds that express the expected behavior for network elements such as nodes, protocols, and routing tables. Specification-based approaches detect intrusions when network behavior deviates from specification definitions. Therefore, specification-based detection has the same purpose of anomaly-based detection: identifying deviations from normal behavior. However, there is one important difference between these methods: in specification-based approaches, a human expert should manually define the rules of each specification. Manually defined specifications usually provide lower false positive rates in comparison with the anomaly-based detection. Besides, specification-based detection systems do not need a training phase, since they can start working immediately after specification setup. However, manually defined specifications may not adapt to different environments and could be time-consuming and error-prone.
4. *Hybrid Approaches*: Hybrid approaches use ideas of signature-based, specification based and anomaly-based detection to maximize their advantages and minimize the impact of their disadvantages.

3. **Security Threats**

The goal of this part is to introduce how various types of attacks have been addressed in the IDS proposals for IoT. Enabling IoT solutions involves a composition of several technologies, services, and standards, each one with its security and privacy requirements. With this in mind, it is reasonable to assume that the IoT paradigm has at least the same security issues as mobile communication networks, cloud services, and the Internet. However, classical security countermeasures and privacy enforcement cannot be directly applied to IoT technologies due to three fundamental aspects: the limited computing power of IoT components, the high number of interconnected devices, and sharing of data among objects and users.

4. **Validation Strategy**

Validation consists of checking that the built model behaves with satisfactory accuracy within the study aims. There are several validation methods, and they may be distinguished by two sources of information: experts and data. While the use of

experts provides a subjective and often qualitative model validation, the use of data may allow a quantitative and more objective validation. The objective of this part is to investigate the validation strategy employed in the intrusion detection methods for IoT. Such criteria could be a starting point for evaluating the maturity level of this domain. For this purpose, the classification of validation methods can be as follows:

- *Hypothetical*: theoretical examples, having unclear relation to actual phenomena and degree of realism.
- *Empirical*: empirical approaches, such as systematic experimental collecting of data from operational settings.
- *Simulation*: Simulation approaches of some IoT scenario.
- *Theoretical*: formal or precise theoretical arguments to support results.
- *None*: no validation approaches are employed.

5 Applying Data Science Methods for Cybercrimes Investigation and Intrusion Detection in Internet of Things

Data science and knowledge discovery methods become significant topics in security domain where they can assist security professionals and digital investigators to detect and investigate cybercrimes as well as introduce solutions to malware and threat prediction, detection, and prevention at an initial stage. Knowledge discovery is known as data mining which refers to the process, in which hidden, unknown and potentially valuable information are extracted from massive, noisy, incomplete, and random data. The extracted information will be used for deriving novel insights, promoting business and scientific events, and speeding up and advancing scientific innovation.

At the present time, the furthermost imperative data mining algorithms mainly cover clustering, classification, regression, association analysis, statistical learning and linking mining. The methods of data science can use in the areas of cybercrimes investigation and intrusion detection in IoT environment to provide effective performance in the investigation, detection, prevention and prediction of IoT-based crimes in a timely fashion manner.

5.1 Data Science Methods for Cybercrimes Investigation

Data science and big data analytics have become significant paradigms to investigate IoT-based cybercrimes. Data science methods can use to analysis generated data from Internet of Things to investigate the crimes as well as predict the new coming severe attacks and crimes in the future. Lately, there are some challenges in digital forensics such as [21]:

- *Visualization of large amounts of data to the tribunal*: Visualization of finding from the analysis of digital evidence is vital for presenting the results in a court of law. Presentation and visualization of large of data is a problem faced digital practitioners and examiners in the digital forensics area so that there is a need for novel methods to deal with the massive size of data that generated from the crime scene in forensically and timely fashion way.
- *Search in a large amount of data*: Search is a commonly used application in digital forensics for extract valuable proof from digital evidence. Classical data is a summarization of structured data, which is enhanced for fast access and well-defined queries. Standard search methods are good for classical data. However, big data is unstructured or semi-structured. Consequently, general search procedures are not applicable to big data, especially for text, images, and videos that are structured for storage and display but not structured according to the content. Big data search objectives to extract convenient evidence from enormous data, and to facilitate decision-making. How to get value out of big data is a big challenge.
- *Storing and rapid indexing of massive amounts of data*: Conventional data storage not suitable for a large amount of data that created as a result of big data idea. This large amount of data that generated from different data sources need high size storage capacity to store for the digital investigation purpose. In recent times, data become big so faster indexing of the data is a challenge for digital investigators. In order to rapid indexing to the analysis of the large size of data, there is a need for faster methods and devices that have the ability analysis data within a given time frame.

From the aforementioned challenges, there is a need for employing data science, data mining and big data analytics methods in cybercrime investigation area because they have many advantages to support the digital investigation. Current digital forensic methods cannot do the extraction and analysis activities for the massive amount of data in an efficient manner so that there is need to scale up these methods to be suitable to the huge size of data. Some advantages of data science methods for digital forensics can be as the following:

- Enhance the analysis of evidential data which extracting from the crime scene.
- Diminish processing time of huge data analysis.
- Improve information quality associated with data analysis.

- Better utilization of existing computing, processing and storage resources.
- Decrease costs and save the time of the digital investigation.

The combining data science and digital forensics is to solve the crucial challenge of analyzing immense amount of data in actionable time while at the same time preserving forensic principles in order for the results to be presented in a court. After introducing digital forensics and data science in the background section explores the challenges to propose how data science methods can be adapted to the unique context of cybercrime investigation, ranging from the evidence managing through Map-Reduce to machine learning approaches for triage and analysis of a large amount of forensic data.

Data science using machine learning techniques can use to handle several complications that currently exist in digital forensics such as extracting and analyzing digital evidence from the crime scene. Using techniques that can automatic extraction of complex data representations or features in digital forensics can enhance the process of analyzing large amount of digital evidence in short time with high quality and accuracy of results. These techniques motivated by digital investigators and examiners to use in the forensic analysis stage. There are a number of topics in cybercrimes investigation in Internet of Things where machine learning techniques can be used as follows:

- *Data Indexing*: Large-scale volume of data such as text, image, video, and audio are being extracted and collected from different sources that can make investigation process harder especially when crimes related to environment such as Internet of Things. These huge amounts of data need semantic indexing rather than being stored as data bit strings. Semantic indexing presents the data in a more efficient manner and makes it useful as a source for knowledge discovery and understanding.
- *Pattern Recognition*: Pattern recognition is an important area in machine learning that working on extracting patterns from input data. Supervised data that trained from labeled data and unsupervised learning that discover unknown patterns. Both of them can use in the pattern recognition. It is used to identify pattern or feature in data through determining and specify types or clusters of data. The pattern recognition can help in digital forensics\ for performing detecting a pattern in an e-mail message which indicates malicious code like spam or virus. Likewise, can be used to discover identities in digital evidence that is extracting from the crime scene.
- *Authorship Identification*: Criminals can use fake emails for performing activities without tracing them through hiding their identity. Authorship Identification is an important technique which used to solve this problem by identifying the authors of these fake e-mails that can help digital investigators and examiners to perform investigation process in a timely fashion manner.
- *Image Region Forgery Detection*: In recent time, the number of tampered images is increased incredible way due to the use of social networks like Facebook, Flicker, and Twitter. These tampered images can be shared easily by

the users that may lead serious consequences so the authenticity of digital images is urgently needed. The presence of tampered images is an important topic in digital forensics.

- *File Fragments*: Detection of data from disks is challenging faced by digital investigators to recover data from disk. The data when deleting from disk is not permanent where they simply mark each block of the file as unallocated and available for use. The process of recovering unallocated data called as 'file carving'. Machine learning introduces approaches to recognizing the file types of file fragments for the purpose of file carving for the reconstruction of partially erased files on disk into whole files.

5.2 Data Science Methods for Intrusion Detection

Intrusion detection refers to the procedure of auditing and analyzing the events occurring in a system to detect malicious behaviors. The intrusion detection process involves detecting a set of nasty actions that compromise available resources. In last years, there is a serious need for new data science methods for analysis of sophisticated attack in Internet of Things environment. Current methods suffer from evaluation, comparison, and deployment which originate from the scarcity of adequate publicly available network trace data sets. Also, publicly existing datasets are either out-of-date or generated in a controlled environment.

Data science involves various analytical techniques such as machine learning, artificial intelligence, and data mining that are useful for extracting features from data sets. There are many techniques which can use to detect unknown new attacks. These techniques such as prediction, classification, clustering and relation rule.

- *Prediction*: It is a technique that predicts the future possibility and trend. Regression analysis is a representative prediction technique. Researchers can predict attack possibilities using regressing analysis. Regressing analysis can predict similar behaviors from collected attack logs.
- *Classification*: It is a technique that predicts the group of a new attack from huge data. Classification helps security administrator to decide the direction of protection and analysis.
- *Clustering*: It is an unsupervised technique where the data set is divided into sub parts sharing same properties. The clustering process is used for finding similarities in data and putting similar data into sets. Clustering partitions a data set into several groups such that the similarity within a group is larger than that among groups. Clustering procedures are used extensively not only to organize and categorize data but are also suitable for data compression and model construction.

- *Relation Rule:* It is a technique that discovers hidden relations among data. The action of discovering relation rule is named association analysis or link analysis. The relation from time flow is named as sequence rule. This analysis technique can determine abnormal behavior by analyzing user or process behaviors.

Data scientists and researchers can make great achievements in this area through designing novel threat detection models that can combine data science, machine learning, and behavioral analysis. They can recognize the underlying purpose of traffic, detect attack behaviors in real time IoT applications. This model can be applied directly to network traffic to expose underlying attack features that unknown. Supervised and unsupervised machine learning algorithms can help in discovering uncover new attack behaviors.

In order to develop and propose new efficient intrusion detection systems based on data science methods, it is required to work on attacks datasets to test and evaluate their innovation detection and prevention models. One of the most common data sets for developing attacks and intrusions detection system is KDD CUP 99 dataset [22]. The KDD CUP 99 has been most commonly used in attacks detection using data mining techniques. The KDD data set contains 10% of original dataset that is approximately 494,020 single connection vectors each of which has 41 features and is labeled with exact one specific attack category. Every vector is labeled as either normal or an attack, with accurately one specific attack category. The simulated attack may be one of the following four categories [23]:

1. *Denial of Service (DOS) Attack*: In this attack, the attacker makes computing or memory resources busy to allow the legitimate request, or deny the access legitimate of users to the system. The DOS involves attacks such as 'land', 'smurf', 'neptune', 'pod', 'back' and 'teardrop'.
2. *Users to Root (U2R) Attack*: In this type of attack, the attacker starts out with access to a normal user account on the system and is able to exploit some vulnerability to obtain root access to the system. The U2R involves attacks such as 'loadmodule', 'rootkit', 'buffer_overflow' and 'perl'.
3. *Remote to Local (R2L) Attack*: In this attack, the attacker sends packets to the system over a network but who does not have an account on that system and exploits a vulnerability to gain local access as a user of that system. The R2L contains attacks such as 'warezclient', 'imap', 'multihop', 'guess_passwd', 'warezmaster', 'spy' 'ftp_write', and 'phf'.
4. *Probing Attack*: In this category, the attacker attempt to gather information about the network of computers for the apparent purpose of circumventing its security. This attack covers the following attacks: 'portsweep', 'satan', 'nmap', and 'ipsweep'.

6 Conclusions and Future Directions

In recent times, data science become a very significant topic that has attracted attention numerous researchers who interest in solving problems of intrusion detection and cybercrimes detection in Internet of Things environment. Therefore, this chapter introduced the principles of Digital Forensics, Intrusion Detection and Internet of Things as well as exploring data science concepts and methods that can help the digital investigators and security professionals to develop and propose new techniques and methods that can be adapted to the unique context of Internet of Things infrastructure for performing intrusion detection and cybercrimes investigation. As future research work, researchers may focus on some issues such as follows:

- To explore advantages and disadvantages of various current intrusion detection strategies.
- To improve the security of alert traffic, alert correlation, and autonomic management systems.
- To develop new/novel detection model for automated risk management through linking machine learning procedures, data science methods, and behavioral analysis.
- To propose invulnerable-based heuristic IDSs using neural and fuzzy methods to control the sensitivity of alerting malicious intrusions to decrease false alarm rate.
- To develop advanced feature extraction and selection algorithms for improving the performance of detection models will be positively affected. And also, help to construct strong and efficient classifier to detect new attacks and threats.
- To use deep learning methods for predictions and classification of attacks.
- To improve the performance of real-time intrusion detection systems.
- Use Big Data analytics tools and platforms such as Apache Hadoop ecosystems and Apache Spark to enhance and increase analysis performance in intrusions and attacks detection.

References

1. Palmer, G.: A road map for digital forensic research. First Digital Forensic Research Workshop, Utica, New York (2001)
2. McKemmish, Rodney: What is forensic computing?. Australian Institute of Criminology, Canberra (1999)
3. Khan, Minhaj Ahmad: A survey of security issues for cloud computing. J. Netw. Comput. Appl. **71**, 11–29 (2016)
4. Oscar Serrano, C.I.S.A.: CISSP CISM, and Luc Dandurand. Big Data Analytics for Sophisticated Attack Detection (2014)
5. Jabez, J., Muthukumar, B.: Intrusion detection system (IDS): anomaly detection using outlier detection approach. Procedia Comput. Sci. **48**, 338–346 (2015)

6. Zarpelão, B.B., et al.: A survey of intrusion detection in internet of things. J. Netw. Comput. Appl. (2017)
7. Kawamoto, Y., Nishiyama, H., Kato, N., Yoshimura, N., Yamamoto, S.: Internet of things (IoT): present state and future prospects. IEICE Trans. Inf. Syst. E **97**(10), 2568–2575 (2014)
8. Gubbi, J., Buyya, R., Marusic, S., Palaniswamia, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Futur. Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
9. Atzori, L., Iera, T., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
10. Li, S., Xu, L.D., Zhao, S.: The internet of things: a survey. Inf. Syst. Front. **17**(2), 243–259 (2014)
11. Andrew, W., Agarwal, A., Xu, L.D.: The internet of things a survey of topics and trends. Inf. Syst. Front. **17**(2), 261–274 (2014)
12. Farooq, M.U., Waseem, M., Mazhar, S., Khairi, A., Kamal, T.: A review on internet of things (IoT). Int. J. Comput. Appl. **113**(1), 1–7 (2015)
13. El-Din, H.E., Manjaiah, D.H.: Internet of things in cloud computing. Internet of Things: Novel Advances and Envisioned Applications. Springer International Publishing, pp. 299–311 (2017)
14. Cady, F.: The Data Science Handbook. Wiley (2017)
15. Giuliano, R., et al.: Security and Privacy in Internet of Things (IoTs): Models, Algorithms, and Implementations (2015)
16. Perumal, S., Norwawi, N.M., Valliappan, R.: Internet of things (IoT) digital forensic investigation model: top-down forensic approach methodology. In: 2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC). IEEE (2015)
17. Zawoad, S., Hasan, R.: FAIoT: towards building a forensics aware eco system for the internet of things. In: 2015 IEEE International Conference on Services Computing (SCC). IEEE (2015)
18. Oriwoh, E., et al.: Internet of things forensics: challenges and approaches. In: 2013 9th International Conference Conference on Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom). IEEE (2013)
19. Qiu, T., Zhang, Y., Qiao, D., Zhang, X., Wymore, M.L., Sangaiah, A.K.: A robust time synchronization scheme for industrial internet of things. IEEE Trans. Ind. Inform. (2017). https://doi.org/10.1109/TII.2017.2738842
20. Sherasiya, T., Upadhyay, H.: Intrusion detection system for internet of things. Int. J. Adv. Res. Innov. Ideas Educ. (IJARIIE) **2**(3) (2016)
21. Uma, M., Salisu, S.: The use of big data in the field of digital forensics investigations (comparative study between digital forensics in uk and nigeria). Int. J. New Technol. Sci. Eng. **2**(4) (2015)
22. Tavallaee, M., et al.: A detailed analysis of the KDD CUP 99 data set. In: IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE (2009)
23. Siddiqui, M.K., Naahid, S.: Analysis of KDD CUP 99 dataset using clustering based Data Mining. Int. J. Database Theory Appl. **6**(5), pp. 23–34 (2013). https://doi.org/10.14257/ijdta.2013.6.5.03

Modelling and Analysis of Multi-objective Service Selection Scheme in IoT-Cloud Environment

Chinu Singla, Nitish Mahajan, Sakshi Kaushal, Amandeep Verma and Arun Kumar Sangaiah

Abstract Internet of Things (IoT) is a heterogeneous ubiquitous network based upon modern computational intelligent techniques. A large scale IoT environment composed of thousands distributed entities and a number of multimedia smart devices. In recent years, due to the improvement of popularity and capability of smart mobile devices, Mobile Cloud Computing (MCC) gains a considerable attention in Internet of Things (IoT) environment. As there are variety of clouds that provides same services, it becomes quite difficult for users to choose an ideal cloud from a variety of clouds for migrating computationally intensive applications. So, selecting the optimal cloud among multiple alternatives which saves resource availability and execution time is a Multi-Criteria Decision Making (MCDM) issue. This chapter introduces an assessment model based on Fuzzy Analytic Hierarchy Process (FAHP) and Fuzzy Technique for Order Preference by Similarity to Ideal Solution (FTOPSIS) which helps the users to select an optimal cloud where uncertainty and subjectivity are parameterized using triangular fuzzy members and is handled by using linguistic values. The proposed computational intelligence decision making model enables decision makers to better understand the whole evaluation process and thus provides more accuracy, systematic and efficient decision support tool.

C. Singla (✉) · N. Mahajan · S. Kaushal · A. Verma
University Institute of Engineering and Technology, Panjab University, Chandigarh, India
e-mail: cheenusingla@gmail.com

N. Mahajan
e-mail: nitish7mahajan@gmail.com

S. Kaushal
e-mail: sakshi@pu.ac.in

A. Verma
e-mail: amandeepverma@pu.ac.in

A. K. Sangaiah
School of Computing Science and Engineering, VIT University, Vellore, India
e-mail: sarunkumar@vit.ac.in

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*, Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_3

Keywords Internet of things · Mobile cloud computing · Fuzzy AHP Fuzzy TOPSIS · Computational intelligence · MCDM · Offloading

1 Introduction

Cloud Computing (CC) directs to accessing, configuring and manipulating the application through the Internet. It provides different resources and services to mobile device such as Software as a Service (SaaS), Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Testing as a Service (TaaS), etc. through various service providers and customers can access these resources through rich-resourceful data centers [1]. With the enhancement of mobile applications and cloud computing support, Mobile Cloud Computing (MCC) is introduced which is defined as the combination of cloud computing into the mobile domain [2]. It provides new services and facilities to mobile users ubiquitously by attaining the benefits of cloud data centres. The Internet of things (IoT) paradigm is based on self-configuring and intelligent nodes that are interconnected in a dynamic network infrastructure. IoT can benefit from the unlimited resources and capabilities of cloud in terms of storage and processing power. The integration of IoT with the cloud solves many issues [1, 3] by providing additional features such as ease-of-use, ease-of-access and decreased deployment costs [4].

Along with the enhancement of MCC [2], offloading has become an ideal solution to enhance energy consumption, battery lifetime and execution time on smart mobile devices. Due to the emergence of a class of clouds providing similar services to mobile users having different CPU utilization, speed and thus migrating the same application may require different amount of computing and execution time due to the difference in cloud's speed, cost and resource availability. So, it is necessary to choose an optimal cloud from a variety of clouds which satisfies the user requirements in less cost and with improved throughput performance.

For selecting an optimal cloud, a large variety of data and parameters needs to be analysed which is a complex process. Various Multiple Criteria Decision Making [5] techniques are used for migrating intensive applications by selecting an appropriate cloud. MCDM is defined as finding the best alternative from all the optimal alternatives usually when multiple conflicting criteria's are evaluated. It consists of a large number of objectives which cannot be achieved simultaneously. One of the most outstanding MCDM techniques is the Analytic Hierarchy Process (AHP) [6, 7] which help in determining the relative weights of different criteria's. Conventional AHP is often criticized as it requires pair-wise comparison which involves uncertainty and impreciseness in decision problems [8]. TOPSIS, another MCDM approach is based on selecting an ideal alternative that has the maximum distance from the negative- ideal alternative and minimize distance from the positive- ideal alternative. Hwang and Yoon [9] described more detailed information about the TOPSIS method.

In most references cases, decision-makers preferences are imprecise and are unable to make accurate numerical comparisons. When an uncertainty exists in a pair-wise comparison, fuzzy decision making is a powerful tool used under fuzzy environment to select an appropriate alternative. Conventional decision-making techniques work only with the exact numerical data which never deals with the uncertain or fuzzy data.

The objective of this chapter is to propose computational intelligence decision making model can enable users to use cloud services during mobility accurately, cost effectively and in systematic way.

The remainder of this study is structured as follows: Sect. 2 discuss the related work. Section 3 describes the different optimization techniques of MCDM model. In Sect. 4, proposed model for cloud path selection is presented by explaining different stages of the proposed approach. Section 5 illustrates how proposed model is used for cloud path selection by considering the real-world application. Section 6 describes the implementation details and experimental results of our proposed work. Section 7 concludes the chapter by giving future directions.

2 Literature Review

In recent period, a lot of research work has been investigated on cloud based IoT technologies to achieve efficient computation offloading process. For effective and accurate evaluation, experts' may require multiple parameters to be considered and huge amount of data to be analysed [10]. The MCDM technique should wait until the experts' and analysts understand the whole problem, objectives, different criteria's, feasible alternatives and level of uncertain data [11]. In [12], researchers had used multiplicative priority rating techniques for the AHP. In this chapter, authors had focused on evaluating the consistency of different decision maker judgements in decision support systems [13]. In [7], authors had described that eigen-vector of pair-wise comparison matrix shows the local priority criteria weights, sub-criteria and alternatives. In [14], sequential decision-making method was used by conducting the question response process and developed a dynamic programming for it. According to [15], authors have proposed AHP and TOPSIS technique to evaluate airlines service quality. An AHP based technique was proposed by Godse and Mulik to select a SaaS service [16]. An ANP based procedure to select IaaS service has been proposed by Menzel et al. [17]. Zeng et al. proposed architecture based on cloud service and an algorithm to select an ideal cloud service [18]. Simple additive weighted-based approaches for ranking different cloud service have been proposed by Saripalli and Pingali [19, 20] discussed different approaches to ERP selection problem using FAHP and Fuzzy TOPSIS method. A greedy technique for a cloud service selection problem using B+ trees has been proposed by Sundareswaran et al. [21]. A MCDM approach was used to evaluate the performance of mobile phone alternatives [22]. Wang et al. described a dynamic model of

cloud service selection by using dynamic learning technique for multi-cloud computing purpose [23].

Although most of the work has been done on cloud path selection for offloading in MCC environment but that is limited only on single criterion. However, in this study, both single and multiple decision analysis approaches are performed by considering different criteria's such as availability, capacity, privacy, speed and cost in selection problem.

3 Optimization Techniques

This section presents a brief overview on optimization techniques to solve various MCDM problems.

3.1 The Analytic Hierarchy Process (AHP) Method

Saaty has developed an Analytic Hierarchy Process (AHP) to determine the relative importance of different activities in a multi-criteria decision problem [7]. AHP makes it possible to incorporate judgement on indefinable qualitative criteria alongside defined criteria [24]. Within a hierarchical structure, AHP separates the complex decision problems into elements and then converts it into mathematical values to select optimal criteria from a set of criteria's. This method has been widely used in solving complicated decision- making problems [25, 26]. It basically consists of six steps to determine the relative weights of the criteria [7]. In the first step, an unstructured problem is defined by clearly defines its goals and outcomes. In the second step, it breaks down a complex multi-criteria decision problem into a hierarchical structure consists of a set of decision elements (criteria, sub-criteria and alternatives). In the decision hierarchy, these decision elements are arranged in hierarchical structure. A hierarchy consists of at least three levels; overall objective at the top, multiple feasible criteria's based on user's preferences at the middle and decision opinions at the bottom [27]. In the third step, pair-wise comparison is performed for constructing the comparison matrix, A_w (n * n). In the fourth step, eigen value method is used where the weights are given by right eigen vector (v) corresponding to the largest eigen value (λ_{max}) as given in Eq. (1). Consistency of the pair-wise comparison matrix is checked in fifth step by using Eq. (2). In the final step, aggregation of relative weights is done to evaluate the overall performance of all the alternatives.

$$\mathrm{A_w} = \lambda_{\max} \mathrm{v} \tag{1}$$

$$\mathrm{CI} = (\lambda_{\max} - \mathrm{n}) \ /(\mathrm{n} - 1) \tag{2}$$

3.2 The TOPSIS Method

Hwang and Yoon was developed the technique named as TOPSIS (Technique for Order Preference by Similarity to Ideal Solution). This technique selects the optimal alternative which is simultaneously close to the positive ideal solution and farthest from negative ideal solution [28]. The positive ideal solution is a solution which minimizes the cost criteria and maximizes the benefit criteria, whereas the negative ideal solution minimizes the benefit criteria and maximizes the cost criteria [29]. There have been lots of literature studies for MCDM problems which use TOPSIS method to obtain final ranking of alternative clouds [30, 31].

3.3 Fuzzy Set Theory

Zadeh has introduced fuzzy set theory to determine uncertainty to complex decision- making problems [32]. It is represented in terms of membership values having values lies between 0 and 1. Using fuzzy MCDM approach, we can convert the existing accurate values to five levels which are represented as Very Low (VL), Low (L), Medium (M), High (H), and Very High (VH). A triangular fuzzy representation is very suitable in fuzzy system as shown in Fig. 1. The function value $\boldsymbol{\mu_h}$ as described in Eq. (3) represents membership function of x and triangular membership function is represented by a triplet (h_1, h_2, h_3) [33].

$$\mu_h(x) = \begin{cases} (y-l)/(n-l), & l \le y \le n \\ (p-y)/(p-n), & n \le y \le p \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Here s and t represents the lower and upper bounds of the fuzzy number and n describes the modal value for h.

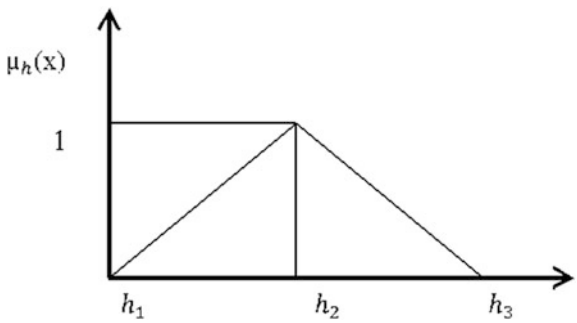


Fig. 1 Triangular fuzzy number

In our work, we have used fuzzy AHP and fuzzy TOPSIS as it helps the users to select an optimal alternative in a fuzzy environment where uncertainty is parameterized using triangular fuzzy numbers and is handled by using linguistic values. Further, our proposed model is presented in next section.

In next section, we have used FAHP and FTOPSIS method to rank the cloud-path pairs for offloading in MCC environment.

4 Proposed Model

This section presents details of proposed algorithm. The algorithm firstly takes input as an evaluation criterion's and gives an optimal output from a set of alternatives. Then fuzzy AHP and fuzzy TOPSIS are applied and presented in Sects. 4.1, 4.2 and 4.3 respectively.

4.1 Evaluation Framework

MCDM is widely used tool for solving the multiple conflicting criteria problems [34–37]. These techniques enable the decision makers to structure the problem systematically and clearly. MCDM problems are assumed to have predetermined and limited number of alternatives are considered. The main aim of our chapter is to select the feasible alternative among a set of clouds in respect to the customers preference orders. For implementation, our proposed algorithm is mainly divided into two parts. Firstly, FAHP is used to determine the priorities of various assessment criteria's and then select the best alternative using fuzzy TOPSIS method. The evaluation framework consists of three main steps as described in Fig. 2. In the first step, we identify the assessment criteria which are considered as the most preferable criteria based on user's preferences. Then, by using FAHP method we determine the weights of criteria after constructing the selection criteria hierarchy and in the final step fuzzy TOPSIS method is used to achieve the final ranking of possible alternatives.

4.2 Fuzzy AHP Model

Due to uncertainty and insufficient information, sometimes it becomes quite difficult for the experts to take the appropriate decision within the decision environment. So, to solve these issues a fuzzy set theory is used based on user perceptions. FAHP is a fuzzy extension of conventional AHP and it consist the following six steps [38].

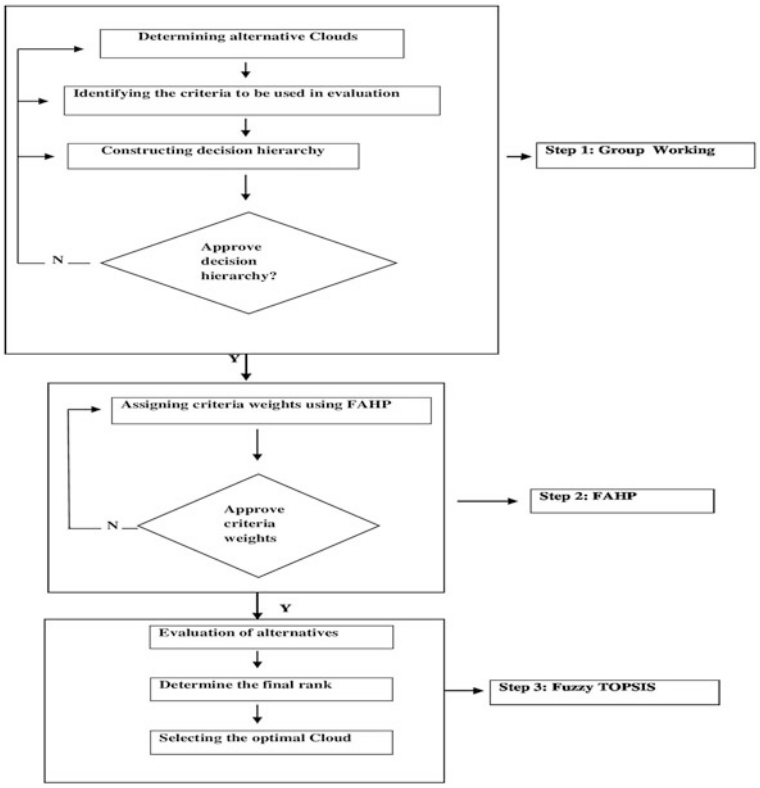


Fig. 2 Evaluation framework

Table 1 9-Point scale and its definition [7]

Intensity of Importance	Definition
1	Equally important
3	Moderately more important
5	Strongly more important
7	Very strongly more important
9	Extremely more important
2, 4, 6, 8	Intermediate values

Step 1: Construct pair-wise comparison matrix among all the elements and express it by using 9-point scale defined by Satty [7] as given in Table 1.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad (4)$$

Here $\mathbf{a_{ii}} = 1$ and $\mathbf{a_{ij}} = \frac{1}{\mathbf{a_{ji}}}$

Step 2: Comparison matrix consistency analysis:

$$\mathbf{A} = \begin{bmatrix} \frac{\mathbf{w_1}}{\mathbf{w_1}+\mathbf{w_1}} & \frac{\mathbf{w_1}}{\mathbf{w_1}+\mathbf{w_2}} & \cdots & \frac{\mathbf{w_1}}{\mathbf{w_1}+\mathbf{w_n}} \\ \frac{\mathbf{w_2}}{\mathbf{w_2}+\mathbf{w_1}} & \frac{\mathbf{w_2}}{\mathbf{w_2}+\mathbf{w_2}} & \cdots & \frac{\mathbf{w_2}}{\mathbf{w_2}+\mathbf{w_n}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\mathbf{w_n}}{\mathbf{w_n}+\mathbf{w_1}} & \frac{\mathbf{w_n}}{\mathbf{w_n}+\mathbf{w_2}} & \cdots & \frac{\mathbf{w_n}}{\mathbf{w_n}+\mathbf{w_n}} \end{bmatrix} = \begin{bmatrix} \mathbf{s_{11}} & \mathbf{s_{12}} & \cdots & \mathbf{s_{1n}} \\ \mathbf{s_{21}} & \mathbf{s_{22}} & \cdots & \mathbf{s_{2n}} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{s_{n1}} & \mathbf{s_{n2}} & \cdots & \mathbf{s_{nn}} \end{bmatrix} \tag{5}$$

If it is consistent it must satisfy:

$$\mathbf{s_{ii}} = 0.5,\ \mathbf{s_{ij}} + \mathbf{s_{ji}} = 1,\ \frac{1}{\mathbf{s_{ij}}} - 1 = \left(\frac{1}{\mathbf{s_{ik}}} - 1\right) \times \left(\frac{1}{\mathbf{s_{ki}}} - 1\right) \tag{6}$$

Step 3: Evaluate positive fuzzy matrix to convert pair-wise comparison matrix into fuzzy numbers having values between 0 and 1. The fuzzy pair-wise comparison scale is given in Table 2 which when compared with Satty's scale (Table 1) must satisfies the following equation.

$$\mathbf{s_{ij}} = \frac{\mathbf{a_{ij}}}{\mathbf{a_{ij}} + 1} \tag{7}$$

Step 4: Calculate the fuzzy weights of decision elements as shown in Eq. (8).

$$\mathrm{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots\ \mathbf{w_n}) \tag{8}$$

$$\mathrm{w_i} = \frac{\mathrm{z_i}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{z_i}} \tag{9}$$

where, $\mathbf{z_i} = \frac{1}{\left[\sum_{\mathrm{j}=1}^{\mathbf{n}} \frac{1}{\mathbf{s_{ij}}} - \mathbf{n}\right]}$

Step 5: Integrate the decision of all experts by taking Geometric mean.

Step 6: Obtain final ranking by evaluating the Consistency Index (CI) as presented in Eq. (10).

Table 2 Fuzzy pair wise comparison scale and its description

Fuzzy pairwise scale	Description
0.5	Equally important
0.55	Slightly important
0.65	Important
0.75	Strongly important
0.85	Very strongly important
0.95	Extremely important

Table 3 Values of RI

Matrix size	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

$$\mathbf{CI} = \frac{\left[\sum_{i=1}^{n} \frac{(AW)_i}{nw_i}\right]}{n-1} \tag{10}$$

And, Consistency Ratio (CR) is obtained according to Eq. (11).

$$\mathbf{CR} = \frac{\mathbf{CI}}{\mathbf{RI}} \tag{11}$$

Where, RI is Random Consistency Index and its values are shown in Table 3.

4.3 Fuzzy TOPSIS Model

Fuzzy TOPSIS method is used for formulating decision problem that are parameterized using triangular fuzzy number and is handled by using Linguistic values [39]. It contains following steps [40].

Step 1: Construct the fuzzy decision matrix for ranking of different alternatives [39].

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ x_{M1} & x_{M2} & \cdots & x_{MN} \end{bmatrix} \tag{12}$$

where N is the number of criteria's and M is the total number of alternatives. Here normalization is not required as belong to [0, 1]. The triangular fuzzy numbers relate with linguistic variables as shown in Table 4.

Table 4 Fuzzy membership functions

Linguistic values	Fuzzy ranges
Very low (VL)	(0, 0, 0.2)
Low (L)	(0, 0.2, 0.4)
Medium (M)	(0.2, 0.4, 0.6)
High (H)	(0.4, 0.6, 0.8)
Very high (VH)	(0.6, 0.8, 1)
Excellent (E)	(0.8, 1, 1)

Step 2: Calculate the weighted fuzzy normalized decision matrix.

$$y_{ij} = x_{ij} \times w_j,\ i = 1, 2 \ldots M \text{ and } j = 1, 2 \ldots N \tag{13}$$

Where w_j represents the weights of jth criterion obtained using FAHP method.

Step 3: Identify the positive ideal (A^+) and negative ideal (A^-) solutions as presented in Eq. (14).

$$A^+ = \{y_1^+,\ y_2^+,\ \ldots,\ y_N^+\} = \left\{\left(\max_i y_{ij}\right)\right\}$$
$$A^- = \{y_1^-, y_2^-,\ \ldots,\ y_N^-\} = \left\{\left(\min_i y_{ij}\right)\right\} \tag{14}$$

We assume fuzzy positive ideal solution as $y_j^+ = (1, 1, 1)$ and negative ideal solution as $y_j^- = (0, 0, 0)$ [22].

Step 4: Calculate distance from A^+ and A^- using Euclidean distance according to Eq. (15).

$$D_i^+ = \sum_{j=1}^{N} d\left(y_{ij}, y_j^+\right)$$
$$D_i^- = \sum_{j=1}^{N} d\left(y_{ij}, y_j^-\right) \tag{15}$$

Step 5: Calculate the closeness to ideal solution.

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{16}$$

Step 6: Finally rank preference order according to S_i in descending order.

So, we have concluded that the cloud having the largest S_i value is considered as the ideal cloud according to the fuzzy TOPSIS calculations.

5 Numerical Analysis

In this section, proposed work is numerically analyzed by considering real-time mobility environment by considering various parameters such as resource availability, privacy, capacity, speed and cost. As real-time mobility environment allows the user to access application, infrastructure and corporate services in a secure and effective manner.

The decision making hierarchical structure of the chapter with the criteria's and decisions is portrayed in Fig. 3.

We assume the priority of importance of criteria's as shown in Table 5 which is ranked as: availability > speed > capacity > cost > privacy. However, the priority of these criteria's is different in different conditions.

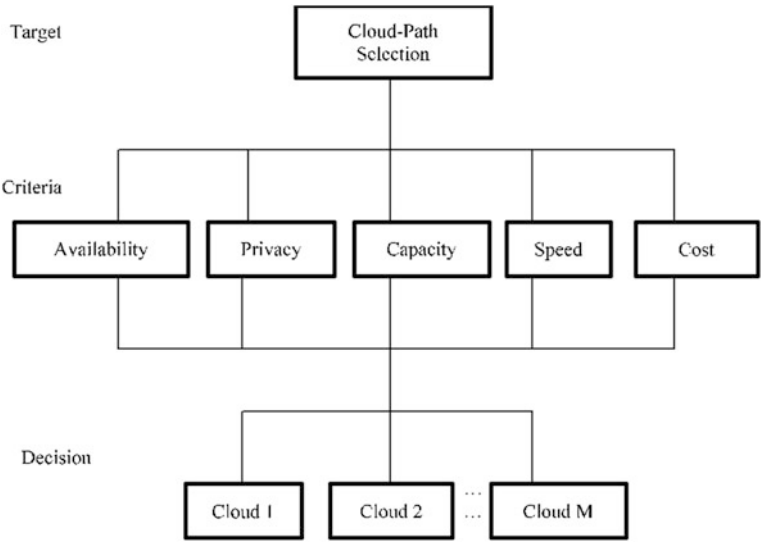


Fig. 3 Decision- making problem hierarchy

Table 5 Fuzzy pair-wise comparison matrix

Criteria	Availability	Capacity	Privacy	Speed	Cost
Availability	0.5	0.7	0.8	0.5	0.9
Capacity	0.3	0.5	0.6	0.3	0.7
Privacy	0.2	0.4	0.5	0.1	0.4
Speed	0.5	0.7	0.9	0.5	0.7
Cost	0.1	0.3	0.6	0.3	0.5

Table 6 Results Obtained using FAHP

Criteria	Weights	Rank	CI, RI, CR
Availability	0.3708	1	CI = 0.056 RI = 1.12 CR = 0.05
Capacity	0.1528	3	
Privacy	0.0609	5	
Speed	0.348	2	
Cost	0.0673	4	

By using the FAHP method, we evaluate the priority weights, CI, RI and CR as shown in Table 6.

From above Table, it can be analyzed that CR = 0.05 < 0.1. It means the weights are consistent under FAHP method.

6 Implementation and Experimental Results

The applications of Cloud have various requirements for configuration and deployment. It is very difficult to analyse the performance of these applications on real Cloud. For the purpose of implementation, CloudSim and MATLAB were used as simulation tools which help in evaluating different application without investing in the purchase of real time Cloud infrastructure. CloudSim allows the users to focus on a specific design or implementation issue without worrying about the detailed workings whereas, MATLAB is used to calculated the weights of a fuzzy decision matrix using FAHP and FTOPSIS method.

Table 7 Fuzzy TOPSIS results

Alternatives	D_i^+	D_i^-	S_i	Rank
Cloud 1	3.761	2.262	0.376	2
Cloud 2	3.724	2.309	0.383	1
Cloud 3	3.900	2.124	0.353	3
Cloud 4	3.981	2.044	0.339	5
Cloud 5	3.919	2.110	0.350	4

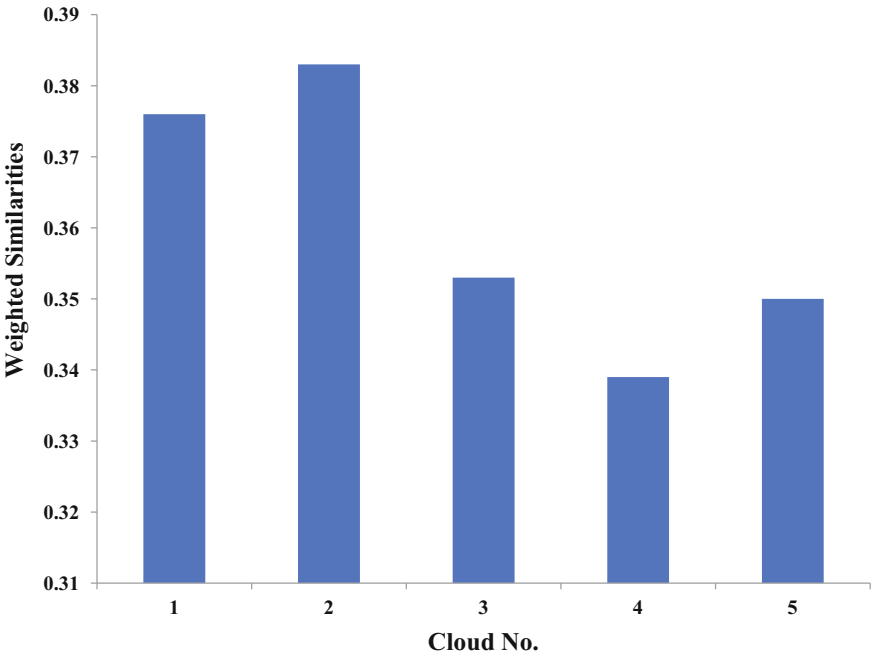


Fig. 4 The decision hierarchy of cloud path selection

In this problem of cloud selection, it can be concluded that privacy is the cost criterion while other criterions are benefit criteria. The results of fuzzy TOPSIS as shown in Table 7 where D_i^+, D_i^- and S_i can be calculated by using Eqs. (15) and (16).

Based on S_i values, the ranking of clouds in ascending order are cloud 4, cloud 5, cloud 3, cloud 1 and cloud 2 as shown in Fig. 4. Therefore, in this case users can choose cloud 2 to offload data and access different types of cloud services when we consider these five criteria's simultaneously.

Thus, when the criteria weights are uncertain and inaccurate, then FAHP and Fuzzy TOPSIS are the preferred techniques which enables makers to better understand the whole evaluation process and thus increases the efficiency of decision making process in cloud path selection for offloading in MCC environment.

7 Conclusion and Future Scope

This study proposes a computational intelligent scheme based on fuzzy AHP and fuzzy TOPSIS to select an optimal cloud for accessing different services of cloud and also offloading data by evaluating the weights of important criteria's and by calculating the final ranking of alternative clouds. This chapter aims in choosing the cloud path when multiple criteria are considered which will be a critical issue for migrating and using the applications in cloud during mobility. When the criteria weights are uncertain and inaccurate, then FAHP and Fuzzy TOPSIS are the preferred techniques. The proposed algorithm allows analysts to better understand the whole evaluation process and thus increases the efficiency of decision making process in cloud path selection for offloading in MCC environment. As a future work, mathematical models can also be integrated using the proposed model path to ensure more integrated or comparative study by addressing different system architecture which combines the mobile cloud and IoT applications that provides better quality of service.

References

1. Fox, G.C., Kamburugamuve, S., Hartman, R.D.: Architecture and measured characteristics of a cloud based internet of things. In: 2012 International Conference on Collaboration Technologies and Systems (CTS), pp. 6–12. IEEE (May 2012)
2. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R.H., Konwinski, A., Zaharia, M.: Above the Clouds: a Berkeley View of Cloud. Electrical Engineering and Computer Sciences, University of California, Berkeley (2009)
3. Suciu, G., Vulpe, A., Halunga, S., Fratu, O., Todoran, G., Suciu, V.: Smart cities built on resilient cloud computing and secure internet of things. In: 2013 19th International

Conference on Control Systems and Computer Science (CSCS), pp. 513–518. IEEE (May 2013)
4. Dash, S.K., Mohapatra, S., Pattnaik, P.K.: A survey on applications of wireless sensor network using cloud computing. Int. J. Comput. Sci. Eng. Technol. **1**(4), 50–55. E-ISSN: 2044-6004
5. Whaiduzzaman, M., Gani, A., Anuar, N.B., Shiraz, M., Haque, M.N., Haque, I.T.: Cloud service selection using multicriteria decision analysis. Sci. World J. (2014)
6. Saaty, T.L.: The Analytic Hierarchy Process. McGraw Hill Company, New York (1980)
7. Saaty, T.L.: Decision Making with Dependence and Feedback: The Analytic Network Process, vol. 4922. RWS publications, Pittsburgh
8. Yu, C.S.: A GP-AHP method for solving group decision-making fuzzy AHP problems. Comput. Oper. Res. **29**(14), 1969–2001 (2002)
9. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications a State-of-the-Art Survey, vol. 186. Springer Science & Business Media
10. Ayağ, Z., Özdemir, R.G.: A fuzzy AHP approach to evaluating machine tool alternatives. J. Intell. Manuf. **17**(2), 179–190 (2006)
11. Mergias, I., Moustakas, K., Papadopoulos, A., Loizidou, M.: Multi-criteria decision aid approach for the selection of the best compromise management scheme for ELVs: the case of Cyprus. J. Hazard. Mater. **147**(3), 706–717 (2007)
12. Stam, A., Silva, A.P.D.: On multiplicative priority rating methods for the AHP. Eur. J. Oper. Res. **145**(1), 92–108 (2003)
13. Aguaron, J., Escobar, M.T., Moreno-Jiménez, J.M.: Consistency stability intervals for a judgement in AHP decision support systems. Eur. J. Oper. Res. **145**(2), 382–393 (2003)
14. Holloway, H.A., White Iii, C.C.: Question selection for multi-attribute decision-aiding. Eur. J. Oper. Res. **148**(3), 525–533 (2003)
15. Feng, C.M., Wang, R.T.: Performance evaluation for airlines including the consideration of financial ratios. J. Air Transp. Manage. **6**(3), 133–142 (2000)
16. Godse, M., Mulik, S.: An approach for selecting software-as-a-service (SaaS) product. In: IEEE International Conference on Cloud Computing, 2009 (CLOUD'09). pp. 155–158. IEEE (2009)
17. Menzel, M., Schönherr, M. Tai, S.: (MC2)2: criteria, requirements and a software prototype for cloud infrastructure decisions. Softw. Pract. Experience **43**(11), 1283–1297 (2013)
18. Zheng, Z., Wu, X., Zhang, Y., Lyu, M.R., Wang, J.: QoS ranking prediction for cloud services. IEEE Trans. Parallel Distrib. Syst. **24**(6), 1213–1222 (2013)
19. Saripalli, P., Pingali, G.: Madmac: multiple attribute decision methodology for adoption of clouds. In: IEEE International Conference on Cloud Computing (CLOUD), pp. 316–323. IEEE (2011)
20. Gupta, R., Naqvi, S.K.: The fuzzy-AHP and fuzzy TOPSIS approaches to ERP selection: a comparative analysis. In: Handbook of Research on Fuzzy and Rough Set Theory in Organizational Decision Making, 188
21. Sundareswaran, S., Squicciarini, A. Lin, D.: A brokerage-based approach for cloud service selection. In: 2012 IEEE 5th International Conference on Cloud Computing (CLOUD), pp. 558–565. IEEE (2012)
22. Işıklar, G., Büyüközkan, G.: Using a multi-criteria decision making approach to evaluate mobile phone alternatives. Comput. Stand. Interfaces **29**(2), 265–274 (2007)
23. Wang, X., Cao, J., Xiang, Y.: Dynamic cloud service selection using an adaptive learning mechanism in multi-cloud computing. J. Syst. Softw. **100**, 195–210 (2015)
24. Badri, M.A.: A combined AHP–GP model for quality control systems. Int. J. Prod. Econ. **72** (1), 27–40 (2001)
25. Chan, F.T., Kumar, N.: Global supplier development considering risk factors using fuzzy extended AHP-based approach. Omega **35**(4), 417–431 (2007)
26. Dağdeviren, M., Yüksel, İ.: Developing a fuzzy analytic hierarchy process (AHP) model for behavior-based safety management. Inf. Sci. **178**(6), 1717–1733 (2008)

27. Albayrak, E., Erensal, Y.C.: Using analytic hierarchy process (AHP) to improve human performance: an application of multiple criteria decision making problem. J. Intell. Manuf. **15** (4), 491–503 (2004)
28. Ertuğrul, Í., Karakaşoğlu, N.: Fuzzy TOPSIS method for academic member selection in engineering faculty. In: Innovations in E-learning, Instruction Technology, Assessment, and Engineering Education, pp. 151–156. Springer Netherlands (2007)
29. Wang, Y.M., Elhag, T.M.: Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment. Expert Syst. Appl. **31**(2), 309–319 (2006)
30. Chen, C.T.: Extensions of the TOPSIS for group decision-making under fuzzy environment. Fuzzy Sets Syst. **114**(1), 1–9 (2000)
31. Chu, T.C., Lin, Y.C.: Improved extensions of the TOPSIS for group decisionmaking under fuzzy environment. J. Inf. Optim. Sci. **23**(2), 273–286 (2002)
32. Zadeh, L.A.: Fuzzy sets. Information and control **8**(3), 338–353 (1965)
33. Gupta, M.M., Saridis, G.N., Gaines, B.R. (eds.): Fuzzy automata and decision processes, vol. 77. North-Holland, New York (1977)
34. Pomerol, J.C., Barba-Romero, S.: Multicriterion decision in management: principles and practice, vol. 25. Springer Science & Business Media (2012)
35. Goyal, R.K., Kaushal, S., Sangaiah, A.K.: The utility based non-linear fuzzy AHP optimization model for network selection in heterogeneous wireless networks. Appl. Soft Comput. (2017). https://doi.org/10.1016/j.asoc.2017.05.026
36. Samuel, O.W., Asogbon, G.M., Sangaiah, A.K., Fang, P., Li, G.: An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. Expert Syst. Appl. **68**, 163–172 (2017)
37. Sangaiah, A.K., Samuel, O.W., Li, X., Abdel-Basset, M., Wang, H.: Towards an efficient risk assessment in software projects–Fuzzy reinforcement paradigm. Comput. Electr. Eng. (2017). https://doi.org/10.1016/j.compeleceng.2017.07.022
38. Singla, C., Kaushal, S.: Cloud path selection using fuzzy analytic hierarchy process for offloading in mobile cloud computing. In: 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS), pp. 1–5. IEEE (December 2015)
39. Dağdeviren, M., Yavuz, S., Kılınç, N.: Weapon selection using the AHP and TOPSIS methods under fuzzy environment. Expert Syst. Appl. **36**(4), 8143–8151 (2009)
40. Olson, D.L.: Comparison of weights in TOPSIS models. Math. Comput. Model. **40**(7–8), 721–727 (2004)

Cognitive Data Science Automatic Fraud Detection Solution, Based on Benford'S Law, Fuzzy Logic with Elements of Machine Learning

Goran Klepac

Abstract Developing fraud detection models always has been challenging area. Low frequency of fraudulent cases within data, indications instead of certainty contribute to very challenging area for data science method applying. Traditional approach of predictive modelling became insufficient, because relaying on few variables as a base of the fraud model are very fragile concept. Reason for that is fact that we are talking about portfolio with low cases of events, and from the other hand it is unrealistic to lean on few variables articulated through logistic regression, neural network or similar method that will be able to detect sophisticated try of fraudulent activities. Chapter gives proposal how to use data science in such situations where there are no solid bases but only potential suspicious regarding fraudulent activities. For those purposes Benford's law in combination with other data science methods and fuzzy logic will be used on sample data set, and will be shown potentials of proposed methodology for fraud detection purposes. Chapter shows case study in domain of finance on public data, where proposed methodology will be illustrated an efficient methodology which can be usable for fraud detection purposes.

Keywords Benford's law · Fuzzy expert system · Cognitive data science · Fraud detection · Machine learning

1 Background

The term "fraud" is commonly used for many forms of misconduct even though the legal definition of fraud is very specific. In the broadest sense, fraud can encompass any crime for gain that uses deception as a principal way of behavior. More specifically, "fraud" is defined as "a knowing representation of truth or concealment of a material fact to induce another to act to his or her detriment." Consequently, fraud includes any intentional or deliberate act to deprive another of property or money

G. Klepac (✉)
Raiffeisen Bank Austria, Magazinska 69, Zagreb, Croatia
e-mail: goran@goranklepac.com

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_4

by guile, deception, or other unfair means [5, 49, 50]. Fraud modeling starts when analysis relies on "critical thinking" skills to integrate the output of diverse methodologies into a cohesive actionable analysis product [35, 49, 50]. Fraud models are used for various purposes, depending on data/information types that are available and the type of analysis that is being performed. The analysis process requires the development and correlation of knowledge. In recent days more and more organizations are depending upon the most effective and efficient tools that can get the job done while new paradigms are introduced to increase efficiency of traditional approaches in fraud detection and management. Each industry adopts internal controls that offer the best protection against fraud and abuse. These safeguards may overlap throughout industries or be totally unique for the sector itself. Financial services industry has been monitoring its controls by incorporating security measures and analytic methods. Government uses internal auditors, along with retaining outsourced assistance to audit its various departments and implement controls. As manufacturers have become more automated, types of controls include inventory management programs, installation of surveillance equipment in plants and loading docks, GPS tracking on delivery trucks, and corporate charge card monitoring for business expenses. Retailers use surveillance, inventory security tags, perpetual inventory systems and sophisticated point-of-sale systems to track purchases and returns to thwart perpetrators. Insurance companies have updated their claims auditing systems to spot suspected instances of fraud. Even with these controls, fraudulent activity still becomes more and more complex, and hard to track. Trivial patterns that may be detected by the model, for instance similar to expert rules, are interesting as they provide some validation of the model. But of course, the key issue is to find the unknown yet interesting and actionable patterns (sometimes also referred to as knowledge diamonds) that can provide added insight and detection power. Once the analytical model has been appropriately validated and approved, it can be put into production as an analytics application (e.g., decision support system, scoring engine). Important to consider here is how to represent the model output in a user-friendly way, how to integrate it with other applications (e.g., detection and prevention system, risk engines), and how to make sure the analytical model can be appropriately monitored and back tested on an ongoing basis [5]. As fraud grows more sophisticated, to fight it company must step up efforts to protect good clients, uncover organized fraud and improve the effectiveness of analytics tools and specialized investigative units. But the most powerful approach in the fight against fraud may be a broad-based and strategic rethink of the overall business process, with a focus not just on the "what" is being done, but also on the "why" and the "how". While typical organization loses 5% of its annual revenues to occupational fraud, some industries are considerably more susceptible than others. Fraud obviously affects individual industries in different ways. For example, the retail industry may experience a greater number of individual fraud occurrences, but the average loss tends to be smaller than other industries. The common denominator is that everyone is susceptible to fraud while almost 40% of fraud cases in the most recent took place in private companies, almost 28% occurred in public offices, 17% in government agencies and 10% in non-profit organizations.

Developing fraud detection models are challenging area. Low frequency of fraudulent cases within data, indications instead of certainty contribute to that traditional approach of predictive modeling became insufficient, because relaying on few variables as a base of the fraud model are very fragile concept. Reason for that is fact that we are talking about portfolio with low cases of events, and from the other hand it is unrealistic to lean on few variables articulated through logistic regression, neural network or similar method that will be able to detect sophisticated try of fraudulent activities. In practice, proved fraud often does not exist in databases as record with binary attribute, or even it is not marked within databases. Sometimes we can talk about indications on fraud, which is not sufficient for predictive model development. This chapter provides new methodological approach by using fuzzy logic along with Benford's law [6] and other methods. It provides solid state for cognitive approach which is able to solve complex, not well defined problems which is challenging regarding their complexity and hidden behind data surface. Proposed methodologsy can be used as well in other business areas for finding deviant and illogical patterns. Synergy of enumerated elements gives solution for complex and challenging problems like fraud detection or finding deviations in business processes. Fraud analysis process is not a series of steps that are processed in a strict order; rather, the processes represent a methodology for accurate and concise analysis and information sharing [49].

As explained by Spann [49], the direction process involves establishing the boundaries of the analysis and what will be discovered during the process. This is also a step where analytical gaps and the effectiveness of the analysis are determined, and the significance of the analysis is established. Problem of low frequency occurrence of fraud event are well known and it is a problem. Alternative approaches than predictive modelling are preferred for fraud detection pattern searching [49] and this chapter gives proposal for such approach. Collection process involves gathering of raw data from which a finished analysis is produced. Collection process seeks to establish a criminal or fraudulent scope with a person or organization. Analysis process also seeks to collect information on trends, patterns, and methods of anomalies that help describe the phenomena of fraud event. Evaluation process involves the conversion of large amounts of data into a final analytical product. The process is done through a variety of methods to ascertain the most effective analysis, including decryption and data reduction. Evaluation includes entering raw data into databases (fraud analysis) where the data will be used in the analysis process. It includes recommendations, findings, and interpretation of information stored in the fraud summary reports, investigative reports, and similar reports. Finally, the collation/description process has four distinct stages in the fraud analytical process and include: evaluating raw data from the information gathered to detect its utility for analysis, examining the validity of raw data for cleanliness, clearly defining the analysis process in order to collect additional resources that will assist in gaining the most accurate raw information for robust analysis and utilizing other activities in the collation/description process. The analysis process is the heart of the methodology. It is essentially the approach to problem solving. It uses established methodologies that are qualitative and quantitative, that seek to integrate correlated variables in a section of raw data in

order to understand their meaning. Finally, the dissemination process is essentially an analytical product that has virtually no value unless the system is able to get the right information.

Idea is creation of cognitive system in conditions where we cannot count on periodical system calibration regarding small data sample and unexpected low frequent events. As in world of fraud detection system, in general we can talk about illogic, extremes; paradoxes proposed model will unite several methodological procedures which have task to find those appearances. For this purposes fuzzy logic, Benford's law and descriptive statistics will be used. We should not neglect fact, that algorithms could not prove fraud with hundred percent certainties, and proposed methodology, also cannot do that. Idea is to recognize few of potential suspicious cases from many of them, in situation where we are not dealing with target variables. Main task is to extract existing suspicious cases from population regarding recognized patterns. Hypothesis is that Benford's law along with descriptive statistics, machine learning algorithms united through Fuzzy expert system can make efficient mechanism for fraud pattern seeking.

2 Basic Tools for Constructing Cognitive Fraud Detection System

2.1 Role of Benford's Law

Benford's law is phenomenon discovered by Simon Newcomb and rediscovered by Benford, [6, 40, 44] which describes probability of the occurrence of the first digits of the number within some data sample. This law says that the probability that the first digit D equals d (which is probability of occurence) is given by [6]:

$$Pr\,(D = d) = \log\left(1 + \frac{1}{d}\right) \tag{1}$$

This empirically proved phenomenon is much more evident in non artificial systems where we are dealing with random events. It was proved in Benford's experiment. In recent period, especially after beginning of 90s Benford's law was renewed mostly as useful tool in area of fraud detection [8, 21, 24, 26, 41]. Purpose of this chapter is not rediscovering of Benford's law, but usage of Benford's law in combination with classical statistical methods and fuzzy expert systems for achieving cognitive automatic fraud detection solution. Fraud detection leans mostly on intuition and educated guess when we are talking of human experts which are investigate fraud. Challenge is to develop artificial system which will have "intuition" as a part of the process. Benford's law give one aspect of intuition, but it is not enough for complete system. Also, we can not claim that it is universally applicable in all situations. It depends on many factors, and Benford's law could not be universal factor

for judgement about fraud. It should be critically observed, taking in consider other factors to have clear picture and serious reasons for claiming that we are dealing with fraudulent cases. Benford's law is in that light, important but not only factor which can help us in determination about potential suspicious fraudulent cases.

2.2 Role of Statistics

As we previously talked about intuition which helps us to recognise fraudulent cases,statistical measures are important factor in acting intuition within artificial systems for fraud detection.

Mean, standard deviation, quartiles, percentiles applied on continuous variables gave an insight on existing attribute characteristics. Standard deviation is useful in finding outliers and extremes within attributes [27]. Extreme values in variables are often milestones for further investigation regarding fraud. Other important thing in data quality check is missing value analysis [32–34]. Missing value analysis gives information about missing values within attributes. It is not the universal rule that attributes with significant percentage of missing values is not usable for model development. Example for that is situation where client/buyer/contractor does not want to provide some piece of information and that information has great impact on aim variable, like providing residential phone number in fraud detection models. Useful technique in data quality check is attributes logical check [36]. This technique controls attribute values by using simple logic checks. Example could be check if working experience is higher than current year minus year of first employment. Another example could be checking how many people within data sample is older than 120 years, or does phone number attributes contains illegal characters. Role of basic statistic methods in fraud detection is important in many different aspects of usage [5, 18, 22–24, 46]. In light of achieving intuition, one of the statistical measure which can be very useful is standardized values. Story about fraud, are often story about extremes. That means if for example thief steal credit card he will try in short period of time to make as many as possible transaction with high limit usage. That is extreme behaviour, and it can be recognised by calculation of extremes and standardized values on client level. Client level means that each owner of credit card has unique behavioural characteristics and calculation of extremes on population level will not give us an answer. As in case of Benford's law taking in consideration only statistical measures are not sufficient for indication regarding fraud. If we make intersections between anomaly between extremes (or other statistical measure which is used taking in consideration specific fraud problem) and digits which significantly deviate from Benford's low we are on the trail to discover something potentially suspicious. This example is simplified idea for intuitive searching of suspicious potential fraud cases within data samples. This approach also needs decision engine which will manipulate with recognised anomalies and will be described later. Traditional approach which leans on machine learning approaches are oriented on usage of models like Bayesian networks, neural networks, decision trees and similar techniques. Problem

is that enumerated methods are not in line with cognitive approach, because they demands existence of target variable. Additional problem is that we are talking of low frequency events, and usage of machine learning algorithms are not efficient in that conditions. Machine learning algorithms could be used as a part of proposed solution, which are much more in line with concept of cognitive system, because it much more relays on intangible events.

2.3 Role of Social Network Metrics

Social metrics measures are powerful measures with which we are in position to discover hidden relations between nodes. Each network metrics plays specific role in discovering relations, as well as for important nodes recognition. During social network analysis stage, analysts are in position to combine different metrics with intention to reveal hidden knowledge [1, 2, 4]. Usage of social metrics combination is not always an easy task, especially in situation where we would like to create filters for network structure based on social network metrics [14]. Interpretation of such defined social network metrics combination could be hard as well as ensuring periodical analysis on same filters.

In conditions where we have social network data; it is valuable source for cognitive computing in light of fraud detection [1–4, 7, 10]. Concentrations of influences within nodes with some detected anomalies can be on help for automated fraud investigation. For example if we have some suspicious activities in multimodal network on node level, where suspicious activities are recognised as frequent events associated with Benford's law or/and extremes and also high value of social network metrics like eigenvector this nodes are for sure worth of further deeper investigation. Social network analysis can be applied only in situation where we are dealing with social network data, and it contributes to higher accuracy in fraud detection. From that perspective, social network analysis contributes also in synergy of different elements which provides solution for complex and challenging problems [11, 12, 14–17]. Usage of social network metrics in fraud is well known in practice and literature. From perspective of proposed solution it is useful element which contribute to cognitive fraud detection solution along with Benford's law, statistical measures, united within fuzzy expert system as holistic system for fraud detection. Generally speaking each social network metrics like degree centrality, betweenes, eigenvector and other measures has great potential as itself in fraud detection. Enumerated metrics by itself can be good milestones for deeper investigation of fraud if we are talking about fraud detection models. As it is case with other mentioned techniques, grounding hypothesis on single, or limited number of techniques or methods are not sufficient, especially if we are dealing with cognitive fraud detection systems. Social network data has great potentials not only for fraud detection, but also for other areas like churn modelling, customer relationship models and other different type of problems [19, 20, 25, 28]. Sometimes it is challenge to make fusion between social network data with other types of data like classical relational data because of different

data structures. In that case pre-processing and ETL processes should be carefully planned. Unfortunately we cannot always count of social network data as integral part of data sample, created for analytical purposes. In case where it is on disposal, it provides great opportunities for building better cognitive systems for fraud detection purposes [30, 31, 38, 39, 43, 45, 47, 48, 52]. Information derived from social networks along with usage of Benford's law, statistical measures contribute to idea of developing cognitive fraud detection system. All mentioned elements provide soft, intuitive information typical for humans when they are faced with similar types of problems. In conditions where we are in situation for seeking patterns without solid benchmark like target variable, where seeking patterns are much more undefined concepts than unsupervised learning process, it is logical that we lean on techniques and its combination which are in line with such type of problems.

3 Automation of the Pattern Seeking by Usage of Fuzzy Logic

Main purpose of the expert fraud models was recognition of suspicious activities on individual client level. Traditional approach, which leans on predictive models as a base for fraud detection models, could not be insufficient. Main reason for that lays in fact that predictive model contains few most predictive attributes as integral part of predictive models. Reasons why they make predictive model imply fact that those attributes shows highest impact on aim variable. That impact by traditional methodology is measured using attribute relevance analysis. Criteria for highest impact on aim variable is statistical significance, and those fact hides pitfall, because if some trend became so obvious that it has statistical significance it is doubtful is it appropriate for early warning sign. That means that some deviant trend has happened during longest period of time and it makes statistically significant data pattern recognizable thought attribute relevance analysis. If some trend or event happens on individual level, and it is fraudulent activity, it is impossible to recognize it with traditional statistical predictive models. For basic trend recognition, and fraudulent pattern recognition which has mass characteristic those methodology is good enough [13, 29], but for early complete fraud detection solution it is not sufficient. It does not mean that predictive models should not be used for fraud modelling; it only means that predictive models should not be only element or base for fraud detection systems. Fuzzy expert systems gives power to fraud detection models to recognize potential suspicions activities based on human expert knowledge which is integrated within automatic solution. Step forward is to improve fuzzy expert systems in way to combine elements like Benford's law with statistical measurements for finding illogical patterns which indicates on suspicious activities. In that way some potential low frequent and unexpected patterns which indicate on fraud can be recognized. This approach gives opportunity for finding non explicit fraud patterns without designing explicit rules for each potential fraudulent case.

A fuzzy set is an extension of a classical set. If X is the set of values and its elements are denoted by x, then a fuzzy set A in X is defined as a set of ordered pairs by flowing formula [9, 20, 25, 37, 42, 51].

$$A = \left\{x, \mu_A(x) | x \in X\right\};, \tag{2}$$

Membership function can be defined for anomalies in digits calculated by usage of Benford's numbers, anomalies for statistical measures, social network metrics. Each variable, can be expressed as linguistic variable and expressed through definition visible in Fig. 1.

Part of linguistic variable definition can be definition on anomaly calculated by usage of Benford's numbers and statistical measures, as well as social network metrics. Such defined linguistic variables can be integrated (combined) through fuzzy rule blocks. In such situation social network metrics as well as other measures can be structured in new categories which describes some new characteristics (Table 1).

Flowing table shows methodology of rule block definition.

As it is visible from the created rule block, it is possible to combine social network metrics into complex structures within one rule block. Connected rule blocks makes fuzzy expert system. Fuzzy expert system can contain linguistic variables based on social network metrics as well as non social network variables. Fuzzy expert system gives opportunity for complex influence calculation constructed from different social network metrics. Such constructed expert system is periodically usable which

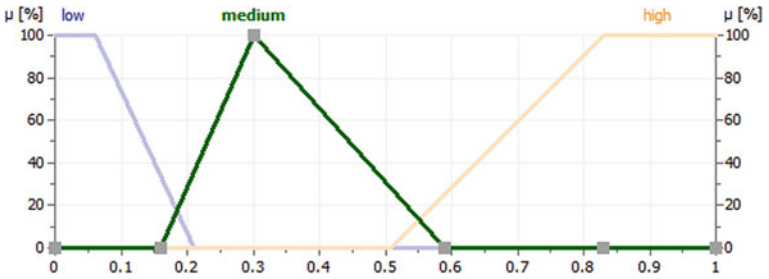


Fig. 1 Example of linguistic variable definition on anomaly calculated by usage of Benford's numbers

Table 1 Fuzzy rules definition

Rule number	IF Benfrod's anomaly	AND anomaly	Extreme THEN suspicious
Rule 1	Low	Low	Low
Rule 2	Medium	Low	Low
Rule 3	High	Medium	Medium
Rule 4	Low	High	Medium

is interesting for social metrics trend observation, even it is complex and constructed with many of them.

Fuzzy expert systems has main role in integration of whole concept of fraud detection pattern seeking. It unites Benford's law, statistical methods, and can unite machine learning elements. Each linguistic variable can contain resulting elements of machine learning algorithm, Benford's law elements, statistical measure connected via set of rules. It is the framework of proposed model for fraud detection.

4 Applying Model on Empirical Data

Generally, holistic solution takes in consideration social network data along with applying Benford's law on data as well as usage of different statistical measures integrated within platform of fuzzy expert system. In this illustrative example on public data Benford's law and standardized values calculated from data sample will be used as illustration of proposed method. Data science is a multidisciplinary area and this example illustrates basics for cognitive solution in domain of fraud detection in which different methodology has been used together for solving specific problem.

4.1 Problem Description

For illustration how proposed methodology can be applied on empirical data, demo data set from which is distributed with SPSS Statistic programming package has been used. This demo set contains 850 applicants of bank loan with attributes which applicant for loan should provide to bank with default flag, which describes was client able in past to service his debt or not. These data sets contains mainly socio-demographic variables like age, level of education, years with current employer, years at current address, household income, debt to income ratio, credit card debt, other debts and default flag. This data set is commonly used for developing predictive models like application scorecards or predictive models for default prediction. Data set does not contain any attribute or data related to fraud. It is usual situation in practice. Our task is to find some hidden patterns which can be indicative regarding fraud activities. For this purposes we will use Benford's law and calculation of statistical extremes. This element should lead us to some suspicious patterns recognition. At the end, those elements can be united within fuzzy expert system which will manage and evaluate suspicious degree. With this step we can get cognitive, universally applied system for any kind of data in which we try to recognise elements on fraud. In that way, we can simulate intuition based on educated guess within data sample. Data contains mostly numerical values, which were used for analysis. Data set do not contain missing data.

4.2 Applying Benford's Law on Data

First step in solution is applying Benford's law on each numerical data within sample. Most interesting attribute is income, because income is in direct connection with loan approval process. This fact is well known to applicant and Bank. Because of that it is variable which should be investigated. It does not mean that other variables should not be observed, contrary, but higher focus should be put on most critical variables which are directly related to credit policies. Variables like years with current employer, years at current address, debt to income ratio, credit card debt, and other debts can also provide useful information regarding potential risk and should be investigated. As it was previously mentioned, Benford's law is not magical method to whom we should trust blindly. It is useful tool for testing some hypothesis which should be challenged, and in this light Benford's law has been used as a first step of analysis. For this purposes Python programming language was uses along with his libraries numpy, pandas, matplotlib, as it is visible in following code.

```
import numpy as np
import pandas as pd
from pandas import DataFrame
import matplotlib as plt

path='C:\Goran\data.csv'
table = pd.read_csv(path, sep=';')

table['B1'] = table.income

table.B1 = table.B1.astype(str)

table['B2'] = table['B1'].str[:1]
table['BL'] = table['B2']
table['BL'] = table.BL.astype(int)
table['Log']=np.log10(1+(1/table['BL']))*100

agr=table.groupby(['B2']).count()
agr ['Benford']=(agr ['B1']/agr ['B1'].sum())*100

table['Z_credcdebt'] = (table.credcdebt -table.credcdebt.mean())/
table. credcdebt.std()

table['Z_othdebt'] = (table. othdebt -table. othdebt.mean())/
table.othdebt.std()

import matplotlib.pyplot as p
table[['Z']].plot(kind='bar')

agr [['Benford']].plot(kind='bar')
```

		Count	Column N %
DIGIT	1	146	17,2%
	2	252	29,6%
	3	149	17,5%
	4	111	13,1%
	5	64	7,5%
	6	54	6,4%
	7	40	4,7%
	8	21	2,5%
	9	13	1,5%

Fig. 2 Occurrence of first digits in sample

Code gives solution for calculating digit occurrence frequency within data, Benford's law calculation and extreme calculation, which is base for planned analysis. Most interesting results we got for variable household income. It is partially understandable because this variable has highest impact on loan approval. First digit frequency is visible in Fig. 2.

As it is visible highest discrepancy is for digits one and two. It still does not mean anything, just a clue, worth of further investigation. Fastest way to prove some suspicious activities is to compare occurrence of first digits with highest discrepancy with defaulted cases. In that way we can have milestone that default is maybe driven by some other unknown or unwanted factors, not only by pure risk. Results are visible on Fig. 3.

If we compare first digit frequency with expectable first digit frequency calculated by Benford's law we can see zones of anomaly in Fig. 3.

As it is visible highest discrepancy is for digits one and two. It still does not mean anything, just a clue, worth of further investigation. Fastest way to prove some suspicious activities is to compare occurrence of first digits with highest discrepancy with defaulted cases. In that way we can have milestone that default is maybe driven by some other unknown or unwanted factors, not only by pure risk. Results are visible on Fig. 4.

After conducting analysis it is visible high concentration in zones with anomaly based on Benford's law. It gives us motivation for further investigation. Obviously there is a hint that something happened in those zones which should be investigated in much more details. Regarding this example, it is evident that Benford's law does not provide answers, it opens new questions and it is his main function. By opening new right questions we are moving closer to the settled aims. For getting some of the answers we should introduce some new methods on existing methodology.

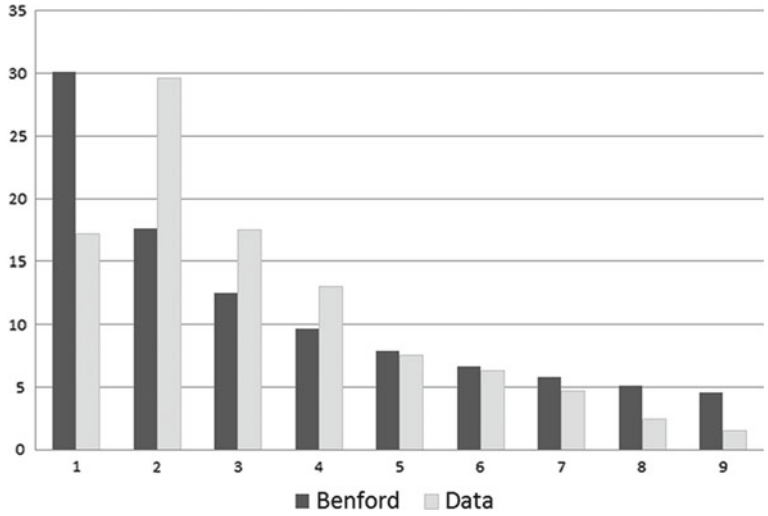


Fig. 3 Comparison of occurrence of first digits in sample with Benford's law

		Previously defaulted			
		No		Yes	
		Count	Column N %	Count	Column N %
DIGIT	1	83	16,1%	38	20,8%
	2	143	27,7%	65	35,5%
	3	104	20,1%	24	13,1%
	4	60	11,6%	24	13,1%
	5	43	8,3%	13	7,1%
	6	38	7,4%	6	3,3%
	7	24	4,6%	8	4,4%
	8	15	2,9%	3	1,6%
	9	7	1,4%	2	1,1%

Fig. 4 Default by occurrence of first digits in sample

4.3 Calculation of Extremes

New methods which can give us answers are statistical methods. There are numerous ways how to use it and which statistical measures to use taking in consideration problem space, aim of analysis, nature of the sample etc. Generally we decide to use extreme calculation and standardized value calculation by formula.

$$z_i = \frac{x_i - \bar{x}}{\sigma} \tag{3}$$

		Zscore: Creditcard debt in thousands (Binned)					
		<= -3,00000		-2,99999 - 3,00000		3,00001+	
		Count	Column N %	Count	Column N %	Count	Column N %
DIGIT	1	0	0,0%	140	16,8%	6	37,5%
	2	0	0,0%	249	29,9%	3	18,8%
	3	0	0,0%	149	17,9%	0	0,0%
	4	0	0,0%	110	13,2%	1	6,3%
	5	0	0,0%	64	7,7%	0	0,0%
	6	0	0,0%	52	6,2%	2	12,5%
	7	0	0,0%	37	4,4%	3	18,8%
	8	0	0,0%	21	2,5%	0	0,0%
	9	0	0,0%	12	1,4%	1	6,3%

Fig. 5 Comparison of credit card debt extremes with occurrence of first digits

Also, naïve Bayesian classificators can also be used as well as other measures. Reason why we choose extreme calculation and standardized value calculation are driven by fact, as it is already explained that story about fraud is mostly story about extremes. On those demo data set short with attributes we will try to put in relation anomaly od first digits by Benford's law with extremes from some attributes. From the list of attributes it is potentially possible to do that for all variables, but most interesting variables for investigation are those one potentially related to credit policy: credit card debt and other debt. Hypothesis is that we potentially have internal fraud if officer did not respect internal regulations regarding credit policy of institution. In other word, someone with extreme high debts got loan, which is in potential discrepancy with credit policy of institution. Figure 5 show that hypothesis is proved regarding credit card debt. Highest concentration of defaults is in zone of leading digits for income one and two where we have also concentration of extreme values of credit card debt.

Figure 5 show that hypothesis is proved regarding credit card debt. Highest concentration of defaults is in zone of leading digits for income one and two where we have also concentration of extreme values of credit card debt.

That leads us to conclusion of potential breaching of credit policy.

Same is for other debts, like it is visible from Fig. 6.

4.4 Usage of Fuzzy Expert System

Fuzzy expert system plays important role in unification of different analytical approaches in way that within system of rules it unites different concepts which cooperate with one aim- finding suspicious patterns. Same is for presented empirical study. Expert system can contain linguistic variables dedicated to Bedford's law where terms can define low, medium or high discrepancy on attribute level. Other

		Zscore: Other debt in thousands (Binned)					
		<= -3,00000		-2,99999 - 3,00000		3,00001+	
		Count	Column N %	Count	Column N %	Count	Column N %
DIGIT	1	0	0,0%	139	16,7%	7	43,8%
	2	0	0,0%	249	29,9%	3	18,8%
	3	0	0,0%	148	17,7%	1	6,3%
	4	0	0,0%	111	13,3%	0	0,0%
	5	0	0,0%	64	7,7%	0	0,0%
	6	0	0,0%	52	6,2%	2	12,5%
	7	0	0,0%	39	4,7%	1	6,3%
	8	0	0,0%	20	2,4%	1	6,3%
	9	0	0,0%	12	1,4%	1	6,3%

Fig. 6 Comparison of other debts extremes with occurrence of first digits

group of linguistic variables can be dedicated to extremes, correlation, distribution flattening and other statistical measures, also declared in manner of linguistic variables like low, medium or high. Another group of linguistic variables can be dedicated to social network metrics if we are dealing with social network data, also declared in manner of linguistic variables like low, medium or high. Within fuzzy expert system we can declare rules like: IF high discrepancy (Benford's Law) for income in zone of first digit 1 AND standardized value for other debts greater than 3 THEN suspicious is HIGH. IF high discrepancy (Benford's Law) for income in zone of first digit 2 AND standardized value for other debts greater than 3 THEN suspicious is HIGH. IF high discrepancy (Benford's Law) for income in zone of first digit 1 AND standardized value for credit card debt debts greater than 3 THEN suspicious is HIGH. . IF high discrepancy (Benford's Law) for income in zone of first digit 2 AND standardized value for credit card debt debts greater than 3 THEN suspicious is HIGH ... ... In that way fuzzy expert system automatize process of seeking suspicious patterns, and makes graduation by suspicious grade. Fuzzy expert system can be designed for particular attributes or combination of attributes on which we can apply proposed methodology. Dependent on problem space, system designer can choose critical attributes which are sensitive regarding fraud and combine them within fuzzy expert system.

4.5 Results from Empirical Data

Empirical research has been done on demo data set which is distributed with SPSS Statistic programming package. It shows in illustrative way how to apply proposed methodology. Generally observed data set on first sight and even on deeper investigation does not contain suspicious cases. Proposed methodology applied on this data sample proved that there is potential reason for doubts that existing data set

contains potentially fraudulent cases. Purpose of the proposed methodology is not explicit declaration of potential fraudulent cases; it only filters out illogical patterns regarding proposed methodology which should be checked by the experts. It is common practice in fraud detection, to recognise narrow scope of cases which should be investigated because of some reasons. Without proposed methodology illogical cases can remain hidden under data surface, and proposed methodology helps to recognise suspicious patterns. Fuzzy expert systems articulate results from different methods and approaches into single cognitive solution which articulate expert knowledge and gives power to each single output to be recognised thorough rule blocks. In that way, we can have automatic, artificial system which is able to seek hidden patterns related to fraud activities and to give an expertise of potentially suspicious cases within data sample. This system can also be improved by introducing elements like social network analysis, or similar elements. Each of used elements at the end contribute in pattern seeking through fuzzy expert system.

5 Conclusion

Proposed methodology gives solution for automatic seeking patterns within data with focus on fraud detection. As it was already mentioned, fraud detection is one of most demanding area in date science area, especially when we are trying to construct system for automatic pattern searching which will be able to recognise suspicious activities. In difference of churn modelling, recommendation system, credit scoring and similar systems, fraud detection modelling is much more complex for automatic solution development. Lack of target variables, hidden infrequent patterns, indications instead of certainty contribute to challenges in this area when we are talking about automatic fraud detection solutions. Solution on empirical data proves efficiency of proposed methodology even with limited usage of methods. This approach simulate human process of educated guess. Cognitive methods like proposed methodology have task to be familiar with human perceptive mechanism with ability to solve specific problems. Usage of some common methods in different ways, stand alone or as integral part of fuzzy expert system, contribute to idea of construction cognitive data science solution for fraud detection. Future research can be focused on integration of new concepts like stability index, social network analysis, additional machine learning algorithms integration into existing infrastructure for achieving better performance in fraud detection patten recognition. Future research can also be focused on finding optimal combination of different techniques and algorithms integrated within proposed architecture for different areas like telecommunication, retail, finance and other business areas.

References

1. Abraham, A., Hassanien, A.-E.: Computational Social Network Analysis Trends. In: Tools and Research Advances. Springer, London, UK (2010)
2. Aharony, N., Pan, W., Cory, I., Khayal, I., Pentland, A.: Social fMRI: investigating and shaping social mechanisms in the real world. Pervasive Mobile Comput. **7**(6), str. 643–659 (2011)
3. Akoglu, L., Vaz de Melo, P.O.S., Faloutsos, C.: Quantifying reciprocity in large weighted communication networks. In: PAKDD 2, Lecture Notes in Computer Science, vol. 7302, pp. 85–96. Springer, str, Berlin, Heidelberg (2012)
4. Altshuler, Y., Pan, W., Pentland, A.: Trends Prediction Using Social Diffusion Models. In: Social Computing, Behavioral-Cultural Modeling and Prediction: Lecture Notes in Computer Science series. vol. 97, p. 104 Springer, str, Berlin, Heidelberg (2012)
5. Baesens, B., van Vlasselaer, V., Verbeke, W.: Fraud Analytics Using Descriptive, Predictive, and Social Network (2015) Techniques, Wiley
6. Benford, F.: The law of anomalous numbers. Proc. Am. Philos. Soc. **78**, 551–572 (1938)
7. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
8. Bolton, R.J., Hand, D.J.: statistical fraud detection: a review. Stat. Sci. **17**(3) 235–249 (2002)
9. Bojadziev, G., Bojadziev, M.: Fuzzy Logic for Business, Finance, and Management, 2nd edn. World Scientific Publishing Co., Inc, River Edge, NJ, USA (2007)
10. Carrington, P.J., Scott, J., Wasserman, S. (eds.): Models and Methods in Social Network Analysis, pp. 248–249. Cambridge University Press, Cambridge (2005)
11. Coser, L.A.: Masters of Sociological Thought: Ideas in Historical and Social Context, 2nd edn. Harcourt, New York, NY (1977)
12. D'Agostini G.D.: Bayesian Reasoning in Data Analysis: A Critical Introduction. World Scientific, New York (2003)
13. Duffield, G., Grabosky, P.: The psychology of fraud. In: Trends and Issues in Crime and Criminal Justice. Australian Institute of Criminology, vol. 199 (2001)
14. Easley, D., Kleinberg, J.: Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge (2010) University Press
15. Erdös, P., Rényi, A.: On random graphs. Publ. Mat. Debrecen **6** (1959)
16. Erdös, P., Rényi, A.: On the Evolution of Random Graphs. Publ. Math. Inst. Hung. Acad. Sci. **5** (1960)
17. Erdös, P., Rényi, A.: On the strength of connectedness of a random graph. Acta. Math. Acad, Sci. Hungar **12** (1961)
18. Fawcett, T., Provost, F.: Adaptive Fraud Detection. Data Min. Knowl. Discov. **13**(3), 291–316 (1997)
19. Freeman, L.C.: The Development of Social Network Analysis: A Study in The Sociology of Science. Empirical Press, Vancouver, BC (2004)
20. Fuller, R., Carlsson, C.: Fuzzy Reasoning in Decision Making and Optimization. Physica-Verlag, Heidelberg, Germany (2002)
21. Giles, D.E.: Benford's law and naturally occurring prices in certain ebay auctions. App. Econ. Lett. **14**, 157–161 (2007)
22. Grabosky, P., Duffield, G.: Red flags of fraud. In: Trends and Issues in Crime and Criminal Justice. Australian Institute of Criminology (2001)
23. Hand, D.: Statistical Techniques for Fraud Detection, Prevention, and Evaluation. Paper presented at the NATO ASI: Mining Massive Data sets for Security, London, England Sept 2007
24. Hill, T.P.: The significant-digit phenomenon. Am. Math. Monthly **102**, 322–327 (1995)
25. Hampel, R., Wagenknecht, M., Chaker, N. (eds.): Fuzzy Control: Theory and Practice. Physica-Verlag, Heidelberg, Germany (2000)
26. Jamain, A.: Benford's Law. Imperial College, London (2001)
27. Jaynes, E.T.: Probability theory. In: The logic of science. Cambridge University Press, Cambridge (2003)
28. Jackson, M.O.: Social and Economic Networks. Princeton University Press, Princeton, NJ (2010)

29. Jennings, C.R., Poston, R.J.: Global Business Fraud and the Law: Preventing and Remedying Fraud and Corruption (May 2016 Edition). Practising Law Institute (PLI), Amazon Digital Services LLC (2016)
30. Jensen, F., Nielsen, T.: Bayesian Networks and Decision Graphs. Springer, New York (2007)
31. Kjarluff, U., Madsen, A.: Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis. Springer, New York (2013)
32. Klepac, G., Kopal, R., Mršić, L.: Developing Churn Models Using Data Mining Techniques and Social Network Analysis, pp. 1–308. IGI Global, Hershey, PA (2015). https://doi.org/10.4018/978-1-4666-6288-9
33. Klepac, G.: Handbook of research on advanced hybrid intelligent techniques and applications. In: Bhattacharyya, S., Banerjee, P., Majumdar, D., Dutta, P. (eds.) Discovering Behavioural Patterns within Customer Population by using Temporal Data Subsets, Hershey, PA, p. 321–348 (2016)
34. Klepac, G., Kopal, R., Mršić, L.: Hybrid soft computing approaches. In: Bhattacharyya, S., Dutta, P., Chakraborty, S. (eds.) REFII Model as a Base for Data Mining Techniques Hybridization with Purpose of Time Series Pattern Recognition, p. 237–270. Springer India (2015)
35. Krambia, K.M.: Corporate Fraud and Corruption: A Holistic Approach to Preventing Financial Crises. Palgrave Macmillan (2016). ISBN 978-1349680818
36. Lauritzen, S.L., Nilsson, D.: Representing and solving decision problems with limited information. Manage. Sci. **47**, 1238–1251 (2001)
37. Leonides, C.: Fuzzy Logic and Expert Systems Applications. Academic Press, New York (1998)
38. Milgram, S.: The small-world problem. Psychol. Today **1**(1), 61–67 (1967)
39. Moreno, J.L.: Sociometry, Experimental Method, and the Science of Society. Beacon House, Ambler, PA (1951)
40. Newcomb, S.: Note on the frequency of use of the different digits in natural numbers. Am. J. Math. **4**, 39–40 (1881)
41. Nigrini, M.: A taxpayer compliance application of Benford's Law. J. Am. Tax. Assoc. **18**, 72–91 (1996)
42. Pedrycz, W., Gomide, F.: Fuzzy Systems Engineering: Toward Human-Centric Computing. Wiley-IEEE Press (2007)
43. Pinheiro, C.A.R.: Social Network Analysis in Telecommunications. Wiley, Hoboken, NJ (2011)
44. Raimi, R.A.: The first digit phenomenon. Am. Math. Monthly **83**, 521–538 (1976)
45. Remer, R.: Chaos theory links to morenean theory: a synergistic relationship. J. Group Psychother. Psychodrama Sociom. **59** (2006)
46. Roselle, B.E.: The Fraud Factor: Recognize It. Overcome It. Leader Press (2016). ISBN 978-0978564629
47. Simmel, G.: How is society possible? In: Levine, D. (ed.) On Individuality and Social Forms Univevrsity of Chicago Press, Chicago, IL (1908/1971)
48. Scott, J.: Social Network Analysis: A Handbook. Sage Publications, London (1987)
49. Spann, D.D.: Fraud Analytics: Strategies and Methods for Detection and Prevention, Wiley (2014)
50. Ward, J., Peppard, J.: The Strategic Management of Information Systems: Building a Digital Strategy. Wiley (2016). ISBN 978-0470034675
51. Zadeh, L.A., Kacprzyk, J. (eds.): Fuzzy Logic for the Management of Uncertainty. Wiley, New York (1992)
52. Zhang, M.: Social network analysis: history, concepts, and research. In: Fuhrt, B. (ed.) Handbook of Social Network Technologies and Applications, pp. 3–22. Springer, New York, NY (2010)

Reliable Cross Layer Design for E-Health Applications—IoT Perspective

P. Sarwesh, N. Shekar V. Shet and K. Chandrasekaran

Abstract Of late, there has been many applications are developed by the aid of IoT technology, such as smart city, e-health, smart home, industrial automation etc. In that, e-health is one of the efficient idea that is decidedly developed for healthcare sectors. IoT devices used in e-health applications are run by battery powered smart objects and low frequency links, which says energy constrained and unreliable nature of IoT network. Thus, providing potent healthcare service (regularly following and reporting the patients' health information) in energy constrained network environment (battery power smart objects and low frequency links) is the prime need in resource constrained networks environment. In this chapter, reliable cross layer design is introduced to prolong the lifetime of IoT devices and reliable data transfer in IoT e-health applications. In proposed cross layer model, network layer and data link layer (MAC layer) are integrated. Reliability related parameter are included in route discover process and later MAC based power control technique make use of routing information, to obtain the suitable transmission power. Our results show that proposed cross layer design is reliable and energy efficient and it is more suitable for IoT e-health applications.

1 Introduction

Internet of Things (IoT) is a network of internet enabled devices that promotes natural resource usage in efficient way [1]. IoT technology plays major role in healthcare service, e-health is highly encouraged by health care sectors. In e-health

P. Sarwesh (✉) · N. S. V. Shet · K. Chandrasekaran
National Institute of Technology Karnataka, Mangaluru, Karnataka, India
e-mail: sarweshpj@gmail.com

N. S. V. Shet
e-mail: shekar_shet@yahoo.com

K. Chandrasekaran
e-mail: kchnitk@gmail.com

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_5

applications, bio-devices (Bio-sensor) are enabled by features such as medical signal sensing and conversion, communication and computation. Bio-devices are the prim component in health care sector, they monitor the patient's health status and they report to medical supervisors. Wearable bio-devices observes the change in behavior of human body and give health care instruction at right time, wearable devices aid the elder people in effectively [2]. Smart devices or Bio-devices used in e-health applications are battery powered and they are connected by low bandwidth radio links. Quality of Low power links changes dynamically, they are prone to error [3]. Health care industry deal with critical data (patients' health information), such scenario need reliable data transfer. Thus, reliability is considered as main objective in e-health applications. In low power wireless network, multi re-transmissions is the major issue that degrades the reliability of the network. Most of the IoT applications (including e-health) uses low power radio links, that are loss and unstable in nature, therefore data transfer through lossy links leads to packet loss and re-transmissions [4]. Obtaining the reliable link information and transmitting data through reliable link can be effective solution for re-transmission issue. Hence, we proposed reliable cross layer design that is suitable for e-health applications.

1.1 IoT Challenges

IoT technology is practiced in many applications such as, health care sectors, power distributions, environmental monitoring applications, industrial sectors etc. In recent decade, it is adopted by private as well as public sectors. IoT world forum reference model (introduced in CISCO White paper) [2] describes the clear view of IoT network architecture, it consists of seven layers, (Query based data processing layers, data storage layer and event based data processing layers). In Fig. 1, lower layers (real time data processing layers) are culpable for collecting real time data, higher layers (Non real time data processing layers) are culpable for collecting and responding for query based data processing.

Devices handled in higher layers are main powered devices, which consumes there required power from main power lines. The major requirement in these layers are efficient data management, since it deals with huge source of data. Where in event based data processing layer, edge devices are used to collect the environmental information, which consumes their power from battery source. Thus, power utilization need to be taken care in these layers. Figure 1, describes the features of IoT world forum reference model and challenges in various layers of IoT network architecture. In low power wireless applications, battery replacement affects the services critically. In critical monitoring applications, reliable data transfer is more important.

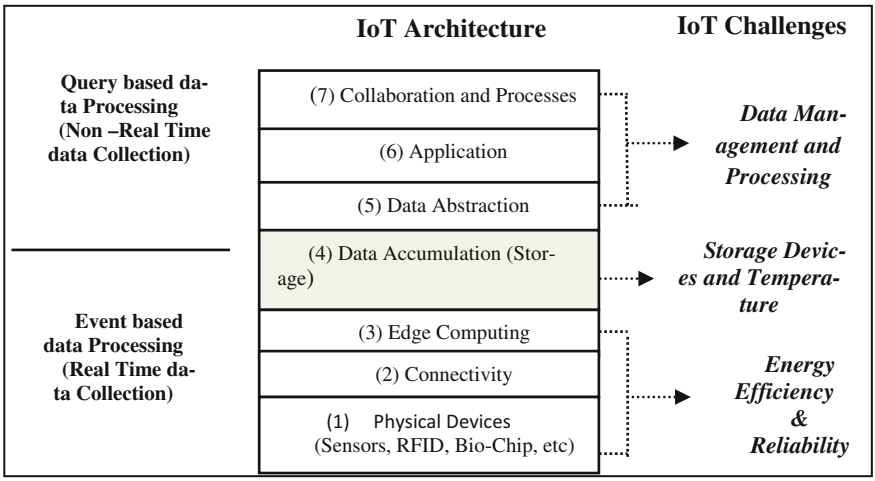


Fig. 1 IoT world forum reference model

1.2 Importance of Reliability and Network Lifetime in Low Power IoT Networks

In Many IoT applications, battery source device is used and they are connected by low bandwidth links, which are lossy in nature and link quality in low power network environment changes dynamically. Thus, energy efficiency and reliability are the major requirements in low power IoT networks. Physical size, cost and power consumption are the major challenges while designing a smart device. MEMS technology is the major reason for low power, low cost and intelligent smart device development. Smart devices work autonomously with its basic capabilities. It is intelligent in nature but resource constrained in nature. In environmental monitoring applications (ex forest fire monitoring) sensor devices will be placed in remote area, when a node runs out its battery, battery replacement is difficult. In commercial applications, frequently battery replacement affects the customer service severely. Hence, efficient energy utilization is the major requirement in IoT networks. Efficient energy utilization highly increases the lifetime of low power wireless networks.

Table 1 describes the characteristic of IoT network and Internet. IoT networks are operated by battery sourced devices and low band width links. low band width links are dynamic in nature. Where in Internet stable links (high power links) are used to connect devices, which are highly stable in nature. Thus, IoT network highly differs from regular Internet. IoT devices are limited by energy, so using low band width links is the only option in IoT network. But low band width links are lossy in nature. Low power radio is characterized by more loss rates. In IoT network, packet loss occurs frequently, due to many reasons, such as interference, weak signal strength, etc. Packet loss leads to multi re-transmission, which severely affects the energy

Table 1 Features of internet and internet of things network

Features	Internet	IoT network
Devices	Higher end devices	Sensor devices
Radio	Stable links	Unstable links
Device	Non-constrained device	Constrained device
Identification	IPv6 address	IPv6 address
Energy source	Main power line	Battery sourced

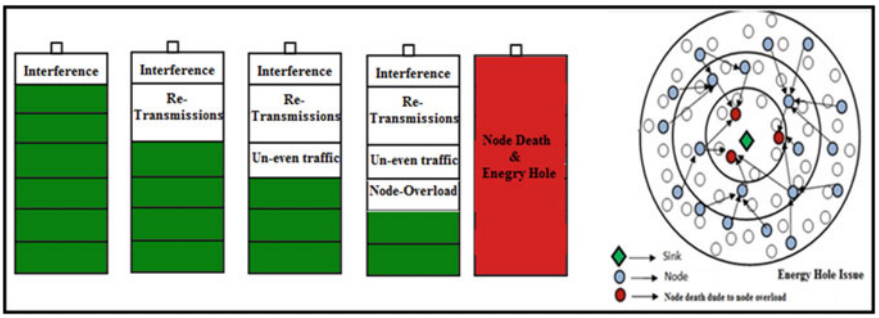


Fig. 2 Factors that affects network performance

efficiency and reliability of the network. The major solution to avoid packet loss is transmitting data in reliable links. Figure 2, describes the energy hole issue and multi retransmission issue. Thus in proposed cross layer model, we concentrated on reliability based parameter to prevent multi re-transmissions.

1.3 Motivation

The aforementioned discussions convey the importance of reliability in IoT e-health applications. Preventing multi re-transmissions decidedly improves the energy efficiency and reliability of the network. In order to prevent the re-transmissions, communication unit need to be regulated in efficient way. Because, in any kind of network architecture, communication unit consumes huge amount of energy (nearly 70–80% of battery power), when compared to other units. Thus, regulating the working process of communication unit decidedly improves the network performance. Communication unit can be optimized by achieving suitable transmission and reception power of nodes. To obtain optimal transmission range for every node, features of various layers need to be integrated and utilized in efficient way. Hence, cross layer design (information exchange and regulating transmission power) can be

a effective solution to achieve optimal transmission range in low power IoT networks.

1.4 Cross Layer Design

Many optimization techniques are introduced to enhance the performance of low power wireless networks. Among them, cross layer design is one of the effective technique that integrates the features of various optimization techniques (routing, power control technique, TCP/IP technique etc.) in single network architecture to satisfy the network requirements. Cross layer design differs from each other based on the layer interfacing. Interfacing for cross layer design can be done in following way [5].

- Upward flow: Information exchange data from lower layers to upper layers [6].
- Downward flow: Data exchange from upper layers to lower layers [6, 7].
- Back and forth flow: Information exchange from lower layer to upper layer as well as from upper layer to lower layer [8, 9]
- Merging two adjacent layers: Merging the features of two different layers without interfacing [8].
- Design coupling without interface: Coupling layers without creating extra interfaces [10].
- Vertical calibration across layers: Design parameters are adjusted that span across layers [11].

In proposed cross layer model, physical layer, MAC (data link) layer and network layer are involved. In this model, reliability related information called Expected Transmission Time (ETT) is included in path computation process. Later, MAC based power control technique uses ETT routing information to obtain the optimum transmission power. AODV [12] routing protocol is used for route discovery process and IEEE 802.15.4 PHY and MAC [13] protocol is used as physical layer and MAC layer. The proposed cross layer design addresses, multi-retransmission issue. Hence, it improves the reliability in IoT e-health networks. In this chapter, Sect. 2 describes the literature survey, Sect. 3 elaborates proposed cross layer design, which includes operational model and routing mechanism. In Sect. 4 results are evaluated and Conclusion is described in Sect. 5.

2 Background

Many research works [14–16] have been carried out in cross layer design. Most of the cross layer models are designed to satisfy specific network requirements. Madan et al. [17] proposed iterative algorithm based on efficient link scheduling, to prolong the lifetime of low power devices. The authors used routing, link transmission

power and link scheduling mechanisms to improve the network performance. Al-Jemeli and [18] introduced cross layer model that reduces control overhead. Authors designed a model that identifies the nodes location information and based on nodes location information transmission power is adjusted and authors showed that their cross layer model outperforms standard IEEE 802.15.4 model and EQSR protocol. Vuran and Akyildiz [19] designed XLP protocol to improve the reliability of the network. XLP integrates the features of transport layer and physical layer, to control the congestion in the network, to maintain reliability and energy efficiency in the network. Felemban et al. [20] proposed SAMAC- Sectored-Antenna Medium Access Control protocol to regulate energy utilization in sectored antennas. SAMAC promotes the energy efficiency of network and reduces delay. Shi et al. [21], proposed the cross layer design based on transmission power control, authors integrated MAC based power control technique and routing technique to balance the energy utilization. ElBatt and Ephremides [9] introduced scheduling and power control scheme related cross-layer model, to overcome multiple access problem in wireless network. Ben-Othman et al. [22] introduced EQSR—Energy efficient and QoS aware multipath Routing protocol that prolongs network lifetime. EQSR is does multipath routing, it computes the path based on parameters such as signal-to-noise ratio (SNR), interface buffer availability and residual energy for better energy utilization and to promote the network QoS. In [23], A service oriented cross-layer operational model has been introduced. In this model, network lifetime is improved by obtaining application requirement.

2.1 Observations Form Literature

Based on our study, it is observed that effectively combining the features of various optimization techniques (various layers of protocol stack) in a single network architecture promotes the network performance. Therefore, it is understood that cross layer model integrates the optimization techniques and satisfies network requirements. Thus, MAC based power control technique and routing technique are effectively integrated in proposed cross layer model, to enhance the network reliability and energy efficiency.

3 Proposed Cross Layer Model

The proposed cross layer model developed with protocols such as, AODV routing protocol, IEEE 802.15.4 PHY and IEEE 802.15.4 MAC protocol, which are more suitable for low power wireless networks. In the proposed cross layer design, effective combination of routing technique and MAC based power control technique improves the reliability in e-health applications. In routing technique, reliability related parameter is included in route discovery process. Later, MAC based

power control technique utilizes the routing information and decides the optimum transmission range of every nodes in the network. Following sub-section elaborates the functions and features of the proposed cross-layer model.

3.1 Network Model

Network Model: The following assumptions have been made in our network scenario.

- Nodes are deployed in M x M network area in random fashion.
- Nodes are stationary.
- Nodes are aware of the Expected Transmission Time (ETT) information.
- Transmission range of the nodes can be varied.
- Energy is limited for Nodes (battery operated).
- To adopt IoT scenario IPv6 and IEEE 802.15.4 are included in the proposed cross layer model.

Figure 3, describes the network model of proposed cross layer design.

3.2 Power Consumption Model

Power consumption model is adopted from standard IEEE 802.15.4 radio. The power consumption model is defined based on transceiver circuit of the device and physical radio. The transceiver circuit operated by IEEE 802.15.4 operates different modes, they are,

1. Transmit state (Ptx): Transmitting packets.

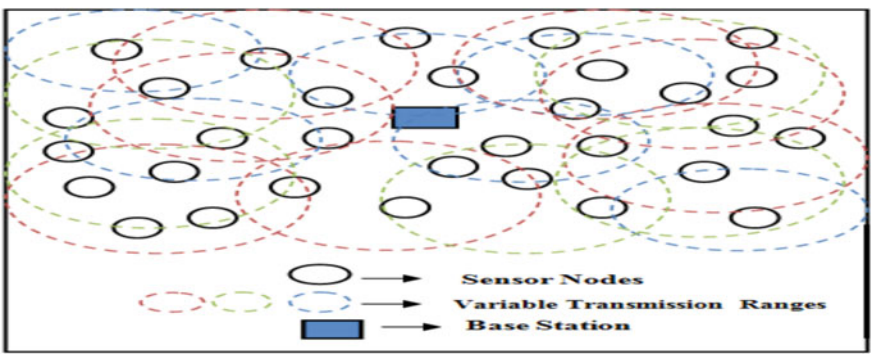


Fig. 3 IoT world forum reference model

2. Receive state (Prx): Receiving packets.
3. Idle state (Pidle): The clock is kept on in this state.

The time need to transmit a packet is Tpacket, The average power consumption is adopted from [24],

$$P_{avg} = \frac{P_{idle} \times T_{idle} + P_{tx} \times T_{tx} + P_{rx} \times T_{rx}}{T_{ib}}$$

Tidle, Trx, Ttx are the time required to transmit the packets in three states, Tib is the inter-beacon period, Tib is defined as,

$$\text{Tib} = \text{Tibmin} \times 2\text{BO}$$

Tibmin is defined as minimum super frame duration and BO is the beacon order, which can be scaled from 0-15. The energy consumption of four states can be said as, Eidle, Etx and Erx.

3.3 Routing Mechanism

Routing is one of the effective technique to satisfy the network requirements. In proposed cross layer model routing and power control technique plays the major role. Reliability related parameter (ETT) is included in standard AODV routing protocol to compute reliable path.

Expected Transmission Time: Finding stable links by link quality measurement parameters such as ETX, ETT, SNR, etc., eminently reduces the number of retransmissions. Expected transmission count (ETT) is efficient routing metric that estimates the link quality. ETT is computed by packet delivery ratios (forward packet delivery ratio and reverse packet delivery ratio) based on bandwidth and packet size. Expected transmission count (ETX) [25] is the basic parameter developed to reduce the number of retransmissions, were ETT [24] is the extended version of ETX is measured based on the forward packet delivery ratio (Df) and reverse packet delivery ratio (Dr).

$$\text{ETX} = 1/(\text{Df} * \text{Dr})$$

Similarly Expected Transmission Time is defined as,

$$\text{ETT} = \text{ETX} * \text{t}$$

$$\text{ETT} = ETX * \frac{\text{LF}}{BL}$$

Type	Flags	Reserved	Hop count
RREQ (broadcast) ID			
Destination IP Address			
Destination Sequence Number			
Original IP Address			
Original Sequence number			
Expected Transmission Time			

Fig. 4 RREQ control packet structure

$$\text{ETT} = ETX * \frac{\text{LF}}{\frac{LL}{TS - TL}}$$

LF is the data packet of fixed-size, BL is bandwidth of link L, LL is data packet of largest-size, TS TL is an interval between the arrivals of two packets. ETT is the time required to successfully deliver the data packets to each neighbor node. Advantage of ETT is it adjusts ETX value based on dynamic link variations and allows variable packet sizes. ETT ignores intra low interference and improves the throughput by measuring loss rate and bandwidth. To add the ETT information in AODV, it should be included in the control packet of AODV routing protocol.

RREQ packet format is described in Fig. 4. After finding the ETT information by RREQ, nodes adjusts its transmission range based on transmission power control techniques. Figure 5 describes the route discovery process y destination node.

3.4 Power Control Technique

Optimizing the communication unit (regulating transmission range and reception range) promotes the network reliability and energy efficiency. Medium access control protocol (MAC) is responsible protocol to regulate radio operations (deciding transmission range) of nodes [26, 27]. Transmission ranges of nodes can be adjusted by the aid of TPC (Transmission Power Control) technique is used to vary RX and CS Thresholds. Transmission range (communication range of the node) is decided based on the RX threshold (the minimum receive signal level the link will work with) and CS threshold (carrier sensing range of the node to sense the sender's transmission) values. When a node receives ETT information, later power control technique it adjusts its transmission power (CS transmit power level of node, RX transmit power level of node) [28, 29] based on ETT information (Fig. 6).

Transmission range can be adjusted between the lower limit and upper limit of the CS and RX transmit power levels.

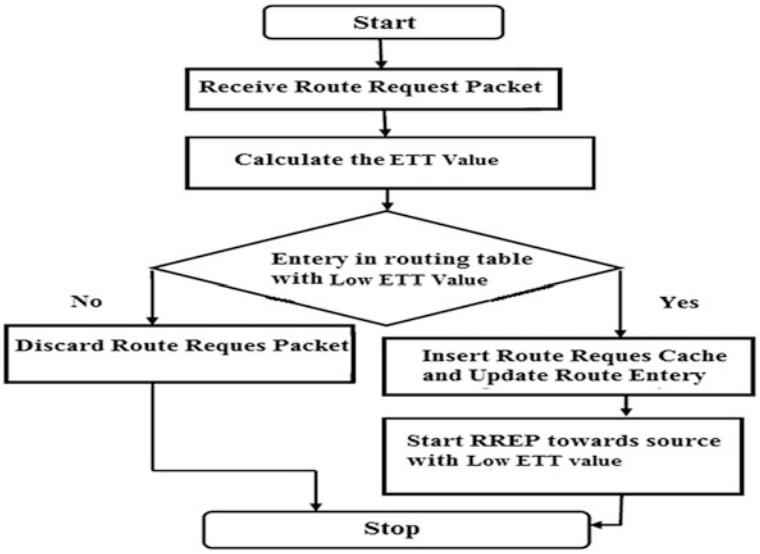


Fig. 5 Route discovery process

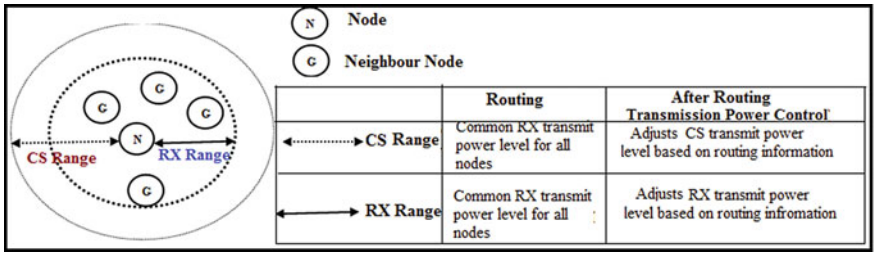


Fig. 6 MAC based power control technique

Rmin (CS and RX min power level) ≤ R (Adjusted transmission range) ≤ Rmax (CS and RX max power level)

Rmin is the minimum transmission range of nodes in network, R is the adjusted transmission range and Rmax is the maximum transmission range of node in the network.

3.5 Operational Flow of Proposed Cross Layer Model

Process of obtaining optimum transmission range

Routing Mechanism

When node deployment is done, Nodes in the network floods, route request packet (RREQ) that contains, ETT Information.

⇩

Every node in network obtains ETT Information with the help of control packet information.

⇩

Node computes its ETT value that is scaled from (0 to 1).

⇩

Based on ETT value, Every node obtains its own threshold value (CS and RX Threshold) from multiple transmission range thresholds of IEEE 802.15.4.

⇩

MAC -Power Control Technique

When every node obtains its ETT value, MAC handles the network operation.

⇩

After Receiving ETT value, nodes adjust its transmission range by varying RX transmit power and CS transmit power.

⇩

When RX and CS power adjustments are done, Every Node obtain their optimum transmission range.

⇩

Physical Layer

Nodes transmit the data in their preferred transmission range.

4 Suitability of Proposed Cross Layer Model for IoT E-Health Application

In wireless network, have been proposed to prolong the network life time and improve the reliability. Many research work concentrates on particular optimization technique for example, routing, MAC based scheme, node placement technique etc. Integrating two different technique in single network architecture, satisfies specific network requirements. Thus, we proposed reliable network architecture, which is more suitable for e-health applications. Proposed cross layer model is designed by IPv6protocol and IEEE 802.15.4 protocol. IPv6 address is more suitable for IoT

applications, due to its large address space. IEEE 802.15.4 radio is commonly used radio protocol for low power IoT applications (e-health). Thus, we included IPv6 protocol and IEEE 802.15.4 PHY/MAC radio in simulation scenario to adopt IoT network scenario.

5 Results

In this chapter, Network Simulator-2 (NS 2.34) is used to measure the performance of proposed network architecture [30]. IPv6 module and IEEE 802.15.4 PHY/MAC radio is implemented in our simulation scenario (Table 2).

5.1 Performance Evaluation

Major objective of proposed cross layer design is maximizing the lifetime of network with better reliability. Thus, we estimated the network performance by network lifetime and throughput (to ensure the reliability of the network) values.

Network Lifetime: Internet of Things applications are run by battery operated sensor devices, therefore prolonging the network lifetime is the major objective in low power IoT networks. Reducing the number of re-transmission highly balances the energy consumption and enhance the network lifetime. In our simulation, we estimated the performance of proposed cross layer model with standard IEEE 802.15.4 PHY MAC (AODV).

Figure 7 and Table 3, elaborates the network lifetime estimation of proposed cross layer model. First node death in standard model (IEEE 802.15.4 & AODV) occur at 510 s and in proposed cross layer model first node death occur at 752 s. In standard model, 11 nodes survived for whole simulation period, and in cross layer model 30 nodes survived. And from Table 3, it is observed that maximum number of nodes (76 nodes) lose their energy from 1500 s to 2000 s, which says proposed cross layer model prolongs the lifetime of nodes. Hence the proposed cross layer model outperforms standard protocol.

Table 2 Parameter specifications

Routing protocol	AODV, AODV (ETT)
MAC/physical layer	IEEE 802.15.4
Channel type	Wireless
Radio propagation	Two ray ground
Traffic type	Constant bit rate
Antenna model	Omni directional
Initial energy (sensors)	50 J
Total number of nodes	200
Packet	IPv6

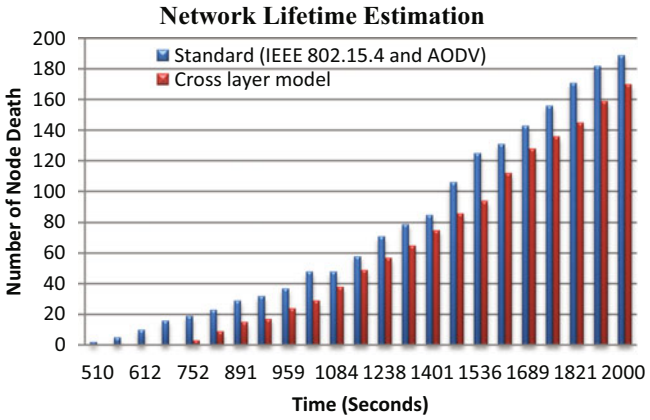


Fig. 7 Network lifetime estimation

Table 3 Network lifetime estimation

Protocols	Standard	Proposed
First node death	510	752
Average node death	77	62
Nodes survived	11	30
1000 s (node death)	48	29
1500 s (node death)	125	94
2000 s (node death)	189	170

Throughput: Network reliability is measured by the throughput of the network. Number of packets successfully transmitted during network operation time is referred as throughput. To achieve a better throughput, data need to be transmitted in stable links (with good transmission power). In the proposed cross layer model, optimum transmission power is achieved by ETT information, which maintains the reliable data transfer.

Figure 8, describes the throughput comparisons of the proposed cross layer model and standard model. In the above graph, it is noticed that the proposed cross layer model achieves better throughput than standard model. From the simulation results it is noticed that the proposed cross layer model performs better in energy efficiency and also maintains the network reliability in efficient way.

5.2 Network Architectures for IoT

Many cross layer designs can be proposed to improve the network performance of low power IoT network, The following architecture, gives the future directions of improve the performance of low power IoT network.

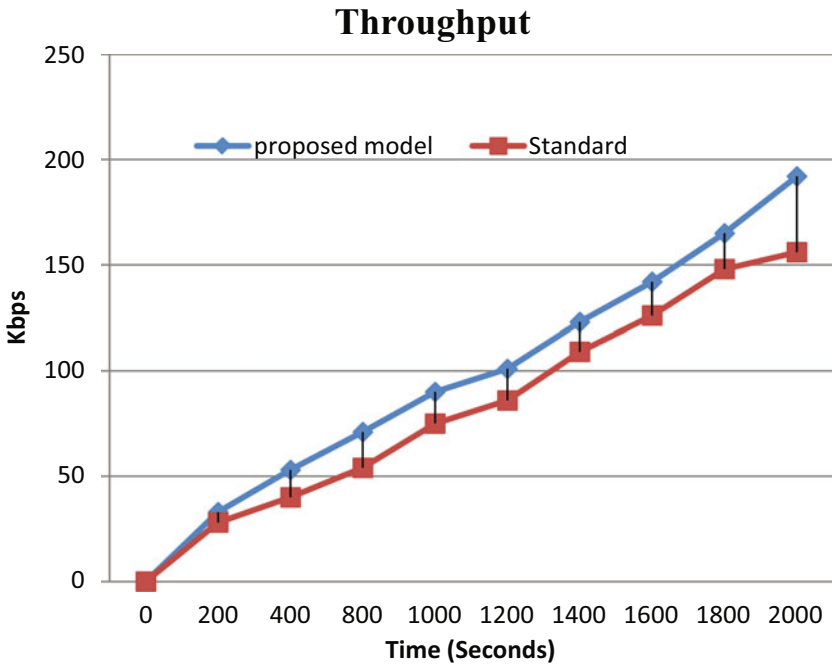


Fig. 8 Throughput

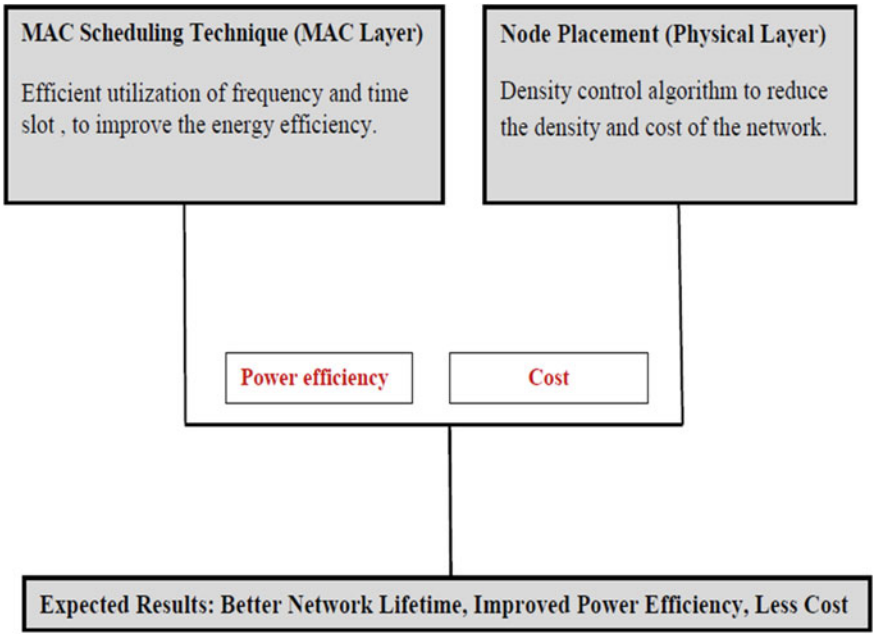


Fig. 9 Energy efficient and cost aware network design

Integrating node placement and scheduling schemes, that improves energy efficiency and network cost (Fig. 9).

Residual energy and data traffic can be improved by integrating routing and TCP estimation technique (Figs. 10 and 11).

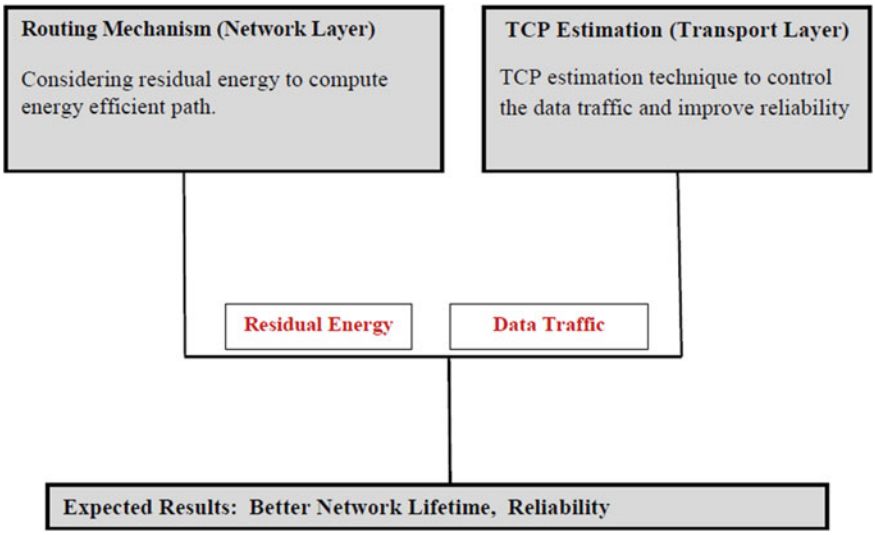


Fig. 10 Reliable and energy efficient network design

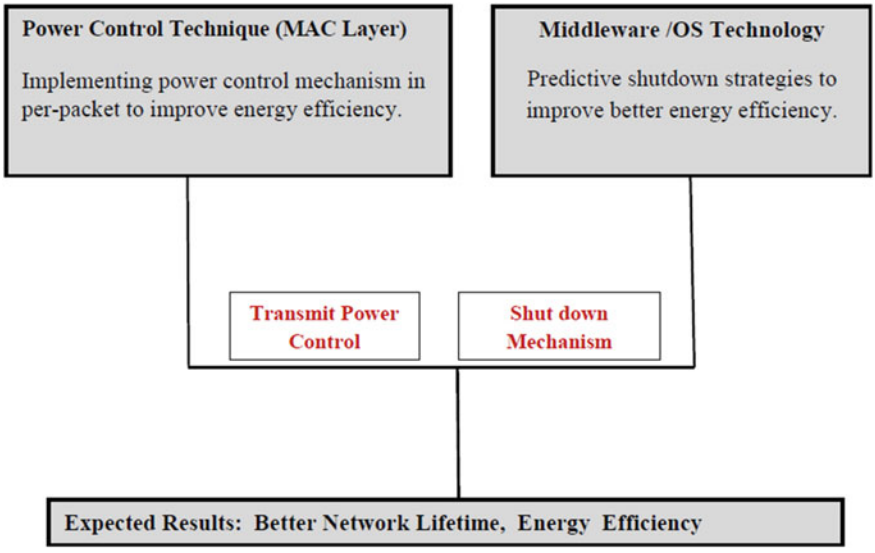


Fig. 11 Network design to improve network lifetime

Similarly, many techniques can be integrated in single cross layer design to improve the performance of low power IoT networks.

6 Conclusion

Internet of things is the network of smart devices that is connected with global network infrastructure. In IoT applications e-health is one of the successful application that provides effective service in health care sectors. IoT devices are resource constrained and IoT links are low power (unstable links). Hence, achieving reliability with better network lifetime is the main challenge in IoT networks. In the proposed cross layer design, network lifetime of nodes is improved with better reliability, by obtaining an optimal transmission range. Optimal transmission range is obtained by effective integration of MAC based power control technique and routing technique. In routing mechanism, ETT information is consider for route discovery process, MAC utilizes the ETT information and it adjust the transmission range of nodes based on ETT value. By obtaining an optimum transmission range of all the nodes in network, reliable data transfer and better network lifetime is achieved. From our result, we conclude that the proposed cross layer model is reliable and energy efficient and it is more suitable for e-health applications.

References

1. ITU Internet reports: The Internet of Things (2005)
2. Xican, C., Woogeun, R., Zhihua, W.: Low Power Sensor Design for IoT and Mobile Healthcare Applications. China Commun. 42–54 (2015)
3. Vasseur, J.-P., Dunkels A.: Interconnecting Smart Objects with IP. Elsevier (2010)
4. Boukerche, A.: Algorithms and Protocols for Wireless Sensor Networks. Wiley-IEEE Press (2009)
5. Srivastava, V., Motani, M.: Cross-layer design: a survey and the road ahead. IEEE Commun. Mag. 112–119 (2005)
6. Xylomenos, G., Polyzos, G.C.: Quality of service support over multi-service wireless internet links. Comput. Netw. (2001)
7. Larzon, L.-A., Bodin, U., Schelen, O.: Hints and notifications. In: Wireless Communications and Networking Conference, pp. 635–641 (2002)
8. Dimi´c, G., Sidiropoulos, N.D., Zhang, R.: Medium access control—physical cross-layer design. IEEE Signal Process. Mag. 40–50 (2004)
9. ElBatt, T., Ephremides, A.: Joint scheduling and power control for wireless ad hoc networks. IEEE Trans. Wirel. Commun. 74–85 (2004)
10. Tong, L., Naware, V., Venkitasubramaniam, P.: Signal Processing in Random Access: A Cross Layer Perspective. http://citeseerx.ist.psu.edu (2004)
11. Liu, Q., Zhou, S., Giannakis, G.B.: Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links. IEEE Trans. Wirel. Commun. 1–10 (2004)
12. Perkins, C., Belding-Royer, E., Das, S.: Ad hoc On-Demand Distance Vector (AODV) Routing. Mobile Ad Hoc Networking Working Group. IETF (2003)

13. IEEE 802.15 Working Group. Standard for Part 15.4: Wireless Medium Access Control Layer (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LR-WPANs). IETF (2003)
14. Chen, C., Liu, X., Qiu, T., Liu, L., Sangaiah, A.K.: Latency estimation based on traffic density for video streaming in the internet of vehicles. Comput. Commun. **111**, 176–186 (2017), ISSN 0140-3664. https://doi.org/10.1016/j.comcom.2017.08.010
15. Qiu, T., Zhang, Y., Qiao, D., Zhang, X., Wymore, M.L., Sangaiah, A.K.: A robust time synchronization scheme for industrial internet of things. IEEE Trans. Ind. Inf. (2017). https://doi.org/10.1109/TII.2017.2738842
16. Medhane, D.V., Sangaiah, A.K.: ESCAPE: effective scalable clustering approach for parallel execution of continuous position-based queries in position monitoring applications. IEEE Trans. Sustain. Comput. (2017). https://doi.org/10.1109/TSUSC.2017.2690378
17. Madan, R., Cui, S., Lall, S., Goldsmith, A.: Cross-layer design for lifetime maximization in interference-limited wireless sensor networks. IEEE Trans. Wirel. Commun. **5**(11), 3142–3152 (2006)
18. Al-Jemeli, M., Hussin, F.A.: An energy efficient cross-layer network operation model for IEEE 802.15.4-based mobile wireless sensor networks. IEEE Sens. J. **15**(2), 684–692 (2015)
19. Vuran, M.C., Akyildiz, I.F.: XLP: a cross-layer protocol for efficient communication in wireless sensor networks. IEEE Trans. Mobile Comput. (2010)
20. Felemban, E., Vural, S., Murawski, R., Ekici, E., Lee, K., Moon, Y., Park, S.: SAMAC: a cross-layer communication protocol for sensor networks with sectored antennas. IEEE Trans. Mobile Comput. 1072–1088 (2010)
21. Shi, L., Fapojuwo, A.: TDMA scheduling with optimized energy efficiency and minimum delay in clustered wireless sensor networks. IEEE Trans. Mobile Comput. **9**(7), 927–940 (2010)
22. Ben-Othman, J., Yahya, B.: Energy efficient and QoS based routing protocol for wireless sensor networks. J. Parallel Distrib. Comput. 849–857 (2010)
23. Wang, J., Li, D., Xing, G., Du, H.: Cross-layer sleep scheduling design in service-oriented wireless sensor network. IEEE Trans. Mobile Comput. **9**(11), 1622–1633 (2010)
24. Draves, R., Padhye, J., Zill B.: Routing in multi-radio, multi-hop wireless mesh network. In: ACM International Conference on Mobile Computing and Networking (MobiCom), pp. 114–128 (2004)
25. de Couto, D.S.: High-throughput routing for multi-hop wireless networks. Ph.D. Dissertation, MIT (2004)
26. Correia, L.H.A., Macedo, D.F., dos Santos, A.L., Loureiro, A.A.F., Nogueira, J.M.S.: Transmission power control techniques for wireless sensor networks. Comput. Netw. 4765–4779 (2007)
27. Monks, J.P., Bharghavan, V., Hwu, W.-m.W: A power controlled multiple access protocol for wireless packet networks. IEEE INFOCOM **2001**, 219–228 (2001)
28. Correia, L.H.A.: Transmission power control techniques for MAC protocols in wireless sensor networks. IEEE Netw. Oper. Manage. Symp. 1049–1054 (2008)
29. Jung, E.-S., Vaidya, N.H.: A power control MAC protocol for ad hoc networks. In: MOBICOM'02, pp. 36–47 (2002)
30. The Network Simulator—NS-2. http://www.isi.edu/nsnam/ns/
31. IoT World Forum Reference Model. http://www.iotwf.com/resources

Erasure Codes for Reliable Communication in Internet of Things (IoT) Embedded with Wireless Sensors

C. Pavan Kumar and R. Selvakumar

Abstract With billions of devices adding up to the internet, broadly termed as Internet of Things (IoT), the need for reliable communication, distributed storage and computation has seamlessly increased. At this juncture, need for reliable communication, distributed storage and computation, adoption of error correcting codes and erasure codes plays a significant role. In this chapter, we give an overview of construction of erasure codes required for reliable communication with emphasis on Internet of Things (IoT) communication which have wireless sensors or sensor networks as its core. Wireless sensors form an integral part of Internet of Things (IoT) devices bridging the virtual world and the real world. Achieving reliability in such networks is highly desirable due to their broad range of applications. The discussed erasure codes in this chapter can be directly employed or with little modification in the context of reliable communication in Internet of Things. In this chapter, the two methods of information transmission, end-to-end transmission and hop-by-hop transmission, prevalent in digital communication scenario is discussed in detail with emphasis on with and without erasure coding. Also, the erasure codes used extensively in the context of achieving reliability in wireless sensor communication namely Reed-Solomon codes, Fountain codes and Decentralized Erasure codes are discussed and compared. This chapter serves as a starting point for researchers interested in working on reliability aspects of communication in Internet of Things embedded with wireless sensors and cyber physical systems.

Keywords Erasure codes · Internet of Things (IoT) · Reliability · Wireless sensors

C. Pavan Kumar
School of Computer Science and Engineering (SCOPE), VIT University, Vellore 632014, India
e-mail: pavankumarc@ieee.org

R. Selvakumar (✉)
School of Advanced Sciences (SAS), VIT University, Vellore 632014, India
e-mail: rselvakumar@vit.ac.in

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_6

1 Introduction

Wireless sensor networks has become integral part of many large scale systems especially of those providing an interface between the cyber world and the real world. Even though the sensors are capable of only sensing and transmitting information to the central sink or immediate next node (acting as relay), the role played by such sensors in diverse atmospheric conditions unsuitable for human intervention is really incredible [1, 2]. Wireless sensor nodes or motes deployed in an area, collaborating to perform a specified task forms a wireless sensor network. Internet of Things (IoT) and Cyber Physical Systems (CPS) will have these wireless sensors or wireless sensor networks as its core for applications ranging from sensing to control applications. Studying and understanding reliable communication in wireless sensor network plays a major role in design of Internet of Things and Cyber physical systems.

Wireless sensor networks deployed in the field will sense the data and transmit it to the sink node via intermediate nodes in the network. All sensed data will be stored in a limited size buffer memory of sensors in the form of packets and transmitted to the sink (or the central node or gateway node) one by one upon receiving acknowledgment (*ACK*) from the sink node. Such acknowledgment will result in network jamming or congestion and increase the latency as every time the nodes have to wait to transmit until *ACK* is received from sink [3]. This will create constraint on node buffer with limited size meant to store the packets temporarily. There are chances that the packets are lost in transmission and sink may also ask sender to resend the packets that further increases the latency [4].

To reliably transmit packets over the network and increase the efficiency, error correcting codes particularly erasure codes were introduced in the literature at the higher layer of communication stack in OSI or TCP/IP reference model [5]. In a typical error correction coding setup, redundant bits r are added at the sender for the information bits k of message m to be transmitted over channel. By addition of redundant bits the length of message will be n (i.e., $n = k + r$) and such messages added with redundant bits are called as codewords [6]. The process of adding redundant bits to the information bits is also called as encoding. Using any suitable decoding technique at the sink or receiver node, received codeword will be decoded to obtain actual message. In case of erasure codes sender divides the packet m to be transmitted into k fragments (i.e., $\frac{m}{k}$) and using any suitable error correction coding technique k fragments are added with r redundancy fragments and encoded into n packets. At the receiver side, accessing or using any k packets out of $n(n > k)$, the decoder will be able to get back the actual information m.

Packets sensed by the sensor nodes will be encoded before transmission and they are decoded at the sink node to obtain the transmitted packages reliably without the need of retransmission. There are various families of error correcting codes introduced in literature and only a few of them are employed successfully in the context of wireless sensor networks [5].

In this chapter, we discuss the various erasure codes introduced in the literature that are suitable for reliable transmission over Internet of Things (IoT) with other

network issues related to them. The chapter is organized as follows: In Sect. 2, preliminaries required for constructing erasure codes and other network techniques are discussed. In Sect. 3, end-to-end transmission and hop-by-hop transmission schemes are discussed with emphasis on erasure coding. In Sect. 4, state of the art erasure codes used in achieving reliability is discussed. In Sect. 5, summary of existing state of the art erasure codes is discussed and future research direction in the context of Internet of Things erasure codes is discussed in Sect. 6. Section 7 deals with the conclusion.

2 Preliminaries

2.1 Data Transmission and Dissemination

Data transmission from source to destination in networks is achieved either through end-to-end transmission or through hop-by-hop transmission [7–9]. In case of end-to-end transmission the data packets are routed through various paths in the network till it reaches the sink or destination node. The intermediate nodes present in the network acts as relay in helping the data packets to reach sink and doesn't alter the contents of the packets. In case of hop-by-hop transmission the intermediate nodes play a major role by making the data packets to reliably reach the destination. In a hop-by-hop network, the path from source to destination will comprise of intermediate nodes with additional processing power termed as hops. At an intermediate hop node data packets can be verified in terms of bits received in error and can ask the previous hop node or source node to resend the packet. Apart from this even packet level encoding and decoding can be performed at these hop nodes similar to network coding to ensure reliability [10].

Figure 1 gives an overview of end-to-end packet transmission scheme. Data packets from source will be received by destination and destination will send an acknowledgement for the packets received. Intermediate nodes will act only as relay and doesn't alter the contents of packets. Sender will wait for the destination to send *ACK* before transmitting the next packet. *ACK* message indicates the source that the packet is received successfully without any error, otherwise the sender will transmit the packet again waiting for a stipulated time assuming that the packet is lost in transmission and has not reached destination.

In other words, in case of end-to-end transmission schemes data packets are combined only at the sink node to obtain the information. If the erasure codes are employed in the setup of end-to-end transmission then all the packets will be decoded only at the receiver. Sensors deployed in the field will sense the parameter of interest and send it to the sink node to further act upon it. This strategy of sensing at source, transmitting over the network and analyzing or acting upon the sensed data remains the same in case of Internet of Things (IoT) devices also.

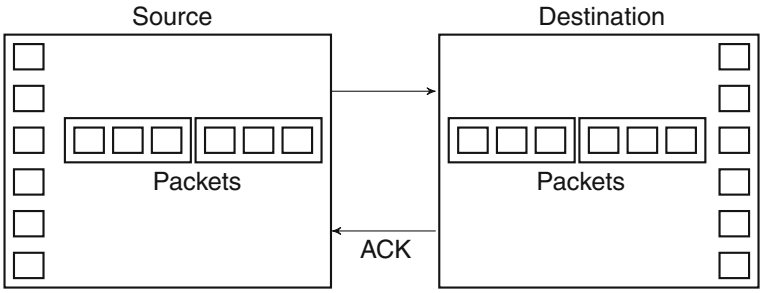


Fig. 1 End-to-end transmission

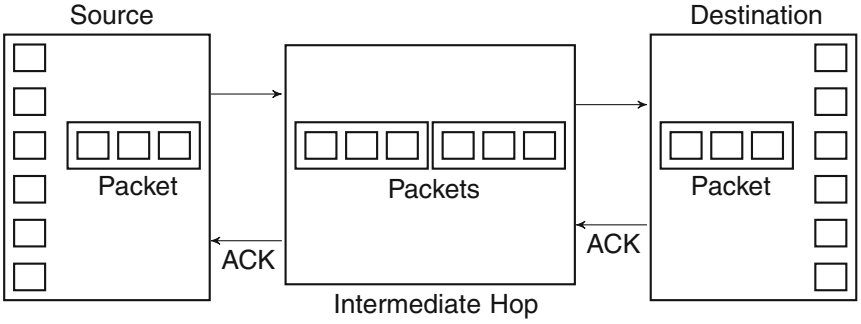


Fig. 2 Hop-by-hop transmission

In case of hop-by-hop transmission scheme, packets can be analyzed whether it is received in error or being correctly received at the intermediate hops. If the erasure codes are employed in the setup of hop-by-hop transmission then the intermediate hops can encode or decode the packets before transmitting further to the sink node. In both end-to-end transmission and hop-by-hop transmission schemes, data packets will be combined only at the sink nodes to obtain the actual message transmitted from source node.

Figure 2 gives an overview of hop-by-hop transmission scheme. Data packets from source will be transmitted to the next hop. Upon receiving the *ACK* from hop, the transmission completes and subsequently the hop forwards the packet to next hop and waits for *ACK*. The process continues until the destination receives all the packets. Similar to the end-to-end transmission, the hops will wait for *ACK*. If it is not received in stipulated time, the hops or sender will retransmit the packet assuming it has been lost in transmission.

On the other hand, Dimakis et al. [11] proposed Decentralized erasure codes wherein the data sensed from the sensors will be stored in the network itself and one can get the data sensed by querying any k out of n nodes. Advantage of such schemes is that the user or data requester can input (or inject) query in any part of the sensor network and obtain actual information sensed by the sensors and stored in

the network by accessing k nodes in the wireless sensor networks. Failure of sensor nodes will be compensated from the Decentralized Erasure coding technique as the data can be obtained by querying any k nodes. We will discuss about such codes in the next section.

2.2 Erasure Codes

Many erasure codes are proposed in the literature for achieving reliability in communication. Tornado codes [12, 13], Fountain codes [14], Raptor Codes [15], Online codes [16], Luby Transform codes [17] and Reed Solomon codes [18] are popular among them. Reed Solomon codes are a class of linear error correcting codes and fountain codes are rate less codes. These two families of codes are extensively used in achieving reliable communication in wireless sensor networks.

2.2.1 Reed Solomon Codes

Reed-Solomon codes [18] introduced by Reed and Solomon in the early 1960s is a class of linear error correcting codes that are extensively used in data storage and communication. Reed-Solomon codes are employed to encode data chunks transmitted at the sender side by adding redundancy and decoding will be performed at the receiver side to get back the actual information. m data chunks will be added with r redundant data chunks. Reed Solomon codes being Maximum Distance separable (MDS) codes, it is possible to get the actual message upon receiving any m out of $n(n = m + r)$ data chunks.

Reed Solomon codes can be constructed with Vandermonde matrix construction approach [19]. Vandermonde matrix V is given as

$$V = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{m-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{m-1} \\ 1 & \alpha_3 & \alpha_3^2 & \cdots & \alpha_3^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_{n-1} & \alpha_{n-1}^2 & \cdots & \alpha_{n-1}^{m-1} \\ 1 & \alpha_n & \alpha_n^2 & \cdots & \alpha_n^{m-1} \end{pmatrix} \tag{1}$$

In a Vandermonde matrix given in Eq. 1, any set of rows is linearly independent. Entries in the matrix V are non zero and distinct from each other such that $V_{ij} = \alpha_i^{j-1} \neq 0$.

Reed Solomon encoding is performed as follows by multiplying vandermonde Matrix V with message M to be transmitted.

$$\begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{m-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{m-1} \\ 1 & \alpha_3 & \alpha_3^2 & \cdots & \alpha_3^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_{n-1} & \alpha_{n-1}^2 & \cdots & \alpha_{n-1}^{m-1} \\ 1 & \alpha_n & \alpha_n^2 & \cdots & \alpha_n^{m-1} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix} \tag{2}$$

Systematic encoding will be performed by substituting any m rows of Vandermonde matrix given in Eq. 2 with $m \times m$ identity matrix. Systematic encoding proves to be advantageous in implementation and by addition of such identity matrix, the property of linear independence of rows of Vandermonde matrix remain unchanged.

Systematic encoding is performed as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{m-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{m-1} \\ 1 & \alpha_3 & \alpha_3^2 & \cdots & \alpha_3^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_{k-1} & \alpha_{k-1}^2 & \cdots & \alpha_{k-1}^{m-1} \\ 1 & \alpha_k & \alpha_k^2 & \cdots & \alpha_k^{m-1} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_k \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_k \\ r_{k+1} \\ r_{k+2} \\ \vdots \\ r_n \end{pmatrix} \tag{3}$$

The above encoding given in Eq. 3 looks similar to a system of linear equation $V \times M = C$ where, V is a Vandermonde matrix, M is message and C is encoded word or codeword. Decoding of Reed Solomon codeword is equivalent to finding M from received codeword R (where, R is $M < R \leq C$), i.e., finding M from $V \times M = R$.

2.2.2 Fountain Codes

The concept of Fountain codes was introduced in the late 1990s [20]. Fountain codes are extensively used ever since its inception in applications involving transfer of files (small or large) in the form of small packets which are received at the sink with small errors or no errors [21]. The flow of packets to receiver is similar to water fountain and hence the name. These kind of codes require no feedback or little feedback from receiver upon receiving packets. Also, an another advantage of rate less codes is that the encoding size of packets need not be known apriori and encoding can be performed on-the-fly, a feature that is not present in linear erasure codes.

Fountain code works as follows: A file generated at source is divided into k fragments and encoded on the fly using a generator matrix G. Such encoded file fragments will be transmitted. Receiver upon receiving any fraction k' (with $k' > k$) of such encoded fragments will decode and get back the actual file generated at source.

Mathematically, the data generated with size k at source is divided into k fragments or packets $S = \{s_1, s_2, \ldots, s_k\}$. It is assumed that each packet is either received without any error or dropped in transmission while proceeding for decoding at receiver.

Encoder generates packets c_n as follows from the source bits by *XOR*-ing the source packets together

$$c_n = \oplus_{k=1}^{K} s_k G_{k \times n} \tag{4}$$

where, $G_{k \times n}$ is the generator matrix. The encoded packets c_n before transmission are bitwise summed with source packets for which $G_{n \times k} = 1$ under $GF(2)$ operation.

On the receiver side, decoding is performed as follows.

Packets will be received either with zero error or not considered at all by the receiver. Decoder at receiver keeps receiving the encoded data packets till the received packets are enough to decode the actual data packets transmitted from the sender. Note that unlike other communication systems receiver will not send *ACK* to sender either to acknowledge the receipt of packet or to resend the packet.

Decoder forms a system of linear equation as follows and solves for data packets $S = = \{s_1, s_2, \ldots, s_k\}$ from $C'_n = \{c'_1, c'_2, \ldots, c'_n\}$ received packets which was transmitted as $C_n = \{c_1, c_2, \ldots, c_n\}$ at the source. Since, all packets will be assumed to be received with zero error C' is always considered equal to C(i.e.,$C' = C$)

$$\begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1k} \\ g_{21} & g_{22} & \cdots & g_{2k} \\ \vdots & \vdots & & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nk} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \tag{5}$$

Success of such decoding relies on the fact that the generator matrix $G_{k \times n}$ used at the sender is invertible, i.e., G'_{nk} exists. Such inverse can be computed using Gaussian elimination method.

Alternatively, decoding can be expressed as

$$s_k = \sum_{n=1}^{N} c_n G_{nk}^{-1} \tag{6}$$

where, G_{nk}^{-1} is the inverted generator matrix, s_k is the symbol input and c_n is the received packets.

Also, it is to be noted that each encoded packet satisfying the following two conditions will only be transmitted:

1. Each encoded packet should be non-zero
2. Each encoded packet should be linearly independent of preceding encoded packets. It is to ensure that, the decoder can form a solvable system of linear equations from the received packets.

The condition to have more number of packets than K at receiver is to increase the success rate of decoding at receiver.

The guarantee that the matrix G_{nk} is invertible is established as follows: Three cases as follows will arise for computing inverse of matrix G_{nk}. Let N be the received packets with additional packets E at the receiver, i.e., $N > K$ or $(N + E) > K$.

1. $\underline{N < K}$:
   Enough packets are not received to proceed for decoding at the receiver.
2. $\underline{N = K}$:
   It is possible for the receiver to perform decoding. But it cannot guarantee that the matrix is invertible with high probability. Since, the condition is to have each packet linearly independent and non zero. The probability of having invertible matrix by keeping $N = K$ is 0.289 for $K > 10$ from following calculation.
   Suppose the first packet (i.e., column) is received as non zero with probability $(1 - 2^{-K})$ and second column is received with probability $(1 - 2^{-(K-1)})$ satisfying the conditions mentioned previously that the subsequent packets are linearly independent with the preceding columns and non zero. Proceeding similarly the probability of matrix being invertible P_{invert} is given as

$$\begin{aligned} P_{invert} &= \prod_{i=0}^{K-1} \left(1 - 2^{-(K-i)}\right) \\ &= \left(1 - 2^{-(K)}\right) \times \left(1 - 2^{-(K-1)}\right) \times \cdots \times \left(1 - 2^{2}\right) \times \left(1 - 2^{1}\right) \end{aligned} \tag{7}$$

   which is 0.289 for any $K > 10$.
3. $\underline{N > K}$:
   By keeping the number of packets required at receiver slightly larger than the source packets S, the success probability of decoding increases.
   Consider δ as the probability that the receiver will not be able to successfully decode the packets from the excess packets E with high probability. Then, $1 - \delta$ will be the probability that receiver will be able to successfully decode from the excess packets with high probability.
   For not so small values of E overhead or excess packets, $\delta(E)$ is roughly 2^{-E} for a random binary fountain code constructed with k source symbols [22]. The probability of failure for any k is bounded above by

$$\delta(E) \leq 2^{-E} \tag{8}$$

   In other words, for each increase in overhead E, the failure probability δ decreases by a factor of two.

2.2.3 Decentralized Erasure Codes

Decentralized erasure codes was introduced by Dimakis et al. [23, 24]. Decentralized erasure codes are a special class of codes proposed mainly for storing data in the wireless sensor network itself rather than depending on an external storage device. Sensors deployed in the field for the purpose of sensing can also be used for storing data. Data sensed by the sensors will be stored distributively in the network and upon querying certain nodes in the network, stored data can be obtained by the query injector.

Distributed erasure codes operates slightly different from the other erasure codes. In typical erasure codes minimum k or slightly greater than k packets need to be stored at a central location (or received) by the decoder to proceed for decoding whereas in case of distributed erasure coding, decoding will be performed by accessing any k encoded packets in the network. Even though fountain codes operates in a slightly similar manner, they are not distributed in nature.

Distributed erasure code setup is constructed as follows: At a given time, k nodes will perform sensing operation and n nodes in the network selected randomly will be used to store the data sensed by k nodes. Sensed data will be stored as packets or encoded packets. Data collector or any entity interested in getting the data sensed by k nodes can inject query to the network and just by accessing any k nodes it should be able to get the data required. There can be overlap among sensing and storage operations, i.e., same nodes can perform both sensing and storage operation and some nodes can be used only for storage operations.

Decentralized erasure codes are constructed as follows [25–27]: Data $M = \{m_1, m_2, \ldots, m_k\}$ sensed from k nodes will be stored by selecting a storage node n_i uniformly at random. Each storage node will have a coefficient $g_{i,j}$ chosen from the Galois Field (GF) with which sensed data will be multiplied and stored. Each storage node will store these field coefficient that increase the overhead on storage nodes. One storage node can encode and store packets of size equivalent to data packet generated at source. If the data packet stored is generated at that source node then $g_{i,j} \neq 0$, otherwise it is 0.

Generator matrix G will be employed to compute the codewords $C = \{c_1, c_2, \ldots, c_n\}$. The elements of M, C and generator matrix G are taken over field $\mathbb{F}$.

Encoding process is defined as follows:

$$C_i = \sum_{i=j=1}^{i=k,j=n} g_{i,j} M_i \tag{9}$$

where, g_{ij} is the element of generator matrix G taken from the field $\mathbb{F}$.

Alternatively, it can be given as

$$\begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & \cdots & m_k \end{bmatrix} \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,n} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,n} \\ \vdots & \vdots & & \vdots \\ g_{k,1} & g_{k,2} & \cdots & g_{k,n} \end{bmatrix} \tag{10}$$

Similarly, decoding is given as a set of linear equation $C_{1\times n} = M_{1\times k} G_{k\times n}$, solving for $M_{1\times k}$ accessing any k storage nodes.

$$\begin{bmatrix} m_1 & m_2 & \cdots & m_k \end{bmatrix} = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,n} \\ g_{2,1} & g_{22} & \cdots & g_{2,n} \\ \vdots & \vdots & & \vdots \\ g_{k,1} & g_{k,2} & \cdots & g_{k,n} \end{bmatrix}^{-1} \tag{11}$$

The success of decoding in decentralized erasure codes depend on the factor that the $k \times k$ sub matrix of generator matrix G is invertible. Decoder only require any k columns of generator matrix and corresponding codeword to get the message $M = \{m_1, m_2, \ldots, m_k\}$.

3 General System Overview

In this section, we overview the general system view employed in data transmission and dissemination in wireless sensor networks employing error correction coding. We discuss about end-to-end transmission and hop-by-hop transmission with emphasis on erasure codes and Decentralized Erasure codes in this section.

3.1 End-to-End Transmission with Erasure Coding

In erasure coding based end-to-end transmission system, data chunks will be encoded and transmitted. An overview of erasure coding based end-to-end transmission scheme is given in Fig. 3. In end-to-end transmission scheme with erasure codes, the data chunks or packets k will be encoded by adding r redundant chunks and $n = k + r$. The encoded packets will be transmitted over the channel. Destination upon receiving any k' packets ($k' > k$) out of n will be able to decode the actual message k. An advantage of using erasure coding in end-to-end transmission scheme is that it does not require *ACK* by the destination node to transfer next packet. Packets will be transmitted over any available path to reach destination. Intermediate nodes present in the path from source to destination will only act as relay in transferring the packets.

Probability of transmission in end-to-end transmission scheme where data packets are sent from source to destination is derived as follows [8]. Suppose L is the number of maximum attempts to be made to transmit message with link delivery probability

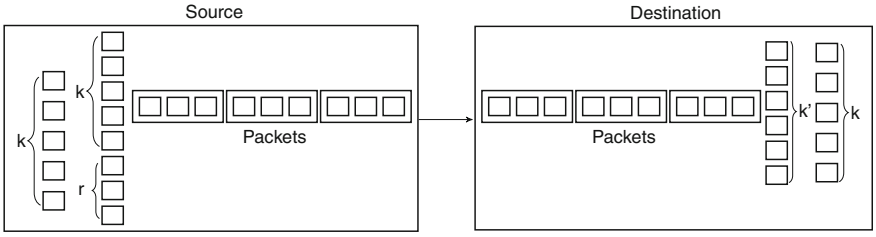


Fig. 3 End-to-end transmission with erasure coding

p from source to destination in end-to-end transmission with intermittent hops not playing any role and acting only as relay hops. N is the number of hops. P_S and P_F are the probability of successful and failed end-to-end transmissions respectively. M is the number of link level transmissions.

If P_1 is the probability of success for one end-to-end transmission over N hops with L attempts then probability of success P_S is

$$P_S = 1 - (1 - P_1)^L \tag{12}$$

Since, the probability of packet link delivery is p as assumed above and packets are transmitted over N hops. Equation 12 is rewritten as

$$P_S = 1 - (1 - p_1^N)^L \tag{13}$$

Similarly, for successful transmission the probability of success for end-to-end transmission with L attempts over N hops is given as

$$\begin{aligned} P_S &= 1 - (1 - P)^L \\ &= 1 - (1 - p^N)^L \end{aligned} \tag{14}$$

If z is the interval with which L takes the value, i.e., $z \in [1, L]$ then probability of successful transmission $P(z)$ is given as

$$P(z) = \begin{cases} (1 - p^N)^{z-1} p^N & 1 \leq z < L \\ (1 - p^N)^{L-1} & z = L \end{cases} \tag{15}$$

3.2 Hop-by-Hop Transmission with Erasure Coding

In erasure coding based hop-by-hop transmission system, data chunks will be encoded at the source and transmitted to the next hop. Hops will be continuously present in the path the packets take from source to destination. Each hop will ensure that the pack-

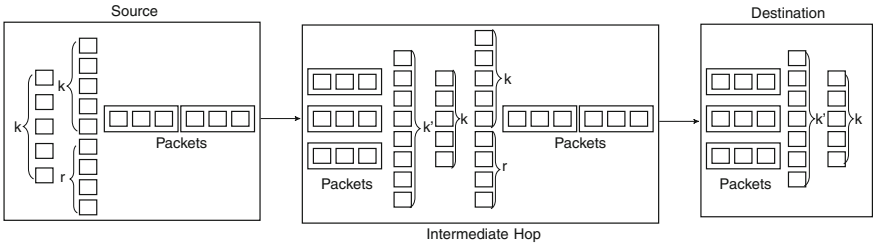


Fig. 4 Hop-by-hop transmission

ets are received in order and if any error in such packets, they will be corrected at the hop itself. This approach is similar to network coding [28]. An overview of erasure coding based hop-by-hop transmission system is given in Fig. 4. Each intermediate hop will receive the packets and if the packets are not enough to be forwarded to the next hop, first they will decode and reconstruct the message. Then again they perform encoding and encoded packets are transmitted to the subsequent node. This will be performed till enough packets will reach the destination. Intermediate nodes play a major role in hop-by-hop transmission with erasure coding. There will no *ACK* from destination to source or intermediate nodes.

In a hop-by-hop transmission setup, data will be transferred from one hop to next hop with link delivery probability p and maximum number of hops present in intermediate path from source to destination [8]. P_S and P_F are the probability of successful and failed hop-by-hop transmissions respectively. If the subsequent hop fails then there will be no transfer of packets. Thus, for successful transmission it is necessary to have N hops availability.

The probability P_S for successful hop-by-hop transmission with L attempts over one hop is given by $1 - (1 - p)^L$.

Thus, the probability for successful hop-by-hop transmission with L attempts over necessary N hops is given by,

$$P_S = \left(1 - (1 - p)^L\right)^N \tag{16}$$

If z is the interval with which N takes the value, i.e., $z \in [1, N]$, then the probability $P(z)$ for successful transmission is given as

$$P(z) = \begin{cases} \left(1 - (1 - p)^L\right)^{z-1} (1 - p)^L & 1 \leq z < N \\ \left(1 - (1 - p)^L\right)^{z-1} & z = N \end{cases} \tag{17}$$

4 Existing State of the Art

In this section, we review the state of the art research in the area of erasure coding based reliability in wireless sensor networks proposed either over end-to-end transmission scheme or hop-by-hop transmission scheme. Employing erasure coding introduce a little computing overhead at the nodes whereas it increases the efficiency in terms of data transmission reliability, network life and reduced network consumption (or traffic) and buffer storage requirement at individual nodes [4]. We have considered only the works that employ redundancy for reliable communication based on specific erasure codes. Reliability methods proposed in [29–31] are not considered as these works does not employ any erasure codes and also in their context the term redundancy is used for sending same packets multiple times. Also, the works [32, 33] are also not considered as they won't use any specific codes and only show that using erasure codes will result in reliable communication.

4.1 Reliable Transfer on Sensor Networks (RTSN)

Kim et al. [19] proposed reliable data transfer on wireless sensor network using systematic Reed - Solomon erasure codes. End-to-end transmission with erasure coding is employed to reliably transmit data from source to destination. The experiments were conducted using real test bed and not by simulations. Systematic Reed-Solomon code is employed to achieve reliability. Code construction is similar to the systematic Reed-Solomon code construction outlined in Sect. 2.2.1. To reduce coding overhead particularly at the decoder of receiver, systematic Reed Solomon coding is employed.

When the packets arrive at the decoder, first it checks whether the received packet is a data packet or a redundant packet. If the received packet is a data packet and non of the original data bits are corrupted, then decoding will not be performed. This will be known by the systematic encoding whether the received data packet contains original data packet or not. If received packet has the redundant bits and if the data bits are missing then systematic Reed-Solomon decoding is performed to reconstruct actual message.

To transmit the message over multiple paths in wireless sensor network, Beacon vector routing protocol is employed, instead of sending or flooding encoded packets on all paths. Transmitting encoded packets over multiple paths in network may increase the network reliability but at the same time it is inefficient as it increases the network overhead.

In beacon vector routing model, a subset of r nodes are selected as "*beacons*" based on network connectivity [19]. These beacons will flood the network to know the distance of nodes in the network to the beacon nodes. Beacons are selected such that they are near to the destination or sink. Distance of each node to beacon node will be stored as $P(p) = \{B_{1p}, B_{2p}, \ldots, B_{rp}\}$ where, B_{ip} is the distance between node

p and B_i. Thus, each node in the network will know the distance to its next neighbor and beacon.

When a packet is to be routed to destination, initially a minimum cost path that is closest to the sink or destination node is deleted based on distance metric defined as

$$\delta(P(p), P(q)) = \sum_{i=1}^{r} |B_{ip} - B_{iq}| \tag{18}$$

To avoid path fail, the algorithm makes use of fallback approach that ensures packets are moving to destination. Thus, reliability is achieved by using systematic Reed—Solomon coding and Beacon vector routing protocol. However, the method doesn't analyze dynamic failure of paths and adaptivity of the proposed method with such failures.

4.2 Optimum Reed Solomon Erasure Coding in Fault Tolerant Sensor Networks

Ali et al. [34] proposed optimum Reed Solomon erasure coding scheme for fault tolerance in wireless sensor networks. This method uses End-to-End transmission with Erasure Coding technique to achieve reliability. The intermittent nodes present in the network through which the data packets are transmitted does not play any role in data transmission and they do act only as relay nodes. On the other hand the system model used proxy sink nodes called *prongs* to collect messages from the sensor nodes via diverse paths. It is assumed that multiple proxy prongs present in the network is connected to the central sink via reliable and sufficient bandwidth path.

Pull based querying is used to collect data at the *prongs* nodes [35]. Source nodes can transmit data over path of its choice. There is no particular way of selecting the path to route packets to *prongs*.

Data packets k generated at the source node will be encoded by adding r redundant packets such that $n = k + r$. The generated data packets per second will be distributed over N parallel path by allocating each path i with a d_i fragment, where, $i = \{1, 2, \ldots, N\}$ are the available number of paths

$$\sum_{i=1}^{N} d_i = d^T \cdot 1 = n = k + r \tag{19}$$

Allocation vector is denoted by $d = [d_1, d_2, \ldots, d_N]$ and 1 to denote a vector of entries with 1.

Destination nodes or prongs will collect the packets over the network and try to decode from any k packets received out of n packets transmitted over N paths. Random variable z_i is used to indicate number of fragments received on path i. This probability is given as

$$P_s = Pr\left[\sum_{i=1}^{N} z_i \geq k\right] \tag{20}$$

It can be observed that if redundant packets r are increased, then the probability of packet loss decreases.

If the loss of consecutive packets on a path, is identical and independent, P_s can be approximated with Poisson distribution [36], i.e.,

$$P \geq Q(d, p, r) \tag{21}$$

and

$$Q(d, p, r) = \sum_{j=0}^{r} \exp^{-\lambda(d)} \frac{[\lambda(d)]^j}{j!} \tag{22}$$

where, $\lambda(d) = -d^T \ln p = -\sum_{i=1}^{N} d_i \ln p_i$ and $p = [p_1, p_2, \ldots, p_n]$ is the probability that data packet is successfully transmitted on each path.

Similarly, probability of packet loss P_l is given as

$$P_l = 1 - P_s = 1 - \sum_{j=0}^{r} \exp^{-\lambda(d)} \frac{[\lambda(d)]^j}{j!} \tag{23}$$

The factors affecting the data reconstruction at the receiver are data fragments n transmitted over N paths, loss of packets in transmission (P_l). However, data fragments loss in a typical network may be due to various factors such as node failure, other underlying protocols associated with transmission.

Cost function considering above factors is defined as

$$\begin{aligned} C &= C_l.P_{loss} + C_t(n) \\ &= C_l.P_{loss} + C_t(k + r) \end{aligned} \tag{24}$$

where, C_l is the cost associated with packet loss and C_t is the cost associated with packet transmission. Since, only Reed Solomon erasure coding approach is to be tested in Wireless sensor network, the normalized value is taken for C_l.

Assigning data packet on each path d_i is the function to be minimized as other parameters associated will be generated from sensor or fixed parameters. So, the constraint $\sum_{i=1}^{N} d_i = d^T 1$ to minimize cost function C.

To minimize cost function, Genetic Algorithms (GA) are used. Genetic algorithm has three distinct phases—selection, crossover and mutation. Roulette wheel is used as a selection operator. For the population of size given as V, probability of selection P_{select} of each ith individual with fitness value f_i is given as

$$P_{select} = \frac{f_i}{\sum_{i=1}^{V} f_i} \tag{25}$$

Uniform crossover is considered and for each subsequent generation, crossover probability is considered.

Mutation value is set to uniform. In each generation the constraint $\sum_{i=1}^{N} x_i = k + r$ is checked. GA halts with specified maximum number of generations is reached. Then, an individual with d vector with lowest C is taken as the solution.

Thus, Ali et al. [34] employ Reed Solomon codes to achieve data reliability and genetic algorithms for path selection out of N to transmit data packets to prongs in turn to sink nodes.

4.3 Reliable Data Transfer Scheme (RDTS)

Srouji et al. [37] introduced reliable data transfer scheme (RDTS) for reliable communication in wireless sensor networks. RDTS approach uses hop-by-hop transmission mechanism with erasure coding for data transfer from source to destination. RDTS employs two schemes namely full erasure coding at hops and partial coding which require coding based on need. RDTS scheme employs systematic Reed-Solomon codes for achieving reliability. Reed-Solomon codes are constructed similar to the methodology outlined in Sect. 2.2.1 using Vandermonde matrix approach.

RDTS scheme employing hop-by-hop encoding works as follows in a n-hop path with $n + 1$ hops and it is assumed that each hop will participate in encoding as well as decoding to ensure that packets are reliably transmitted. Each hop in the path will receive data from previous path with a probability p_i, also termed as successful data arrival probability. Thus, for the N_{i+1}th hop, the successful probability of receiving data from N_i th hop is p_i. Computing in the same manner, the probability of successful delivery P_{path} through n hops is

$$P_{path} = \prod_{i=1}^{n} p_i \tag{26}$$

where, n is the number of hops, p_i is the probability of success of data transfer.

If n packets (i.e., $n = k + r$) are transmitted from one hop, the probability that the same is received by the next hop is given by,

$$N_{received} = p_i \times N_{sent}$$

For successful decoding at hop, the number of received packets must be greater than or equal to k, i.e., $k' \geq k$. Also, number of received data packets must be greater than or equal to $\frac{k}{p_i}$ for successful decoding, i.e.,

$$k + r \geq \frac{k}{p_i}$$

Since, each hop will transmit with probability $\frac{k}{p_i}$, the number of fragments transmitted is given by

$$Total_{RDTS} = \sum_{i=1}^{K} \frac{k}{p_i} \tag{27}$$

In hop-by-hop full erasure coding RDTS scheme, each intermediate hop will encode or decode based on requirement (calculated using $\left\lceil \frac{k}{p_i} \right\rceil$). If the received data fragments k' has missing x original data packets out of k, then the received data fragments are decoded to get original k data fragments first. Then, again from the k (reconstructed from k' received packets) data packets at hop additional r packets are generated by encoding and n packets are transfered to next hop.

In hop-by-hop partial erasure RDTS scheme, hop n upon receiving any k packets will proceed for decoding. First, it will check whether the received packets are enough for the decoding at the next hop (i.e., $k' > k$) calculated using $\left\lceil \frac{k}{p_i} \right\rceil$. If it is enough then the packets are transmitted as it is to the next hop without any encoding. If the received packets $k' = k$, then r redundant packets are constructed (or encoded) from received packets and transmitted to next hop. If the received packets are less than k then full erasure coding scheme is employed first to decode the lost packets and then encode to get r redundant packets to send it to the next hop based on requirement.

RDTS scheme is compared with end-to-end transmission scheme and showed that by employing RDTS scheme better performance can be achieved in terms of network traffic, network lifetime and energy consumption by nodes. However, RDTS scheme require apriori knowledge (to compute $\left\lceil \frac{k}{p_i} \right\rceil$) of the path to reliably transmit packets over that path. Also, computation cost of simultaneous encoding and decoding at the hops is not analyzed by the scheme.

4.4 FBCast

Kumar et al. [38] used Fountain codes for the data transmission in wireless sensor network with hop-by-hop transmission with erasure coding scheme. The novelty of the work lies in employing erasure codes particularly fountain codes to update software installed in the wireless sensor nodes deployed in the field. This kind of software update requires the data flow from sink node to source nodes whereas the data flow is in reverse direction in usual data transmissions, i.e., data flow will be from nodes to sink.

Sensor nodes in the network needs the periodic update of software to cope up or adopt with the new architecture or system update. This also poses another challenge that if software has to be updated then it has to be updated in all the nodes in the

Table 1 Summary

Scheme	End-to-end erasure coding	Hop-by-hop erasure coding	Employed erasure code
Kim et al. [19]	✓		Systematic Reed Solomon codes
Ali et al. [34]	✓		Systematic Reed Solomon codes
Srouji et al. [37]		✓	Fountain codes
Kumar et al. [38]		✓	Fountain codes

network and not only partial update. To reliably transmit the packets from sink to source nodes FBcast method was proposed.

Software update in the form of k data packets at sink node will be encoded into n data packets to be transmitted over the network to reach source nodes. Fountain codes are constructed as given in Sect. 2.2.2. The approach of broadcasting encoded data packet over the network is termed as FBcast. Probabilistic broadcasting approach is used over simple flooding.

Using of fountain codes in the firmware update on wireless sensor nodes is motivated by several facts that these rate less fountain codes provide over other linear erasure codes. To list a few:

1. Fountain codes provide flexibility of choosing the encoded packet length based on requirement. It can change arbitrarily.
2. On-the-fly encoding and no fixed size on packet length.
3. Simple encoding and decoding operations that gives users the flexibility of deciding encoding packets on-the-fly.

Authors propose two models of transmission: Rebroadcast and Retransmission for evaluation of proposed FBcast. Rebroadcast is broadcasting the packets received from another node and retransmission is transmission of same packets again in the network.

In a simple FBcast model, data packets are transmitted from sink to the next hop. In FBcast, probabilistic rebroadcast model, data packets are transmitted based on link success probability, there by achieving reliability over simple FBcast model. To achieve more reliability in terms of packet delivery over different topologies FBcast model with repeaters was introduced. In FBcast model repeaters are selected nodes that receive enough packets in a window period and upon satisfying a threshold value these nodes can act as repeaters rather than employing all nodes in the network as repeaters. Repeater nodes first reconstructs (or decodes) the original data from k' (which should be greater than k) received packets and then again encodes the k packets (constructed from received k' packets) into n packets and broadcasts in the network to achieve higher reliability. This scheme achieves adaptability for the changing network conditions and higher reliability but at a higher computation overload as each node has to decode and then again encode.

5 Summary and Discussion

In the previous sections, we have given an overview of erasure codes employed and the works from literature that employs the erasure codes for achieving reliability in wireless sensor networks. Table 1 gives the overview of schemes, erasure codes employed, whether they are based on end-to-end erasure coding transmission scheme or hop-by-hop erasure coding transmission scheme.

From the Table 1, we can see that majority of schemes has used only systematic Reed Solomon codes and fountain codes for achieving reliability in wireless sensor network data transmission. Only FBcast approach has analyzed the reliability in terms of updating the software firmware in the wireless sensor nodes whereas other schemes have concentrated only on disseminating data reliability from source nodes to sink nodes along with other routing protocols. Also, we can observe that except Kim et al. [19] all other schemes use simulations to validate their approaches.

Only reliability aspect has been investigated by the researchers and not on security aspect. Many devices are connecting to internet and the number is ever increasing. Employing any security parameter that will authenticate the device and as well as secure reconstruction of data at the receiver will be helpful in attacks against intruders, side channel attacks or spoofing kind of attacks.

6 Future Research Direction

Internet of Things communication offers a plethora of challenges that can be investigated. In this section we discuss the possible research directions that can be taken up in the context of communication in Internet of Things domain particularly reliability and security aspects.

6.1 Channel Model Analysis

The erasure codes used in the reliable communication are selected in random to improve the performance rather than information theoretically analyzing the channel on which the devices will communicate. Some models assume Binary Erasure Channel (BEC) for communication wherein received data may be corrupted or it may be received in error. But analyzing channel model in the Internet of Things (IoT) communication information theoretically provides new insights about the channel nature. It is advantageous to have the understanding of channel model [39] so that more robust and adaptive erasure codes can be used from the existing ones or new erasure codes can be proposed.

6.2 New Coding Techniques

Due to the new challenges in general digital communication scenario, new erasure codes are proposed in literature. The existing methods in Internet of Things (IoT) devices and wireless sensor networks use the Reed-Solomon codes and Fountain codes which were proposed more than two decades ago. Analyzing the other erasure codes understanding the nature of channel and communication requirement (as the communication channel may have wired or wireless or mobile technologies such as 2G/3G/4G as backbone network) in Internet of Things (IoT) devices will be useful as it involves many other underlying protocols. Also, it is worth to investigate performance employing joint erasure and error correcting codes at higher layers of communication stack and at physical layer respectively.

6.3 Security in Addition to Reliable Communication

Existing communication protocols and erasure codes concentrates mainly on reliable communication and not on secure communication. Using code based cryptosystems will be beneficial as it provides both reliability and security. Instead of using any higher end, computation intensive security protocols, it will be interesting to analyze code based cryptosystem that achieves the need of security at less computation overhead. Design of light weight cryptography or code based cryptosystem will be an interesting domain to explore.

6.4 Authentication of Devices

With many devices connecting to the Internet and such devices are used in critical applications, there is a need for authenticating the devices as many intruder devices can pose as actual devices and disrupt communication or may make the system to behave adversely. To overcome such challenges using authentication methods in addition to reliable communication will be beneficial. Authentication of devices can be made in several ways and using data hiding codes [40] or similar techniques will provide reliability in communication as well as security in terms of authenticity. One such method of employing code based cryptosystem to authenticate RFID tags is proposed by Maurya et al. [41] using binary codes.

7 Conclusion

In this chapter, we have given an overview of existing erasure codes in the wireless sensor networks which are integral part of Internet of Things communication. Construction methods of extensively used Reed-Solomon codes and Fountain codes are provided in addition to Decentralized erasure codes. Also, the basic communication paradigm of information transmission namely end-to-end transmission and hop-by-hop transmission are discussed in detail with and without emphasis on erasure codes. Many future research directions are discussed based on the erasure codes. This chapter serves as starting point for the researchers interested in working on reliability in Internet of Things communication with emphasis on erasure coding and wireless sensor networks.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. **38**(4), 393–422 (2002)
2. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Comput. Netw. **52**(12), 2292–2330 (2008)
3. Zheng, X.L., Wan, M.: A survey on data dissemination in wireless sensor networks. J. Comput. Sci. Technol. **29**(3), 470–486 (2014)
4. Albano, M., Chessa, S.: Replication vs erasure coding in data centric storage for wireless sensor networks. Comput. Netw. **77**, 42–55 (2015)
5. Mahmood, M.A., Seah, W.K., Welch, I.: Reliability in wireless sensor networks: a survey and challenges ahead. Comput. Netw. **79**, 166–187 (2015)
6. Moon, T.K.: Error Correction Coding: Mathematical Methods and Algorithms. Wiley-Interscience (2005)
7. Heimlicher, S., Karaliopoulos, M., Levy, H., May, M.: End-to-end vs. hop-by-hop transport under intermittent connectivity. In: Proceedings of the 1st International Conference on Autonomic Computing and Communication Systems, Autonomics '07, pp. 20:1–20:10. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium (2007). http://dl.acm.org/citation.cfm?id=1365562.1365589
8. Heimlicher, S., Nuggehalli, P., May, M.: End-to-end vs. hop-by-hop transport. SIGMETRICS Perform. Eval. Rev. **35**(3), 59–60 (2007)
9. Jing, C., Lixiang, L., Xiaohui, H., Fanjiang, X.: Hop-by-hop transport for satellite networks. In: Aerospace Conference, 2009, IEEE, pp. 1–7. IEEE (2009)
10. Ahlswede, R., Cai, N., Li, S.Y., Yeung, R.W.: Network information flow. IEEE Trans. Inf. Theory **46**(4), 1204–1216 (2000)
11. Dimakis, A.G., Prabhakaran, V., Ramchandran, K.: Decentralized erasure codes for distributed networked storage. IEEE/ACM Trans. Netw. (TON) **14**(SI), 2809–2816 (2006)
12. Luby, M.G., Mitzenmacher, M., Shokrollahi, M.A., Spielman, D.A.: Efficient erasure correcting codes. IEEE Trans. Inf. Theory **47**(2), 569–584 (2001)
13. Luby, M.G., Mitzenmacher, M., Shokrollahi, M.A., Spielman, D.A., Stemann, V.: Practical loss-resilient codes. In: Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing, pp. 150–159. ACM (1997)
14. MacKay, D.J.: Fountain codes. IEE Proc.-Commun. **152**(6), 1062–1068 (2005)
15. Shokrollahi, A.: Raptor codes. IEEE Trans. Inf. Theory **52**(6), 2551–2567 (2006). https://doi.org/10.1109/TIT.2006.874390

16. Maymounkov, P.: Online codes. Technical report, New York University, Tech. rep. (2002)
17. Luby, M.: Lt codes. In: The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings, pp. 271–280 (2002). https://doi.org/10.1109/SFCS.2002.1181950
18. Reed, I.S., Solomon, G.: Polynomial codes over certain finite fields. J. Soc. Ind. Appl. Math. **8**(2), 300–304 (1960)
19. Kim, S., Fonseca, R., Culler, D.: Reliable transfer on wireless sensor networks. In: 2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004, pp. 449–459. IEEE (2004)
20. Byers, J.W., Luby, M., Mitzenmacher, M., Rege, A.: A digital fountain approach to reliable distribution of bulk data. ACM SIGCOMM Comput. Commun. Rev. **28**(4), 56–67 (1998)
21. Stockhammer, T., Shokrollahi, A., Watson, M., Luby, M., Gasiba, T.: Application layer forward error correction for mobile multimedia broadcasting. In: Handbook of Mobile Broadcasting: DVB-H, DMB, ISDB-T and Media Flo, pp. 239–280 (2008)
22. Shokrollahi, A., Luby, M., et al.: Raptor codes. Found. Trends® Commun. Inf. Theory **6**(3–4), 213–322 (2011)
23. Dimakis, A.G., Prabhakaran, V., Ramchandran, K.: Distributed data storage in sensor networks using decentralized erasure codes. In: Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004, vol. 2, pp. 1387–1391. IEEE (2004)
24. Dimakis, A.G., Prabhakaran, V., Ramchandran, K.: Ubiquitous access to distributed data in large-scale sensor networks through decentralized erasure codes. In: Proceedings of the 4th International Symposium on Information Processing in Sensor Networks, p. 15. IEEE Press (2005)
25. Al-Awami, L., Hassanein, H.: Energy efficient data survivability for WSNs via decentralized erasure codes. In: 2012 IEEE 37th Conference on Local Computer Networks (LCN), pp. 577–584. IEEE (2012)
26. Al-Awami, L., Hassanein, H.S.: Distributed data storage systems for data survivability in wireless sensor networks using decentralized erasure codes. Comput. Netw. **97**, 113–127 (2016)
27. Lin, H.Y., Tzeng, W.G.: A secure decentralized erasure code for distributed networked storage. IEEE Trans. Parallel Distrib. Syst. **21**(11), 1586–1594 (2010)
28. Fragouli, C., Soljanin, E., et al.: Network coding applications. Found. Trends® Netw. **2**(2), 135–269 (2008)
29. Seah, W.K., Tan, H.P.: Multipath virtual sink architecture for wireless sensor networks in harsh environments. In: Proceedings of the First International Conference on Integrated Internet Ad hoc and Sensor Networks, p. 19. ACM (2006)
30. Wu, C., Ohzahata, S., Kato, T.: An adaptive redundancy-based mechanism for fast and reliable data collection in wsns. In: 2012 IEEE 8th International Conference on Distributed Computing in Sensor Systems (DCOSS), pp. 347–352. IEEE (2012)
31. Yang, Y., Wählisch, M., Zhao, Y., Kyas, M.: Raid the wsn: packet-based reliable cooperative diversity. In: 2012 IEEE International Conference on Communications (ICC), pp. 371–375. IEEE (2012)
32. Marchi, B., Grilo, A., Nunes, M.: DTSN: distributed transport for sensor networks. In: 12th IEEE Symposium on Computers and Communications, 2007. ISCC 2007, pp. 165–172. IEEE (2007)
33. Wen, H., Lin, C., Ren, F., Zhou, J., Yue, Y., Huang, X.: Retransmission or redundancy: transmission reliability study in wireless sensor networks. Sci. Chin. Inf. Sci. **55**(4), 737–746 (2012)
34. Ali, S., Fakoorian, A., Taheri, H.: Optimum reed-solomon erasure coding in fault tolerant sensor networks. In: 4th International Symposium on Wireless Communication Systems, 2007. ISWCS 2007, pp. 6–10. IEEE (2007)
35. Gehrke, J., Madden, S.: Query processing in sensor networks. IEEE Pervasive Comput. **3**(1), 46–55 (2004)
36. Trivedi, K.S.: Probability & Statistics with Reliability, Queuing and Computer Science Applications. Wiley (2008)
37. Srouji, M.S., Wang, Z., Henkel, J.: Rdts: A reliable erasure-coding based data transfer scheme for wireless sensor networks. In: Proceedings of the 2011 IEEE 17th International Conference on Parallel and Distributed Systems, pp. 481–488. IEEE Computer Society (2011)

38. Kumar, R., Paul, A., Ramachandran, U., Kotz, D.: On improving wireless broadcast reliability of sensor networks using erasure codes. In: International Conference on Mobile Ad-hoc and Sensor Networks, pp. 155–170. Springer (2006)
39. Shannon, C.E.: A mathematical theory of communication. ACM SIGMOBILE Mob. Comput. Commun. Rev. **5**(1), 3–55 (2001)
40. Moulin, P., Koetter, R.: Data-hiding codes. Proceedings of the IEEE **93**(12), 2083–2126 (2005)
41. Maurya, P.K., Pal, J., Bagchi, S.: A coding theory based ultralightweight RFID authentication protocol with CRC. Wireless Personal Communications, pp. 1–10 (2017). https://doi.org/10.1007/s11277-017-4546-z. http://dx.doi.org/10.1007/s11277-017-4546-z

Review: Security and Privacy Issues of Fog Computing for the Internet of Things (IoT)

Binara N. B. Ekanayake, Malka N. Halgamuge and Ali Syed

Abstract Internet of Things (IoT), devices, and remote data centers need to connect. The purpose of fog is to reduce the amount of data transported for processing, analysis, and storage, to speed-up the computing processes. The gap between, Fog computing technologies and devices need to narrow down as growth in business today relies on the ability to connect to digital channels for processing large amounts of data. Cloud computing is unfeasible for many internet of things applications, therefore fog computing is often seen as a viable alternative. Fog is suitable for many IoT services as it has enabled an extensive collection of benefits, such as decreased bandwidth, reduced latency, and enhanced security. However, Fog devices that are placed at the edge of the internet have met numerous privacy and security threats. This study aims to examine and highlight the security and privacy issues of fog computing through a comprehensive review of recently published literature of fog computing and suggest solutions for identified problems. Data extracted from 34 peer-reviewed scientific publications (2011–2017) were studied, leading to the identification of 49 different issues that were raised, in relation to fog computing. This study revealed a general agreement among researchers about the novelty of Fog computing, and its early stages of development, and identifies several challenges that need to be met, before its wider application and use reaches its full potential.

Keywords Fog computing · Edge computing · Mobile edge computing
Cloud computing · IoT · Security · Privacy · Big data

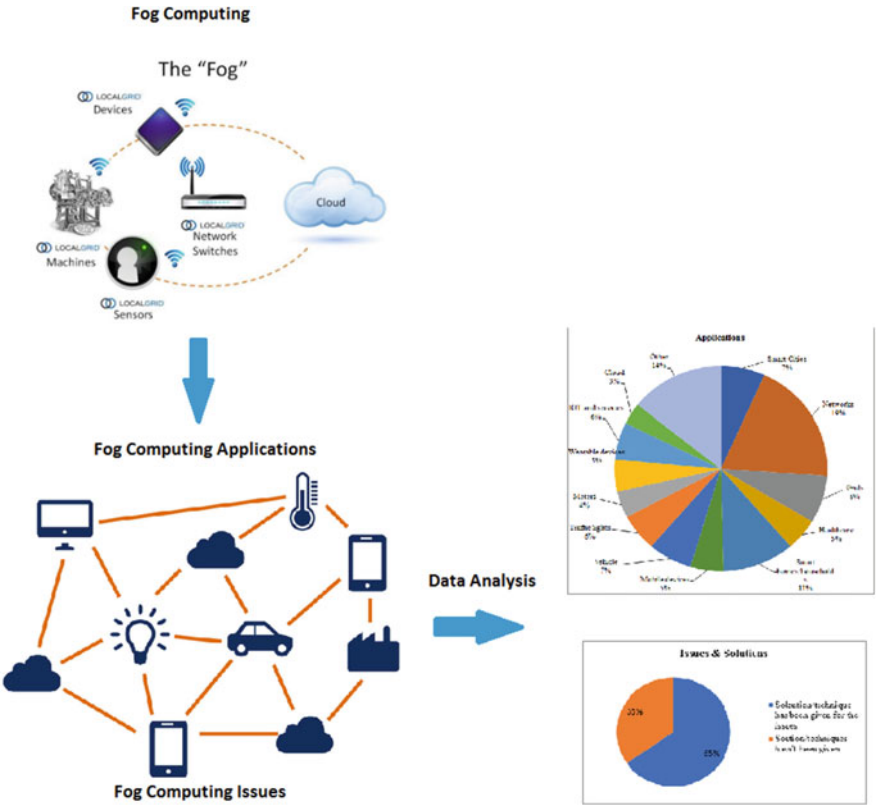
B. N. B. Ekanayake · M. N. Halgamuge (✉)
School of Computing and Mathematics, Charles Sturt University, Melbourne,
VIC 3000, Australia
e-mail: malka.nisha@unimelb.edu.au

B. N. B. Ekanayake
e-mail: ekanayake_navinda@yahoo.com

A. Syed
Department of Electrical and Electronic Engineering, The University of Melbourne,
Parkville, VIC 3010, Australia
e-mail: ASyed@studygroup.com

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_7

Graphical Abstract



1 Introduction

Fog computing, also known as fog networking or fogging, refers to a system of distributed computing infrastructure managed by smart devices, although it may still manage some in the cloud it still operates from smart phones. Thus, Fog can be viewed as a middle layer between cloud and the hardware system. Having a middle layer enables more efficient data searching, analysis, and storage, and minimizes the amount of data that needs to be transported to cloud.

Fog computing initially introduced a concept to describe the technology that forms a link between remote data centers and Internet of Things (IoT) devices. In "fog computing" or "edge computing," its system operates on the network ends, instead of hosting and working entirely from a centralized cloud. It is designed to facilitate the storage of data, computing, and other services between end devices, mostly IoT devices, and cloud computing data centers [1]. Thus, the fog proposes a

smart substitute that enables the processing of data locally, minimizing cloud involvement and enabling a smarter, more self-sufficient space between the data center and the network edge. Through a steep increase in the number of endpoint devices that communicate with the data center, it works as a latency that will escalate the network routine.

Fog computing offers many other advantages, including more efficient real-time processing, rapid and affordable scaling, and local content and resource pooling. The ability to link the IoT with the existing Internet computing infrastructure has attracted considerable interest among academics, as well as of the IT industry. Fog supports (IoT) are growing rapidly, as the traditional cloud will never be enough. Applications that use fog computing presently will be in greater demand in the future due to its storage capacity. The areas that rely on this system include, connected vehicles, smart cities, smart grids, mobile healthcare systems, wireless sensor, actuator networks, connected manufacturing, connected oil, gas systems, autopilot vehicles, smart homes are names of a few [2]. In terms of mobile big data analytics, big data processing is still a new subject to the big data architecture in the cloud and mobile cloud. Fog computing is able to deliver flexible resources with huge capacity data process system, without being the drawback of cloud's high latency [3]. Another advantage of fog computing is data security which helps with confidentiality issues. Through the adoption of user behavior profiling, it can help to mitigate insider data theft attacks in cloud and combs both technologies that can be used for masquerade detection [4].

At present, Fog computing is going through its early stages of development, and there are some challenging aspects that need to be addressed. Some of the concerns comprise issues that relate to privacy and programming models and abstracts; fog architecture; IoT support; storage constraints, network, and computing; resource provisioning and management; and allocated fog computing centers [2]. For example, when we consider data protection, many IoT device applications have issues in fog computing. Handling of messages created from IoT devices, and sent to the closest fog nodes is a difficult task. To overcome this difficulty, the data is usually separated into few parts, and sent to numerous fog nodes, as there is a greater chance of assuring reliability. Since processed data is a compound that makes it more secure, nonetheless its challenge is to decrypt or encrypt data on IoT device as there are inadequate resources [5]. In fog computing security applications such as Smart Traffic Lights and Connected Vehicle, Wireless Sensor and Actuator Networks, IoT and Cyber-physical systems and Software Defined Networks, have some issues such as Verification at the smart meters which are installed in consumers' house as well as numerous levels of the doorways. All the smart appliances and the smart meters have an IP address. A mischievous user can either report incorrect readings, interfere with the own smart meter or tricky IP addresses [6]. in the case of privacy in fog computing, there are some issues that have been identified as regarding as mobile device applications such as, places which have very weak surveillance and protection systems implemented into the fog node devices. They are local to the end user which makes it weak to security challenges. In contemplation of comprehending mischievous intentions become obtainable to fog

computing framework such as; eavesdropping, data hijack and in-the-middle-attack, etc. [7].

Recent research focuses on fog computing platform and applications. Researchers have briefly introduced Fog computing and after analyzing similar concepts of fog computing they have given a broader definition of fog computing [8]. Studies with surveys in fog computing concepts, applications, and issues that have been discussed the definitions of fog computing with similar concepts are given numerous types of issues that may find design and implement with fog computing systems [3]. Some authors have researched about research opportunities of the Fog and IoT, as their research indicates some future fog and its application in multiple industries and driving revolution through network operators like AT&T, IBM, and Huawei etc. [9].

New technologies such as Fog computing have the potential to offer many benefits, privileges, conveniences, and efficiencies to the users. However, technological advances often present new problems, and one major concern is the protection of privacy and security of data.

This study undertook a content analyses of the literature published in the area of security and privacy issues in fog computing, and the study is presented in four main parts: introduction, materials, and methods, results, discussion and conclusion.

2 Materials and Method

The overarching research question for this study is: What exactly can be done to resolve security and privacy issues of fog computing? The data for this research is gathered from peer-reviewed scientific studies, on Fog computing security and privacy issues, published in scientific journals, during the 2011–2017 period.

2.1 Raw Data Collection

The raw data collected for this study is from 34 peer-reviewed scientific publications which were published between the years of 2011–2017. The raw data presented in Table 3 specifies the variables used for the analysis including application, security issues, and algorithm.

2.2 Data Inclusion Criteria

In our analysis, we considered attributes such as author, applications, security and privacy issues and techniques. Nonetheless, in our analysis, we excluded studies when no complete attributes are disclosed and the publication is not published in peer-reviewed scientific publications.

2.3 Raw Data Analysis

Furthermore, the review articles were gathered according to the Fog computing security and privacy issues, applications and techniques used. We pooled and analyzed the used studies based on the influence of variables that they used for their studies. Then the comparison of issues, application and techniques was investigated to detect the trend of the security and privacy issues in fog computing for the Internet of Things (IoT) . The methodology implemented allows solving these issues, as the algorithm selected is purely for this purpose.

Table 1 provides a summary of the applications, of the issues addressed, by respective techniques and solutions, from 34 publications were outlined thoroughly.

3 Results

Table 2 shows the considerations and applications according to the year.

Table 4 is obtained from the data from articles that published between 2015 and 2017. Within this table, only the application headings have been mentioned, and the table above explains it in a broader application.

Figure 1, based on the data from Table 4, indicates that 19% of the issues were found from the network regarding applications, while smart homes/household's applications accounted for 11% of the issues. Figure 2 shows the percentages of issues in fog computing that have either been/not been suggested a solution of the total number of 49 issues, only 32 (65%) have been given solutions or techniques.

4 Discussion

This study reviewed 34 research papers published from 2011 to 2017, about security and privacy issues associated with Fog computing. An important observation is the relatively low number of articles published on Fog computing issues during the said period, highlighting the need for more research done in this important area.

Table 1 Fog computing security and privacy issues using data extracted from 34 peer-reviewed scientific publications (2011–2017)

No	Author	Consideration	Application	Security and privacy issues	Technique
1	Elkhati et al. (2017) [10]	Fog Infrastructure	Smart cities Home automation Data-driven industries	• Cloud has the advantages of cost-effectiveness and scalability • However, it isn't suitable for hosting all the applications • As a solution, off-loading some computations to the edge is proposed • The potential edge infrastructure is not well understood • There is no clarity on the types of applications that can be off-loaded	• Using micro-clouds which are collections of Raspberry Pis, to host a range of fog applications • They are particularly useful in network-constrained environments • The startup latency, I/O overhead, serving latency and hosting capability have been experimentally tested for several different applications
2	Hao et al. (2017) [11]	Heterogeneity	Universities Corporations Commonwealth organizations Personal households.	• There are heterogeneous of Fog nodes • There are no guarantees that the nodes will include comparable resources • Privacy and security issues are tied with the heterogeneity • They are mostly cast aside to accomplish interoperability and general functionality • When exchanging data to random devices, strict privacy policies and encryption create more complications	• A flexible software architecture, incorporating different design choices and user-specified polices, is described • Also presented in a design of WM-FOG. A computing framework for fog environments that makes use of the proposed software architecture, and an evaluation of their prototype system
3	Alrawais et al. (2017) [1]	Secure and Efficient Protocols	Smart grids Health care systems Wireless sensor networks Smart homes	• Most of the existing protocols such as time synchronization use wireless packet transmissions • They aren't suitable for resource-constrained IoT devices • Wireless transmissions and security computations utilize the major part of the energy budget	

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
4	Alrawais et al. (2017) [1]	Authentication	Smart grids Health care systems Wireless sensor networks Smart homes	• Authentication in IoT has several problems related to scalability and efficiency • Traditional authentication methods are inefficient • Therefore, a secure, scalable, efficient, and user-friendly solution to cope with resource-constrained IoT devices is required	• The application of a lightweight encryption algorithm between fog nodes and IoT devices will improve the efficiency of the authentication process • Fog has the capacity to create an opportunity for authentication solutions in IoT devices, particularly wearable devices
5	Hu et al. (2017) [7]	Privacy	Mobile devices	• Places which has very weak surveillance and protection are used to implemented the fog node devices. They are local to the end users • Therefore, they are weak to security challenges • In contemplation of comprehending mischievous intentions become obtainable to fog computing framework such as; eavesdropping, data hijack and in-the-middle-attack, etc.	
6	Alrawais et al. (2017) [1]	Updating Internet of things Devices	Smart grids Health care systems Wireless sensor networks smart homes	• Many IoT devices continue to be vulnerable to attacks • Management of security updates requires the design of remote software update capabilities • Vulnerable firmware may expose IoT devices to attacks that may not be protected by traditional security solutions such as firewalls	• Fog computing can be helpful in finding a solution • Fog can be used to identify vulnerabilities and to track firmware updates in IoT devices • However, updating billions of IoT devices is an unwieldy task

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
					• Security updating supply to IoT devices could be helped by the geo-distribution characteristic of fog computing
7	Wen et al. (2017) [12]	Reliability	Tablets Smart Phones Desktop PCs	• A large part of IoT applications is made of physical systems. Therefore, the assumptions made about fault and failure modes are weaker than those for Web-based applications • IoT applications are vulnerable to crash and timing failures due to low-sensor battery power, high network latency, environmental damage etc. Also, the uncertainty caused by potentially unstable and mobile system components increases difficulties in predicting and capturing system operation	• Therefore, methods are needed to measure an IoT application workflow's reliability, and also apply ways of enhancing it
8	Wen et al. (2017) [12]	Security Criticality	Sensors Computer chips Communication devices	• Multiple sensors, computer chips, and communication devices are integrated to enable overall communication, in the IoT environment • There may be multiple components in a given service, with each deployed in its own geographic location, This makes each a separate attack target • Fog nodes can be easily attacked, especially those in network enabled IoT systems, where attack vectors can include human-caused sabotage of network infrastructure, malicious programs	• Accurate evaluation of the security and risks to obtain a holistic measure of security and risk susceptibility, is critical • This becomes challenging when workflows are changing and adapting to runtime

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
				provoking data leakage, or even physical access to devices	
9	Wen et al. (2017) [12]	Dynamicity	IOT Services	• Software upgrades via fog nodes or the frequent join-leave behavior of network objects, will cause changes to internal properties and performance, which in turn can change the overall workflow execution pattern • Software and hardware aging is a concern with hand-held devices because that too will change workflow behavior and device properties • Changes in application performance are due to their transient and/or short-lived behavior within the system	• There is a need for automatic and intelligent reconfiguration of the topological structure and assigned resources within the workflow, and importantly, that of fog nodes
10	Wen et al. (2017) [12]	Fault Diagnosis and Tolerance	IOT Services	• Scaling a fog system increases the probability of failure • Certain software bugs or hardware faults that may be harmless at smaller scale or in testing environments, such as stragglers, can be devastating on system performance and reliability • Different fault combinations may occur at the scale, heterogeneity, and complexity	• Developers should incorporate redundant replications and user-transparent, fault-tolerant deployment and execution techniques in orchestration design
11	Xiao et al. (2017) [13]	Wireless Networking (Vehicular Fog Computing)	Vehicles	• Because of changes to the connectivity of vehicles and frequent changes in the network topology, the reliability of vehicle networking still causes problems,	• In vehicular fog computing scenarios where D2D and WLAN-based approaches co-exist, can be used. The co-scheduling must take into account the

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
				although a recent field test showed that 802.11p performs reliably most of the time • The existing performance analysis of D2D-based vehicular networking is based on simulation • Also, there are concerns that the cellular networks may not fully cover all the urban and suburban areas and the bandwidth resources of cellular networks are limited	performance requirements (i.e. bandwidth, latency) and the available capacity. Co-scheduling mechanisms have still not gained sufficient attention
12	Xiao et al. (2017) [13]	Application Provisioning (Vehicular Fog Computing)	Vehicles	• Vehicular and latency-sensitive mobile applications can be deployed on central cloud, cellular fog nodes and/or vehicular fog nodes • Pham et al. [14] discussed task scheduling between rented cloud nodes and owned fog nodes, and proposed a heuristic-based algorithm as a way of balancing between the make-span and the monetary cost of cloud resources • Mobility of fog nodes pose new challenges such as those arising from the simultaneous mobility of both vehicular fog nodes and their data sources and users. Another arises from the complexity of coordinating the scheduling of computing and communications resources	• Huang et al. [15] proposed an adaptive content reservation scheme to reserves the resources on cloud and fog nodes for real-time streaming to mobile devices. This takes account of the mobility of mobile devices • Lin et al. [16] idea was to develop a Cloud-fog that utilizes the idle machines of game players. It depends on the organization of fog nodes for rendering game videos and streaming them to nearby players

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
13	Xiao et al. (2017) [13]	Security and Privacy	Vehicles	• Another avenue of security and privacy challenge involves the mobility of fog nodes and the dynamic vehicular network topology • The distributed vehicular fog nodes serve as gateways to the hybrid cloud consisting of fog nodes and central cloud • A hacker who gets access to any of the fog nodes, can send malicious messages and illegitimate commands that can seriously harm the reliability of the network services • Attackers can also duplicate the personal data of the clients by hacking into the vehicular fog nodes, seriously threatening the clients' privacy	• As a way of protecting against malicious attacks, Mtibaa et al. [17] implemented Honey Bots to detect and track the activities of malicious communications in D2D network. Public key infrastructure [18] and Diffie-Hellman key exchange [19] have been elaborated to enhance the security of authentication in smart grid networks • However, the high expansion rate of the vehicle fog platform may continue to cause more security problems in the future • More effective encryption methods and powerful middle-ware need to be developed for security-ware of fog computing to address such challenges
14	Niranjanamurthy et al. (2016) [20]	Compute/ Storage limitation	Wireless access points	• There are attempts currently to expand storage capabilities with devices that are smaller, have better energy-efficiency and power • For example, a present day phone has more power than many of the desktops used 15 years ago • For non-consumer devices more and more improvements are being made	

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
15	Niranjanamurthy et al. (2016) [20]	Management	Wireless access points Smart traffic lights	• IoT/ubiquitous computing nodes and applications running on top must be put in place and configured to operate as required, to set up communication routes across end nodes • FOG be dependent much on distributed (scalable) management mechanisms • Possibly there are billions of small devices that need such configuration • As fog and asymptotic declarative configuration techniques become more common, it is unlikely that there will be complete control achieved of the whole • They are to be examined at this unparalleled scale	
16	Niranjanamurthy et al. (2016) [20]	Standardization	Wireless access points Smart traffic lights Smart cities	• Presently, there are no standardized mechanisms • The availability of the different members of the network (terminal, edge point…) could be announced, so that others can send their software to be run	
17	Niranjanamurthy et al. (2016) [20]	Accountability/ Monetization	Wireless access points Smart traffic lights	• The facility for users to share their spare resources to host applications will help in the development of new business models around the fog concept • It needs the creation of an incentive scheme • Inducements can be financial (e.g. unlimited free data rates)	

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
				• It is difficult to regulate if a given device is hosting a section (droplet) or not at a given time, in the absence of a central controlling entity in the fog	
18	Niranjanamurthy et al. (2016) [20]	Discovery/Sync	Wireless access points Smart traffic lights	• For applications using devices will need agreed, central point (e.g. if there are too few peers in the storage application to establish an upstream‖ backup)	
19	Khalid (2016) [21]	Network fortification	Smart home/ cities Smart meters connected vehicles Generous scale remote sensor networks	• Because of the extensive use of wireless networking in fog computing, remote framework or wireless network security is a major concern in fog computing • The examination range of wireless networks can be subjected to attacks such as jamming and sniffing • Fog nodes are located at the edge of Internet, which certainly passes on an overpowering load to the network management • As cloud servers are scattered all through the framework/network edge, there is a high cost associated with their maintenance	• The control of software defined networks can enhance the execution and management • Techniques to introduce adaptability of network and lessening expenses, applicable to fog computing • Application of SDN technique in fog computing will offer fog computing security novel capabilities and prospects
20	Khalid (2016) [21]	Access Control	Smart home/ cities Smart meters connected vehicles	• Standard access control is ordinarily handled in the same trust region • For out-sourced data in cloud computing, the access control is, in general, cryptographically realized	• Some open-key based approaches for action are proposed as an effort to fulfill fine-grained access control • Other authors [22] (e.g.: [23]) a fine-grained data access control

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
			Generous scale remote sensor networks	• Symmetric key based course of action is not versatile in key management	preparation generated on attribute-based encryption (ABE) • Furthermore, to support secure joint exertion and interoperability among various resources [23], Dsouza et al. [22] proposed a policy-based resource access control in fog computing
21	Kumar et al. (2016) [24]	Network Security	Smart metering system Smart wearable devices Smart cities	• Fog computing is affected by attacks like sniffer, spoofing, jamming etc. • Usually such attacks are targeted between the fog node and the centralized system • Fog nodes are at the edge of the network. Therefore, it increases the burden for the network manager	• SDN (Software Defined Networking) can be used as an approach for network managers to work at the low level of abstraction for network services • It can help in management, increase scalability of network as well as reduce costs of fog computing • To watch the traffic, we can use Intrusion Detection System and Network Monitoring. To prevent attack Prioritization system and Traffic Isolation can be used by shared resources, Network resource access control system helps to get access control on SDN (Open Control), Network Sharing System can help the fog node router to be open to guests considering the security issues as well
22	Kumar et al. (2016) [24]	Data Security	Smart metering system Smart wearable devices	• It is difficult to maintain data Integrity, as data may be lost or be modified • The data uploaded to the fog node can also be used by a third party	• To provide data verifiability, confidentiality and Integrity combination of searchable encryption techniques and homomorphic encryption can be used

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
			Smart cities		• These techniques will be useful for ensuring that clients do not store data on untrusted servers
23	Kumar et al. (2016) [24]	Access Control	Smart metering system Smart wearable devices Smart cities	• Access control is significant as a tool to provide system's security and ensure user privacy. Usually access control is labeled to the same domain, but better to implement cryptographically, because of the distributive nature of cloud computing	• Many solutions have been proposed for these problems • One of them by Yu et al. recommends that the access control is based on Attribute-based encryption (ABE) • Others proposed solutions are theory-based with suggestions for policy based access control mechanism is applied to handle the heterogeneous nature of fog computing
24	Petac et al. (2016) [25]	Security	Routers Switches IP based video cameras	• Security of stored data is one of the major concerns in the fog computing security area • All data, in this environment, is stored within a third party • This makes the implementation of traditional security solutions A more difficult proposition • The use of cryptographic methods for data storage, although is better in terms of security, creates problems for the users, because the latter will not have any control over their own data.	• Proposed, Adaptive Fog Computing Node Security Profile (AFCNSP) based on security Linux solutions • In the case of fog nodes, the decision about what kind of security is necessary is important. Without authorization, a fog node application cannot access the services, the data and the network of other application • Simplified Mandatory Access Control Kernel –, Discretionary Access Control – DAC, Cynara and net filter are Linux kernel security modules that protects data and process interactions from malicious manipulation by using a set of custom mandatory access control rules

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
25	Wang et al. (2015) [26]	Geographical distribution	Wireless Sensor and Actuator Networks Decentralized Smart Building Control Software Defined Networks IoT and Cyber physical systems	• A huge number of nodes and sensors are needed In Fog computing environment • It helps to release the restrictions of bandwidth for advanced communicational speed in the real-time interactions and network	
26	Lee et al. (2015) [5]	Data Protection	Internet of things devices Smart watches	• Messages which are created from IoT devices were sent to the closest fog nodes • To manage many data on IoT devices are challenging • To handle the data, it is separated into few parts as well as sent to numerous of fog nodes • Without revealing the substances of data should be analyzed • The reliability of data must be assured, while it's been distributed • Processed data is compound • It is challenging to decrypt or encrypt data on IoT device as there are inadequate resources	
27	Lee et al. (2015) [5]	Data Management	Internet of things devices Smart watches	• To make it challenge to find data's location, the fog nodes are geographically allocated	

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
				• In the other areas also, the users require equal services • It is very challenging to find out whether the node delivers the equal service • By having replicated files, it will cause a waste of resources • Result from a wrong approach by a mischievous user fog nodes got distributed on security issues of personal information	
28	Yue Shi et al. (2015) [27]	Attack Detection	Mobile devices Wi-Fi access points	• Fog computing provides new opportunities to detect unusual behavior or to identify malicious attacks • A detection system can be signature- or anomaly-based • A newly detected pattern can be checked against existing or possible patterns	• Cloudlet mesh architecture through the collaboration of cloudlet members to monitor and detect malware, malicious attacks, and other threats, is proposed • Such a collaborative intrusion-detection technique will be applicable between fog nodes to monitor IoT environments and their surroundings
29	Stojmenovic et al. (2015) [28]	System security	Smart grids Smart traffic lights in vehicular networks Software defined networks	• Man-in-the-middle attack • Fog computing is typically vulnerable to the man-in-the-middle attack • Gateways serving as Fog devices may be altered or replaced completely, in such attacks • For an example, Star Bar or KFC customers being connected to mischievous access points which provide misleading service set identifiers as public legitimate ones. once the attackers	• Complete avoidance and defending against man-in-the-middle attacks are difficult tasks • However, building an anti-tampering mechanism in the Fog device could be done as a potential solution

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
				take control of the gateways, users' private communications will be stolen	
30	Stojmenovic et al. (2015) [28]	Authentication and authorization	Smart grids Smart traffic lights in vehicular networks Software defined networks	• The connection between Fog and Cloud is fragile, and as such can be easily disrupted. In such instances, authentication of users could be a problem as it is deployed on the Cloud server	• A new mechanism is introduced for user authentication when there is no connection to the Cloud server. It is known as Stand-Alone Authentication (SAA)
31	Yi et al. (2015) [29]	Secure Data Storage	Wireless sensor networks Smart home/ cities Smart meters	• Just like in cloud computing security threats are there in fog computing too because user data is outsourced and user's control over data is handed over to fog node • Maintenance of data integrity is a problem because the outsourced data may be lost or changed • Furthermore, the data may be vulnerable to abuse by unauthorized parties for fraudulent	• As a means of countering such threats, for auditable information storage service there have been proposals • Techniques being searchable encryption and homomorphic encryption can be used • Public auditing for data stored in cloud, which relies on a third-party auditor (TPA), using random mask and technique homomorphic authenticator has been proposed to ensure privacy
32	Yi et al. (2015) [29]	Data Privacy	Wireless sensor networks Smart home/ cities Smart meters	• Privacy-regulating algorithms in the fog network are run between the fog and cloud, while those algorithms at the end devices are usually resource prohibited • Sensitive data generated by sensors and end devices are usually collected by fog nodes at the edge	• Privacy-assurance aggregation at the local gateways without decryption can be enhanced by techniques such as homomorphic encryption

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
33	Chiang et al. (2015) [30]	Trustworthiness and security	5G, home/ personal networking Internet of Things	• Fog may be useful in enhancing security in some cases. However, there is also the possibility of new security challenges, at times • Since it is rather easier to hack into client software, perhaps consideration of security at hardware level on devices could be worthwhile	
34	Yi et al. (2015) [8]	Security and Privacy	Health Data Management Smart Home Smart Grid Smart Vehicle	• Admittedly, ensuring security and privacy is one of the biggest challenges in every stage of fog computing platform design	• These problems can be overcome by applying intrusion detection system, and access control which require assistance from every layer of the platform
35	Yi et al. (2015) [8]	Network Management	Health Data Management Smart Home Smart Grid Smart Vehicle	• One of the major issues in fog computing is network management which can be tackled by using SDN and NFV techniques • However, the integration of SDN and NFV into fog computing is no easy task • The need to re-design the north-bound and south-bound, the east-west-bound APIs to include necessary fog computing primitives pose many challenges	
36	Yi et al. (2015) [8]	Fight with Latency	Health Data Management Smart Home Smart Grid Smart Vehicle	• Since fog computing is aiming delay-sensitive applications and services, high inactivity will have negative effects on user satisfaction and experience	

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
				• Many factors would introduce high prospective into service or application implementation on fog computing platforms	
37	Yi et al. (2015) [3]	Fog networking	Wireless network virtualization Privilege traffic reservation Frequency hopping communication	• Located at the edge of Internet, fog network is heterogeneous. Fog network's task is to connect all its components. However, managing such a network, maintaining connectivity and providing services, especially in the scenarios of the Internet of Things (IoT) , is a challenging task	• Software-defined networking (SDN) and network function virtualization (NFV) are two of the emerging techniques that have been proposed for use to create flexible and easy maintaining network environment
38	Yi et al. (2015) [3]	Connectivity (Quality of Service)	ad-hoc wireless sensor networks Smart phones	• Opportunities for cost-cutting, data trimming and connectivity expansion are provided by network relaying, partitioning and clustering. For example, due to the coverage of rich-resource fog nodes an ad-hoc wireless sensor network can be partitioned into several clusters (cloudlet, sink node, powerful smartphone, etc.)	• Work [31] proposes an online AP association strategy that not only achieves a minimal throughput, but efficiency in computational overhead. Similarly, the selection of fog node from end user will heavily impact the performance • A subset of fog nodes can be selected as relay nodes for optimization goals of maximal availability of fog services limited to a certain area or a single user. It can also include constraints such as delay, throughput, connectivity, and energy consumption

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
39	Yi et al. (2015) [3]	Reliability (Quality of Service)	Clustering computing Grid computing Cloud Sensor networks	• Periodical check-pointing to resume after failure, rescheduling of failed tasks or replication to exploit executing in parallel can improve reliability. But check pointing and rescheduling may not suit the highly dynamic fog computing environment since there will be latency period, and adaptation to changes may be slow	• Replication appears to be more promising, but relies on multiple fog nodes working together
40	Yi et al. (2015) [3]	Capacity (Quality of Service)	Cloud Sensor networks	• Capacity has two folds: network bandwidth and storage capacity • To achieve high bandwidth and efficient storage utilization, it is important to investigate how data are placed in fog network. In computation, data locality is very important. There are similar works in the context of cloud, and sensor networks • This is an issue that can cause problems in fog computing. For example, a fog node may need to compute on data that is distributed in several nearby nodes. Since computation requires the finish of data aggregation, it can delay the services	• The problem can be solved by leveraging user mobility pattern and service request pattern to place data on suitable fog nodes to either minimize the cost of operation, the latency or to maximize the throughput • Fog computing offers the possibility of dynamic data placement and large overall capacity. Therefore, it may need to redesign search engine which can process search query of content scattered in fog nodes • It would be of interest to redesign cache on fog node to exploit temporal locality and broader coverage to save network bandwidth and reduce delay, while there is existing work of cache on end device and cache on edge router

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
41	Yi et al. (2015) [3]	Delay (Quality of Service)	Streaming Complex event processing	• Latency-sensitive applications, such as streaming mining or complex event processing, are typical applications which need fog computing to provide real-time stream processing rather than batch processing	• According to Hong et al. an opportunistic spatio-temporal event processing system can be used to address the latency issue. This system predicts future query region for moving users, and the event processing is done early to make information available when the user reaches a different location
42	Yi et al. (2015) [3]	Interfacing and programming model	Developers Internet applications	• To ease the developers' efforts to port their applications to fog computing platform, unified interfacing and programming model are needed • Application-centric computing will be a useful fog computation model. Here, the components will allow suitable optimizations and application-aware for different kinds of applications • It is hard for a developer to put together heterogeneous resources, and hierarchical dynamic to build well-matched applications on diverse platforms	• Hong et al. [35] suggest a high-level programming model for future Internet applications with on-demand scaling. They are large-scale geo-spatially distributed and latency sensitive • However, their system depends on a tree-based network hierarchy where fog nodes have fixed locations. Also needed are more general schemes for diverse networks with fog nodes of dynamic mobility
43	Yi et al. (2015) [3]	Computation Offloading	Mobile devices	• Dealing with dynamics is one of the main challenges in offloading in fog computing, The dynamics are three fold (1) radio/wireless network access, (2) nodes in the fog network, (3) resources in the fog • The combining of fog and cloud actually would give a three-layering construction:	

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
				device-fog cloud. However, computation offloading in such infrastructures can lead to both new challenges and opportunities • To choose the type of granularity for offloading at different hierarchy of cloud and fog; how to make offloading conclusions to adjust active changes in resources, fog devices and network, etc. also how to partition application to offload on cloud and fog with dynamism are the major questions	
44	Yi et al. (2015) [3]	Accounting, billing and monitoring	Cloud service providers End users Internet service providers	• The success of Fog computing enterprise depends on a sustainable business model Following parties are required in fog computing service: – Cloud service providers – End users – Internet service providers or wireless carriers • Therefore, there are many issues to be resolved for the introduction of a "Pay-as-you-go" system • For example, in billing, decisions have to be made on how to set the price for different resources and how to set the fraction of the payment to go to different parties of fog • Methods of accounting and monitoring the fog in different granularity are	User Incentives • "Join Fog computing with private local cloud at the edge" is an innovative business model in this field • Local private clouds with computation and storage capacity., can be deployed at the edge of Internet, although private cloud is aiming at service to private parties only • It will also be possible to lease spare computation and storage facilities to fog service provider for a fee which will also help in reducing the costs

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
				required for enforcing such pricing models • Another interesting feature will be the ability to do dynamic do pricing in fog computing services to maximize revenue and utilization, just similar to traditional industries do in airline ticketing, car rental or hotel bookings	
45	Yi et al. (2015) [3]	Intrusion detection		• There are new opportunities to investigate how fog computing can help with the centralized cloud side and the interruption recognition on both client side • However, high mobility fog computing environment, large-scale, implementing intrusion detection in geo-distributed pose some challenges	
46	Yi et al. (2015) [3]	Privacy	Cloud Smart grid Wireless network Online social network	• Internet users are anxious about the risk of privacy leakage (data, location or usage) • There are many suggested techniques for preserving privacy for different scenarios such as cloud, smart grid, wireless network and online social networks • In fog network, Privacy-preserving algorithms can be located between the cloud and fog since storage and computation are acceptable for both sides while at the end devices those algorithms are usually resource-prohibited	• Privacy-preserving aggregation at the local gateways without decryption. can be facilitated by techniques like homomorphic encryption. For aggregation and statistical queries, differential privacy can be applied to ensure non-disclosure of an arbitrary single entry in the data set, without affecting privacy

(continued)

Table 1 (continued)

No	Author	Consideration	Application	Security and privacy issues	Technique
				• Data is produced by end devices and sensor, fog node at the edge generally gather data from there	
47	Yi et al. (2015) [3]	Access control	smart devices cloud	• Because of its ability ensure security, access control is a reliable tool on smart devices, and cloud • The way to design control spanning client-fog-cloud, to meet the goals and resource constraints at different levels, is another issue in fog computing	
48	Stojmenovic (2014) [6]	Security	Smart Traffic Lights and Connected Vehicle Wireless Sensor and Actuator Networks IoT and Cyber-physical systems Software Defined Networks	• Verification at the smart meters which are installed in consumer's house as well as numerous levels of the doorways • All the smart appliances and the smart meters has an IP address • A mischievous user can either report incorrect readings, interfere with the own smart meter or tricky IP addresses	
49	Bonomi (2012) [32] Bonomi (2011) [33]	Programmability	Programmers Advertising filed Entertainment and other applications	• The control of application lifecycle is one of the challenges in cloud environments	• As there are many small functional units (droplets) distributed over many devices, they require the correct placing of the abstractions • It may not be necessary for programmers to deal with these issues

Table 2 Considerations and applications according to the year

Year	Consideration	Application
2017	Fog Infrastructure	• Smart cities • Home automation • Data-driven industries
	Heterogeneity	• Universities • Corporations • Commonwealth organizations • Personal households
	Secure and Efficient Protocols	• Smart grids • Health care systems • Wireless sensor networks • Smart homes
	Authentication	• Smart grids • Health care systems • Wireless sensor networks • Smart homes
	Privacy	• Mobile devices
	Updating Internet of things Devices	• Smart grids • Health care systems • Wireless sensor networks • Smart homes
	Reliability	• Tablets • Smart Phones • Desktop PCs
	Security Criticality	• Sensors • Computer chips • Communication devices
	Dynamicity	• IOT Services
	Fault Diagnosis and Tolerance	• IOT Services
	Wireless Networking (Vehicular Fog Computing)	• Vehicles
	Application Provisioning (Vehicular Fog Computing)	• Vehicles
	Security and Privacy	• Vehicles
2016	Compute/Storage limitation	• Wireless access points
	Management	• Wireless access points • Smart traffic lights
	Accountability/Monetization	• Wireless access points • Smart traffic lights
	Standardization	• Wireless access points • Smart traffic lights • Smart cities
	Discovery/Sync	• Wireless access points • Smart traffic lights
	Network Fortification	• Smart home/cities • Smart meters connected vehicles

(continued)

Table 2 (continued)

Year	Consideration	Application
		• Generous scale remote sensor networks
	Access Control	• Smart home/cities • Smart meters connected vehicles • Generous scale remote sensor networks • Smart metering system • Smart wearable devices
	Network Security	• Smart metering system • Smart wearable devices • Smart cities
	Data Security	• Smart metering system • Smart wearable devices • Smart cities
	Security	• Routers • Switches • IP based video cameras
2015	Geographical Distribution	• Wireless Sensor and Actuator Networks • Decentralized Smart Building Control • Software Defined Networks • IoT and Cyber physical systems
	Data Protection	• Internet of things devices • Smart watches
	Data Management	• Internet of things devices • Smart watches
	Attack Detection	• Mobile devices • Wi-Fi access points
	System Security	• Smart grids • Smart traffic lights in vehicular networks • Software defined networks
	Authentication and Authorization	• Smart grids • Smart traffic lights in vehicular networks • Software defined networks
	Secure Data Storage	• Wireless sensor networks • Smart home/cities • Smart meters
	Data Privacy	• Wireless sensor networks • Smart home/cities • Smart meters
	Trustworthiness and Security	• 5G, home/personal networking • Internet of Things
	Security and Privacy	• Health Data Management

(continued)

Table 2 (continued)

Year	Consideration	Application
		• Smart Home • Smart Grid • Smart Vehicle • Cloud • Wireless network • Online social network
	Network Management	• Health Data Management • Smart Home • Smart Grid • Smart Vehicle
	Fight with Latency	• Health Data Management • Smart Home • Smart Grid • Smart Vehicle
	Fog networking	• Wireless network virtualization • Privilege traffic reservation • Frequency hopping communication
	Connectivity (Quality of Service)	• Ad-hoc wireless sensor networks • Smart phones
	Reliability (Quality of Service)	• Clustering computing • Grid computing • Cloud • Sensor networks
	Capacity (Quality of Service)	• Cloud • Sensor networks
	Delay (Quality of Service)	• Streaming • Complex event processing
	Interfacing and Programming Model	• Internet applications • Developers
	Computation Offloading	• Mobile devices
	Accounting, Billing and Monitoring	• Internet service providers • Cloud service providers • End users
	Access Control	• Smart devices • Cloud
2014	Security	• Smart Traffic Lights and Connected Vehicle • Wireless Sensor and Actuator Networks • IoT and Cyber-physical systems • Software Defined Networks
2011	Programmability	• Programmers • Advertising filed • Entertainment and other applications

Table 3 Category of applications, and their description: Data from 2015 to 2017

Category	Description
Smart Cities	Smart cities
Networks	Wireless sensor networks, Wireless access points, Wireless Sensor and Actuator Networks, Software Defined Networks, Wireless networks, Wireless network virtualization, Wi-Fi access points, Generous scale remote sensor networks, 5G, home/personal networking, ad-hoc wireless sensor networks
Grids	Smart grids Grid computing
Healthcare	Healthcare systems Health Data Management
Smart homes/ Households	Smart homes Personal households/Home automation Decentralized Smart Building Control
Mobile devices	Mobile devices Tablets/Smart Phones
Vehicles	Vehicles, Smart meters connected vehicles, Smart Vehicle
Traffic lights	Smart traffic lights in vehicular networks Smart traffic lights Privilege traffic reservation
Meters	Smart meters Smart metering systems
Wearable devices	Smart wearable devices Smart watches
IOT	IOT Services Internet of Things Internet applications IOT and Cyber physical systems
Cloud	Cloud
Other	Data-driven industries, Universities, Corporations, Common wealth organizations, online social network, Desktop PCs, Sensors, Computer chips, Communication devices, Routers, Switches, IP based video cameras, Frequency hopping communication, Clustering computing, Streaming, Complex event processing, Developers

As seen from the distribution of issues (Fig. 1), shows that 19% of the research is related to network or in other words the communication between applications. Smart homes/household's applications are ranked second in frequency, as 11% of the papers are related to research in frequency. Thus, most of the issues highlighted in the papers are related to the network section, and the reason for this is because fog computing is still in its early stages.

There are 49 issues that emerge from the publications that were reviewed. The highlighted issues include lack of clarity on types of applications [10] and problems of scalability and efficiency [1]. The major concern, however, is related to privacy and security of data storage eg: [13, 17, 18, 41]. Some authors draw attention on

Table 4 Number of applications has been used in each year (2015–2017)

Application	Number of applications	
Smart Cities	**Year**	**Number**
	2017	2
	2016	4
	2015	2
	Total	8
Networks	**Year**	**Amount**
	2017	3
	2016	7
	2015	13
	Total	23
Grids	**Year**	**Amount**
	2017	3
	2015	6
	Total	9
Healthcare	**Year**	**Amount**
	2017	3
	2015	3
	Total	6
Smart homes/Households	**Year**	**Amount**
	2017	5
	2016	2
	2015	6
	Total	13
Mobile devices	**Year**	**Amount**
	2017	3
	2015	3
	Total	6
Vehicles	**Year**	**Amount**
	2017	3
	2016	2
	2015	3
	Total	8
Traffic lights	**Year**	**Amount**
	2016	4
	2015	3
	Total	7
Meters	**Year**	**Amount**
	2016	3
	2015	2
	Total	5

(continued)

Table 4 (continued)

Application	Number of applications	
Wearable devices	**Year**	**Amount**
	2016	3
	2015	3
	Total	6
IOT and services	**Year**	**Amount**
	2017	3
	2015	4
	Total	7
Cloud	**Year**	**Amount**
	2015	4
	Total	4
Other	**Year**	**Amount**
	2017	8
	2016	3
	2015	6
	Total	17
Total no of applications		**119**

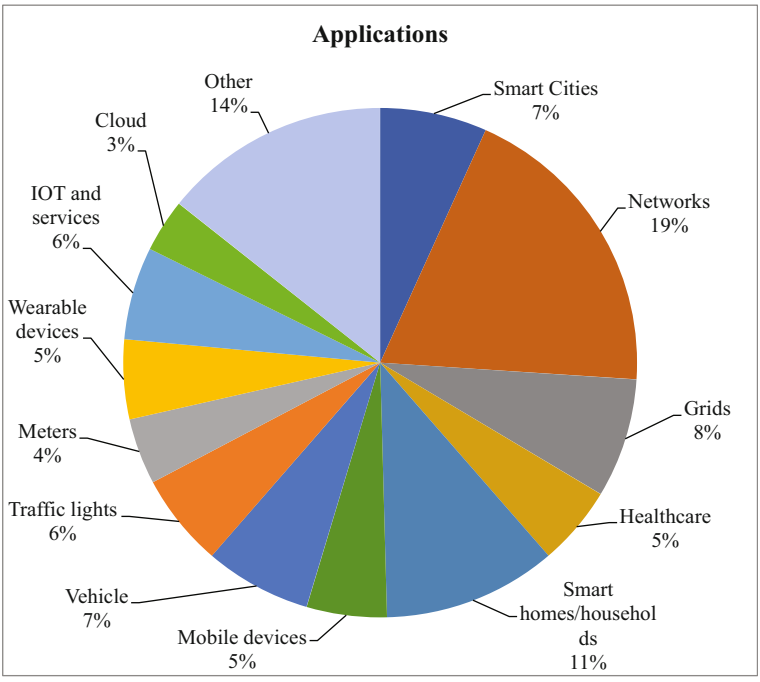


Fig. 1 The percentages of the applications that has been used in fog computing

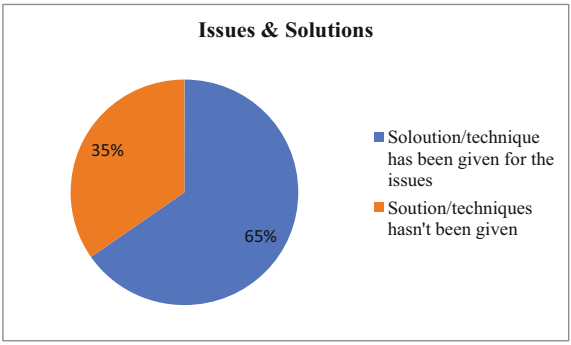


Fig. 2 The percentages of solutions and techniques have not been related to the issues in fog computing

attacks and crashes that occur because of privacy as well [1, 12, 42, 43]. However, it has also been suggested that Fog computing itself may be capable of offering solutions to such attacks [1]. Many authors have drawn attention to the need of configuration and topological structural changes to Fog nodes [12, 13, 15, 16]. The current efforts to expand storage capabilities have also been discussed in the bulk of the papers examined [20].

The 32 out 49 of the issues raised, have a solution or have been suggested a particular technique, while the rest of the issues have not been priorities with a solution. Some articles failed to suggest any solutions for one-third of the identified problems and highlight the need for more research in this area. A similar observation made by Baccarelli et al. [34] discussed the energy-efficient networking system and computing architectures, after carrying out a statistical search on Google, using keywords such as Internet of Everything (IOE) and Fog. The search was limited to the research contributions published during 2011–2016, and the authors drew attention to two main findings. Firstly, there has been a move towards a combination of the Fog and Internet of Everything paradigms. Secondly, as for the actual integration of the Fog and Internet of Everything pillar paradigms, there is still no up-to-date research or technical contribution that has been published.

Research opportunities related to the Fog and IoT was highlighted with the focus Mobile Cloud Computing (MCC) and emphasizes that the Mobile cloud computing is an infrastructure where both data storage and processing occur in an exterior platform for mobile devices, by sending computations and data storage from mobile phones to cloud [5]. Chiang et al. [9] From their analysis, the authors concluded that in the future, fog will be applicable in multiple industries, creating a revolution through the entire industry, including network operators like AT&T, Network equipment vendors like Huawei, System integrators like IBM, and end user experience providers like Toyota etc.

Considering these emerging themes analyzed in the current review, the summary of the papers seen in (Table 1) shows that authors have discussed a multitude of issues regarding fog computing, and each paper discusses one or two aspects of a particular issue, some taking an extra step, and suggests some solutions to the identified problems, while others have limited views of the issue only. For example,

Yi et al. [8] have discussed fog computing platform and applications. They have briefly introduced fog computing, and following the analyzes, several concepts on fog computing have emerged as the authors gave a broader definition of fog computing. In the paper, they have discussed the challenges and design goals for fog computing. Then they designed an implementation of prototyping platforms for fog computing. Finally, they tested prototypes and platforms in smart home applications.

Yi et al. [3] have done a survey of fog computing concepts, applications and issues. The survey has been discussed about the definitions of fog computing with similar concepts and has given numerous issues that may arise when designing and implementing fog computing systems. The authors have discussed the challenges, new opportunities and techniques in fog computing and issues such as quality of service, privacy, security, resource management and interfacing, have been emphasized. As they indicate, fog computing will evolve with rapid development in the underlying Mobile cloud, NFV, radio access techniques, IoT, SDN, VM and edge devices. Besides, studying diverse techniques for fog computing, Big Data databases [36, 37], security [38, 39] prediction and pattern analysis [40] could be an attractive avenue to explore in future.

5 Conclusion

In summary, the review and analysis of data obtained from 34 published research articles (papers regarding issues in fog computing) from 2011 to 2017, have been evaluated, and 49 distinct issues have been identifying and addressed, relating to fog computing. From 49 issues, 32 issues have been provided with a solution technique and the remaining of the 17 issues have not been given a solution or a technique, suggesting the need for greater attention of researchers to these areas. This study also found that 19% of the fog computing-related to applications issues of through networks. Additionally, 11% of the issues were found from smart homes/household's applications as they are the second-biggest issues in those articles. Consequently, it is clear that most of the issues are regarding the network section is from the application of fog computing. When reviewing the research articles, it was quite clear that most of the studies focused on security or privacy and there were no issues that were concerned with networks applications. Furthermore, past research papers have not shown much concern about issues of fog computing applications. All the authors have only discussed a particular aspect of fog computing, but no one discusses the issues in a single document. This study has highlighted the areas that need attention and it is evident that Fog computing is still an under-researched today, and is still not well understood and researched, despite the significant role it plays. There are challenges in establishing solutions to the issues that are raised, but finding solutions require urgent attention.

Author Contribution Binara N. B. Ekanayake and Malka N. Halgamuge conceived the study idea and developed the analysis plan. Binara N. B. Ekanayake analyzed the data and wrote the initial paper. Malka N. Halgamuge helped to prepare the figures and tables, and finalizing the manuscript. All authors read the manuscript.

References

1. Alrawais, A., Alhothail, A., Hu, C., Cheng, X.: Fog Computing for the Internet of Things: Security and Privacy Issues. IEEE Computer Society (2017)
2. Chen, S., Zhang, T., Shi, W.: Fog Computing. IEEE Computer Society (2017)
3. Yi, S., Li, C., Li, Q.: A survey of fog computing: concepts, applications and issues. In: MobiHoc Mobile and Ad Hoc Networking and Computing. ACM, New York (2015)
4. Stolfo, S.J., Salem, M.B., Keromytis, A.D.: Fog computing: mitigating insider data theft attacks in the cloud. In: IEEE Symposium on Security and Privacy Workshops (2012)
5. Lee, K., Kim, D., Ha, D., Rajpu, U., Oh, H.: On Security and Privacy Issues of Fog Computing supported Internet of Things Environment. IEEE Computer Society (2015)
6. Stojmenovic, I., Wen, S.: The Fog Computing Paradigm: Scenarios and Security Issues. Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1–8 (2014)
7. Hu, P., Ning, H., Qiu, T., Song, H., Wang, Y., Yao, X.: Security and Privacy Preservation Scheme of Face Identification and Resolution Framework Using Fog Computing in Internet of Thing. IEEE Computer Society (2017)
8. Yi, S., Hao, Z., Qin, Z., Li, Q.: Fog computing: platform and applications. In: Third IEEE Workshop on Hot Topics in Web Systems and Technologies (2015)
9. Chiang, M., Zhang, T.: Fog and IoT: An overview of research opportunities. IEEE Internet Things J. **3**(6) (2016)
10. Elkhatib, Y., Porter, B., Ribeiro, H.B., Faten Zhani, M., Qadir, J., Rivière, E.: On Using Micro-Clouds to Deliver the Fog. IEEE Computer Society (2017)
11. Hao, Z., Novak, E., Yi, S., Li, Q.: Challenges and Software Architecture for Fog Computing. IEEE Computer Society (2017)
12. Wen, C., Yang, R., Garraghan, P., Lin, T., Xu, J., Rovatsos, M.: Fog Orchestration for Internet of Things Services. IEEE Computer Society (2017)
13. Xiao, Y., Zhu, C.: Vehicular fog computing: vision and challenges. In: IEEE International Conference on Pervasive Computing and Communications (2017)
14. Pham, X.-Q., Huh, E.-N.: Towards task scheduling in a cloud-fog computing system. In: 18th Asia-Pacific Network Operations and Management Symposium (APNOMS), pp. 1–4, Oct (2016)
15. Huang, C.Y., Xu, K.: Reliable realtime streaming in vehicular cloud fog computing networks. In: IEEE/CIC International Conference on Communications, (ICCC), China, pp. 1–6, July 2016
16. Lin, Y., Shen, H.: Cloud fog: towards high quality of experience in cloud gaming. In: The 44th International Conference on Parallel Processing (ICPP), pp. 500–509, Sept 2015
17. Mtibaa, A., Harras, K., Alnuweiri, H.: Friend or foe? Detecting and isolating malicious nodes in mobile edge computing platforms. In: IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom), pp. 42–49, Nov 2015
18. Law, Y.W., Palaniswami, M., Kounga, G., Lo, A.: Wake: Key management scheme for wide-area measurement systems in smart grid. IEEE Commun. Mag. **51**(1), 34–41 (2013)
19. Fadlullah, Z.M., Fouda, M.M., Kato, N., Takeuchi, A., Iwasaki, N., Nozaki, Y.: Toward intelligent machine-to-machine communications in smart grid. IEEE Commun. Mag. **49**(4), 60–65 (2011)

20. Niranjanamurthy, M., Kavitha, P.B., Priyanka, K., Vishnu, S.N.: Research study on Fog computing for secure data security. Int. J. Sci. Technol. Manage. **5** (2016)
21. Fakeeh, K.A.: Privacy and security problems in fog computing. Commun. Appl. Electron. (CAE) **4**(6) (2016). ISSN: 2394-4714
22. Dsouza, C., Joon, G., Taguinod, M.: Policy-Driven Security Management for Fog Computing: Preliminary Framework and A Case Study. IEEE Computer Society (2014)
23. Yu, S., Wang, C., Ren, K., Lou, W.: Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing. IEEE INFOCOM (2010)
24. Kumar, P., Zaidi, N., Choudhur, T.: Fog computing: common security issues and proposed countermeasures. In: IEEE Conference, 5th International Conference on System Modeling and Advancement in Research Trends, 25–27 Nov 2016
25. Petac, E., Petac, A.O.: About Security Solutions in Fog Computing. Ovidius University Press (2016)
26. Wang, Y., Uehara, T., Sasaki, R.: Fog Computing: Issues and Challenges in Security and Forensics. IEEE Computer Society (2015)
27. Shi, Y., Abhilash, S., Hwang, K.: Cloudlet mesh for securing mobile clouds from intrusions and network attacks. In: Proceedings of 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, pp. 109–118 (2015)
28. Stojmenovic, I., Wen, S., Huang, X., Luan, H.: An overview of Fog computing and its security issues. Concurr. Comput.: Pract. Exp. **28**(10), 2991–3005 (2016)
29. Yi, S., Qin, Z., Li, Q.: Security and privacy issues of fog computing: a survey. In: Xu, K., Zhu, H. (eds.) Wireless Algorithms, Systems, and Applications. WASA. Lecture Notes in Computer Science, vol. 9204. Springer (2015)
30. Chiang, M., Doty, A.L.: Fog Networking: An Overview on Research Opportunities. Electrical Engineering Princeton University (2015)
31. Xu, F., Tan, C.C., Li, Q., Yan, G., Wu, J.: Designing a practical access point association protocol. In: INFOCOM. IEEE (2010)
32. Bonomi, F.: Fog computing and its role in the internet of things. In: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, pp. 13–16. ACM (2012)
33. Bonomi, F.: Vehicles, the internet of things, and fog computing. In: The Eighth ACM International Workshop on Vehicular InterNetworking (VANET), Las Vegas, USA (2011)
34. Baccarelli, E., Naranjo, P.G.V., Scarpiniti, M.S., Shojafar, M., Abawajy, J.H.: Fog of Everything: Energy-Efficient Networked Computing Architectures, Research Challenges, and a Case Study. IEEE Computer Society (2016)
35. Hong, K., Lillethun, D., Ramachandran, U., Ottenwalder, B., Koldehofe, B.: Mobile fog: a programming model for large-scale applications on the internet of things. In: ACM SIGCOMM Workshop on Mobile Cloud Computing (2013)
36. Vargas, V., Syed, A., Mohammad, A., Halgamuge, M.N.: Pentaho and Jaspersoft: a comparative study of business intelligence open source tools processing big data to evaluate performances. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **7**(10), 20–29 (2016)
37. Kalid, S., Syed, A., Mohammad, A., Halgamuge, M.N.: Big-data NoSQL databases: comparison and analysis of "Big-Table", "DynamoDB", and "Cassandra". In: IEEE 2nd International Conference on Big Data Analysis (ICBDA'17), pp. 89–93, Beijing, China, 10–12 Mar 2017
38. Kaur, K., Syed, A., Mohammad, A., Halgamuge, M.N.: Review: an evaluation of major threats in cloud computing associated with big data. In: IEEE 2nd International Conference on Big Data Analysis (ICBDA'17), pp. 368–372, Beijing, China, 10–12 Mar 2017
39. Pham, D.V., Syed, A., Mohammad, A., Halgamuge, M.N.: Threat analysis of portable hack tools from USB storage devices and protection solutions. In: International Conference on Information and Emerging Technologies (ICIET'10), pp. 1–5, Karachi, Pakistan, 14–16 June 2010
40. Gupta, A., Mohammad, A., Syed, A., Halgamuge, M.N.: A comparative study of classification algorithms using data mining: crime and accidents in Denver City the USA. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **7**(7), 374–381 (2016)

41. Li, X., Niu, J., Kumari, S., Wu, F., Sangaiah, A.K., Choo, K.K.R.: A three-factor anonymous authentication scheme for wireless sensor networks in internet of things environments. J. Netw. Comput. Appl. https://doi.org/10.1016/j.jnca.2017.07.001 (2017)
42. Hu, P., Ning, H., Qiu, T., Xu, Y., Luo, X., Sangaiah, A.K.: A unified face identification and resolution scheme using cloud computing in Internet of Things. Future Gener. Comput. Syst. https://doi.org/10.1016/j.future.2017.03.030 (2017)
43. Xiao, X., Chen, C., Sangaiah, A.K., Hu, G., Ye, R., Jiang, Y.: CenLocShare: a centralized privacy-preserving location-sharing system for mobile online social networks. Future Gener. Comput. Syst. https://doi.org/10.1016/j.future.2017.01.035 (2017)

A Review on Security and Privacy Challenges of Big Data

Manbir Singh, Malka N. Halgamuge, Gullu Ekici and Charitha S. Jayasekara

Abstract Big data has a growing number of confidentiality and security issues. New technology doubtlessly brings people benefits, privileges, convenience and efficiencies, with confidentiality issues. Additionally, technological advances are accompanied with threats that can pose dangerous privacy risks that can be more detrimental than expected. Privacy of data is a source of much concern to researchers throughout the globe. A question that remains unanswered is the question that "what exactly can be done" to resolve confidentially and privacy issues of big data? To answer some questions, data collected for this chapter has been through the analyze of 58 peer reviewed articles collated from 2007 to 2016 in order to find some resolutions for Big Data confidentiality issues. The articles range from different industries that include healthcare, finance, robotics, web applications, social media, and mobile communication. The selected journal articles are aimed used to make comparative analysis of security issues in different areas to cast solutions. This chapter consists of four main parts: introduction, materials and method, results, discussion and conclusion. This inquiry aimed to find different security issues of big data in various areas and gives solutions by analyzing results. The results of the content analysis suggest that the internet applications and financial institutions are dealing with specific security problems, whereas social media and other industries deal with confidentiality issues of sensitive information which have heightened privacy concerns. Both these issues are addressed in this study, as retrieved results from the data, highlights the gaps that can be further researched for development. The method used to gather data for this chapter is through the analysis of studies that deal with a particular confidentiality issue. After the analysis and evaluations, the suggestions of confidentiality issues are displayed

M. Singh · M. N. Halgamuge (✉) · G. Ekici · C. S. Jayasekara
School of Computing and Mathematics, Charles Sturt University, Melbourne, VIC 3000, Australia
e-mail: MHalgamuge@studygroup.com

G. Ekici
e-mail: GEkici@studygroup.com

C. S. Jayasekara
e-mail: subhashi@ieee.org

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*, Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_8

by using a different algorithm method. This research has addressed gaps in the literature by highlighting security and privacy issues that big companies face with recent technological advancements in corporate societies. By doing this, the research could shed revolutionary light on issues of big data and provide futuristic research directions to solve them.

Keywords Big data · Security · Privacy · Security issues · Data analysis

1 Introduction

A significant portion of Information technology research efforts goes into analyzing and monitoring data regarding events on the servers, networks and other connected devices. Big data is a fairly new concept in modern technological world. In recent times, there has been an increasing usage of big data, as the problem of security has become very important [1]. This chapter covers different aspects of big data security, in particular challenges related to big data variety, velocity and volume. The amount of sensitive information that needs to be protected is constantly increasing [2]. However, in the era we live in, information is required to be protected from hackers. Insufficient protection can bring various security challenges [3–5]. The notion of big data came to use not long time ago, as it relates to big amounts of information that companies produce and need to store. This information is then used to analyze and then predict the emerging trends in future sales by analyzing annual trends. Growing number of informative data is subjected to storage issues as there are no security measures yet established. In fact, any given amount of information that is generated is an issue of security and privacy. Confidentiality of sensitive information becomes a perplexing issue for companies if they do not take considerable amount of time, effort, and resources to deal with confidentiality issues. Big Data is used to manage great number of datasets, as all the vast amount of data is often not structured and have been stored from different sources [6, 7]. Traditional access control mechanisms to ensure privacies are insufficient in recent times with the growth of demand which brings the need of a fine granular access control mechanism to make sure every aspect of privacy is reflected. This framework is called an ontology-driven XACML context [5]. On the other hand, providing privacy in cloud is much more complex than one thinks [8]. The majority of data preservation techniques are targeted at small levels as they often fall short; an algorithm is designed with MapReduce to gain high scalability by performing computing in parallel. This method is called local record anonymization [8]. In this case, the hybrid cloud is a very different approach and difficult to implement. The idea is to separate the Sensitive data from non-sensitive data and store them in different trusted clouds. This method of isolation is best suited for processing the image files in an entirely different approach, as it is used to deal with data at rest [4]. A multilevel identity encryption method is used both at the

file level and at block level to satisfy data protection. This process helps to leverage cloud provider because of transparency [9].

Preserving privacy of information is one big issue; nonetheless providing security to IT is another huge matter, nonetheless this matter needs to be taken into hands urgently. There are a lots of security risks in big data, as the main one is: Privacy leakage, this is one of the most dangerous issues that has already caused many problems for companies [5]. Therefore, a whole range of data needs different computational techniques to make it secure and safe. The first step in cloud security is to ensure the entry points. This helps to detect possible attacks [10–13], and alert users to use the instruction detection system [14]. Encryption is another way of making data safe in the local area network, and VPN encryption is used to safeguard data [15]. On the other hand, in the case of web application, randomization is based on Random4 encryption algorithm which is similarly used [14]. This prompts Big cooperates to use MuteDB architecture, where they incorporate data encryption into, key management, authorization, and authentication from a new MuteDB architecture. This architecture assures scalable solution to guarantee the confidentiality of information in the database [6].

Security of medical data falls into the same category, as both cloud-based technologies and attribute-based encryptions are used for storage and retrieval [16]. Another way is to use a pocket-sized computer, which is called Raspberry Pi. This computer makes sure that regional data is collected and kept isolated [17].

2 Data Processing Method

The Knowledge Discovery from Data (KDD), is often treated, as a synonym for 'Data mining'. This is a method used to discover information from data to avoid leakage. Every day millions of bytes of data are generated throughout the world. This process in return allows companies to grow and remain in the competitive market by identifying seasonal trends and launching products in the peak seasons. Therefore, the research in the area is significant and requires development. There are usually three steps involved in the method of KDD, which is performed in an iterative way. They are discussed as below:

Step 1. **Data processing**: Data processing method is a step that selects inconsistencies of missing data fields and removes them while reintegrating the data pool. It is presented in a form so that it can be read quickly, to generate and reach potential results.

Step 2. **Data Transformation**: This step is to transfer data into appropriate forms for mining. The Data is not presented in its proper form, and therefore, it must be sorted out to represent a type that can generate some useful information.

Step 3. **Data Mining**: In data mining, various methods are employed to extract the information from the data, as algorithms are used to extract information from the data pool.
Step 4. **Pattern Evaluation**: After the data is extracted, patterns are then evaluated to obtain knowledge on trends.

2.1 Privacy

Every day, large amounts of data is generated and processed in an array of industries. Thus, the privacy of data can be established by methodical procedures. In fact, there are four different types of steps involved: (i) Data provider, (ii) Data Collector, (iii) Data Miner, and finally (iv) decision makers, as these are people who are involved in the processing of data that is collected and derived from knowledge from groups of data. Each one has different challenges to privacy protection, as these are discussed below.

Approaches to privacy protection by:

Data Provider:

Data providers can provide data voluntarily according to the demands of the Data Collector. The data collectors can retrieve data from the providers of customers' daily activities. However, there are many ways to limit the data collectors' access to this data and this could be done in several ways. Internet companies now have a strong motivation to track users' movement over the Internet to ensure that the valuable information can be extracted from the data produced by users' online activities. These can block the advertisements on the sites, and also kill the script, for example, AdBlock, Encryption tools are used to encrypt data and transfer them into Cypher-text which is not in a readable form and so it can be transferred in a safe way.

Data Collector:

The original data retrieved from data provider generally obtains sensitive information. The sensitive information needs specific precautions before being passed onto the data mining process. Before sensitive information can be disclosed to the public, this process must be done otherwise it will trouble companies. The process of enables to replace some value with a parent value, as this method is a good way to hide sensitive information. Permutation de-associates the relationship between quasi-identifier and numerical attributes dividing the data sets into groups and shuffles information among groups. Perturbation can also operate data with some false value to hide it from collectors. This includes adding noise, swapping data, and generating synthetic data.

Data Miner:

The data miner uses an algorithm to obtain data from the data collectors. However, there are two types of privacy issues that can risk confidentiality in this process. Firstly, when data is directly observed, the information could be leaked. At times, even the data mining results may also leak private information. Some approaches are then helpful, such as, the Heuristic distortion approach that helps to resolve how to select the appropriate data sets for data modification. This method works by replacing certain attributes of data items with a particular symbol. Probabilistic distortion approach distorts data through random numbers generated from a pre-defined probability distribution function. The reconstruction based approach generates a database from scratch that is compatible with a set of non-sensitive rules.

Decision Maker:

The ultimate goal of data mining is to provide information to the decision makers', however, to achieve its objectives it is compulsory to meet the confidentiality rules and regulation to protect people. At first glance, it seems that the decision maker has no responsibility, nonetheless in actuality they must have a duty of care. If the results are disclosed to competitors, the policy makers will suffer the loss, as the openness, freedom and anonymity of the Internet, as Data Provenance poses great challenge for seeking the provenance of information. Also, the decision maker must look at five aspects of information including authority, accuracy, objectivity, currency and coverage.

2.2 Security

Security has always been an issue however in recent times with the growth of conventional security mechanisms that are used to secure small scale or static data are no longer adequate, as far as big data is concerned. There are often loopholes in the system that allows intruders to exploit services. There are some security challenges, which is prevalent in the vicinity. The majority of organizations are dealing with sensitive information at the threat of data theft. Table 1 displays types of security challenges and threats that need attention.

2.3 Internet of Things

In society today, technological gadgets are the new modern-day slaves. The public uses the word "internet" to do shopping, reading, paying bills, basically running all their errands from the tip of their fingers. Existence is at ease through novel inventions in the perception of multi purposeful Internet. There has been countless

Table 1 Different security challenges and solutions

	Security challenges	Solution for data security
1.	***Real-Time Monitoring***: Real-time monitoring has always been a big issue on account of the number and frequency of security alerts that generates. It is now easy to rectify any loopholes or dangers nonetheless it takes lots of effort to find such threats [23]	***Layered Protection***: In computer hardware, the security of information system is formed through expansion of layers. The securities of outer layers rely on the security of inner layers. The more the layers the better security is [24]
2.	***Granular Audits***: In case the real-time monitoring system, it does not capture the attacks, therefore, audits are needed [25]	***Protection of Different Domains***: DNS can be divided into local region, network perimeter, network transmission, and infrastructure. Therefore, different technologies are used in various procedures to secure areas in order to establish distributed security system [26]
3.	***Secure Computations in Distributed Systems***: In this system, parallel computation and storages are used to process massive amounts of data. Securing the mapper and data in the presence is not a trusted mapper and becomes a primary concern [27]	***Hierarchical Protection***: Since the importance of the same information is different in different institutes. Thus, hierarchical protection is needed, and in this case, different access control measures are used so that a particular user that covers the only specific parameter [28]
4.	***Secure Data Storage and Transactions Logs***: Data and transactions are stored in multi layers. Moving data manually among levels is not an issue, nonetheless given the amount of data that is generated; auto tiring for management is needed. Nonetheless it does not keep track where the data is stored. Thus maintaining 24/7 availability is a big issue [27]	***Time-Sharing Protection:*** Information security in big data is a dynamic process. Taking time into consideration of securing of Big Data can be incredibly enhanced [29]
5.	***Endpoint Evaluation***: Many large organizations require data collection from various sources. A key challenge here is to validate the input, and this indeed is the validation and filtering of data which can be daunting as the challenges are posed by untrusted data sources [30]	***3KDEC Algorithm***: A symmetric key block encipherment algorithm is used to present the practical solution to the problem where numeric data is converted to alphanumeric type and thus encrypted data is not stored in existing numeric fields [31]

research done to assess the daily use of the internet by people and predictions have been made for future use of the internet.

Despite the usefulness of the internet, it is still an under researched area and facts are insufficient in the area. Therefore, this chapter aims to go through 12 research papers to analyze emerging themes to seize research development. The correlation among these 12 research papers will draw data to analyze emerging themes of internet usage. These discoveries will allow gaining information to futuristically

look in the direction of automated systems. This will also allow looking at focal benefits in using technologically advanced plans.

The "Internet of everything" is the greatest significant idea established in the real world to be used by individuals globally. Novelties under this motif is accommodating in the pursuance to ease lives of people in their daily dealings. Internet development has been gradually slowly, nonetheless momentarily improvements are at a noteworthy frequency. The internet emerged in 1969 with the ARPANET (Advance Research Project Agency Network) with few web sites [18]. Currently, it is figured that, in 2020 around 20 billion gadgets will be joined to the internet. More innovative work will be on how to connect people who are unconnected to the internet.

Many new researches have undertaken studies in exchange to find futuristic measures of internet usage. This pursuance to draw data from 12 researches in the same topic under different areas will prove the connection between each research that revolutionizes in connecting the unconnected to the internet. This document can be referred to once a problematic issue is encountered in a certain area under the topic and knowledge can be issued from it. Cloud computing also known as fog computing, is an essential area once the topic is connectivity between each and every device in household [19]. Cloud computing emerges in a superior haste, in several ways, that can be used to connect the unconnected. Fog computing and cloud computing is utilized as an average construction of the "Internet of Everything".

"Internet of Everything" comprises of numerous crucial notions, that is imperative in learning and teaching [20]. However, the core variance amid the "Internet of Everything" and "Internet of things" that they are both physical substances physically present. However, "Internet of Everything" is a method outside this as it is amongst the utmost extensively documented concerns as Internet associated devices are liaised through an Internet reminder, for example, consequently alters the temperature on a Nest indoor controller. People might feel that they can operate the "Internet of Everything" to examine information from these devices. However, considering that, even today, the Internet pages have cyber spaces that are intangible perhaps we are unaware that gadgets are not only made from physically tangible gadgets nonetheless also have intangible cyber ones too. This is evident as of the world's supreme site, "Google", is not a physical tangible tool. It occurs in a cyber space amid the wires. This is the legal administrations that is used each day, for example, Dropbox or Instagram is also virtual spaces. This is the main expansion of the "Internet of Everything"; it does the administration that you cannot put a finger on it and say that it exists in a physical space. Although, the Internet is furthermore covered of significantly greater number of things, however we are only using the online administrations utilized. This shows that people need to understand that the Internet consists of information streams and relations of associations. Similarly, one can claim that the Internet contains clients that are most of the general population as associates.

The "Internet of Everything" links up different concepts into one secure idea. This method allows devises to communicate with each other. It is obvious that the "Internet of Things" might as well be called a rail street line, which includes the tracks and the associations, while the "Internet of Everything" is most of the trains, ticket machines, staff, clients, climate conditions, and so forth. This research, elaborates ideas on whether and how the theme of Internet of Things and Internet of Everything is developing, as this is a new area that comes under these two themes. What are the major developments done so far? will be figured out in this research. The gaps will also be identified and will be pointed out and clearly explained.

3 Material and Method

This chapter has categorized different security and privacy issues of big data in accordance to types of issues and some parameters. The data that has been collected through a content analyses, which is retrieved by content analysis of 24 peer-reviewed scientific studies from 2007 to 2016 and found answers and analyzed the data of each context in different industries including healthcare, finance, robotics, web applications, social media, and mobile communications. An agile approach is used for this project because data that depends on IT has a massive computation [21]. Data about security issues and solutions was collected from different journal articles, and data was surmized through thorough analysis of the collected data, then it was structured, and analyzed. The data is displayed in a table for better comprehension that shows the area of security issue, solution, algorithms and then used for solution and also general and technical remarks accompany. After the analyzes of data, the results and conclusion were drawn objectively (Fig. 1).

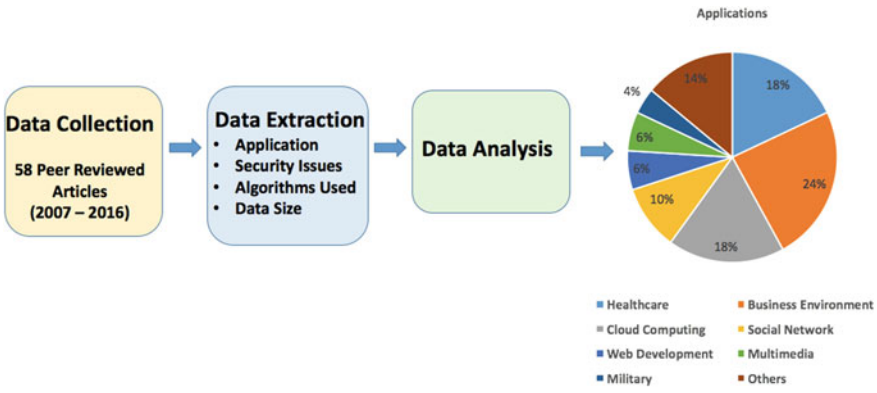


Fig. 1 Graphical abstract of the analysis: Security and Privacy Challenges of Big Data

3.1 Data Collection Method

This analysis has pooled the data of the published articles from 58 peer-reviewed scientific studies which were published between the years of 2007–2016. The raw data presented in Table 3 specifies the variables used for the analysis including application, security issues, and algorithm (Table 2).

3.2 Data Analyses Method

The data is analyzed by categorizing the collected data and displaying emerging themes in a table. The data sets include attributes such as volume of data,

Table 2 Description of algorithms and other key terms used

Algorithm/Key terms	Description
MapReduce [4]	MapReduce Programming model is for generating and processing data sets with distributed and parallel algorithm on clusters
AES [8]	AES (Advanced Encryption Standard) algorithm is a symmetric block cipher used to protect the sensitive and confidential information by encrypting it to an unreadable form
Multilevel identity encryption [9]	This is an extension of normal encryption. Here the identity of the data on Cloud is encrypted and protected by multiple layers
ORAM [32]	Oblivious RAM (Random Access Memory), allows clients to access their data on a remote server
XACML framework [5]	Extensible access control markup language, is used to implement attributes-based on access control policies
Data masking [33]	The method of creating inauthentic nonetheless structurally similar versions of data for testing and training purposes
Random 4 [34]	Application specific encryption algorithm, which is used to prevent SQL (Structured Query Language) injection
StarLight [3]	This tool is used to gather information from different sources like visual intelligence, geospatial to alert staff on sea ports
3DES [2]	Triple data encryption algorithm is a symmetric key block cipher, which applies DES (Data Encryption Standard) three times on same data
VPN [15]	The virtual private network extends private network across the public network. Establishing virtual P2P connection through dedicated connections and traffic tunneling creates VPN (Virtual Private Network)
OBEX [35]	The objective exchange is a communication protocol that helps the exchange of binary objects between devices
MuteDB Architecture [6]	This architecture devise incorporates data encryption, authentication, authorization and key management to assure confidentially of data in cloud

application area, and issues that are prevailing. The methodology adopted allows solving these issues, as the algorithm selected is merely for this purpose.

4 Results

This study has collected and analyzed data from 58 peer reviewed articles published from 2007 to 2016 in order to find answers for confidentiality issues of Big Data. The observations clearly show that, industries that have kept sensitive data of customers are trying to preserve their privacy whereas, the industries, which have their computations in real time, are working hard to keep it as a secret. Moreover, it is significant to say that most of solutions are based on encryption algorithms. Furthermore, in different areas, as for example Finance and Health (role-based access control), can be applied as a solution. In addition, most seeable solutions are directed to protect the access of big data, as there is not much concern on the security aspects of big data.

The percentage of mostly used application is given in the Fig. 2. After analyzing the Data from Table 3 the following conclusion can be drawn.

a. **Healthcare**: Preserving the privacy of data is a critical issue. Healthcare information should not be disclosed or retrieved easily, or leaked to hackers, as it is sensitive information that may cause financial detriment. To avoid this, there are some techniques used that include the anonymization of the records by

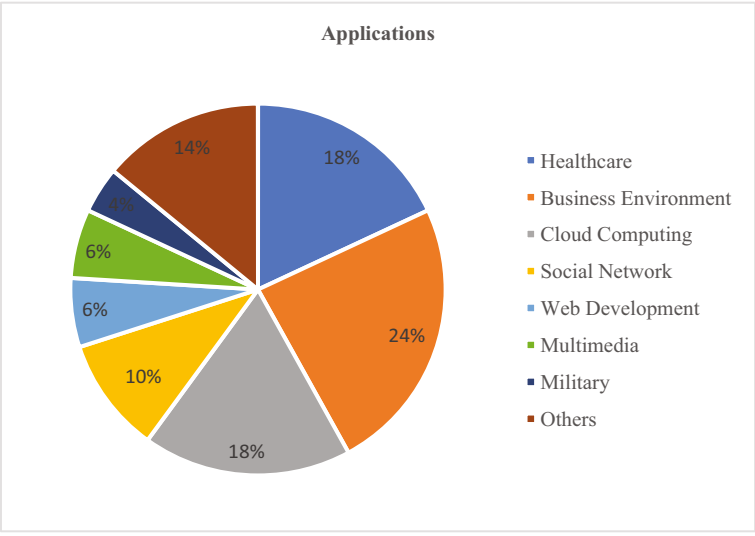


Fig. 2 Overview of the different application, using the data from the 58 peer reviewed scientific articles published in 2007–2016 of Security and Privacy Challenges of Big Data

Table 3 Security issues in big data: algorithms used in published papers

Article	Application	Security issues	Algorithms used	Data size	Method used
1. Nair et al. (2016) [35]	Health and Finance	Privacy of customers' and employees' information	FPE (Format Preserving Encryption) algorithm that combines: one algorithm to encrypt, another algorithm to decrypt and one algorithm to sample	Dataguise	This helps to decrease data breaches risk Dataguise is created to detect, protect and also handle compliance to regulatory mandates
2. Moura et al. (2016) [36]	Cloud computing	Users' privacy and management of big data	Functional encryption algorithm consisting of Key Generation, Encryption, Decryption, and Evaluation	Homomorphic encryption	Keeps private information secure This forms encryption that allows certain computations to be executed cipher text and create encrypted results
3. Swarna et al. (2016) [37]	Cloud environment	Privacy of transmitting and stored information	Ring signatures include only two algorithms: Sign and Verify	Ring signature	This type of signature can be executed by any group member Ring signature uses PSA algorithm
4. Moura et al. (2016) [36]	Social network	Privacy of users' information	Rendering algorithm	User rights management	Secure storage of information The user identifies rights and limits the content; all users are registered securely
5. Mengke et al. (2016) [15]	Local area networks	Interface security	VPN encryption	–	Background linkage for managers including remote lock, data wiping, and automatic alarm
6. Charishma et al. (2015) [38]	Business organizations	Privacy of company's and employees' information	Cryptographic algorithms	Analytic tool Splunk	Provides log management by taking and analyzing the logs using certain patterns

(continued)

Table 3 (continued)

Article	Application	Security issues	Algorithms used	Data size	Method used
7. Chandankere (2015) [39]	Cloud storage	Privacy of stored information	Group signature algorithms consist of four algorithms: KeyGen, Sign, Verify, and Trace. KenGen and Sign are randomized, Verify and Trace are deterministic	Dynamic encryption and group signature	Enables secure sharing of information Dynamic encryption enables transmitting encrypted data to group member by adding members to admin. Group signature enables disclosure of identity by admin in case of dispute
8. Zeng (2015) [22]	Social network	Privacy of users	K-anonymity algorithm	Anonymity protection	Anonymity protection is used to protect data that can include relationship, attributes and identity anonymities
9. Inuwa (2015) [40]	Business organizations	Privacy of company's and employees' information	Symmetric encryption algorithm	AES encryption	This type of encryption makes data unreadable for attackers It contains operations, including substitutions and permutations.
10. Zeng (2015) [22]	Multimedia	Privacy of transmitting media files	Watermarking detection algorithms (cryptographic algorithms)	Data Watermarking	"Data Watermarking" is applied to protect Copyrighting "Data Watermarking" relates to the information identification that is inserted imperceptibly

(continued)

Table 3 (continued)

Article	Application	Security issues	Algorithms used	Data size	Method used
11. Kaur et al. (2015) [41]	E-Commerce	Privacy of customers' information	Rivest Shamir-Adleman (RSA) algorithm	A new E-banking security system	In the new system activities and functions are s This new system is grounded on a strong access control
12. Jyothirmai et al. (2015) [42]	HealthCare	Privacy of patients' information	access control algorithms (cryptographic algorithms)	Role-Based Access Control	Role-Based Access Control is a tool for managing access of data and making data safe Role-Based Access Control is able to manage different policies of access control which is grounded on role hierarchies
13. Nair et al. (2016) [35]	Mobile phone	Intrusion detection in Mobile phones	BLA, Bluetooth object exchange (OBEX) Protocol	Covers 60% of Bluetooth market	Using Bluetooth logging agent, and also using database rules to authenticate
14. Zeng (2015) [22]	Commercial companies	Privacy of company's and employees' information	Access control algorithms (cryptographic algorithms)	Role-based access control	Roles are able to be generated and behavior of each user can be checked. It is grounded on authorization of "Users-Object" and role optimization
15. Feng et al. (2015) [17]	Healthcare	Regional secure data process/collect to limit issues in future health care	Raspberry Pi	All digital healthcare industry	Raspberry Pi is a pocket-sized computer used in forensic medicine, and forensic etymology

(continued)

Table 3 (continued)

Article	Application	Security issues	Algorithms used	Data size	Method used
16. Raji et al. (2015) [43]	Social networking	Privacy challenges in online social networks (OSN)	P2P-ONS Architecture	Can include Facebook, Twitter, Messenger, etc.	Architecture composed of privacy enabled start-up for user's social communication and adaptive replica for ensuring availability of shared data
17. Zeng (2015) [22]	Business environment	Privacy of company's and employees' information	Clustering algorithm	Access control (risk adaptive)	Risk adaptive access control is appropriate when it is not clear which data is accessible for users This method uses information theory and statistical methods to identify quantization algorithm
18. Zeng (2015) [22]	Commercial organizations	Privacy of company's and employees' information	Provence graph algorithms	Data Provence	Data Provence is used through labelling, so it is able to check if the results are correct, to differentiate the data in the table or to update the data
19. Kepner et al. (2014) [44]	Bioinformatics and social media	Privacy of patients' information bioinformatics, personal information of users in social media	Graph algorithms that are created using associative arrays	Computing on masked data	Computations are allowed to be executed directly on masked data, as the authorized recipients are allowed to unmask data. CMD (Computing on Masked Data) includes methods of cryptographic encryption and associative arrays that represents big data

(continued)

Table 3 (continued)

Article	Application	Security issues	Algorithms used	Data size	Method used
20. Wagh et al. (2014) [45]	Educational organizations	Privacy of education resources, and personal information of users, integrity	Digital signature: a randomized KeyGen algorithm, a randomized Sign algorithm, and a deterministic Verify algorithm.	Digital signature, data encryption, access control	Access control is achieved through verification by transmitting secure data using data encryption and digital signature
21. Merkel (2014) [46]	Health and Finance	Privacy of customer's information	Mathematical algorithms	Bash tool	It simplifies a complex process It is used mostly as an intermediary
22. Hsu et al. (2014) [47]	Group communication in social media	Privacy of users' information	Changed RSA algorithm, that relies on NP class	Group key transfer protocol	This protocol protects from attacks that decreases system implementation overhead. This protocol is based on LSSS (Large Scale Survey System and DH key agreement and does not have online KGC (Key Generator Center)
23. Pace (2014) [48]	Scientific computing	Privacy of scientists' works	ARC algorithm for storage pool	Data management	Data management simplifies data workflow and makes it secure Data management assigns to pool service management
24. Bertine et al. (2014) [49]	Web application	Encrypted data is not secure	PPDM and PPDA	Sheer amount of data in cloud	Using Cryptography to encrypted data

(continued)

Table 3 (continued)

Article	Application	Security issues	Algorithms used	Data size	Method used
25. Ferretti et al. (2014) [6]	Cloud database services	Scalable solution to guarantee confidentially of information in the database	MuteDB architecture	Cloud providers	Data encryption, authentication and authorization to form new MuteDB
26. Huang et al. (2014) [8]	Web application and Hybrid cloud service providers	Privacy of image data stored in public cloud	AES algorithm	Image Data	Dividing images to blocks and shuffle them, to make them unrecognizable
27. Mehak (2014) [50]	Cloud storage	Privacy of stored information	Map Reducer Algorithms including Master Key Algorithm	Hadoop	Enables processing of big data sets It parallelizes processing of data across computers in a cluster
28. Mirarab et al. (2014) [51]	E-Commerce	Privacy of customers' information	Stenography: least significant bit algorithm (LSB)	Encryption, Steganography	Encryption and Steganography makes E-Commerce more safe and secure Encryption is executed though Elliptic Curve Cryptography. LSB (Least Significant Bit) Steganography is used for image compression
29. Syed et al. (2014) [16]	Cloud services including medical data, confidential defense records, etc.	Encryption of data in the database is not sophisticated enough to provide enhanced security	Encryption for frequent access node	Potential to enhance the security of all cloud providers	Combining cloud-based technologies and attribute based on encryption for secure storage and retrieval

(continued)

Table 3 (continued)

Article	Application	Security issues	Algorithms used	Data size	Method used
30. Raghuwansi et al. (2014) [52]	Cloud service providers	Privacy of data at rest in cloud	Multilevel identity encryption	Cloud vendors and consumers	By using encryption and verification services both at file and block storage level
31. Li et al. (2014) [32]	Google Drive, Dropbox, Amazon S3, SkyDrive, iCloud, gnyte, OneDrive, etc.	Privacy preserving data access to cloud	ORAM algorithm	Geo-distributed cloud sites	Using ORAM (Oblivious Random Access Memory) for load balancing thus revealing the access patterns
32. Tan et al. (2014) [14]	Any cloud-based application e.g. one drive, cloud	Cloud security	Secured entry points	Terabytes	Sensing attacks and alerting the user by intrusion detection system and data leak prevention system
33. Islam et al. (2014) [2]	Multimedia like email, music, and images	Dealing with structured data and unstructured data	Statistical learning algorithms	Infinite	Text analysis by filtering, clustering and classification, building security node
34. Abawajy et al. (2014) [53]	Robotics and control system	Malware detection	LIME classifier	Massive and expected to grow exponentially	
35. Islam et al. (2014) [2]	Financial system	Security to unstructured data	3DES	1200–1400 Exabytes	Digital certificate, using hash functions
36. Tankard (2012) [54]	Multi-silo environment	Privacy and law of data protection	Symmetric encryption key algorithms	Vormetric	Vormetric manages data access control that combines storage elements, policy management and data encryption

(continued)

Table 3 (continued)

Article	Application	Security issues	Algorithms used	Data size	Method used
37. Kaplan (2012) [55]	Business organizations	Privacy of company's and employees' information	Cryptographic algorithms	Encryption	It protects the system from attacks Encryption protects information by encoding messages
38. Pramila et al. (2012) [56]	Healthcare	Current system to diagnose the patient has slight range and is not secure	Using location tracking technology, telediagnosis, Access using PKC	Covers the person suffering from Alzheimer's Disease	A new method proposed with the long-range outdoor environment with GPS (Global Positioning System) and fine-grained distributed data access control
39. Eler (2012) [57]	Web applications or websites	Web application is suffering security attack especially SQL injection attack	Random4 encryption algorithm	Have potential to provide security to web traffic	Using an encryption algorithm based on Randomization
40. Faulkner et al. (2011) [3]	Military	Port has very serious security issues; they lack port specific security technologies to alert security personnel in case of any hazard or danger	StarLight uses Visual intelligence, entity detection, and intrusion	Commercial and military ports	StarLight combines information from different sources and integrates text, geospatial and temporal data to alert security staff
41. Motiwalla et al. (2010) [33]	Healthcare	Privacy preserving for healthcare data	Data masking	Spent $39.4 billion in 2008	Changing the data values by using noise perturbation, data aggregation, and data swapping
42. Wang (2010) [58]	Database or Data Warehouse of any company	Information security structure for database processor	Following Plan-Do-Check-Act cycle to implement nine principles established by OECD	Cover all organization having a database	Using a structure that complies with laws requirements and conflicts between consumer and database processor

(continued)

Table 3 (continued)

Article	Application	Security issues	Algorithms used	Data size	Method used
43. Chang et al. (2009) [7]	Mobile Phones	RFID enabled credit cards lacks sophisticated computation mechanism for authentication	A new RFID system based on mobile phones	Have Potential to expand it in credit card related computations	Proposed an efficient and secure mechanism using mobile devices like RFID (Radio-Frequency Identification) readers together with credit cards
44. Abou-Tair et al. (2007) [5]	Any information system	Privacy in enterprises	Ontology-driven XACML Framework	Any amount of computational data	Using ontology to focus on generating access control policies to provide fine granular access on diverse data force
45. Skinner et al. (2007) [59]	Virtual community including gaming, and multimedia	Privacy issues in the virtual communities are ineffective and pose a threat to data integrity	Hippocratic Security Method	Virtual world	Monitoring the use of personal information through Hippocratic database principle to enforce Hippocratic policies
46. Dimitropoulos (2007) [60]	Healthcare	Privacy of patients' information	Probabilistic Matching algorithms that are made to match patients and records	Identification Management System (IMS) and Master Patient Index (MPI)	IMS (Information Management Software) and MPI (Multiple Protocol Interface), PMA (Parallel Multithreaded Architecture) are concentrated on provider and patient identification making the service secure. IMS (Information Management Software) and MPI (Multiple Protocol Interface) are created for functioning within regional health organization or HIE

(continued)

Table 3 (continued)

Article	Application	Security issues	Algorithms used	Data size	Method used
					(Health Information Exchange)
47. Agrawal and Srikant (2000) [61]	Enterprises with "Bring your own" environment	Privacy and data protection of employees' information and company's information	RSA (Rational Software Architect) algorithm public key Cryptographic algorithm applied by the issuing CA (Certificate Authority)	SonicWall	SonicWalls contain email security, mobile access security and network security. It collects input from a lot of sensors and defends and informs against threats

using a tool called MapReduce. Additionally, data masking, which is a unique computer device that is called Raspberry Pi, and this devise collects data in a secure manner, by using location tracking technology to treat patients.

b. **Web Application**: Web applications suffer the majority of data breaches. The violations occur not only at the backend nonetheless also during the data transmission and data collection. The best solution is to assure and identify the owner by encrypting data before the transmission thru encrypting the frequently accessed nodes securing entry points are also uses specific architectures for authentication, authorization, and key management.
c. **Mobile Devices**: In mobile devices, the first issue is to verify the user as this verification can be a problem. Another fundamental problem is to secure the transaction from all devices. The possible solutions include using a concrete architecture like OBEX, to monitor the behavior of data and network for detecting intrusions by using RFID-based mobile phones.
d. **Social Networking**: The Privacy and access to shared data causes the majority of the problems. A two-way authentication of a user is one of the possible privacy preserving solutions that need to be done.
e. **Finance**: This application area has been under the scanner, and the majority of attacks and threats are the target for monetary benefits. Both privacy of sensitive information and data storage in cloud or databases are at risk. Authentication, authorization of the user is one of the first challenges of underlying issues that need immediate attention. Banks use an intrusion detection system to detect threats, ergo, it is not safe enough. Various encryption algorithms used makes computations and transactions secure.

5 Discussion

Big data is a fairly new concept in IT, and it is obvious that research in the area is not thorough enough and more research needs to be done. However, there is an important gap in many articles suggesting that research is bias towards traditional methods and Big Data is under-researched. In most of the articles it is easily seen that research findings are one-sided and incomplete. This disadvantage refers to all articles used except article of [22] where different areas and different solutions for security issues are represented in his study. For instance, social network, multi-media, commercial organizations, companies, business environment, anonymity protection, data watermarking, data Provence, role-based access control and risk-adaptive access control are all aspects that can be looked into by scholars. To draw more conclusive results, a wider range of information needs to be considered in most articles. Moreover, there is lack of comparison of different solutions and why or how they could apply is another gap in these articles. The literature chosen for this study also has gaps in the area of application as to how certain security measures could be applied. Secondly, in most articles there is no detailed

information about solutions including algorithms that are used for particular problems, for example case studies would be valuable to understand and gain solutions in a particular area. Thirdly, in many articles security issues are represented in a general form without specific information related to a certain problem of an area. In particular, most common security and privacy issues do not include details about what privacy was meant and what kind of information needs to be protected. Technological development with variety of benefits can also bring threats that can pose a danger and result in the breach of privacy and if important information is made public by companies then they can be facing hefty fines. Big data is a new area referred to the vast amount of information that needs to be analyzed and stored in order to eliminate confidentiality breaches. There are many security issues in different areas of big data and sensitive information needs to be protected. Though a big research in big data needs to be done, as this area is still new, and more research is demanded because many questions need to find answers. Results clearly show that security issues are similar in different areas, solutions can be the same in different areas and many solutions are grounded on encryption algorithm. Moreover, protection of access is very significant in big data. To summarize this study, this research has added knowledge about big data security issues and highlighted research gaps in the area. However, this area is neglected and requires a continuing research covering different aspects of big data.

6 Conclusion

In conclusion, this chapter has analyzed data obtained from 58 peer-reviewed scientific publications from 2007 to 2016. This study has highlighted certain gaps in the literature to evaluate possible solutions to a rising problem in various privacy and security issues in different areas of big data. The company-provider needs to ensure security for a safer infrastructure and protection of customers' information and this data has to comply with confidentiality standards. In some areas as Health and Finance solving, security issues is the key point of effective and successful work of the company. Many different technologies are created to protect against securities issues, however, the existing technology is not able to completely solve security issues and research in this area is poor but continuing. Although a big research has been done regarding big data issues, it is still a fairly new advancement in IT and a lot of questions and aspects of security problems of big data are not answered and covered enough. In particular, there is a lack of comparative analysis of both security issues in different areas, as to draw solutions by comparing and contrasting studies and finding solutions for them. Therefore, this work is aimed at finding and comparing important security issues in big data in different areas and also evaluated solutions that can solve security issues. This analysis is also important in a sense of providing the grounds for further research and enriching existing information about big data. In order to address these gaps and highlight issues in regards to some security and privacy issues of big data, certain tools and

techniques, have been used to find possible answers to particular alarming issues of Big Data. Data has been categorized, and then the second step was to group them under different parameters. The revelation concluded that web applications and financial institutes are dealing with security problems, and each problem is resolved in varying ways. Social media and other industries dealing with sensitive information have individual privacy concerns, which are treated with a uniform approach. This research has addressed gaps in the literature by highlighting security and privacy issues that big companies face with recent technological advancements in corporate societies. By evaluating these gaps there may be some light shed on these issues of big data and provide future researcher directions to solve them.

Author Contribution Manbir Singh and Malka N. Halgamuge conceived the study idea and developed the analysis plan. Manbir Singh analyzed the data and wrote the initial chapter. Malka N. Halgamuge helped to prepare the figures and tables, and finalizing the manuscript. All authors read the manuscript.

Acknowledgements The authors acknowledge the data collection support provided by Mariia Talalaeva.

References

1. Kim, S., Kim, N., Chung, T.: Attribute relationship evaluation methodology for big data security. In: 2013 International Conference on IT Convergence and Security (ICITCS) (2013)
2. Islam, M., Islam, M., Shawkat Ali, A.: An approach to security for unstructured big data. Rev. Socionetwork Strat. **10**, 105–123 (2016)
3. Faulkner, L., Kritzstein, B., Zimmerman, J.: Security infrastructure for commercial and military ports. In: OCEANS'11 MTS/IEEE KONA (2011)
4. Zhang, X., Dou, W., Pei, J., Nepal, S., Yang, C., Liu, C., Chen, J.: Proximity-aware local-recoding anonymization with MapReduce for scalable big data privacy preservation in cloud. IEEE Trans. Comput. **64**, 2293–2307 (2015)
5. Abou-Tair, D., Berlik, S., Kelter, U.: Enforcing privacy by means of an ontology driven XACML framework. In: Third International Symposium on Information Assurance and Security (2007)
6. Ferretti, L., Pierazzi, F., Colajanni, M., Marchetti, M.: Scalable architecture for multi-user encrypted SQL operations on cloud database services. IEEE Trans. Cloud Comput. **2**, 448–458 (2014)
7. Chang, A., Tsai, D., Tsai, C., Lin, Y.: An improved certificate mechanism for transactions using radio frequency identification enabled mobile phone. In: 43rd Annual 2009 International Carnahan Conference on Security Technology (2009)
8. Huang, X., Du, X.: Achieving big data privacy via hybrid cloud. In: 2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (2014)
9. Raghuwanshi, D., Rajagopalan, M.: MS2: practical data privacy and security framework for data at rest in cloud. In: 2014 World Congress on Computer Applications and Information Systems (WCCAIS) (2014)
10. Pham, D., Syed, A., Mohammad, A., Halgamuge, M.: Threat analysis of portable hack tools from USB storage devices and protection solutions. In: 2010 International Conference on Information and Emerging Technologies (2010)

11. Li, X., Niu, J., Kumari, S., Wu, F., Sangaiah, A.K., Choo, K.K.R.: A three-factor anonymous authentication scheme for wireless sensor networks in internet of things environments. J. Netw. Comput. Appl. (2017). https://doi.org/10.1016/j.jnca.2017.07.001
12. Hu, P., Ning, H., Qiu, T., Xu, Y., Luo, X., Sangaiah, A.K.: A unified face identification and resolution scheme using cloud computing in Internet of Things. Future Gener. Comput. Syst. (2017). https://doi.org/10.1016/j.future.2017.03.030
13. Xiao, X., Chen, C., Sangaiah, A.K., Hu, G., Ye, R., Jiang, Y.: CenLocShare: a centralized privacy-preserving location-sharing system for mobile online social networks. Future Gener. Comput. Syst. (2017). https://doi.org/10.1016/j.future.2017.01.035
14. Tan, Z., Nagar, U., He, X., Nanda, P., Liu, R., Wang, S., Hu, J.: Enhancing big data security with collaborative intrusion detection. IEEE Cloud Comput. **1**, 27–33 (2014)
15. Mengke, Y., Xiaoguang, Z., Jianqiu, Z., Jianjian, X.: Challenges and solutions of information security issues in the age of big data. China Commun. **13**, 193–202 (2016)
16. Syed, S., Teja, P.: Novel data storage and retrieval in cloud database by using frequent access node encryption. In: 2014 International Conference on Contemporary Computing and Informatics (IC3I) (2014)
17. Feng, X., Onafeso, B., Liu, E.: Investigating big data healthcare security issues with Raspberry Pi. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (2015)
18. Hussain, F.: Internet of Things. Springer International Publishing (2017)
19. Johnson, S.: The Internet Changes Everything: In: Revolutionizing Public Participation and Access to Government Information through the Internet. http://www.jstor.org/stable/40709905
20. Kopetz, H.: Real-Time Systems. Springer, New York (2011)
21. Valentinova, B.: The 5 Methodology Milestones for Big Data. https://icrunchdata.com/blog/480/the-5-methodology-milestones-for-big-data/
22. Zeng, G.: Research on privacy protection in big data environment. Int. J. Eng. Res. Appl. 46–50 (2015)
23. Kizza, J.: Guide to Computer Network Security. Springer International Publishing, Cham (2017)
24. Kuhn, D., Walsh, T., Fries, S.: Security considerations for voice over IP systems. In: U.S. Dept. of Commerce, Technology Administration, National Institute of Standards and Technology, Gaithersburg, MD (2005)
25. Lee, W., Stolfo, S., Chan, P., Eskin, E., Fan, W., Miller, M., Hershkop, S., Zhang, J.: Real time data mining-based intrusion detection. In: Proceedings DARPA Information Survivability Conference and Exposition II. DISCEX'01, vol. 1, pp. 89–100
26. Stouffer, K., Falco, J., Scarfone, K.: Guide to Industrial Control Systems (ICS) security. National Institute of Standards & Technology, Gaithersburg (2011)
27. Montlick, T.: Method and apparatus for wireless remote information retrieval and pen-based data entry (2017)
28. Zissis, D., Lekkas, D.: Addressing cloud computing security issues. Future Gener. Comput. Syst. **28**, 583–592 (2012)
29. Vashist, R.: Big Data in Complex Systems. Springer International Publishing AG, Cham (2015)
30. Goel, S., Hong, Y.: Security Challenges in Smart Grid Implementation. In: SpringerBriefs in Cybersecurity, pp. 1–39 (2015)
31. Kaur, K., Dhindsa, K., Singh, G.: Numeric To numeric encryption of databases: using 3Kdec Algorithm. In: 2009 IEEE International Advance Computing Conference, pp. 1501–1505 (2009)
32. Li, P., Guo, S.: Load balancing for privacy-preserving access to big data in cloud. In: 2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (2014)

33. Motiwalla, L., Xiaobai, L.: Value added privacy services for healthcare data. In: 2010 6th World Congress on Services, pp. 64–71 (2010)
34. Avireddy, S., Perumal, V., Gowraj, N., Kannan, R., Thinakaran, P., Ganapthi, S., Gunasekaran, J., Prabhu, S.: Random4: An Application Specific Randomized Encryption Algorithm to Prevent SQL Injection, pp. 1327–1333 (2012)
35. Nair, K., Helberg, A., Van der Merwe, J.: An Approach to Improve the Match-on-Card Fingerprint Authentication System Security. http://hdl.handle.net/10204/8879 (2016)
36. Moura, J., Serrão, C.: Security and privacy issues of big data. In: Handbook of Research on Trends and Future Directions in Big Data and Web Intelligence, pp. 20–52 (2016)
37. Swarna, S., Maryam, S.: Increasing security level in data sharing using ring signature in cloud environment. J. Eng. Res. Appl. 01–06 (2016)
38. Charishma, P., Venkatesh, K.: Big data security analytic solution using Splunk. Int. J. Eng. Res. Appl. (IJERA) (2015)
39. Chandankere, R., Begum, M.: Secure data sharing in an untrusted cloud. Int. J. Eng. Res. Appl. **5**, 49–54 (2015)
40. Inuwa, U.: The risk and challenges of cloud computing. Int. J. Eng. Res. Appl. **5**, 5–10 (2015)
41. Kaur, K., Pathak, A., Kaur, P., Kaur, K.: E-Commerce privacy and security system. Int. J. Eng. Res. Appl. **5**, 63–73 (2017)
42. Jyothirmai, M., Ashwitha, M.: A survey of information security in healthcare sector. Int. J. Eng. Res. Appl. (IJERA) (2015)
43. Raji, F., Miri, A., Davarpanah Jazi, M.: Preserving privacy in online social networks. In: Foundations and Practice of Security, pp. 1–13 (2012)
44. Kepner, J., Gadepally, V., Michaleas, P., Schear, N., Varia, M., Yerukhimovich, A., Cunningham, R.: Computing on masked data: a high performance method for improving big data veracity. In: 2014 IEEE High Performance Extreme Computing Conference (HPEC) (2014)
45. Wagh, K., Jathar, R.: Securing data transfer in cloud environment. Int. J. Eng. Res. Appl. **4**, 91–93 (2014)
46. Merkel: Privacy. https://www.merkleinc.com/privacy
47. Hsu, C., Zeng, B., Zhang, M.: A novel group key transfer for big data security. Appl. Math. Comput. **249**, 436–443 (2014)
48. Pace, A.: Technologies for large data management in scientific computing. Int. J. Mod. Phys. C **25**, 1430001 (2014)
49. Bertine, H., Faynberg, I., Lu, H.: Overview of data and telecommunications security standardization efforts in ISO, IEC, ITU, and IETF. Bell Labs Tech. J. **8**, 203–229 (2004)
50. Mehak, G.: improving data storage security in cloud using Hadoop. Int. J. Eng. Res. Appl. **4**, 133–138 (2014)
51. Mirarab, A., Kenari, A.: A new framework for secure M-Commerce. Int. J. Eng. Res. Appl. **4**, 163–167 (2014)
52. Raghuwanshi, S., Nigoti, R.: Modified active monitoring ant clustering based load balancing over public clouds. Int. J. Comput. Appl. **167** (2017)
53. Abawajy, J., Kelarev, A., Chowdhury, M.: Large iterative multitier ensemble classifiers for security of big data. IEEE Trans. Emerg. Top. Comput. **2**, 352–363 (2014)
54. Tankard, C.: Big data security. Netw. Secur. **2012**, 5–8 (2012)
55. Kaplan, B.: How should health data be used? Camb. Q. Healthc. Ethics **25**, 312–329 (2016)
56. Suji Pramila, R., Shajin Nargunam, A., Affairs, A.: A study on data confidentiality in early detection of Alzheimer's disease. In: 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET) (2012)
57. Eler, A.: A Look Into 3 Social Video Apps: Socialcam, Viddy & Klip—ReadWrite. https://readwrite.com/2012/05/02/a-look-into-3-social-video-apps-socialcam-viddy-klip/ (2012)
58. Wang, Y.: Image enlargement engine based on one-dimensional data interpolation algorithm. J. Jilin Univ. (Inf. Sci. Ed.) **4**, (2010)

59. Skinner, G., Chang, E., McMahon, M., Aisbett, J., Miller, M.: Shield privacy Hippocratic security method for virtual community. In: 30th Annual Conference of IEEE Industrial Electronics Society, 2004. IECON (2004)
60. Dimitropoulos, L., Rizk, S.: A state-based approach to privacy and security for interoperable health information exchange. Health Aff. **28**, 428–434 (2009)
61. Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data—SIGMOD '00 (2000)

Recent Trends in Deep Learning with Applications

K. Balaji and K. Lavanya

Abstract Deep learning methods play a vital role in Internet of things analytics. One of the main subgroups of machine learning algorithm is Deep Learning. Raw data is collected from devices. Collecting data from all situations and doing pre-processing is complex. Monitoring data through sensors continuously is also complex and expensive. Deep learning algorithms will solve these types of issues. A deep learning method signifies at various levels of representation from lower level features to very higher level features of data. The higher level features provide more abstract thoughts of information than the lower level which contains raw data. It is a developing methodology and has been commonly applied in art, image caption, machine translation, natural language processing, object detection, robotics, and visual tracking. The main purpose of using deep learning algorithms are such as faster processing, low-cost hardware, and modern growths in machine learning techniques. This review paper gives an understanding of deep learning methods and their recent advances in Internet of things.

Keywords Deep learning · Convolutional neural networks · Internet of things Restricted Boltzmann machines · Autoencoder · Sparse coding

1 Introduction

One of the main subgroups of machine learning [1] algorithm is Deep Learning. Today deep learning is found everywhere. Internet and cloud include image classification, speech recognition, sentiment analysis, language processing, and language translation. Medicine and biology include cancer cell detection, diabetic

K. Balaji (✉) · K. Lavanya
School of Computer Science and Engineering, VIT University,
Vellore, Tamil Nadu, India
e-mail: balaji.2016@vitstudent.ac.in

K. Lavanya
e-mail: lavanya.sendhilvel@gmail.com

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_9

grading, and drug discovery. Media and entertainment include video captioning, video search, and real-time translation. Security and defense include face detection, video surveillance, and satellite imagery. An autonomous machine includes pedestrian detection, lane tracking, and recognizes traffic sign. Figure 1 shows that input vector representation is converted into very high-level features of data representation.

The deep learning methods are divided into four classes of representation such as Convolutional Neural Networks (CNNs), Restricted Boltzmann Machines (RBMs), Auto-encoder and Sparse Coding. The classification of deep learning methods is presented in Fig. 2.

The influence of deep learning algorithms is having the ability to classify or predict nonlinear data using a different number of simultaneous nonlinear phases. A deep learning algorithm learns at multiple levels from the input vector features of raw data to the classification features of those data. Each layer will extract features from the output of preceding layer. The Deep Neural Network consists of an input

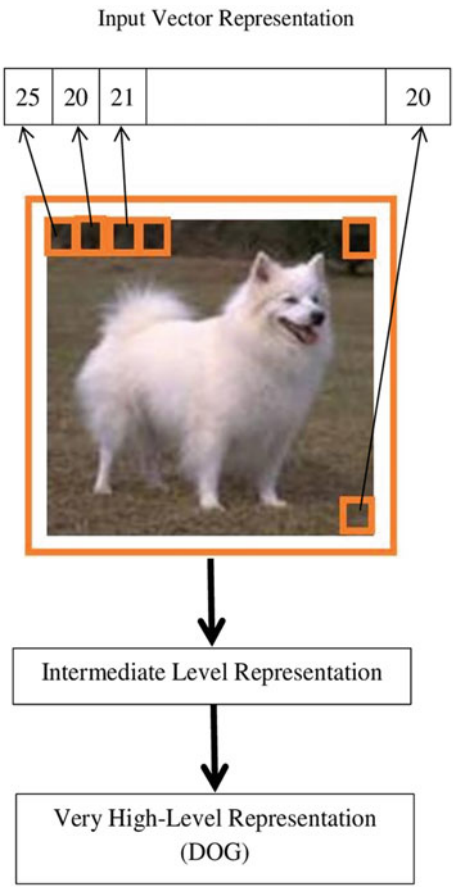


Fig. 1 Input vector representation converted into very high-level features of data representation

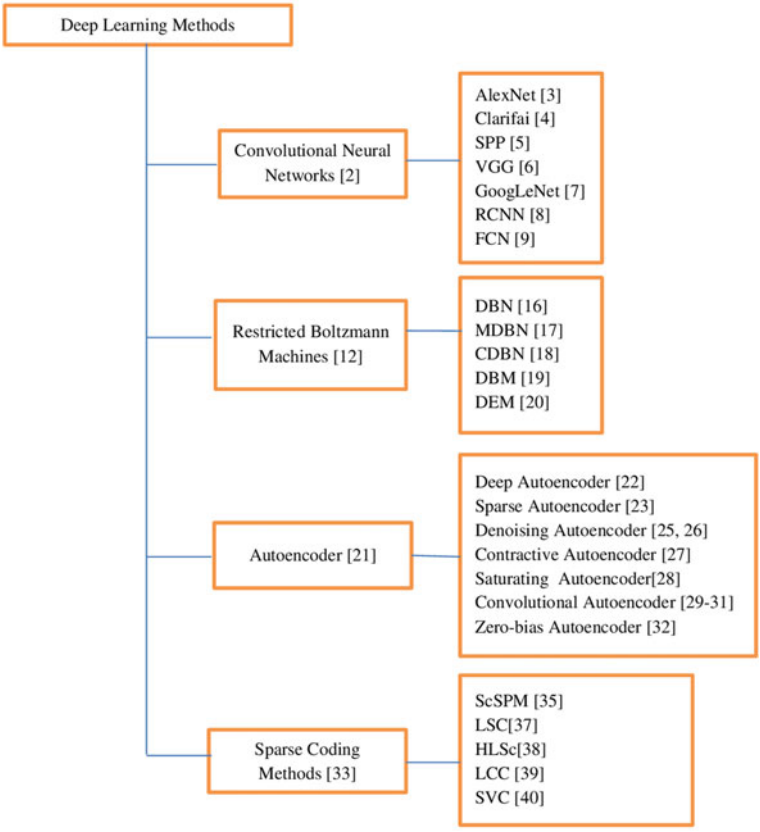


Fig. 2 Classification of deep learning methods

layer, a number of hidden layers and an output layer. It is similar to a multi-layer perceptron but having multiple hidden layers with numerous connected neurons in each layer. The most difficult transformation of features is the hidden layers. They get the input from input layers, make processing, and finally pass the classification of the object to the output layers. During the training process, deep neural networks learn from weight adjustments in order to get correct outcomes.

The methodology of deep learning is shown in Fig. 3.

Fig. 3 Methodology of deep learning

The main objectives of this research can be summarized as follows,

1. To provide an existing literature of numerous deep learning methods and their applications,
2. To classify the consequences of this domain and
3. To track the recent advances of research in this domain

The rest of this paper is organized as follows. The various deep learning methods are described such as Convolution Neural Networks in Sect. 2, Restricted Boltzmann Machine in Sect. 3, Autoencoder in Sect. 4, Sparse Coding in Sect. 5, and their applications in Sect. 6. Discussion about various models is in Sect. 7. Finally, Sect. 8 concludes the paper.

2 Convolutional Neural Networks (CNNs)

In Convolutional Neural Networks, the multiple layers are processed in a powerful way [2]. It is one of the most common efficient methods for visual image classifications [3]. The architecture of CNNs consists of three layers such as convolutional layers, pooling layers, and fully connected layers. Each layer will play every important role in their parts. The CNN structure for classification of images is shown in Fig. 4. The operation of convolution layer is shown in Fig. 5. The training of CNN contains two phases: forward phase and backward phase. Forward phase contains the representation of input images with weights and bias values in each layer. The actual output is computed and compared with the target output. The weight adjustments are done in each layer to get target output during the backward phase. After completing enough amounts of iterations for forward and backward, the learning of network will be stopped. In convolution networks, the first parameter to the network is input, and the second parameter is referred as the kernel. The output vector is called as feature map. In machine learning, the input data is represented in the form of multi-dimensional arrays and kernel is also multi-dimensional arrays. Each input and kernel vectors are stored individually. An input image may consist of a large number of pixels. We convert the input vector into kernel images by detecting only the meaningful features. So the space complexity will be reduced and computation will become faster. If there are n inputs and m outputs, then we need nm arguments and the time complexity is O (nm). If we reduce it to k inputs and m outputs, then we require only O (km) time for

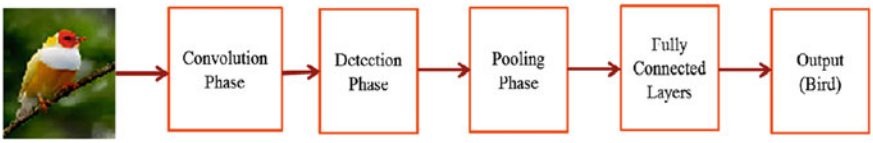


Fig. 4 CNN architecture for classification of images

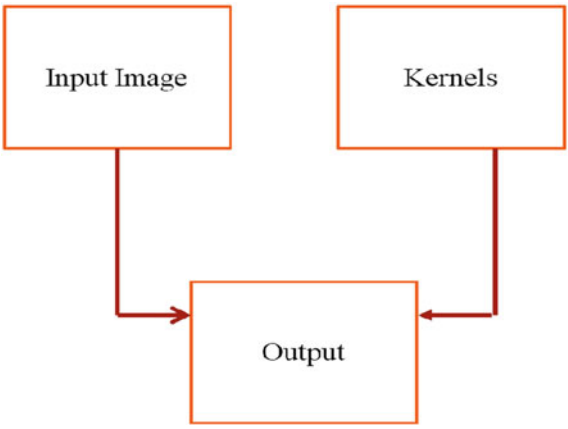


Fig. 5 Operation of convolutional layer

execution which will improve the performance. To get good image classification results using convolution layer, we can use the method of Network in Network [4]. The convolution network consists of three phases such as Convolution Phase, Detection Phase, and Pooling Phase. Convolution Phase is responsible for producing a group of presynaptic activations. The presynaptic activations are passed via rectified linear activation function during Detector Phase.

A Pooling layer is used for reduction of dimensions of input parameters from the convolutional layer. If we convert the input into smaller values, then these values become invariant. Two general techniques are available in pooling layers such as average pooling method and max pooling method [5]. The max pooling method converts the given feature map into reduced dimensions. Finally, fully-connected layers are responsible for converting two-dimensional feature maps into a one-dimensional feature vector. The comparison of CNN methods is summarized in Table 1.

3 Restricted Boltzmann Machines

In deep probabilistic models, one of the best structures is Restricted Boltzmann Machines (RBM), proposed by Hinton et al. [12]. RBM is an undirected graph model with observable variables and latent variables. RBMs are used for forming deeper models. The bipartite graph is formed with visible and hidden units. This constraint allows training effective algorithms [13]. RBM can be used to represent data of labeled as well as unlabeled pictures, documents, video, and speech recognition. RBM is an undirected graph mode contains hidden units X and the visible units Y_1 are conditionally independent. So,

Table 1 Comparison of CNN methods

Method	Application	Structure	Characteristics	Advantages
AlexNet [3]	Image classification	5 convolutional layers with 3 fully connected layers	1. It has 60 million inputs with 650,000 neurons 2. It uses non-saturating neurons	1. Implemented with GPU operation of convolution 2. Dropout method is used to reduce overfitting
Clarifai [6]	Image classification	5 convolutional layer with 3 fully connected layers	1. A new method for clear understanding of intermediate layers and their improvement	1. Implemented with GPU operations of convolution 2. Perform well with large training sets 3. Dropout method is used for performance improvement
SPP [7]	Image classification and object detection	5 convolutional layers with 3 fully connected layers	1. Spatial pyramid pooling method to produce fixed-length image for input of any image size	1. Powerful method for object deformations 2. Flexible resolution for managing various aspect ratios, sizes and scales
VGG [8]	Image classification	16 convolutional layers with 3 fully connected layers	1. Assessment of networks with growing depth	1. Improvement with 16–19 weight layers
GoogLeNet [9]	Image classification	21 convolutional layers with one fully connected layer	1. Improving depth and width of the network without additional computation resources	1. Optimizing the quality using Hebbian principle 2. Perception of multi-scale computation
RCNN (Regions with CNN features) [10]	Object detection	5 convolutional layers with one fully connected layer	1. Special category of region proposals 2. Fixed-length feature map is represented from each region 3. Group of classes from SVM	1. Improves mean average precision
FCN [11]	Semantic segmentation	19 convolutional layers with 3 fully connected layers	1. Any arbitrary sized input is taken to produce consistent output	1. Fine-tuning is performed from input classification to output segmentation for getting predictions for each network

$$P(XY_1) = P(X_1Y_1)P(X_2Y_1)\ldots P(X_nY_1)$$

The Boltzmann distribution function is satisfied by hidden units X and visible units Y_1. From the given visible units, we can get hidden units through probability distribution function. By doing weight adjustments in hidden units, we can get the best representation of a feature of visible units Y_1. Hinton [14] gives a clear practical representation for training RBMs and training issues is discussed in [15]. The RBM is composed into three different categories such as Deep Belief Networks (DBNs), Deep Boltzmann Machines (DBMs), and Deep Energy Models (DEMs).

3.1 Deep Belief Networks

Deep Belief Network (DBN) [16] is a non-convolutional structure. It will solve the issues found in the objective function of neural networks. It will outperform kernelized SVM architecture. The architecture of deep belief networks is shown in Fig. 6. It consists of different layers with latent variables. The latent variables are represented using binary and the visible units are represented using either real or binary. Instead of intra-layer communication between layers, each layer communication is done with every neighboring layer. The topmost two layers are undirected and all other layers are directed. Initialization of network architecture in each layer is performed by the greedy approach and proper weight adjustments are done to get target output.

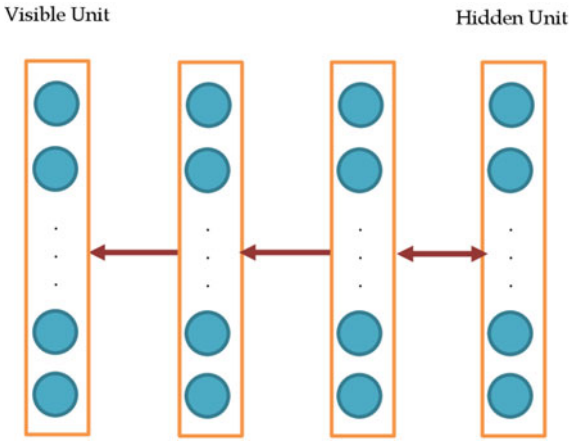


Fig. 6 Architecture of deep belief networks

3.2 Modified Deep Belief Networks

Modified Deep Belief Networks (MDBN) is proposed by Nair et al. [17]. The top-layer is made-up of the third-order Boltzmann machine. The model consists of the visible unit, hidden unit, and label unit in every clique. The model performs better than support vector machine and convolutional neural network. The oldest model limits the modification of biases on hidden units. There is no communication between the visible and label units. This model makes direct communication between all the three units. But still, there are no direct interactions between two units of the same category.

3.3 Convolutional Deep Belief Networks

The main issue of Deep Belief Network is scalability and high-dimensional images which are rectified by Convolutional Deep Belief Networks (CDBN) [18]. It is a hierarchical model which is scalable to real inputs and conversion invariant. Deep Belief Network avoids two-dimensional architecture of images and every location learns weight adjustments individually. But Convolutional DBN shares weight adjustment to every location of the image and scalable due to interpretation using convolution. Convolutional DBN is made up of a classified procreative model for full-sized images and structure consists of various max pooling convolutional RBMs stacked on top of one another. Initialization of network is done with the greedy method. After training is completed for a given layer, the weight adjustments are fixed and its activation function is given as input to the next layer. Convolutional DBN has an undirected connection between layers.

3.4 Deep Boltzmann Machines

Deep Boltzmann Machines (DBM) [19] consists of various layers of hidden units. The entire network contains undirected connections between layers. The architecture of deep Boltzmann machine is shown in Fig. 7. The connection has been made in the network from visible neurons to hidden neurons and from hidden neurons to hidden neurons, but there is no connection within the layer. Deep Boltzmann Machine has initial learning of input representation of images which becomes gradually difficult at higher layers. High-level feature map must be made with the large inputs of unlabelled data and smaller inputs of labeled data. The initialization of units in all layers becomes robust if DBM learning is made in a right way. The uncertainty conditions in intermediate layers are solved with knowledge of higher-level layers. During training process of the network, joint training of all layers is performed. Deep Boltzmann Machine has multiple layers of hidden units,

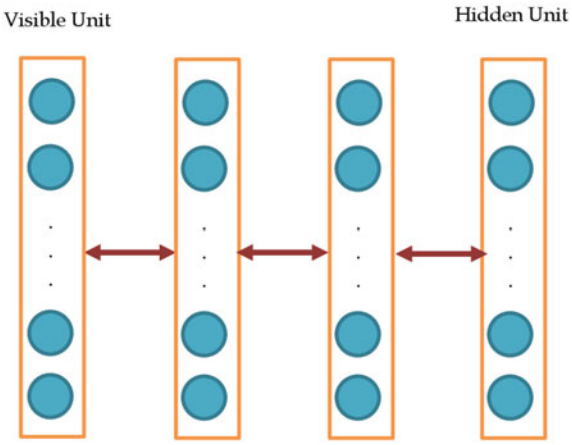


Fig. 7 Architecture of deep Boltzmann machine

in which each unit of odd-numbered layers are conditionally independent of even-numbered layers and vice versa. While training a DBM, the parameters are chosen to increase the lower bound on the probability.

3.5 Deep Energy Model

The main issue of DBN [16] and DBMs [19] is that they share the property of requiring multiple stochastic hidden layers which create interpretation and learning difficulties. Therefore, hidden units in the network become inflexible. Deep Energy Model (DEM) [20] is based on features of having a particular layer of hidden units for effective learning and interpretation. The architecture of deep energy model is shown in Fig. 8. It uses deep feedforward networks to model the energy settings that describe probabilistic prototypes. All the layers are trained concurrently from lower layers which modify the learning of higher layers to get better models. In Deep Energy Model, the input is converted by deterministic unit and output of the network with stochastic hidden units. The performance of this network is tested with natural images and shown that it produces better results by using the greedy approach of training. It also performs better results on object recognition. For training, the model Ngiam et al. [20] uses Hybrid Monte Carlo (HMC) method (Table 2).

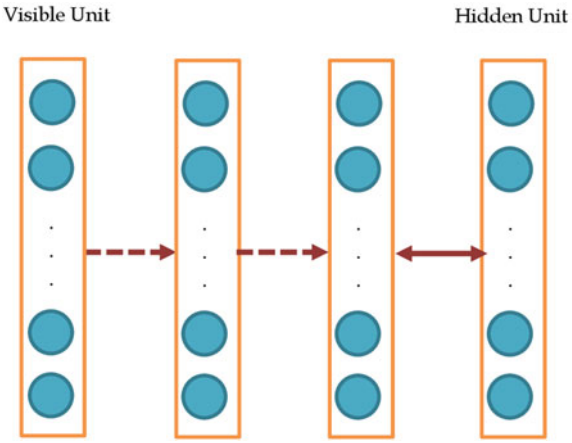


Fig. 8 Architecture of deep energy model

Table 2 Comparison of RBM methods

Method	Structure	Advantages	Disadvantages
DBN [16]	Top two layers are undirected and lower layers are directed	1. Initialization of network is done with greedy method which avoids poor local optimum values 2. Unsupervised training	1. DBN model is computationally expensive 2. Optimization of model is not clear
MDBN [17]	Top layer is made up of third-order Boltzmann machine	1. Initialization of network is done with greedy method which produces recognition of object with more accuracy 2. Unsupervised training	1. Interaction between two units of the same type is not possible
CDBN [18]	Undirected connection between layers	1. Initialization of network is done with greedy method 2. Unsupervised training	1. Decide on input parameters 2. Optimization of model is not clear
DBM [19]	Entire network contains undirected connection between layers	1. Effective method for initialization of hidden layer in feedforward network 2. Robustly compacts with uncertain inputs	1. Time complexity is higher for joint optimization of larger datasets
DEM [20]	Lower layer consists of deterministic hidden units and top hidden layer consists of stochastic hidden units	1. All the layers are trained concurrently from lower layers which modify the learning of higher layers to get better models	1. Initial weight does not have good stopping criterion

4 Autoencoder

Autoencoder [21] is a neural network feedforward approach which is trained to predict the input itself in the output. The input layer to hidden layer belongs to the encoder, whereas hidden layer to output layer belongs to the decoder. Autoencoder is an unsupervised learning technique. The optimization of autoencoder is performed by reducing the reconstruction error and learning from the code feature. The General framework of an autoencoder is shown in Fig. 9.

4.1 Deep Autoencoder

The encoder and decoder of autoencoder are constructed with multiple hidden layers is called Deep Autoencoder [22]. The encoder is used to convert high-dimensional information into low-dimensional code and decoder is used to reconstruct the data from the code. Beginning with initial weights, encoder and decoder networks are trained by reducing the difference between the original information and its reconstruction. Gradient descent method is used for fine-tuning the weights. The performance of deep autoencoder is better than principal component analysis. The main complexity faced by the deep encoder is that gradient becomes too small when it passes through many hidden layers.

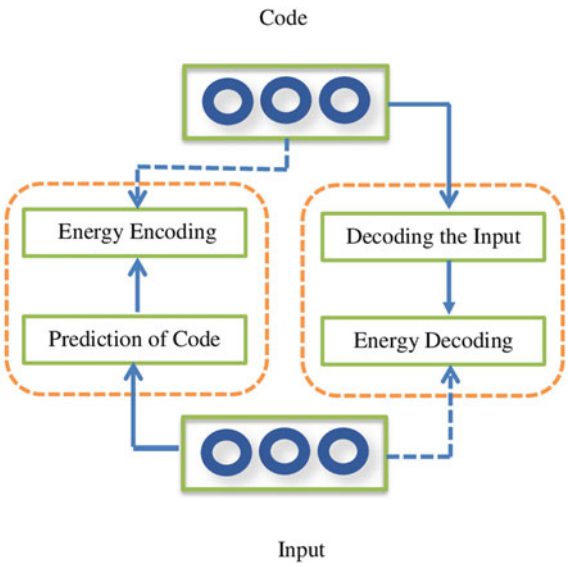


Fig. 9 General framework for autoencoder

4.2 Sparse Autoencoder

The main objective of sparse autoencoder is to extract sparse features from input dataset [23, 24]. The sparse autoencoder uses encoder and decoder managed by a sparsity constraint which converts a code vector into sparse code vector. The sparse encoder has various essential advantages. By maximizing the probability in high-dimensional representations, image classifications are feasibly linearly separable. Interpretation of input data becomes easier by using sparse representations which extract feature map in the hidden layer. Sparse representation becomes as considerable benefits in biological vision. The general framework for sparse autoencoder is shown in Fig. 10.

4.3 Denoising Autoencoder

Denoising autoencoder (DAE) [25, 26] aims to recover from corrupted input to proper input. The robustness of the network architecture will be improved. It is possible to detect noise inputs more efficiently. The general framework for denoising autoencoder is shown in Fig. 11.

4.4 Contractive Autoencoder

Contractive autoencoder (CAE) [27] is based on the concepts of denoising encoder. In denoising encoder, the noise will be removed from encoder part, whereas in contractive autoencoder, the noise will be removed from whole network

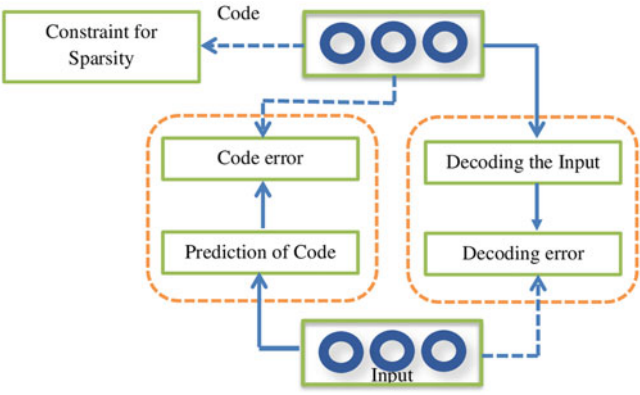


Fig. 10 General framework for sparse autoencoder

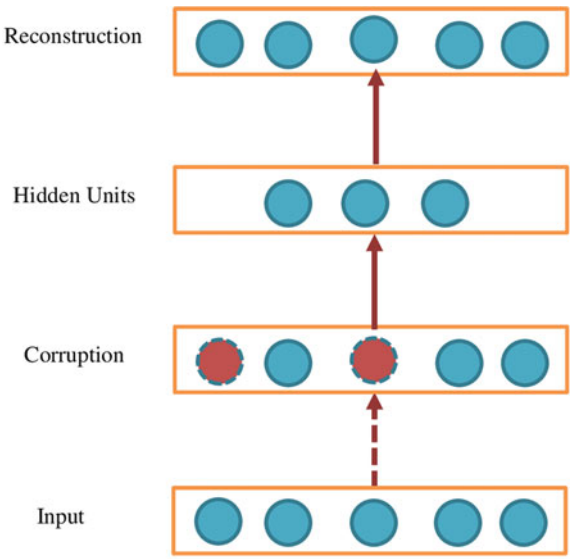


Fig. 11 General framework for denoising autoencoder

architecture. The reconstruction is performed as a whole instead of only on encoder part. It is an unsupervised learning process.

4.5 Saturating Autoencoder

Saturating autoencoder [28] is a simple autoencoder consists of at least one saturated region in the activation functions of hidden units. It has the ability to reconstruct inputs which are not near the data manifold. The network connections of saturating encoders are created with sparse autoencoders and contractive autoencoders.

4.6 Convolutional Autoencoder

The main issue in autoencoder [24] and denoising autoencoder [25, 26] is that both will not support two-dimensional images and also parameters having redundancy. Today's trends of object recognition and vision require best network structure to find localized features in their input. Convolutional autoencoder [29–31] structure consists of weights which are shared by all of the inputs. It supports three-dimensional image structures. The greedy method is applied for training the network and gradient descent approach is applied for training each layer.

Table 3 Comparison of autoencoder methods

Method	Characteristics	Advantages
Deep autoencoder [22]	Encoder and decoder with multiple hidden layers	Conversion of high-dimensional into low-dimensional data
Sparse autoencoder [23]	Extract sparse features	Linearly separable Converts complex data into meaningful Beneficial in biological vision
Denoising autoencoder [25]	Recovery from corrupted input to proper input	Learn more robust features with noisy input
Contractive autoencoder [27]	Remove noise from whole network architecture	Learn more robust features throughout the network
Saturating autoencoder [28]	Reconstruct inputs which are not near the data manifold	Limits the capability of reconstructing the input
Convolutional autoencoder [29–31]	Weights are shared among all of the inputs in the network	Supports three-dimensional image architecture
Zero-bias autoencoder [32]	A new activation function for hidden units	More robust for classification of images

4.7 Zero-Bias Autoencoder

Training of autoencoder usually outcomes in large negative values of hidden units. But hidden units are more responsible for sparse representation as well as reconstruct the correct input. Zero-bias autoencoder [32] introduces a new activation function to solve the issues of hidden units with high dimensional representation. It is more robust in the classification of images. The comparison of autoencoder methods is summarized in Table 3.

5 Sparse Coding

Sparse coding [33] is a method of describing input data set from large data sets. The sparse features provide properties of images. It is linearly separable. It is beneficial for biological vision. In sparse coding method, updating the weight is done with gradient descent algorithm [34]. The general framework for sparse coding is shown in Fig. 12.

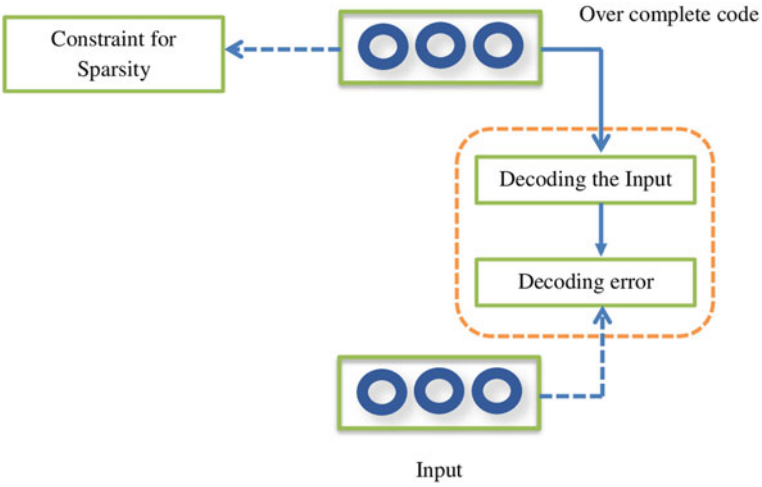


Fig. 12 General framework for sparse coding

5.1 *Sparse Coding SPM*

Sparse coding SPM (ScSPM) [35] is based on the concept of Spatial Pyramid Matching (SPM) technique [36]. Vector quantization technique is used by SPM [36] for image classification. ScSPM uses sparse coding (SC) coding method in feature extraction and followed by multi-scale spatial max pooling. The local features of images are described and detected by Scale Invariant Feature Transform (SIFT) sparse code algorithm. The sparse coding method creates features of smaller datasets from larger datasets. The reconstruction error rate is minimized in sparse coding technique when compared to vector quantization. The main issue of the ScSPM method is that it treats local features individually which avoids related dependence between them. The sparse codes possibly will contrast a percentage even for related representation.

5.2 *Laplacian Sparse Coding*

To overcome the issue of ScSPM [35], Laplacian Sparse Coding (LSC) [37] technique is proposed. LSC method will solve this issue of dependence between local features in image classification. The k-nearest neighbor algorithm is used to create a Laplacian matrix for describing similarity of local features and also realm stability in a sparse representation. The local features are not independent by using sparse codes. The error rate of local features is considerably reduced. It guarantees that similar cluster centers are selected and similar features are assigned to them. The performance of LSC is better than ScSPM in image classification. The LSC

algorithm can also be used for face recognition problem. Hypergraph Laplacian Sparse coding (HLSc) [38] technique is based on the concept of LSC and similarity of local features are represented by using a hypergraph. It improves the robustness of sparse coding. The problem of semi-auto image tagging is solved by HLSc.

5.3 Local Coordinate Coding

Local Coordinate Coding (LCC) [39] algorithm is a new technique of nonlinear learning of high dimensional data which is distributed on manifolds. It obviously inspires the coding to be local. The LCC algorithm shows that locality is more important than sparsity. The locality can improve sparsity. The sparse coding is beneficial for training only when codes are local. The codes have similar non-zero dimensions with their similar data. Optimization of the algorithm is time-consuming.

5.4 Super Vector Coding

Super Vector Coding (SVC) [40, 41] is based on the concept of vector quantization method and extends its features. SVC algorithm is composed of three steps such as descriptor coding, spatial pooling and image classification. In the first phase, it transforms non-linear features on descriptors. A new nonlinear coding method called Super Vector Coding is used. In the second phase, coding from all descriptors is combined into a single vector. The image-level feature vector is formed by combining different regions vectors. In the final phase, normalizing the

Table 4 Comparison of sparse coding methods

Method	Characteristics	Advantages
ScSPM [35]	Uses SIFT sparse codes	Reconstruction error rate is minimized
LSC [37]	The k-nearest neighbor algorithm is used for creating Laplacian Matrix	Similar features are mutually dependent by using sparse coding More robust in characterizing the local features Also used for face recognition problem
HLSc [38]	Similarity of local features are defined by hypergraph	More robust in characterizing the local features
LCC [39]	Uses L1-norm regularization	Improvement in computation against classical sparse coding
SVC [40]	Improves the local features by using a smoother coding technique	Extends the vector quantization technique

Table 5 Comparison characteristics of deep learning methods

Characteristics	Convolutional neural networks	Restricted Boltzmann machine	Autoencoder	Sparse coding
Simplification	Possible	Possible	Possible	Possible
Method of unsupervised training	Impossible	Possible	Possible	Possible
Feature mapping	Possible	Possible	Possible	Impossible
Dynamic training of networks	Impossible	Impossible	Possible	Possible
Dynamic prediction of outputs	Possible	Possible	Possible	Possible
Beneficial in biology	Impossible	Impossible	Impossible	Possible
Validation of theory	Possible	Possible	Possible	Possible
Invariance	Possible	Impossible	Impossible	Possible
Works with smaller datasets	Possible	Possible	Possible	Possible

image-level features and provide for a classifier. The comparison of sparse coding methods is summarized in Table 4. The comparison characteristics of deep learning method are summarized in Table 5.

6 Applications

6.1 Image Caption

Deep learning methods are helpful in describing image caption automatically, which is the main issue in artificial intelligence that interacts with computer vision and natural language processing. Deep learning networks are able to generate caption describing images. This application is significantly difficult compared to image classification or object recognition in computer vision. The description of image caption describes not only the content of images but also specifies how objects and attributes are related together, and actions involved with them. The main motivation of this application is the recent developments in machine translation. The input image is processed by convolutional neural network and attention over the image is done with the recurrent neural network. The final outcome presents the generation of sentences of a given input image. The general framework for image caption is shown in Fig. 13. The examples of image captions are shown in Fig. 14.

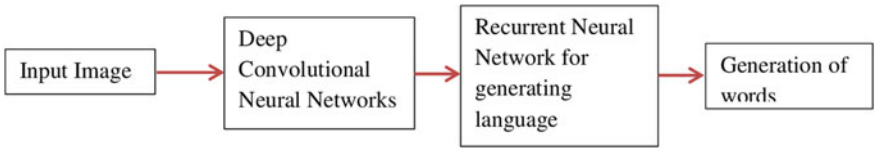


Fig. 13 General framework for image caption

Fig. 14 Examples of image caption

6.2 Object Detection

Deep Neural Networks shows an outstanding performance in object detection. Object detection is different from image classification. The entire image is given as input in image classification. But object detection is classifying the localization of objects. First, object locations are computed from the input image. After that, it computes convolutional neural network features and classifies images using output classifier. The general framework for object detection is shown in Fig. 15.

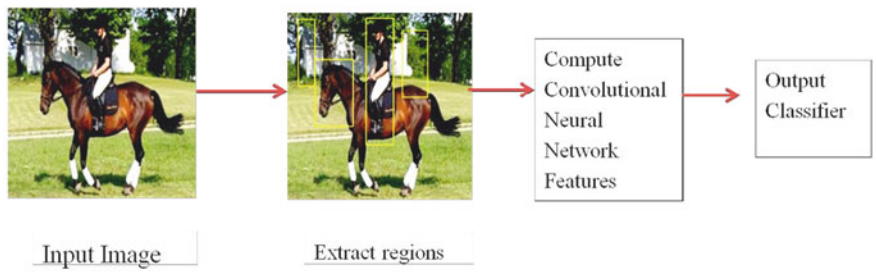


Fig. 15 General framework for object detection

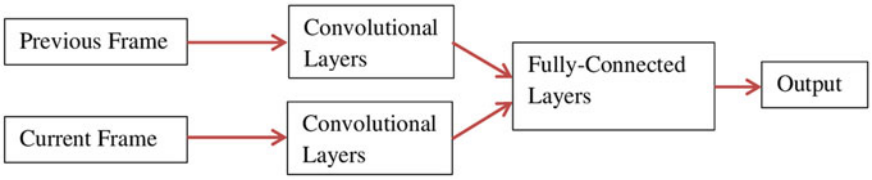


Fig. 16 General framework for visual tracking

Fig. 17 Examples of visual tracking

6.3 Visual Tracking

Visual Tracking is the automatic computation of the path of an object which moves from one place to another in a video. It has various applications which include security, analysis of sports video or human-computer interaction. When a particular problem needs multiple objects to be tracked while moving, proper setting is to be done for each object individually. Once object tracking is identified in the first frame of video, subsequent frames of an object are easy to identify. It is a very challenging problem in IOT analytics. The general framework for visual tracking is shown in Fig. 16. The examples of visual tracking are shown in Fig. 17.

7 Discussion

The main purpose of this review paper is to understand all the four deep learning algorithms and compare their features; we summarize their benefits and limitations with respect to different properties, as listed in Table 5. The simplification property shows whether the particular method will be effective in text, audio, images and their application, or not. Unsupervised Method shows that learning without a teacher. Feature Mapping is the capability to learn structures spontaneously from the

given dataset. Training and prediction of networks dynamically shows the efficiency of models. Beneficial in Biology, as well as, Validation of Theory shows whether the model has biological foundations or theoretical underpinnings. Invariance property shows whether the model has the ability of robust to transformations. Final property shows whether the model has the ability to work even with small datasets or not.

8 Conclusion

This paper reviews the concepts of deep learning methods and presents the comparison characteristics of all those methods. The deep learning methods are divided into four classifications such as Convolutional Neural Networks, Restricted Boltzmann Machines, Autoencoder and Sparse Coding. The applications with respect to Internet of things such as image caption, object detection, and visual tracking are also discussed.

Author Contribution Authors whose names appear on the submission have contributed sufficiently to the scientific work and therefore share collective responsibility and accountability for the results.

References

1. Deng, L.: A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans. Signal Inf. Process. **3**, e2 (2014)
2. LeCun, Y., Bottou, L., Bengio, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
3. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Image net classification with deep convolutional neural networks. In: Proceedings of the NIPS (2012)
4. Lin, M., Chen, Q., Yan, S.: Network in network. In: Proceedings of the ICLR (2013)
5. Boureau, Y.L., Ponce, J., LeCun, Y.: A theoretical analysis of feature pooling in visual recognition. In: Proceedings of the ICML (2010)
6. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional neural networks. In: Proceedings of the ECCV (2014)
7. He, K., Zhang, X., Ren, S., et al.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Proceedings of the ECCV (2014)
8. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the ICLR (2015)
9. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: Proceedings of the CVPR (2015)
10. Girshick, R., Donahue, J., Darrell, T., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the CVPR (2014)
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the CVPR (2015)
12. Hinton, G.E., Sejnowski, T.J.: Learning and Relearning in Boltzmann Machines, vol. 1, p. 4.2. MIT Press, Cambridge, MA (1986)

13. Carreira-Perpinan, M.A., Hinton, G.E.: On contrastive divergence clearing. In: Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, pp. 33–40. Society for Artificial Intelligence and Statistics, NP (2005)
14. Hinton, G.: A practical guide to training restricted Boltzmann machines. Momentum **9**(1), 926 (2010)
15. Cho, K.H., Raiko, T., Ihler, A.T.: Enhanced gradient and adaptive learning rate for training restricted Boltzmann machines. In: Proceedings of the ICML (2011)
16. Hinton, G., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
17. Nair, V., Hinton, G.E.: 3D object recognition with deep belief nets. In: Proceedings of the NIPS (2009)
18. Lee, H., Grosse, R., Ranganath, R., et al.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the ICML (2009)
19. Salakhutdinov, R., Hinton, G.E.: Deep Boltzmann machines. In: Proceedings of the AISTATS (2009)
20. Ngiam, J., Chen, Z., Koh, P.W., et al.: Learning deep energy models. In: Proceedings of the ICML (2011)
21. Liou, C.Y., Cheng, W.C., Liou, J.W., et al.: Autoencoder for words. Neuro-computing **139**, 84–96 (2014)
22. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**(5786), 504–507 (2006)
23. Poultney, C., Chopra, S., Cun, Y.L.: Efficient learning of sparse representations with an energy-based model. In: Proceedings of the NIPS (2006)
24. Jiang, X., Zhang, Y., Zhang, W., et al.: A novel sparse auto-encoder for deep unsupervised learning. In: Proceedings of the ICACI (2013)
25. Vincent, P., Larochelle, H., Bengio, Y., et al.: Extracting and composing robust features with denoising auto encoders. In: Proceedings of the ICML (2008)
26. Vincent, P., Larochelle, H., Lajoie, I., et al.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. **11**, 3371–3408 (2010)
27. Rifai, S., Vincent, P., Muller, X., et al.: Contractive auto-encoders: explicit invariance during feature extraction. In: Proceedings of the ICML (2011)
28. Goroshin, R., LeCun, Y.: Saturating auto-encoders. In: Proceedings of the ICLR (2013)
29. Masci, J., Meier, U., Cireşan, D., et al.: Stacked convolutional auto-encoders for hierarchical feature extraction. In: Proceedings of the ICANN (2011)
30. Baccouche, M., Mamalet, F., Wolf, C., et al.: Spatio-temporal convolutional sparse auto-encoder for sequence classification. In: Proceedings of the BMVC (2012)
31. Leng, B., Guo, S., Zhang, X., et al.: 3D object retrieval with stacked local convolutional autoencoder. Signal Process (2014)
32. Memisevic, R., Konda, K., Krueger, D.: Zero-bias auto encoders and the benefits of co-adapting features. In: Proceedings of the ICLR (2015)
33. Olshausen, B.A., Field, D.J.: Sparse coding with an over complete basis set: a strategy employed by V1? Vis. Res. **37**(23), 3311–3325 (1997)
34. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533–536 (1986)
35. Yang, J., Yu, K., Gong, Y., et al.: Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of the CVPR (2009)
36. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the CVPR (2006)
37. Gao, S., Tsang, I.W., Chia, L.T., et al.: Local features are not lonely–Laplacian sparse coding for image classification. In: Proceedings of the CVPR (2010)
38. Gao, S., Tsang, I.W.H., Chia, L.T.: Laplacian sparse coding, hypergraph Laplacian sparse coding, and applications. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 92–104 (2013)

39. Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. In: Proceedings of the NIPS (2009)
40. Zhou, X., Yu, K., Zhang, T., et al.: Image classification using super-vector coding of local image descriptors. In: Proceedings of the ECCV (2010)
41. Nan, X., Bao, L., Zhao, X., Zhao, X., Sangaiah, A.K., Wang, G.G., Ma, Z.: EPuL: an enhanced positive-unlabeled learning algorithm for the prediction of pupylation sites. Molecules **22**(9), 1463 (2017)

High-Level Knowledge Representation and Reasoning in a Cognitive IoT/WoT Context

Gian Piero Zarri

Abstract This chapter presents an overview of the Generalized World Entities (GWEs) paradigm, used to add a semantic/conceptual dimension to the ordinary IoT/WoT procedures. Its purpose is to expand the range of entities to be considered when describing a sensor-monitored environment by allowing, in particular, to seamlessly model in a unified way (i.e., within the same representation framework) physical entities like objects, humans, robots, etc. and higher levels of abstraction structures corresponding to general situations/actions/events/behaviours. The unifying factor is provided by the conceptual representation of the world used for modelling the GWEs of both types. This is ontology-based and general enough to take into account both the "static" (background information about, e.g., common notions like robot, person or physical object) and the "dynamic" (foreground information concerning, e.g., a robot or a person moving in real time towards a given object) characteristics of the different entities to deal with. After having presented a short state of the art in the cognitive/semantic IoT/WoT domain, we will specify the notion of GWE by describing its implementation under NKRL (Narrative Knowledge Representation Language) format. NKRL is a high-level modelling language, whose main characteristic concerns the use of two ontologies, an ontology of standard concepts and an ontology of events, this last dealing with the representation of the dynamic and spatio-temporal characterized information denoting behaviours, complex events, situations, circumstances etc. We will show, using several examples, that this dichotomy allows us to effectively model, in a seamlessly way, all the different entities managed by the usual IoT/WoT procedures.

Keywords Physical entities · High-level abstraction entities · Ontologies NKRL · Seamless conceptual modelling · Interoperability · Inference rules

G. P. Zarri (✉)
STIH Laboratory, Sorbonne University, 75005 Paris, France
e-mail: zarri@noos.fr

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_10

1 Introduction

This chapter is based on the assumption that, to accelerate the development of effective SWOT (Semantic web of things) applications able to make a concrete use of high-level Cognitive Science/Artificial Intelligence techniques, it is necessary to go beyond the shortcomings of the present cognitive/conceptual IoT/WoT proposals, see also in this context, e.g., [1: 308–310]. These are characterized, in fact, by (at least) two important types of limitations:

- The first consists in identifying largely the "T" of IoT/WoT with "*physical things*" and can be seen as a heritage of the first years of development of the new IoT/WoT disciplines, when these last were identified, basically, with some extensions of the RFID technology. This limitation to the physical domain *is not acceptable* in the current IoT/WoT conception where the "things" are, in reality, "entities" of a general kind and when we must take fully into account also the *domain knowledge* that represents both the general context and the inner support of all the possible IoT/WoT operations. We need to make use, then, of a *genuine cognitive/conceptual knowledge representation approach* able to denote in a unified way (i.e., within the same conceptual representation framework) *physical entities* like objects, humans, robots, etc. and *higher levels of abstraction structures* corresponding to general situations/actions/events/behaviors… As we will see more in detail below (Sect. 2.2), even in the most recent and advanced definitions of IoT/WoT that illustrate, at least in principle, the integration of "physical and virtual things", these "virtual things/entities" correspond in reality to nothing else than simple *digital world* images of traditional *physical world* entities.
- Even when the exigence of a unified and general representation—domain knowledge included—is taken into account, the tools chosen to implement this general representation are *often inadequate*. Looking in fact at the so-called semantic/conceptual IoT/WoT applications—in short, the "SWOT domain"—implemented so far, we can note that they have now adopted, essentially, a W3C/Semantic Web (SW)/Linked Data (LD) philosophy. However, the W3C/SW/LD tools are affected by several *theoretical and practical limitations* (mainly, *limited expressiveness*) as we will see more in detail in Sect. 2.3. We may wonder whether these tools can really represent the ultimate support for the development of (general) SWOT proposals.

The solution advocated in this chapter to make use of efficient Cognitive Science/Artificial Intelligence techniques in a SWOT context is centered around the use of the *GWEs paradigm* [2, 3], proposed as a powerful and general *conceptual framework* that extends the standard IoT approaches described in [4]. Its aim is to broaden the range of entities to be taken into account when describing a sensor-monitored environment by offering a *unified, coherent and seamless way to deal with both:*

- *Conceptual representations* of all the observable, *real world elementary entities* like physical objects, humans, robots, sensors, actuators, low-level signals etc.—i.e., all the entities dealt with by the usual IoT/WoT procedures.
- *Higher level of abstraction structures* corresponding to *general actions/events/behaviors/scenarios involving the previous lower level entities and their relationships*. These high-level entities are normally left aside or dealt with in a sweeping way in an IoT/WoT context, in spite of their pervasiveness within any sort of real-world situations.

The *unifying factor* that allows us to deal with both *physical* and *higher-level entities* using a coherent set of computational tools is represented by the single conceptual/ontological representation of the world used for modelling the GWEs that correspond to both types of entities. The use of this unified approach implies, among other things, the opportunity of:

- Getting rid in a general way of the *interoperability problems* that notoriously concern the IoT/WoT domain—see, e.g., the *silos flaw* drawback that denotes the development of IoT/WoT applications under the appearance of *independent vertical systems*. The use of a unique knowledge representation language, ontology-based, for IoT/WoT entities of any origin and level of conceptual complexity *assures strong semantic scalability* by permitting the easy integration of information coming from multiple, distributed and heterogeneous sources.
- Carefully filling the *semantic gap* between values gathered at the sensor level (i.e., at the *sub-symbolic level*) and their representation in conceptual semantic format (i.e., at the *symbolic level*). Actually, the unified representation framework will be able to supply a *target symbolic picture* of the original sensors outputs independently of their level of conceptual complexity. We can switch easily, e.g., from the simple instantiation of a concept like emergency_situation[1] originated from the pressure on an emergency button to the formal modelling of a structured command given to a robot that moves and acts in a specific environment.
- Implementing *advanced inference techniques*, endowed with an amount of deductive power higher than that existing within the usual SWOT applications. This is associated with the opportunity of representing in a fluently integrated way—as GWEs entities characterized by different levels of formal complexity but that are part of the same conceptual model—both the *objects* and the *contexts* (facts, events, scenarios, environments, circumstances etc.) where these objects/entities are involved. It becomes then easier to generalize from prior knowledge/experiences, to understand the environment and to draw meaningful deductions about actions, corrections, commands and strategies.

[1]In this chapter, concepts (i.e., general notions, like emergency_situation, proper of several IoT/WoT specific applications) are represented in lower case while individuals (instances of concepts, like EMERGENCY_SITUATION_7) are in upper case. Moreover, the symbolic labels of concepts and individuals always include at least one underscore symbol.

In the following, we will present in Sect. 2 a quick picture of the present state of the art in the SWOT domain; Sect. 2.3, in particular, will discuss the shortcomings associated with the uniform use of W3C/SW solutions in this domain. Section 3 illustrates the main principles that support the GWEs approach and the use in this context of NKRL, the Narrative Knowledge Representation Language. Section 3.1, in particular, will introduce a simple example concerning the construction of GWEs of different levels of complexity; Sect. 3.2 will emphasize the practical aspects of the use of the GWEs paradigm making use of the inferential tools associated with this paradigm. Section 4 will describe briefly some architectural/implementation aspects of the GWEs approach, referring the readers to other publications for further details. Section 5 is a short "Conclusion".

2 An Overview of the SWOT (Semantic Web of Things) Domain

From the publication of "historic" papers like [5, 6], a consensus has rapidly emerged that the best solution for associating *semantic features* to the usual IoT/WoT data and procedures could be identified in the use of *ontologies*. Ontologies are supposed, in fact, to represent a powerful abstraction technology, able to hide the heterogeneity of IoT/WoT entities, to act as a mediator between IoT/WoT application providers and consumers, and to support their semantic matchmaking. The GWEs paradigm is no exception in this context even if, as explained in Sect. 3, the *meaning* we associate to the notion of ontology is wider and more complete than the standard one.

2.1 Some Relevant Examples of General Ontologies

Two pre-2010 examples of general IoT/WoT ontologies are commonly mentioned in the literature, OntoSensor [7] and CSIRO [8]. OntoSensor was a sort of *general knowledge base* including a significant amount of sensor-related notions. It combined definitions of concepts and properties adopted from SensorML, the Sensor Model Language,[2] some extensions of the IEEE SUMO [9] upper ontology classes, references to ISO 19115[3] (a set of metadata used to describe geographic information and services) and allowed the use of simple queries using Protégé 2000 [10] and Prolog. CSIRO—CSIRO is the Australian Commonwealth Scientific and Industrial Research Organization—was an OWL [11, 12] ontology created for the

[2] https://live.osgeo.org/fr/standards/sensorml_overview.html (accessed May 20, 2017).

[3] https://www.iso.org/standard/53798.html (accessed May 20, 2017).

semantic representation of sensors and for *reasoning* about sensors and observations (operations, processes and results). It was built around three sets of concepts, namely conceptual entities, domain concepts, and abstract and concrete sensor properties, and gave the user the possibility to use sophisticated forms of structural and sequencing composition.

CSIRO is considered as a direct precursor of the *Semantic Sensor Network (SSN) ontology* [13]. SSN has been developed in 2010–2011 by a specific W3C Semantic Sensor Network Incubator group with the aim of overcoming the limits of pre-existing XML-based attempts and the fragmentation of sensor ontologies into specific domains applications. The SSN's architecture is structured according to the *Stimulus-Sensor-Observation (SSO) paradigm* [14], which links sensors, what they sense, and the resulting observations. Stimuli are detectable changes in the physical world that act as *triggers* for sensors. They can be either directly or indirectly related to *observable properties* (features of interest) and can also be actively produced by a sensor to perform observations. The same types of stimulus can trigger different kinds of sensors and be used to reason about different properties. Examples for stimuli include the expansion of liquids or sound waves emitted by a sonar where, e.g., expansion of mercury can be used to draw conclusions about the temperature of a surface that is in close contact. *Sensors* are *physical objects that perform observations*, i.e., that transform an incoming stimulus into another (often digital) form of representation. They implement a *method* (an abstract description) that describes the transformation of stimuli to results. *Observations* represent the context that bring sensors and stimuli together, and are then the sticking items of the SSO patterns. They define how a sensor should be realized and deployed to measure a given *observable property*; they are defined then by *procedures* that determine how a certain observation has to be carried out. From an implementation point of view, and to facilitate reuse and interoperability, the main classes of SSN have been aligned with classes in the DOLCE-UltraLite (DUL)[4] upper level ontology. SSN can be considered, then, *as an extension* of this last ontology.

SSN is currently considered as *a sort of standard* for describing sensors and the resources in the sensor networks in terms of capabilities, measurement processes, observations and deployments processes. However, it does not include *any modelling facilities* for many features of interest, like units of measurement and, mainly, for *domain knowledge in general*. To make an actual use of SSN it is then necessary to associate it with other *domain ontologies* (e.g., for meteorology, agriculture, commercial products, environmental data, health monitoring, safety services, military applications, emergency applications, etc.).

Several IoT ontologies developed after 2010 have been implemented as adaptations and improvement of the SSN ontology. For example, the IoT-Ontology [15] is an extension of SSN supporting the *automated deployment of applications in heterogeneous IoT environments*. It adds to SSN the notions of actuating, identity, and embedded devices provided by associated software agents. The extension is

[4] http://ontologydesignpatterns.org/ont/dul/DUL.owl (accessed May 20, 2017).

realized by including two new ontology layers, namely the layer for representing IoT entities (*IoT entities layer*) and, mainly, the layer for representing IoT entities alignment (*IoT entities alignment layer*)—this last intends to provide a true interoperability of IoT/WoT entities with minimal human involvement. The OpenIoT ontology [16] has been developed by integrating SSN with several existing sensor-oriented tools, such as some libraries of the Global Sensor Networks (GSN) open source project [17]. Moreover, the ontology exploits the Linked Data concept of related sensor data sets, see Sect. 2.2. From our GWEs paradigm point of view, an interesting characteristic of OpenIoT is its generalized understanding of the notion of "*sensors*", which are assimilated to *anything that can estimate/calculate the value of a phenomenon*. Thus, either a device or a computational process, or a combination thereof, could play the role of a sensor; the representation of a sensor in OpenIoT links what it measures (the domain phenomena), the physical sensor (the device) and its functions and processing (the models). Another interesting SSN-based work is an ontology for water quality management developed to support the classification of water quality based on different regulation authorities [18]. This chapter contains significant considerations about the problems triggered in a sensor context by the choice of OWL/OWL 2 (the inner support of SSN) and SWRL, the Semantic Web Rule Language [19] options, because of their *backing of monotonic inference and open world assumption*—see also, in this context, Sect. 2.3 below.

A recent, interesting system developed independently from an SSN context is the Smart Appliances REFerence (SAREF) OWL-based ontology [20]. SAREF is a *shared model of consensus* that facilitates the matching of existing assets (standards/protocols etc.) in the smart appliances domain. Its modular architecture makes use of *pre-defined building blocks* that allow separation and recombination of different parts of the ontology depending on specific needs. The notion of device (saref:Device) is central in the SAREF's world. Examples of devices are a light switch, a temperature sensor, an energy meter, or a washing machine; devices are then *tangible objects* designed to accomplish particular tasks in households, common public buildings or offices. The modular structure of SAREF is used to allow the definition of any device by associating, according to the function(s) and purpose (s) of this device, some of the pre-defined building blocks of the ontology.

SSN is a quite complex tool to use; moreover, given its layered implementation strategy (OWL2, DOLCE, DUL…) it can also be *quite ineffective* from a running point of view. The IoT-Lite ontology [21] is then a *lightweight instantiation* of the SSN ontology. This version allows us to represent and use IoT platforms without consuming excessive processing time when querying the ontology. Moreover, it represents a sort of *meta-ontology* that can be extended to represent IoT concepts in different domains. In IoT-Lite, IoT devices are classified into three main classes: sensing devices, actuating devices and tag devices; presently, IoT-Lite focuses only on sensing. A more complete (and recent) restructuring of SSN is now proposed as W3C Recommendation [22]. This new version is characterized by a *modularized architecture* based on a lightweight but self-contained *core ontology* called SOSA (Sensor, Observation, Sample, and Actuator) that includes the SSN elementary

classes and properties and that can be independently used to create elementary conceptual annotations without a too important ontological commitment. Given their different scope and degree of axiomatization, SSN and SOSA can support together a wide range of applications and use cases.

2.2 *Projects Carrying Out a "Semantic" Approach in the IoT/WoT Domain*

Several projects have been developed in these last years in a generic SWOT context, some making use of the ontologies mentioned above (especially SSN) and others employing their own ontological/conceptual tools. Due to the weighty investments of the European Commission in the IoT/WoT domain, and in its semantic/conceptual variants in particular, Europa is playing a central and internationally recognized role in this type of developments,[5] see also, e.g., [23].

The Semantic Sensor Web (SSW) applications. These systems represent a first attempt to introduce semantic features in the common IoT/WoT practice by implementing *a sort of merge* between the Sensor Web [24, 25] and the Semantic Web technologies. Sensor Webs have been conceived to deal with sensors in an *interoperable, platform-independent and uniform way*. They consist of wireless-communicating, spatially distributed sensor platforms (*pods*) deployed to monitor and explore environments using Web services and database tools. *Sensor Webs work as autonomous, stand-alone, sensing entities capable of interpreting and reacting to the data measured*; they can perform intelligent autonomous operations, such as responding to changing environmental conditions and carrying out automated diagnosis and recovery. The Sensor Web Enablement (SWE) initiative of the Open Geospatial Consortium (OGC) standardizes the web service interfaces and data encodings that can be used as building blocks for a Sensor Web [26].

In the SSW systems, the Sensor Web/Semantic Web integration is implemented as an extension of Sensor Webs that introduces an additional *semantic layer*. In this, the *semantics* of the sensor data is specified by annotating them with *semantic metadata* according to well-defined conceptual schemas (i.e., ontologies) and formal languages. In the Sheth et al. 2008 paper already mentioned [5], the semantic layer is implemented by annotating sensor data in the weather domain with spatial, temporal, and thematic semantic metadata. To this aim, the Authors utilize one of the languages proposed in a Semantic Web context, RDFa [27], an RDF variant often used for annotation purposes. To derive additional knowledge from the semantically annotated sensor data, they make use of SWRL *antecedent* → *consequent rules*. Follow-up of this and related work is now developed at the Kno.e.sis

[5] http://www.internet-of-things-research.eu/index.html (accessed May 20, 2017).

Center, Department of Computer Science and Engineering Wright State University in Dayton, USA.

The European SemSorGrid4Env project[6] investigates the use for *environmental management* of open, large-scale SSW tools based on a *semantically coherent view* of several heterogeneous sensor networks as a *global data resource*. In this context, Koubarakis and Kyzirakos [28] agree with Sheth and colleagues on the opportunity of avoiding the use of a pure RDF approach for modelling SSW metadata. They argue that RDF is only able to represent *thematic metadata* in a correct and exhaustive way, but that it needs to be extended when spatial and temporal information must also be modelled. Rather than using an existing SW language like RDFa, they have chosen to develop a specific data model called stRDF and its corresponding query language stSPARQL, an extension of SPARQL [29]. The main idea underpinning stRDF's development stems from the constraint databases domain [30] and consists in representing *spatial and temporal objects* as *quantifier-free formulas in a first-order logic of linear constraints* so that spatial and temporal data can be represented in RDF using constraints. In a SemsorGrid4Env context, a stSPARQL query evaluation module, called Strabon, has been built up in order to manage thematic, spatial and temporal metadata about environmental monitoring that are stored in stRDF format in a PostGIS DBMS.[7]

Again in a SSW context, the paper by Calbimonte et al. [31] deals with the problem of *coherently searching, correlating and combining sensor data while taking into account the heterogeneous characteristics of the sensing environments*. For the querying of sensor data within a federated sensor network, the Authors propose to make use of the SSN ontology, properly integrated with *domain-specific ontologies* for effectively modelling the underlying heterogeneous sensor data sources. The query processing utilizes a semantic-enriched mechanisms in the semantic expansion style. If, for instance, two sensors types are named differently (temperature and thermometer for example), the query processing recognizes they belong to the same type and include them in query results. A prototype has been built up and runs as the backbone of the Swiss Experiment platform,[8] a large-scale federated sensor network.

The problem concerning the "best possible choice" for the symbolic knowledge representation language to be used for the SSW semantic layers *is far from being definitely solved*. Dietze and Domingue [32], e.g., propose to make use of Conceptual Spaces (CS) [33] as a means of representing knowledge under the form of geometrical vector spaces, to enable then computation of similarities between knowledge entities by means of distance metrics. In the context of the VSAIH project, Molina and Sanchez-Soriano [34] describe a SSW application, implemented in Prolog, for interpreting and analyzing sensor data that summarize the behavior of hydrologic networks. To represent sensor knowledge, they use a

[6] http://www.semsorgrid4env.eu/ (accessed May 20, 2017).

[7] http://postgis.net/ (accessed May 20, 2017).

[8] http://www.swiss-experiment.ch/ (accessed May 20, 2017).

component-based approach formalized in many-sorted first order logic terms [35]. Another solution for knowledge representation has been adopted in SEMbySEM [36]. In this project, following a critical analysis about the use of the Semantic Web solutions for semantic-based industrial applications [37], both a specific μConcept Knowledge Representation Language and a specific μConcept Rule Language have been defined by ignoring any possible SW/W3C suggestions. An illustrative and implemented use-case of SEMbySEM concerns the management of sensors in a railway station.

Recent high-level semantic/conceptual IoT projects. Semantic Sensor Web (SSW) applications have been particularly popular in the first decade of 2000. They are now increasingly criticized as being *too sensor-centric*—or, in some specific cases, *too knowledge-centric*—without, anyway, providing *comprehensive, integrated abstractions*, in a GWEs style, for "things" and their high-level states. As noticed by Pfisterer and his colleagues [38], an employee looking for a free room for a meeting is mainly interested in real-world entities (meeting rooms) and their high-level states (e.g., free room) rather than in sensors (e.g., sensor 536) and their raw output (e.g., room temperature detected at time *T*). Moreover, this search must refer not only to the output of sensors, but also to additional knowledge like company maps and meeting schedules. Search engines should integrate, then, these different *static and dynamic data sources in a seamless way* by providing comprehensive abstractions, linked to the possibility of an efficient search, for all sort of *concrete and virtual entities* and their high-level states. *Integrated abstractions* and *virtual entities* are the present catchwords in the global SWOT domain.

The IoT-A FP7 project aimed at creating a *European architectural reference model* for the Future IoT. The output of IoT-A was then a proposal called ARM, Architectural Reference Model [39], which consisted of four parts:

- The *vision*, which summarizes the rationale for an architectural reference model.
- The *business scenarios*, i.e., the requirements issued by the stakeholders.
- The *IoT-A Reference Model*, including an IoT Domain Model as a top-level account of the architecture, an IoT Information Model explaining how to shape the IoT information, an IoT Communication Model illustrating the communication procedures between heterogeneous IoT devices and the Internet as a whole, etc.
- The *IoT-A Reference Architecture*, which supplies views and perspectives on different architectural aspects that are of concern to the IoT stakeholders, focusing on abstract sets of mechanisms rather than concrete application architectures.

IoT-A ARM aims then at providing *best practices* to the organizations so that they can create compliant IoT/WoT architectures in different application domains that are seen as i*nstances* of the above Reference Architecture. When the application domains are overlapping, the compliance to this architecture ensures the interoperability of the different solutions and allows the formation of new synergies across the domains.

Within the above general framework, the IoT-A "things" dealt with are understood as *augmented entities* formed by the association of *physical entities* with *virtual entities* see, e.g., [40: 121]. Physical entities correspond to sensors, actuators and any sort of *possible physical devices*. The introduction of virtual entities *could be* interpreted like the addition of a *new abstraction layer* capable, as in the GWEs approach, of augmenting the generality level of the IoT/WoT applications. However, in an IoT-A context, these *virtual entities* are simple *virtual* (i.e., computer-usable) *counterparts of the physical ones*, see "Physical Entities are represented in the digital world by a Virtual Entity" [40: 120]. In more detail, Bauer and colleagues state that virtual entities are "…synchronized representations of a given set of aspects (or properties) of the Physical Entity" [40: 121] and that "… each Virtual Entity must have one and only one ID that identifies it univocally" [40: 120]. Any Virtual Entity has attributes such as a name, a type and one or more values to which meta-information, like time and location, can also be associated [40: 128]. This vision of *virtual entities as simple computer usable, digital images of physical entities* is largely shared in the IoT/WoT domain. See, e.g., De et al. [41], where the IoT-A physical things are called *entities* and the virtual things *resources* (software components)—this particular terminology derives from that introduced in the SENSEI project[9]—and the Authors state: "A resource is the core software component that represents an entity in the digital world".

Several recent high-level European projects are of interest from a general SWOT point of view. For example, the iCore (Internet Connected Objects for Reconfigurable Ecosystems) project also introduces *Virtual Objects (VOs)* as the *digital world representations of real world entities* such as sensors, devices, or everyday objects [42]—see the interactions between physical and virtual entities in IoT-A. Each VO element in the model represents then an Information and Communication Technologies (ICT) object, which in turn is associated to a (non-ICT) Real World Object (RWO). An abstract, general definition of a specific VO is called a VO *template*; a live replica of a VO template is called a VO *instance*. VOs give us the possibility of obtaining a semantically enriched description of the RWOs by providing a basic set of functionalities that represent the actual functions of these Real World Objects. This includes fostering the RWOs reuse and making them to behave more autonomously, e.g., by generating events, notifications and streamed sensed data that can be tailored to the needs of different applications.

The aim of the COMPOSE (Collaborative Open Market to Place Objects at your Service) project was the specification and implementation of a general platform for the creation of new services *integrating real and virtual worlds* through the convergence of the *Internet of Services (IoS)* with the *Internet of Things (IoT)*. At its lower level, the platform deals with Physical External Resources—in practice, physical objects. To be managed, these physical objects should be submitted to an *abstraction process* that, in the COMPOSE case, was structured around the notions

[9] http://www.surrey.ac.uk/ics/research/internet-of-things/projects/completed/sensei/index.htm (accessed May 20, 2017).

of *Web Objects* and *Service Objects* [43]. Web Objects are the key elements that provide data flows into the COMPOSE platform and communicate with the platform following a standard web-based protocol. Within the COMPOSE platform, they take on a *virtual identity* under the form of Service Objects. These last can be combined for data processing related tasks, defining then "Composite Service Objects" in order to be integrated in high-level applications and services.

The BUTLER (uBiquitous, secUre inTernet-of-things with Location and contExt-awaReness) project has produced a prototype of Context-Aware Information system, able to operate *seamlessly* across various scenarios within a unified smart urban environment. Applications concern several use cases like, e.g., SmartParking and SmartTransport trials, SmartHealth, SmartShopping, etc. The interest of this project from a SWOT point of view lies in the use of advanced semantic techniques in the *spatio-temporal reasoning* and *behavior exploration* domains [44].

In the first case, BUTLER enriches the usual geo-localization contextual information with *semantic annotations* making use of custom operators that exploit the background knowledge stored in a semantic database. For example, the transformation of coordinates to symbolic names (e.g., living_room) involves the use of an operator in the location(x,y) style that, translated into a SPARQL query, calls the semantic database to infer and retrieve the corresponding name and enables rich spatial reasoning. The semantic reasoner is able, then, to automatically build up semantic links between locations and activities (linking, e.g., watching_TV with living_room). To this aim, the system makes use of a GeoSPARQL [45] enabled storage backend—such as Parliament, a high-performance triple store designed for the Semantic Web, or Strabon—able to supply, e.g., all possible rooms where a given user can be present.

Behavioral models are semantically powerful tools that can improve the system smartness by recognizing the user contexts. To create these models, BUTLER makes use of algorithms based *on both symbolic and probabilistic approaches* that are integrated within a SAMURAI [46] framework. SAMURAI is a scalable and event-based stream mining architecture that combines and displays *machine learning* (discovering co-occurrences of events and spatio-temporal correlations) and *knowledge representation* (linking positions with semantic locations and activities) results as RESTful [47] services software building blocks for complex event processing (feature extraction, information fusion, notification). Each building blocks is integrated in a loosely coupled fashion, allowing easy deployment of multiple instances of the architecture in the cloud. A detailed description of the behavioral component of BUTLER is given in [44].

A recent, interesting non-European project [48] introduces the concept of *Semantic Gateway as Service* (SGS) as a bridge between the physical world and the high-level layers of an IoT/WoT system. According to the SGS architecture, raw sensor data are transferred from external *sink nodes* to the central *gateway node* via a multi-protocol proxy. Before being forwarded, data are *annotated* using SSN and other domain specific ontologies. Semantic annotation of sensor data provide semantic interoperability between messages and supply higher-level actionable

knowledge for implementing, e.g., powerful inference procedures. The sink nodes (or base stations) act as low-level data collectors: all the sensor nodes send data to the different sink nodes, characterized by low computational resources, stringent energy constraints and limited communication resources. The gateway nodes provide connectivity among the sink nodes: they have more computing resources compared to the sink nodes and are then able to support the annotation procedures.

Among the ongoing IoT/WoT-related projects developed in the framework of the new HORIZON-2020 applied research programme[10] of the European Commission, we mention here the FIESTA-IoT project[11] because it reflects well the new trends of the IoT/WoT European research. FIESTA-IoT stands for Federated Interoperable Semantic IoT/cloud Testbeds and Applications. Based upon cutting edge semantics-based solutions like the use of a general FIESTA-IoT ontology and making use of previous work done in European projects such as, e.g., OpenIoT, IoT-A, iCore, Sensei, etc., it aims at realizing the *interconnection and interoperability* of all the diverse, existing IoT/WoT platforms, testbeds, and specific/isolated ("*silo-like*") applications. The FIESTA-IoT infrastructure aims then at representing the only entry point for the European researchers in the IoT/WoT domain, by supplying them with a unique capability for accessing and sharing IoT/WoT datasets across multiple testbeds in a testbed-indifferent way.

Application of the linked data principles in the SWOT domain. Unlike the traditional Web, where documents are crawled by following hypertext links, in the Linked Data (LD) Web [49, 50] they are crawled by following *RDF links* to gather information stating that *one piece of data has some kind of relationship to another piece of data*. One popular framework is represented by LOD, the Linked Open Data diagram.[12] By September 2011, this diagram had already grown to 31 billion RDF triples, interlinked by around 504 million RDF links; an updated version has been published in April 2014. From a sensor and IoT/WoT perspective, sources of geospatial information such as GeoNames[13] and LinkedGeoData[14] are of particular importance in a LOD context. For example, following the linked data principles, ontology instances in different independent domains can refer to GeoNames place names through their unique URIs to enable then semantic reasoning about their relative positioning.

From a general, architectural point of view, the LD approach has been criticized because of the necessity of associating all the exploitable data with HTTP URLs that point at RDF descriptions. Taken to its extreme consequences, this could mean that all the real world entities should be characterized by HTTP URLs supplying RDF data when fetched. This criticism does not nullify the practical utility of this approach in many domains, IoT/WoT included, where it is necessary to operate in a

[10] https://ec.europa.eu/programmes/horizon2020/ (accessed May 20, 2017).

[11] http://fiesta-iot.eu/ (accessed May 20, 2017).

[12] http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/ (accessed May 20, 2017).

[13] http://www.geonames.org/ (accessed May 20, 2017).

[14] http://linkedgeodata.org/About (accessed May 20, 2017).

big data context by trying to achieve useful results through the application of shallow surface techniques. These are opposed to a symbolic approach used, e.g., in a GWEs context and characterized by a careful exploration of the meaningful characteristics of *(relatively limited) conceptual domains*.

An interesting analysis of the reasons that can suggest the use of LD techniques in the Sensor Web/IoT/WoT domain is given in [51], which supplies also the Kno.e.sis position on this matter. A number of government, corporate, and academic organizations are collecting enormous amounts of data provided by environmental sensors. However, this data is too often enfolded within these organizations, processed locally according to specific domain descriptions and their specific properties and *underutilized* then by the greater community. A strategy to make this sensor data openly accessible consists in publishing them on the Linked Open Data (LOD) Cloud. This requires, however, the execution of a set of relatively complex operations that consist, mainly, in converting raw sensor observations to RDF and in linking the new RDF dataset with other datasets on LOD to allow querying and analysis over collected sensor descriptions and observations. The paper describes also a concrete experiment using observations about hurricanes like Katrina, Wilma, Charley, etc. The original data, in O&M (Observation and Measurement) format,[15] are converted to RDF using a specific API and stored on a Virtuoso open source triple store.[16] Virtuoso provides a SPARQL endpoint to query the datasets.

In the Pfisterer et al. [38] paper that describes the European SPITFIRE project, the Authors evoke some *basic possibilities* concerning the application of LD techniques in the IoT domain by using simple examples concerning the search for parking spots in Berlin. Supposing an elementary RDF graph composed of two RDF triples, [Sensor3 is-in ParkingSpot41] and [ParkingSpot41 is-in Berlin], and utilizing standard ontological knowledge stating that is-in is a transitive property, it is possible to infer that Sensor3 is in Berlin. Assuming then that sensors are described as RDF triples, a search service based on SPARQL queries can find sensors based on meta-data such as sensor type, location, or accuracy. This sort of application still refers, however, to a (limited) Semantic Sensor Web (SSW) world. To pass to more complex SWOT applications, SPITFIRE suggests (i) to integrate semantic descriptions with the LOD cloud to support semantic reasoning; (ii) to semi-automatically create semantic descriptions for sensors and things to allow end-users to use this technology at scale; (iii) to provide abstractions for things, their high-level states, and assure their integration with sensors; and (iv) to allow the search for things using a given current state. By dynamically linking arbitrary datasets on the LOD to describe complex real-world processes and to detect facts and correlations it is possible, for example, to use the parking spot sensors to provide also information about environmental pollution.

[15] http://www.opengis.net/doc/IS/OMXML/2.0 (accessed May 20, 2017).

[16] http://virtuoso.openlinksw.com/ (accessed May 20, 2017).

The University of Surrey (UK) hosts an important research group, the Institute for Communication Systems (ICS),[17] heavily involved in the development of linked data applications in the IoT/WoT domain; many of their publications are popular among people working on the relationships between LD and IoT/WoT techniques. For example, [52] is a general paper that provides an overview of recent developments in applying semantic technologies on various aspects of IoT/WoT by recalling, among other things, that "interoperability" represents the main reason for applying linked data principles to this domain. Barnaghi et al. [53] describes a linked-data platform to annotate sensors data that enables users to publish their sensor data as RDF triples, to associate any other existing RDF sensor description data, to link existing resources on publicly available linked-data repositories and make descriptions available for linked-data consumers through SPARQL endpoints. The platform, called Sense2Web, employs graphical user interfaces for annotation, which is performed using concepts obtained from open linked data sources (e.g., DBPedia and GeoNames) and other local domain ontologies. To demonstrate the linked-data usage and the integration of data from different sources, a mash-up application has been created by using Google Maps API. A *validation tool for the SSN ontology*—i.e., a tool allowing ontologies and linked-data descriptions to be validated against the concepts and properties used in the SSN model—is described in [54]. It generates validation reports and collects statistics about the most commonly used terms and concepts within the ontologies. The tool can also be used to validate linked-data and ontological descriptions against other reference ontologies.

2.3 *Problems Linked to the Use of W3C/SW Techniques in the SWOT Domain*

As already stated in Sect. 1 above, and as illustrated in the previous paragraphs, the SWOT domain seems to have now *resolutely embraced* a W3C/Semantic Web/Linked Data philosophy. This trend has undoubtedly favored the quick development of a new breed of semantic/conceptual IoT/WoT applications, more powerful and easy to use than their XML-based ancestors. However, the W3C/SW tools are affected notoriously by *theoretical and practical limitations*—as now openly admitted by some of the SW's fathers see, e.g., [55]—and some doubts can be raised about the opportunity of using these tools as the *sole vehicle* for the development of (general) SWOT applications.

On the theoretical level, an important source of frustration, as explicitly recalled by the founders mentioned above, finds its origin in the *heavy formal constraints* imposed by the strict Description Logic [56] orientation adopted by the Semantic Web and by the obligation to strictly respect the quite limiting "*decidability*" [57, 58] dogma. Other theoretical difficulties are associated with *conception weaknesses*

[17] https://www.surrey.ac.uk/ics/ (accessed May 20, 2017).

of some SW tools like the use of the Open World Assumption (OWA) or the lack of variables in the initial OWL specifications see, e.g., [3: 119–121, 37, 59].

From a knowledge representation point of view, however, the most important problem concerns the *reduced expressiveness* that affects the SW/W3C languages like RDF(S), OWL, SWRL etc.; many papers have been published about this topic. All these languages are, in fact, *binary languages*: this implies that their properties can only denote *binary relationship used to link two individuals or an individual and a value*. Consequently, there are no facilities in OWL etc. for implementing the (neatly more expressive) *n-ary relationships*: for example, it is not possible to represent directly in OWL as a single, structured and coherent conceptual unit an *n-ary (ternary) situation* represented by a simple event like "John gives a book to Mary". An interesting analysis of the different (and largely unsuccessful) efforts developed these last years to transform W3C binary tools into *n*-ary ones—RDF reification, named graphs, RDF blank nodes plus rdf:value properties, quad-tuples, n-ary design patterns, F-events etc.—is provided in [60]. The W3C languages represent, therefore, an *inadequate solution* for modelling *complex situations and events* and for dealing with any type of *dynamic and real-time information*.

With respect now to the *practical drawbacks*, we recall here that, because of their *limited expressiveness* (see above) on the one hand and of the adoption of a *quite limited inference paradigm* (inference by inheritance) derived from Description Logic on the other hand, SW/W3C languages are unable to provide concrete assistance *for building up rules*—see again the previous papers for a detailed analysis. This in spite of the existence of rule languages expressly created for use in a W3C/SW context like RuleML, TRIPLE, and SWRL—SWRL [19] enhances OWL by permitting the creation of if-then rules written in terms of OWL classes, properties and individuals. In concrete application environments, then, the SW scholar often prefer to develop rule-based applications that make use of *commercial business rules proposals* like, e.g., Ilog JRules BRMS—now IBM Operational Decision Manager[18]—and Oracle's Business Rule Language,[19] or homebased solutions installed on top of JESS[20] or DROOLS.[21] This state of uncertainty about the actual availability of reliable and user-friendly W3C/SW computational tools has *strongly contributed* to the very cautious adoption of the W3C/SW approach by the big international corporations and to the consequent scarcity of running, exemplary and large scale W3C/SW applications.

The GWEs paradigm is implemented according to a *full n-ary approach*. This does not mean that this paradigm rejects as a matter of principle any contributions coming from the W3C/SW world. For example, it is evident that all the effective

[18] https://www.ibm.com/cloud-computing/products/business-process-management/business-rules/ (accessed May 20, 2017).

[19] http://docs.oracle.com/middleware/1213/bpm/index.html (accessed May 20, 2017).

[20] http://www.jessrules.com/ (accessed May 20, 2017).

[21] http://drools.org/ (accessed May 20, 2017).

SWOT applications cannot make abstraction of the results obtained in the modelling of the most common IoT notions achieved in a SSN (and similar work) context.

3 The Foundations of the GWEs Paradigm

GWEs are entities proper to the *digital world*, i.e., they are created using one of those Knowledge Representation Languages (KRL) whose aim concerns the *modelling in computer-usable form* of large aspects of the real world and the *concrete exploitation* of the resulting formal representations. GWEs represent then the *digital counterparts* (the knowledge representation images) of *any possible* (at least in principle) *entity that it is possible to recognize within the real world*.

Fundamental constitutive elements of any ontological-oriented KRL are *concepts* and *instances*. A concept corresponds to a *particular notion* about the real world that we must represent in digital format to be able to create and run computer applications in a specific domain. These notions can correspond to very broad-spectrum concepts (like "human being", "event", or "artefact")—proper, then, to several application domains—or to concepts specifically associated with a particular application/set of applications (like "control room operator", "level of temperature", "valve" or "heat exchanger" in some IoT applications in an industrial domain). *Instances* correspond simply to *specific, single examples* of the notions represented by the concepts.

GWEs are then, in practice, *concrete examples* of specific notions of any possible sort/origin that can be recognized/described in the real world and that can be denoted in the digital world through their association to single/multiple concepts. Considering for example a possible, industrial/commercial company called "Acme", the GWE ACME_, digital image of the real world entity "Acme", can be created as an instance of the general concept company_ thanks to the insertion of an InstanceOf link associating ACME_ with this concept. Note that GWEs are *more general* than *virtual entities* as they are normally conceived in an IoT-A, iCore etc. context, i.e., as *digital images of chiefly physical entities*—see again [40, 61, 62] etc. GWEs represent on the contrary, in the digital world, *all the possible abstract and concrete entities that can be identified (and then named) in the real world*. These correspond to physical objects, but also to humans, actions, events, (the President Trump's speech in front of the Congress), scenarios (going to the restaurant), and even imaginary entities (e.g., Gandalf, or that fire-breathing green-spotted dragon).

3.1 A Simple Example

Let's assume we have to represent in GWEs terms a fragment of an IoT/WoT scenario concerning an Ambient Assisted Living (AAL) application; in this, an

ageing person, Mary, is monitored at home by a distributed control system that interrelates with Mary via a mobile robot. This fragment reads, "On a date corresponding to April 11, 2017, at half past nine p.m., the robot reminds Mary, via audio warning, of the obligation to take her pills".

In this fragment, we can remark the presence of several sorts of entities that are to be translated into GWEs of different levels of complexity:

- An *animate entity*, Mary.
- Some *physical entities*, the robot and the pills.
- A *modality*, the audio warning.
- Two *elementary events*, the first corresponding to the warning expressed by the robot—this event is identified by the surface natural language verb "to remind", see in this context, e.g., [63: 2366–2367]—the second to the information about the necessity of taking the pills (surface verb "to take").
- The *logical relations* between the above events. Being able to represent these links means to be able to represent correctly the global scenario fragment in GWEs terms.

Additional information, such as the date and the obligation, must also be represented, although they do not give rise to specific GWEs directly.

The entities included in the first three categories do not pose particular problems in order to be encoded in digital format. They correspond, in fact, to a class of *stable, self-contained and basic notions*—a sort of *background terminological/definitional knowledge*—that can be considered, at least in the short term, as *a-temporal* (or *static*), *universal and definable* a priori. Translated into digital terms, they give rise to four "*simple GWEs*", MARY_, ROBOT_1, AUDIO_WARNING_1 and MEDICAL_TABLET_1, obtained as instances of *plain concepts* like human_being, robot_, audio_warning and medical_tablet. Concepts like these can be formalized using the simple *binary model* usually employed for the creation of the present *standard ontologies*. In this, all the properties are denoted as a *binary* (i.e., accepting only two arguments) relationship linking two individuals or an individual and a value. Several ontologies corresponding to this type of *background, terminological/definitional knowledge* can be obtained freely from the Web see also the SSN ontology [13].

Let us examining now the features of those GWEs ("*structured GWEs*") that represent the two *elementary events* comprised in the fragment. Their original data refer to a *particularly complex, dynamic and structured (foreground) information* representing the *interpersonal, dynamic, unpredictable and spatio-temporal characterized behaviors* proper to the *terminological/definitional entities* (background knowledge) corresponding to "Mary", "robot", "audio warning" and "pills". The conceptual model to be employed to formalize this sort of foreground knowledge must necessarily utilize:

- *Conceptual predicates*, corresponding to surface verbs as "remind" and "take" and used to denote the *basic kind of information* conveyed by the two events.

Table 1 Some examples of high-level, "structured" GWEs

aal9.g1)	MOVE	SUBJ	ROBOT_1
		OBJ	#aal9.g2
		BENF	MARY_
		MODAL	AUDIO_WARNING_1
		date-1:	11/4/2017/21:30
		date-2:	

Move:StructuredInformation (4.42)

On 11/4/2017, at 21:30, the robot reminds Mary through an audio message of what is described in the GWE aal9.g2.

aal9.g2)	PRODUCE	SUBJ	MARY_
		OBJ	intake_
		BENF	MARY_
		TOPIC	(SPECIF MEDICAL_TABLET_1 (SPECIF cardinality_ several_))
		{ oblig }	
		date-1:	11/4/2017/21:30
		date-2:	

Produce:PerformTask/Activity (6.3)

On 11/4/2017, at 21:30, Mary must necessarily, deontic modulator oblig*(ation), take her pills.*

- The notion of *functional role* [64], used to denote the logical/semantic function of the *background terminological/definitional entities* (that represent the arguments of the predicate) involved in the elementary events. In our example, the (simple GWE) ROBOT_1 is the SUBJ(ect) of the action of sending, AUDIO_WARNING_1 the OBJ(ect) and MARY_ the BEN(e)F(iciary)—see Table 1 for the complete representation. SUBJ(ect), OBJ(ect), BEN(e)F(iciary) are functional roles.
- An appropriate, peculiar formalism to denote the *temporal and location information* and its connections with the overall representation of the elementary events.
- A way of *reifying* the resulting "structured GWEs" *to be able to refer to them within larger, complex scenarios/events/narratives* etc.—i.e., within GWEs *of the highest level of complexity*. In our example, e.g., we must specify that the "internal" structured GWE describing the warning is included in an "external" one denoting the transmission of the message by the robot.

As mentioned in Sect. 2.3 above, the standard *binary* model is *fairly inefficient* for representing the foreground, dynamic/temporally-characterized knowledge. Therefore, an *n-ary* representation must be used. Formal representations of this type will allows us, in fact, to assemble coherently within a *single* symbolic structure information (*predicate, arguments of the predicate*, *functional roles*…) that is

different from a syntactic and semantic point of view even if *conceptually related*. Using NKRL, the Narrative Knowledge Representation Language [65], a tool often employed for the conceptual modelling of *high-level, structured and spatio-temporally denoted information*, the GWEs-based global picture of the above fragment is showed in Table 1. NKRL is both a KRL and a *fully operational environment*, implemented in Java and built up thanks to several European projects. It includes powerful inference engines able to carry out complex inference procedures based, e.g., on *analogical* and *causal* reasoning.

According to Table 1, the two components of the scenarios' fragment are represented by *two structured GWEs* corresponding to *instances* of conceptual entities that, in this case, do not denote simple *concepts* but *multilayered templates*. NKRL's templates—collected in a *Hierarchy of Templates* (HTemp)—formally designate *classes* of elementary events/states/situations like, e.g., displacement of a physical object, production of a supporting service, messages sent or received, the state of an entity is changed, etc.

As it appears from the two structured GWEs of Table 1, templates (represented implicitly in this table under the form of *templates instances*) are *n-ary structures* formed of several triples of the "predicate—functional role—argument of the predicate" form. The triples are *indissolubly connected* and have the predicate in common—MOVE and PRODUCE in Table 1, but also BEHAVE, EXIST, EXPERIENCE, OWN, and RECEIVE [65: 56–68]. Extra formal features of the *determiners/attributes* types are used to supply *additional information* with respect to the (ternary) basic structure of templates and their instances. For example, the deontic modulator oblig(ation) has been employed in aal9.g2 to denote the *absolute necessity* of taking the pills. The *temporal attributes* date-1/date-2 introduce the *temporal information* proper to the original elementary events, see [65: 80–86] in this context. Note that, in this example, we have reified under the form of a unique instance, MEDICAL_TABLET_1 (a *simple GWE*) the undetermined (in number and quality) set of pills that Mary must take—this will allow us of referring again to the same set of pills for, e.g., any sort of checking operations. intake_ (a specific term of personal_activity), cardinality_ (a specific term of quantifying_property) and several_ (a specific term of logical_quantifier) are *concepts* included in HClass —the "hierarchy of classes" that correspond to the *standard ontology* of NKRL. (SPECIF cardinality_ several_) is the normal modality for expressing *plurality* in NKRL. SPECIF(ication), the "*attributive operator*", is one of the four NKRL operators used to create *complex arguments* (expansions) of the predicate. The attributive operator is used to add, in a recursive way, some additional information to the term that represents the first argument of the SPECIF lists, MEDICAL_TABLET_1 and cardinality_ in the example.

In the template instances (i.e., the structured GWEs), *semantic labels* like, e.g., aal9.g1 in Table 1, *reify* the global formal structures giving them a *name*. The semantic labels can be used, then, to *associate together* several independent structured GWEs, allowing then the symbolic representation of (very complex in case) *real world scenarios*. Looking at Table 1 for example, the transmission of the message to Mary is represented by assuming the symbolic label aal9.g2,

which denotes *indirectly* the content of the message, as the OBJ(ect) of the transmission of information represented by the aal9.g1 GWE. This associative modality, which utilizes *Higher Order Logic (HOL) structures*, is called *completive construction* in NKRL [65: 87–91].

A second HOL linking modality is the *binding construction* [65: 91–98]. According to this, several *symbolic labels* denoting elementary events are collected into a list as *arguments* of a particular *binding operator*. Conceptual tools of this type are, e.g., (CAUSE s_1 s_2), denoting that the event (structured GWE) specified by the label s_1 originates from the event denoted by s_2, and (GOAL s_1 s_2), meaning that the goal of the s_1 is the creation of the situation denoted by s_2. For example, a new *high-level GWE*, labelled as aal9.g4 and consisting in a *binding construction* as (CAUSE aal9.g1 aal9.g3) could be used to specify that the warning has been caused by Mary's failure to comply with her obligations. aal9.g1 still denotes the transmission of the message; aal9.g3 is a structured GWE similar to aal9.g2, where (i) the *deontic modulator* oblig(ation) has been replaced by the *modal modulator* negv (negated event) denoting that the ingestion of the pills did not take place, and (ii) a *temporal interval* ending with 11/4/2017/21:30, the date of the message, must be inserted in place of the exact date indicated in aal9.g2.

3.2 The GWEs Paradigm and the Reasoning Procedures

To illustrate the practical use of the GWEs paradigm, we supply here some basic information about the use of this paradigm in association with the *automatic reasoning* capabilities of NKRL see, e.g., [59, 65: 183–243, 66].

The general framework of the GWEs' use in an inference context. *Reasoning* in NKRL ranges from the *direct questioning* of a knowledge base of *structured GWEs*—making use of search patterns p_i (formal queries) that unify information in the base thanks to the use of a Filtering Unification Module (*Fum*), see [65: 183–201]—to *high-level inference procedures* that utilize powerful *InferenceEngine*(s). Search patterns p_i are very important in an NKRL context. Apart from giving the user the possibility of asking directly, in an information-retrieval style, some questions to a GWEs knowledge base, they can be also *automatically generated* by the *InferenceEngine*(s) as the final forms of the different *reasoning steps* that constitute the high-level inference procedures. Formally, these patterns correspond basically to those "*templates*" that are used to build up *structured GWEs*. For example, in Table 1, the GWE aal9.g1 has been obtained by instantiating the Move:StructuredInformation template included in the (HTemp) hierarchy that includes the (about 150) easily adjustable and customizable NKRL templates.

The NKRL high-level inference procedures involve mainly two classes of rules, "*transformations*" and "*hypotheses*", see [65: 201–239].

The transformation rules try to *adapt*, from a semantic/conceptual point of view, an *unsuccessful* search pattern p_i (i.e., a pattern that was unable to find a match

within the knowledge base) to the actual contents of the base using a sort of *analogical reasoning*. Transformations attempt, then, to automatically *transform* p_i into one or more different p_1, p_2 … p_n that are not *precisely equivalent* but only *semantically close* to the original one. Therefore, a transformation rule is made of a left-hand side, the *antecedent* (the search pattern to be transformed) and of one or more right-hand sides, the *consequent*(s)—the one/more search patterns to be used to replace the given one. Denoting with A the antecedent and with Cs_i all the possible consequents, these rules can be expressed as:

$$A(var_i) \Rightarrow Cs_i(var_j), var_i \subseteq var_j \quad (1)$$

The restriction $var_i \subseteq var_j$ is the usual *safety condition* that guarantees the logical congruence of the rule by asserting that all the variables declared in the antecedent A appear also in the consequent Cs_i complemented, in case, by additional variables. As an informal example, we will mention here a transformation rule used in the context of a NKRL application concerning the management of *storyboards* in the oil/gas industry [67]. Suppose we want to ask whether, in a knowledge base including all the GWEs related to the activation of a gas turbine, we can recover the information that a specific oil extractor is running. *In the absence of a direct answer*, we can trigger the transformation inference engine to reply *indirectly* by providing information that is only *somewhat related* (analogical reasoning) to the subject of the original query, e.g., a GWE stating that the site leader has heard the working noise of the extractor. This result can be rephrased as: "The system cannot assert that the oil extractor is running, but it can attest that a site operator has heard its working noise".

With respect to the hypothesis rules, these allow us to create automatically a kind of *causal explanation* for an event (structured GWE) found within the knowledge base. These rules are formalized as *biconditionals* of the type:

$$X \text{ iff } Y_1 \text{ and } Y_2 \ldots \text{and } Y_n, \quad (2)$$

where the head X of the rule corresponds to a structured GWE g_j to be "explained" and the reasoning steps Y_i—the "*condition schemata*"—must all be satisfied. This implies that, for each of them, at least one *successful* search patterns p_i must be created by the hypothesis *InferenceEngine* in order to find, using *Fum* (see above), a positive match with some information of the knowledge base. In this case, the set of g_1, g_2 … g_n structured GWEs retrieved by the condition schemata Y_i thanks to their conversion into p_i may be interpreted as a context/causal explanation of the original GWE g_j (X).

We can now suppose we have directly retrieved, in a querying-answering mode, the information: "Pharmacopeia, a USA biotechnology company, has received 64,000,000 dollars from the German company Schering in connection with an R&D activity"; this corresponds to g_j (X). Using a hypothesis rule, we can *automatically construct* a sort of *causal explanation* for this event by retrieving information like: (i) "Pharmacopeia and Schering have signed an agreement concerning the production by Pharmacopeia of a new compound", g_1 (Y_1) and (ii) "in this framework,

Pharmacopeia has actually produced the new compound", g_2 (Y_2). Note that these "explications" correspond only to some of the *possible causes* of the original event: a hypothesis rule must always be understood as a member of a (potentially large) set of possible explication procedures.

An important development of NKRL is related to the possibility of using the two modalities of inference in an *integrated* way, see Zarri [65: 216–234]: this means, in practice, the possibility of making use of "*transformations*" when working in a "*hypothesis*" context. Therefore, every time a pattern p_i is obtained from a hypothesis condition schema Y_i, we can use it *as it is*—i.e., as it was been originally created by *InferenceEngine* from its "father" condition schema—but also in a *transformed form* if a suitable transformation rules exist. In this way, a hypothesis that was supposed to fail can now continue if a new p_i, derived by transformation, will find a successful unification within the knowledge base, obtaining then new values for the hypothesis variables.

An actual example of use of the GWE/NKRL inference procedures. This use case has been introduced to: (i) give some precisions about the *specific formats* used by the GWE/NKRL inference rules and the *use of variables* in particular; (ii) supply further information on the use of the *InferenceEngine*(s); (iii) show the importance of an *integrated use* of transformations/hypotheses. The example—concerning again the management of storyboards in the oil/gas industry—is based on the use of a simple hypothesis (*h1*, Table 3). This will be activated after the retrieval in the storyboard knowledge base (*KB*) of a specific GWE relating the stop a technical/industrial procedure—in our example, the stop of the start of the GP1Z_TURBINE, see the virt3.g14 in Table 2.

Hypothesis *h1* tries to verify whether this stop is linked with the discovery of an industrial accident *in the general environment of the specific tool* (i.e., the GP1Z_TURBINE) concerned by the stopped technical/industrial procedure.

When *h1* (Table 3) is launched after the (successful) unification of its premise with the structured GWE of Table 2—this unification is needed to check that the selected hypothesis is able *in principle* to supply an explication of the event related by the original GWE—its processing *halts* when trying to find successful unifications in the *knowledge base KB* with search patterns p_i derived from the first condition schema (*cond1*). No one of these patterns unifies in fact the structured

Table 2 A GWE relating the stop of the activation of the GPIZ_TURBINE

virt3.g14)	PRODUCE	SUBJ	INDIVIDUAL_PERSON_102: (GP1Z_MAIN_CONTROL_ROOM)
		OBJ	activity_stop
		TOPIC	(SPECIF turbine_startup GP1Z_TURBINE)
		date-1:	1/11/2016/10:20
		date-2:	
Produce:PerformTask/Activity (6.3)			
The production activities leader ends the start-up of the GP1Z_TURBINE			

Table 3 A hypothesis rule to explain the halting of an industrial activity

h1: "*halting procedure*" *hypothesis*		
premise:		
PRODUCE	SUBJ	*var1*
	OBJ	activity_stop
	TOPIC	(SPECIF technical/industrial_procedure *var2*)
var1 = human_being		
var2 = technical/industrial_tool		
An individual has stopped a technical procedure concerning a given industrial tool		
first condition schema (*cond1*):		
PRODUCE	SUBJ	*var3*
	OBJ	detection_
	TOPIC	(SPECIF *var4* (SPECIF *var5 var2*))
var3 = human_being		
var3 ≠ *var1*		
var4 = industrial_accident		
var5 = spatial_relationship, relational_property		
A different individual had discovered an accident in the environment of the stopped tool		
second condition schema (*cond2*):		
BEHAVE	SUBJ	*var1*
	MODAL	*var6*
var6 = industrial_site_operator		
This first individual is an industrial site operator		
third condition schema (*cond3*):		
BEHAVE	SUBJ	*var3*
	MODAL	*var7*
var6 = industrial_site_operator		
The second individual is an industrial site operator too		

GWEs included in the *KB*; this means that information in the *KB* base *is not able to prove directly* that some sort of accident has concerned the GP1Z_TURBINE.

To try to overcome this impasse, the system explore its rule base to see whether it can find a transformation rule whose antecedent corresponds to *at least one* of the patterns p_i originated from the condition schema where the failure occurred. The transformation rule of Table 4 corresponds to this requirement—i.e., *Fum* can immediately check that the antecedent of *t8* corresponds to at least one of those p_i, see also Table 5—and, in this case, the consequent of *t8* can be activated. This corresponds to state that detecting an accident, which implies a *component of the tool affected by the stop procedure or another tool deeply associated with the stopped one*, can be assumed as *equivalent* to the detection of an accident regarding the specific object of the halting procedure (i.e., the GP1Z_TURBINE in our case).

Note that the main problem concerning the execution of transformations in a hypothesis context concerns the possibility of discovering a *correspondence*

Table 4 Transformation rule concerning 'related' accidents

t8: "*part of, linked with*" *transformation*		
antecedent:		
PRODUCE	SUBJ	*var1*
	OBJ	detection_
	TOPIC	(SPECIF *var2* (SPECIF *var3* *var4*))
var1 = individual_person		
var2 = industrial_accident		
var3 = relational_property, spatial_relationship		
var4 = technical/industrial_tool		
first consequent schema (*conseq1*):		
PRODUCE	SUBJ	*var1*
	OBJ	detection_
	TOPIC	(SPECIF *var5* (SPECIF *var6* *var7*))
var5 = industrial_accident		
var6 = relational_property, spatial_relationship		
var7 = technical/industrial_tool		
var7 ≠ *var4*		
second consequent schema (*conseq2*):		
OWN	SUBJ	*var7*
	OBJ	property_
	TOPIC	(SPECIF *var8* *var4*)
var8 = part/whole_relationship, binary_relational_property		
Being unable to identify directly an accident that concerns a particular industrial tool, we can (i) check if an accident involving another tool can be discovered, and (ii) try to show then that this second tool is either a component of the original tool, or it is strictly associated with this last one		

between the variables var_h initially present in the hypothesis condition schema from which the search pattern to transform has been obtained (see the variables named as *var2, var3, var4* and *var5* in the *cond1* schema of Table 3) and those, *totally unrelated*, var_t, that appear in the (*application independent*) transformation rules to be used (see the corresponding variables named, in contrast, as *var1, var2, var3* and *var4* in the *conseq 1* schema of Table 4). To let the hypothesis at hand to resume after the failure, new values must be associated to var_h thanks to the execution of the consequent part of the transformation; if this last is successful, some values will be found in fact for the transformation variables, *but these concern* var_t *instead of* var_h. Finding a correspondence implies the execution of a complex set of comparison among the values already stored for var_h and the new ones retrieved for var_t: this procedure is explained in detail in, e.g., [65: 219–221].

Table 5 Partial results concerning the application of *t8* of Table 4

The (instantiated) search pattern to be transformed		
PRODUCE	SUBJ	human_being
	OBJ	detection_
	TOPIC	(SPECIF industrial_accident (SPECIF spatial_relationship GP1Z_TURBINE))
GWE retrieved making use of *conseq1*		
PRODUCE	SUBJ	INDIVIDUAL_PERSON_104: GP1Z_COMPLEX
	OBJ	detection_
	TOPIC	(SPECIF lubrication_oil_leakage (SPECIF around_AUXILIARY_LUBRICATION_PUMP_M202))
	date-1:	1/11/2016/10:10
	date-2:	
INDIVIDUAL_PERSON_104 *has discovered the presence of a lubrication oil leakage around the lubrication pump* M202		
GWE retrieved making use of *conseq2*		
OWN	SUBJ	AUXILIARY_LUBRICATION_PUMP_M202
	OBJ	property_
	TOPIC	(SPECIF related_to GP1Z_TURBINE)
	{ obs }	
	date-1:	1/11/2016/10:10
	date-2:	
On November 11, 2016, at 10h10, we can observe (temporal modulator obs*(serve)) that the auxiliary lubrication pump is related to the* GP1Z_TURBINE		

The results concerning the application of transformation *t8* of Table 4 in the context of the processing of the first condition schema *cond1* of hypothesis *h3* are partially reproduced in Table 5. Trying to construct *automatically* a context/causal explanation for the GWE virt3.g14 of Table 2 shows then, eventually, that the discovery of an industrial accident in the environment of the GP1Z_TURBINE turbine can be brought back to (i) the detection of an *oil leakage* that concerns the AUXILIARY_LUBRICATION_PUMP_M202, and (ii) the proof of a *relationship* between this pump and the turbine.

Several (successful) experiments concerning the association of the GWE paradigm with the NKRL inference tools have been carried out in these last year at the LiSSi (*Laboratoire Images, Signaux et Systèmes Intelligents*) of the Paris-Est/Creteil (UPEC) University, in domains like IoT/WoT, AAL (Ambient Assisted Living) and Ubiquitous and Cloud Robotics—see, e.g., [68, 69]. In [69: 165–171] we can find, e.g., the detailed description of a complex scenario (implying the integrated utilization of several hypotheses and transformations) where a robot detects the absence of food in the refrigerator of an aged person and identifies a relative of this person close to a supermarket and able then to run an errand.

Note, to conclude, that complex situations analogous to the above scenarios are routinely managed in the context of important IoT/WoT applications. A use case in BUTLER concerns, e.g., the NFC car parking examples [70]; an analogous scenario is described in CALIPSO [71]. See also the manufacturing and traceability scenarios in EBBITS [72: 42–43], some very extended narratives that describe the IoT-A logistic and health at home scenarios [73], the weather observation experiments summarized in [5], etc. Very often, however, and at the contrary of what we have seen in the previous examples, the representations of the IoT scenarios are only *partially formalized* and largely expressed in *natural language*. This implies, among other things, *reduced possibilities* of adequate inference activity.

4 Practical Aspects Concerning the GWE Paradigm

In this section, we will supply some *essential information* about the practical aspects concerning the implementation of concrete GWEs applications, like the architectural features and the procedures needed to create and utilize knowledge bases of (simple and structured) GWEs. See, for more details [2, 3].

4.1 The Architectural Aspects

FIWARE[22] and oneM2M[23] are well-known and fully operational examples of open platforms able to both integrate hardware and to enrich IoT/WoT installations with further capabilities and flexibility through components like brokers and enablers. At a higher structural and conceptual level, systems that take into account issues like *semantic integration and interoperability of interconnected devices* are now *slowly emerging*—this delay is not so surprising since even general IoT/WoT tools that seem to have been unanimously adopted like the SSN ontology find it difficult to be concretely used, see Sect. 2.1 above. In a SWOT domain, *advanced architectural proposals* have been proposed recently in the context of EC-supported projects. Interactions among the different BUTLER platforms to achieve access and communication among embedded devices, servers and mobile terminals are described in [40: 8]. A Service-Oriented Architecture (SOA) and its use for distributed intelligent service structures in the EBBITS platform is illustrated in [72]. Bassi et al. [39] describe the architectural aspects of the IoT-A project. Several architectural solutions for advanced IoT/WoT applications are mentioned in [74], see projects like OpenIoT, iCORE, COMPOSE, SmartSantander, etc.

[22] https://www.fiware.org/ (accessed May 20, 2017).

[23] http://www.onem2m.org/ (accessed May 20, 2017).

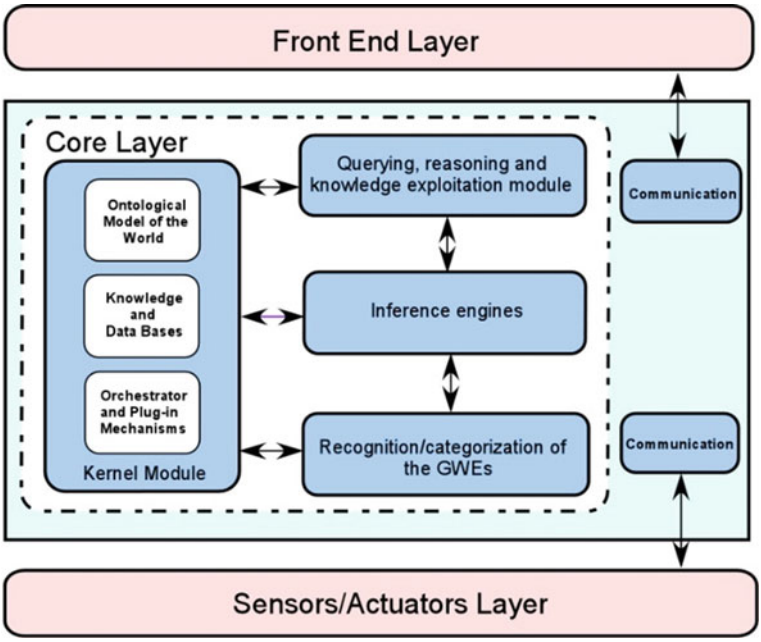


Fig. 1 A schematic view of the architecture of a GWEs-based IoT platform

With respect to the GWEs paradigm, we have already stated that this allows us to deal in a logically coherent *unified way* with both *physical* entities detected at sensor level and with *higher level of abstraction* structures. From an architectural perspective, this approach *must be reflected* in the design of the *middleware layer* of any IoT/WoT system that implements this paradigm. This layer must be realized, then, according to a strongly *cognitive-oriented design*. A high-level representation of this sort of platform's architecture—structured into three main layers—is represented in Fig. 1, where the central function of the middleware layer is emphasized. The layers are:

- The *Front End* layer, used to connect with the users and the stakeholders in order to monitor and/or to interact with the applications through *views* that represent parts of the global system. Moreover, in the set-up phase of a specific application, the Front End addresses the users' needs by providing a set of services and features like, e.g., the definition of the associated authorizations. This layer is a web-server component; XUL-compatible web-browsers[24] can be used, e.g., for the connection.
- The *Middleware* (or *Core*) layer handles the *different applications* according to the GWEs paradigm. It contains the modules to be used to build up and to

[24]https://developer.mozilla.org/en-US/docs/Mozilla/Tech/XUL (accessed May 20, 2017).

manage the GWEs, the inference engines, the ontological tools etc., and it is devised as a plug-in modular set of mechanisms allowing the inclusion/deletion of supplementary modules. The platform should then be *open* to allow, with a limited personalization effort, the easy plug-in of: (i) new modules that could be needed to improve the functioning of the platform; (ii) domain-specific GWEs applications.

- The *Sensors/Actuators* layer permits the communication *between the monitoring and actuation entities* (sensors, robots, services mounted on these entities) and the *core layer*. An external entity can be accessed only through this layer, which notifies the core about the state changes concerning the input entities. Conversely, in case the inference operations imply some modifications of the external world (close/open a door, remove an obstacle, activate an alarm…), the sensors/actuators layer receives the state changes and operations decided in the core model and translates them into data/information/commands intelligible by the external world.

The tasks accomplished by the main modules of the core are as follows:

- *Kernel module*: deals with the GWEs semantic/conceptual model and provides the *interfaces* for accessing this model. It includes several *functional components*, e.g.:
  - *Tools for the ontological/conceptual model of the world* (see Sect. 4.2). This component provides the tools for *storing and maintaining* the different GWEs ontological structures; it assures the correct instantiation of these components to build up GWEs of different levels of complexity. It must also be able to update the model according to the evolutions of the external world and the results of the inference operations.
  - *Knowledge and databases*. This component manages a set of *permanent data and knowledge structures* storing both metadata about the maintenance of the world model and knowledge about the various contexts. An important role is played by the data persistency and security/privacy functions that assure, among other things, model recovery in case of failure.
  - *Orchestrator and Plug-in mechanism*. This component is in charge of the *orchestration of all the core/middleware modules*. It provides the interfaces for accessing the model of the world and deals with the notification of the changes in this model; note that this component must wait for the results produced by the inference engines before issuing any modification message. It also enables the inclusion of other modules (safety, resilience, security…) into the core.
- *Recognition/categorization of the GWEs module*: This module is in charge of the operations regarding GWEs of different degrees of complexity that correspond to *new entities and events detected at the Sensor/Actuator Layer level*. These operations use the conceptual representation of the world stored and managed in the kernel module and of a set of inference rules, see Sect. 4.2.

- *Inference engines module*: Inference procedures of diverse degrees of complexity are used in practically all the phases of the GWEs processing. *Several sorts of inference engine will then be employed*. These will range from tools in the SW style as RACER, Pellet, Fact++ or Hoolet to be used in all the operations of the subsumption type like those, e.g., corresponding to the automatic/semi-automatic updating of the ontology, to the backtracking based, multiple reasoning-steps engines needed to deal with complex uses of GWEs like those mentioned in Sect. 3.2.
- *Querying, reasoning and knowledge exploitation module*: This module loads and handles the user's queries coming from the Front End layer. Moreover, it makes use of different sorts of inference engines to execute the rules proper to a given application. "Knowledge exploitation" means that, starting from the conceptual representation of the world and from the results of the inference engines, it is possible to improve both the *breadth* and the *depth* of the knowledge of the environment. The module can get, e.g., new/richer information about detected events, or pro-actively move the sensors to acquire new significant information.
- *Communication module*: This enables *asynchronous communications* among the layers. Some Core's internal components can however communicate directly via synchronous services or internal messaging systems.

4.2 Creating and Utilizing the GWEs Structures

See again, for more detailed information [2, 3].

Input entities. We deal here with the *identification and the accurate characterization* of all the possible input entities coming into a GWEs-based system from (an) external data stream(s) and to be transformed into GWEs. These entities can correspond to static objects, persons, multimedia information but also to spatio/temporally bounded events/situations/circumstances to be converted into structured GWEs. The original data stream(s) can be generated from a variety of different hardware-based sensors of different levels of complexity, including RFIDs, contact switches and pressure mats, cameras, LIDARs, radars, Wireless Sensor Networks (WSNSs), etc. Infrared sensors and 3-D video tracking systems can be used for detecting and tracking the presence and activities of given entities; audio signals can prove useful in order to classify the interactions among people. All the original entities—including those corresponding to complex situations/events—must:

- be endowed with a *(provisional) identifier* (e.g., URI-like), to be transformed into a specific instance label, like aal9.g1 or ROBOT_1 above, during the following recognition/categorization phase;
- be provided with a *set of features/properties* to be computed in real time;

- be equipped with some *interface* allowing them to communicate/be integrated with other entities.

Extracting the initial characteristic features (identifiable attributes/properties) of the input entities by analyzing the output of several sensors—for temperature, motion, localization (RFID, GPS…), weight etc.—is a complex activity that can benefit of various integrated techniques. For example, with respect to the physical objects, all sorts of analytical tools like first- and second-order derivative expressions, Haar transforms, auto-regressive models and Canny-Deriche etc. filters (for edge detection), local colour descriptors, global color histograms, (syntactic) pattern recognition techniques and discriminant factor analysis (for identify movements) etc. can be used. For the detection and tracking of human beings, traditional symbolic approaches grounded, e.g., on the use of *geons* (simple 3-D geometric primitives for the torso, the head, upper and lower arms etc.) are now replaced by the use of *up-to-date commercial systems based on RGB-D (Red Green Blue + Depth) and time-of-flight sens*ors and enabling advanced gesture, facial and voice recognition—such as Kinect [75], SoftKinetic, and Leap Motion.

Independently from the techniques used to identify the input entities and their type it is evident that, for real scenarios beyond a certain level of complexity, the sensor layer module will never be able to deliver *perfect and complete information*. What this layer can realistically provide will be, in many cases, some *low-level partial and inaccurate information* (about positions, pressures, etc. and low-level events in general) with an *associated confidence or probability distribution*. This information can be processed using tools like Bayesian networks and Dempster-Shafer methods whilst delaying the selection of a final solution to the utilization of *high-level inference techniques*. Note also that Kalman filters [76] are today largely used for dealing with uncertainty in IoT-related applications like time series analysis in signal processing and motion planning and control and trajectory optimization in robotics [77].

To conclude, we remark that all the above techniques can be successfully used *one by one* to identify *independent entities* corresponding to *single* GWEs and their characteristic features—e.g., by reconstructing the constitutive elements, edges, corners, interest points, curvatures etc. of a squared_object like a table. The situation is different when we must identify the characteristics of complex events, situations and circumstances and their behavioural properties that correspond to *structured GWEs*—e.g., by describing at the feature level a complex event that will give rise to a GWE representing an entity_ that MOVE(s) towards a squared_object. In the most complex cases the *execution of inference operations is then required*, implying the use of the modelling of events and situations components of the world model.

Ontological/conceptual representation of the world. The ontological/conceptual model utilized for the representation of the world must be *general enough* to describe *both the static/background and the dynamic/foreground characteristics/*

features/properties… of the GWEs. Taking into account the discussion developed in Sect. 3.1 above, this model must be able to represent, *at the same time*:

- *Those stable, self-contained,* a priori *and basic notions (terminological knowledge) that can be assumed, at least in the short time, as a-temporal and universal*. A characteristic of these notions concerns the fact that their associated definitions/descriptions in terms of properties *are not subject to change*, at least within the framework of a given IoT application. In the following, for the sake of simplicity, we will refer to them as the *background knowledge*.
- The complex and structured (dynamic knowledge) information designating the *multiple, interpersonal, unpredictable and strongly spatio-temporal characterized behaviours proper to sets of interrelated, specific background knowledge entities* like Mary, robot, message and pills in the example of Sect. 3.1. This knowledge is typically structured into *elementary events* like "The robot sends a message to Mary" or "Mary must take her pills". We refer to this temporally characterized and in progress information as the *foreground knowledge*.

The modelling of the simplest, background GWEs like physical objects as tables, cars, bottles or vegetables—but also temperatures, light levels, pressures etc.—can be easily realized by using a *standard ontology*. Several ontologies of objects can be found on the Web, see the SSN ontology; the majority of them are in RDF/OWL format or have bridges towards a SW format. Even if a W3C/SW format is not mandatory—and its adoption could cause logical consistency problems with respect to the *n*-ary representation principles generally adopted in a GWEs context—the adoption of an RDF(S)/OWL solution for the background knowledge could be quite reasonable. This choice will allow us to profit from the several W3C/SW tools (RACER, Pellet, FaCT++…) existing on the market and committed to facilitating the set up and the coherence checking of standard ontologies. Note the, as already stated, NKRL is endowed with its own "standard" ontology, HClass (ontology of classes).

With respect to the formalization of the *foreground knowledge* (events/situations/circumstances etc.) where the above static entities are involved, *standard ontologies and RDF(S)/OWL solutions are not sufficient* (see the discussion in Sect. 2.3). More complex and powerful conceptual structures should then be used to describe the interrelationships of the elementary, basic entities involved in the events/situations/…, see the relations between the robot and the ageing person, or the ageing person and the pills. Such conceptual structures correspond, e.g., to NKRL's templates—see those used in the encoding of the Sect. 3.1 example. As already stated templates—collected into a HTemp hierarchy—correspond to *formal descriptions of general classes of dynamic structured entities denoting events/situations/circumstances/behaviors* that are based on the notions of semantic predicate (like MOVE) and functional roles (like SUBJ(ect), OBJ(ect)…).

This solution provides also the advantage of allowing the *structured GWEs* (instances of general templates) to be linked in turn into *more complex conceptual structures* making use of Higher Order Logic (HOL) tools. An example is given in

Table 1; see also the "*completive*" and "*binding*" constructions in Sect. 3.1. Note that this sort of HOL mechanisms is largely used in NKRL to represent complex scenarios of the *intention/willingness/behavior type*.

Some simple *auto-evolving possibilities* of the "representation of the world" model—in practice, the possibility of progressively and semi-automatically adding new conceptual entities to the HClass-like hierarchies—must also be foreseen. These possibilities could be (at least partially) based on the facilities for dealing with the "subsumption" phenomena proper to the W3C/SW "reasoners", see [3: 136–137] in this context.

Full recognition and categorization of the GWEs. The conceptual representation of the world mentioned above will be used in association with a set of inference rules to *recognize/categorize* the GWEs corresponding to the input entities.

In a GWEs context, recognizing/categorizing signifies to *create a correspondence* (recognizing as instances) between the *real entities* (objects, events, relationships, situations, circumstances etc.) coming from the external environment and the *high level conceptual and ontological representations* proper to the (digital) world model proposed in the previous sub-section. Note that being able of implementing a full recognition and categorization of GWEs represents a particularly difficult task. In fact, there is no warranty in general that the information available in the original external environment could be sufficient to execute the recognition task in a sufficiently complete way—and this independently from the level of completeness of the ontologies. For instance, sensing and inferencing operations might not be capable of fully categorizing a table, leading then to the instance SQUARE_OBJECT_1 for an incoming GWE instead of the correct TABLE_1 (or TOFFEE_BOX_1) instance. When stored knowledge is not sufficient for defining a complete state of the environment, sensors can be *pro-actively directed to get missing information*—when this is possible of course, e.g., when sensors are mounted on a mobile robot.

The recognition/categorization activities will be implemented in two *subsequent phases*. In a first one, the *raw descriptions* of the input entities that concern some (extended) physical objects category will be *unified* with the *conceptual entities* included in the background (i.e., standard) component of the world model introduced previously. This will be realized making use, mainly, of a semantic-based reasoning system able to *match*, in the best possible way, the *low-level features* (properties/attributes) attached initially to the input entities with the *semantic properties* of the general concepts included in this background component. This conceptual unification activity must be supported utilizing the usual, *algorithmic machine learning techniques* used to recognize an object through a comparison of the associated features with those of standard objects stored in a database, see methodologies like SIFT, Scale Invariant Feature Transform [78, 79], SURF, Speeded-Up Robust Features [80], HOG, Histograms of Oriented Gradients [81], etc. Note also that this first phase procedure *is equivalent to semantically annotate input entities (input sensor data) with concepts associated with standard ontologies*. This topic has been dealt with in several Computer Science domains;

in a specific IoT/WoT context, a particularly interesting solution has been proposed in [48]. As mentioned in Sect. 2.2, this approach is based on the Semantic Gateway as Service (SGS) as a bridge between the physical world and the high-level layers of an IoT/WoT system, and on semantic annotation procedures of sensor data using the Semantic Sensor Network (SSN) ontology. See also, in an (IoT/WoT-oriented) Semantic Web annotation context, well-known works like [5, 53], etc.

The procedures utilized for recognizing the external entities represented by contexts, events, situations, circumstances—i.e., for building up the corresponding *structured GWEs* making use of the dynamic/foreground component of the world model—consist in a multi-steps process based mainly on the results of the previous process of recognition of the physical/static/background GWEs. In the presence, e.g., of an event of the Move:StructuredInformation type, and before being able to add in the dynamic component of the world model a new GWE corresponding to an event of this type (see, e.g., the GWE aal9.g1 of Table 1), we must:

- Identify the possible concept/instances that, in the case of this example, are *candidate* to fill the SUBJ(ect), OBJ(ect), BEN(e)F(iciary), MODAL(ity)—and others in case—functional roles associated with the MOVE predicate.
- Verify that these potential fillers *satisfy* the constraints associated with the above roles. This means to verify that, e.g., a GWE labelled as ROBOT_1 in the previous, background categorization phase, really correspond to an instance of the human_being/robot_ concept in the static/background component of the world model. A constraint of this type is associated, in fact, to the filler of the SUBJ(ect) role in the NKRL Move:StructuredInformation template. In the same way, the filler of the OBJ(ect) role must be a structured GWE label (like aal9.g2 in Table 1), the filler of the BENF role an instance of the human_being concept, etc.
- Verify the *global coherence* of the new structured, dynamic/foreground GWE against the global situation dealt with. This means verifying that, as in the example of Table 1, this GWE corresponds really to the action of addressing a message to Mary or, in a following phase, to an action performed by Mary in consequence of the previous message.

It is also possible that, when dealing with categorization of foreground entities, a pure cognitive-driven approach *could not be sufficient* to decode the correspondence problem in the most complex situations. Robotics-oriented solutions built up, e.g., on temporal constraint propagation principle—see, e.g., those used in PEIS, Physically Embedded Intelligent Systems [82], for situation recognition—could then be combined with the cognitive ones to give rise to an optimized result.

Reasoning about the full recognized situations. When all the GWEs (corresponding to objects, agents, situations, circumstances, complex events and scenarios, behaviors, contexts etc.) have been created, we can use the general world description enriched with all of them to *take decisions* (such as event processing to choose which goal to pursue, or to change the current plan). In case, *physical actions* (like opening/closing gates/doors, allowing/disallowing switches etc.)

on the external environment must also be foreseen. In a GWEs context, these reasoning activities—realized, mainly, under the form of inference procedures—can be seen as the carrying out of a set of *services*.

In an IoT/WoT context, inferencing is often implemented according to formalisms based on probabilistic principles as the Bayesian Inference or the Dempster-Shafer theory. Techniques of this type are also used in a GWEs context to solve specific problems (see above in this section). However, in agreement with GWEs' grounding on a high-level ontological approach, *"reasoning" and "inferencing" are implemented here, mainly, according to a symbolic approach*, see Sect. 3.2 above. GWEs' inference procedures make use, then, of a system of *generalized "if/then rules"* similar to the rule sets utilized in the commercial Business Rules Engines (BREs) solutions. An overview of the general problems associated with the development of (mainly symbolic) advanced rule-based tools can be found in [83].

Examples of "services" implemented under the form of *inference scenarios* can concern, among many others:

- *Avoiding and managing critical situations*. This type of inference refers to situations like that schematically illustrated in [2], where the goal of the GWE-based application consists in preventing a dependent person with vision troubles (or a robot or a baby) to collide with potentially dangerous objects. The same type of inference can be easily generalized in a context of homeland security, of driving control, of exploration of unknown territories by a rover, of butler robots, of crisis situations etc.
- *Planning*. These reasoning activities could concern the optimization, e.g., of the supervision tasks of an ageing person, or the creation of a buying path within a supermarket, or the simple provision a cold drink. Planning includes prioritizing the goals, establishing when goals are complete, defining when the system is required to re-plan, etc. This type of activity concerns a wide set of GWEs applications.
- *Monitoring*. Monitoring concerns as well a large class of possible applications, from those concerning an elderly person in homecare after hospitalization to the anticipation of terrorism activities, decontamination of lands and buildings, identity management, gas/oil plants control, etc.; see, e.g., [84]. In all the monitoring cases, different independent GWEs of different degrees of complexity must be identified/categorized, and then aggregated/correlated to denote complex events/situations.
- *Intentions/behaviors detection*. Inferences of this type have normally (but not necessarily) GWEs of the human type as central characters. They can be associated with monitoring activities when, in an elderly monitoring situation for example, it is necessary to infer from the actions of the old person her/his (may be risky) real intentions, or when the hostile purposes of an intruder must be detected. They also concern a wide range of "sociological" applications like

detecting particular behaviors in young people, deducing attitudes towards health related fields like leisure, exercise or diet, implementing perspectives studies concerning the behavior of shoppers or intentions of (human/automatic) drivers, etc.

5 Conclusion

The GWEs paradigm concerns an *innovative understanding* of the general SWOT aims where the possibility of: (i) interpreting the environmental and context information, (ii) detecting information related to human intentions/behaviors, (iii) enabling human-like inferences and multi-modal interactions, and eventually (iv) acting on behalf of the users purposes are *particularly important*. The key property of GWEs is linked to the fact they are not limited to *physical objects* (as it is still normal within an IoT context). Rather, this paradigm supplies a *uniform formalism* (a uniform context) for describing objects, agents, events, situations, circumstances behaviors etc. and their evolution in time, as well as the relations among all these entities.

In this chapter, we have presented first a *quick picture* of the present state of the art in the so-called SWOT (Semantic Web of Things) domain by discussing briefly, in particular, the *shortcomings* associated with the uniform use of W3C/SW solutions in this domain. We have then illustrated in some detail, using a simple example, the main principles supporting the GWEs approach, and the use in this context of NKRL, the Narrative Knowledge Representation Language. Examples of the practical utilization of the GWEs paradigm under NKRL format have allowed us to discuss the *inferential aspects* of this paradigm. An architecture for GWEs-based systems has been displayed, along with a sketchy explanation of the steps required to pass from information collected at the raw sensor level to GWEs described at semantic/conceptual level and exploitable then for advanced inference procedures.

References

1. Gyrard, A., Serrano, M., Patel, P.: Building interoperable and cross-domain semantic web of things applications. In: Sheng, Q.Z., Qin, Y., Yao, L., Benatallah, B. (eds.) Managing the Web of Things—Linking the Real World to the Web, pp. 305–324. Morgan Kaufmann Elsevier, Cambridge, MA (2017)
2. Zarri, G.P.: Generalized world entities as a unifying IoT framework: a case for the GENIUS project. In: Bessis, N., Xhafa, F., Varvarigou, D., Hill, R., Li, M. (eds.) Internet of Things and Inter-Cooperative Computational Technologies for Collective Intelligence. Studies in Computational Intelligence, vol. 460, pp. 345–367. Springer, Heidelberg (2013)
3. Amarilli, F., Amigoni, F., Fugini, M.G., Zarri, G.P.: A semantic-rich approach to IoT using the generalized world entities paradigm. In: Sheng, Q.Z., Qin, Y., Yao, L., Benatallah, B.

(eds.) Managing the Web of Things—Linking the Real World to the Web, pp. 105–147. Morgan Kaufmann Elsevier, Cambridge, MA (2017)
4. Atzori, L., Iera, A., Morabito, G.: The Internet of Things: a survey. Comput. Netw. **54**, 2787–2805 (2010)
5. Sheth, A., Henson, C., Sahoo, S.: Semantic sensor Web. IEEE Internet Comput. **12**(4), 78–83 (2008)
6. Toma, I., Simperl, E., Hench, G.: A joint roadmap for semantic technologies and the Internet of Things. In: Proceedings of 3rd STI Roadmapping Workshop, co-located with 6th European Semantic Web Conference (ESWC09). Semantic Technology Institute (STI), Wien (2009)
7. Russomanno, D.J., Kothari, C.R., Thomas, O.A.: Building a sensor ontology: a practical approach leveraging ISO and OGC models. In: Proceedings of 2005 International Conference on Artificial Intelligence (IC-AI), vol. 2, pp. 637–643. CSRA Press, Athens, GA (2005)
8. Compton, M., Neuhaus, H., Taylor, K., Khoi-Nguyen Tran, K.-N.: Reasoning about sensors and compositions. In: Proceedings of 2nd International Workshop on Semantic Sensor Networks (SSN09), co-located with 8th International Semantic Web Conference (ISWC-2009), pp. 33–48. CEUR Workshop Proceedings, vol. 522, Aachen (2009)
9. Niles, I., Pease, A.: Towards a standard upper ontology. In: Smith, B., Welty, C.A. (eds.) Proceedings of 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), pp. 2–9. ACM, New York (2001)
10. Noy, N.F., Fergerson, R.W., Musen, M.A.: The knowledge model of Protégé-2000: combining interoperability and flexibility. In: Dieng, R., Corby, O., (eds.) Knowledge Acquisition, Modeling, and Management—Proceedings of EKAW 2000, LNCS, vol. 1937, pp. 17–32. Springer, Heidelberg (2000)
11. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., Stein, L.A. (eds.): OWL Web Ontology Language Reference (W3C Recommendation 10 February 2004). http://www.w3.org/TR/owl-ref/. Accessed 20 May 2017
12. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S. (eds.).: OWL 2 Web Ontology Language Primer (W3C Recommendation 27 October 2009). http://www.w3.org/TR/owl2-primer/. Accessed 20 May 2017
13. Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W.D., Le Phuoc, D., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K.: The SSN ontology of the W3C semantic sensor network incubator group. Web Semant. Sci. Serv. Agents World Wide Web **17**, 25–32 (2012)
14. Janowicz, K., Compton, M.: The stimulus-sensor-observation ontology design pattern and its integration into the semantic sensor network ontology. In: Proceedings of 3rd International Workshop on Semantic Sensor Networks, pp. 64–78. CEUR Workshop Proceedings, vol. 668, Aachen (2010)
15. Kotis, K., Katasonov, A.: Semantic interoperability on the Internet of Things: the semantic smart gateway framework. Int. J. Distrib. Syst. Technol. **4**(3), 47–69 (2013)
16. Soldatos, J., Kefalakis, N., Hauswirth, M., Serrano, M., Calbimonte, J.-P., Riahi, M., Aberer, K., Jayaraman, P.P., Zaslavsky, A., Podnar Žarko, I., Skorin-Kapov, L., Herzog, R.: OpenIoT: open source Internet-of-Things in the cloud. In: Podnar Zarko, I., Pripuzic, K., Serrano, M. (eds.) Interoperability and Open-Source Solutions for the Internet of Things. LNCS, vol. 9001, pp. 13–25. Springer, Heidelberg (2015)
17. Aberer, K., Hauswirth, M., Salehi, A.: Infrastructure for data processing in large-scale interconnected sensor networks. In: Becker, C., Jensen, C.S., Su, J., Nicklas, D. (eds.) Proceedings of International Conference on Mobile Data Management (MDM 2007), pp. 198–205. IEEE Computer Society, Washington (DC) (2007)
18. Ahmedi, L., Jajaga, E., Ahmedi, F.: An ontology framework for water quality management. In: Proceedings of 6th International Workshop on Semantic Sensor Networks (SSN13), co-located with 12th International Semantic Web Conference (ISWC-2013), pp. 35–50. CEUR Workshop Proceedings, vol. 1063, Aachen (2013)

19. Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: A Semantic Web Rule Language Combining OWL and RuleML (W3C Member Submission 21 May 2004). http://www.w3.org/Submission/SWRL/. Accessed 20 May 2017
20. Daniele, L., den Hartog, F., Roes, J.: Assets for Smart Appliances Interoperability (D-S4: Final Report). Netherlands Organisation for Applied Scientific Research (TNO), The Hague (2015)
21. Bermudez-Edo, M., Elsaleh, T., Barnaghi, P., Taylor, K.: IoT-Lite Ontology (W3C Member Submission 26 November 2015). https://www.w3.org/Submission/iot-lite/. Accessed 20 May 20 2017
22. Haller, A., Janowicz, K., Cox, S., Le Phuoc, D., Taylor K., Lefrançois, M. (eds.); Atkinson, R., García-Castro, R., Lieberman, J., Stadler, C. (contributors): Semantic Sensor Network Ontology (W3C Recommendation 19 October 2017). http://www.w3.org/TR/vocab-ssn/. Accessed 2 December 2017
23. Vermesan, O., Friess, P. (eds.): Building the Hyperconnected Society—IoT Research and Innovation Value Chains, Ecosystems and Markets. River Publishers, Aalborg (2015)
24. Delin, K.A., Jackson, S.P.: The sensor web: a new instrument concept. Paper Presented at the SPIE's Symposium on Integrated Optics, 20–26 January 2001, San Jose, CA. http://www.sensorwaresystems.com/historical/resources/sensorweb-concept.pdf. Accessed 20 May 2017
25. Gibbons, P.B., Karp, B., Ke, Y., Nath, S., Seshan, S.: IrisNet: an architecture for a world-wide sensor web. IEEE Pervasive Comput. **2**(4), 22–33 (2003)
26. Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., Lemmens, R.: New generation sensor web enablement. Sensors **11**, 2652–2699 (2011)
27. Herman, I., Adida, B., Sporny, M., Birbeck, M. (eds.): RDFa 1.1 Primer, 3rd edn.—Rich Structured Data Markup for Web Documents (W3C Working Group Note 17 March 2015). http://www.w3.org/TR/xhtml-rdfa-primer/. Accessed 20 May 2017
28. Koubarakis, M., Kyzirakos, K.: Modeling and querying metadata in the semantic sensor web: the model stRDF and the query language stSPARQL. In: Proceedings of 7th Extended Semantic Web Conference, ESWC-2010 (Part 1). LNCS, vol. 6088, pp. 425–439. Springer, Heidelberg (2010)
29. Harris, S., Seaborne, A. (eds.): SPARQL 1.1 Query Language (W3C Recommendation 21 March 2013). http://www.w3.org/TR/sparql11-query/. Accessed 20 May 2017
30. Revesz, P.Z.: Introduction to Constraint Databases. Springer, New York (2002)
31. Calbimonte, J.-P., Jeung, H., Corcho, O, Aberer, K.: Semantic sensor data search in a large-scale federated sensor network. In: Proceedings of 4th International Workshop on Semantic Sensor Networks (SSN11), co-located with 10th International Semantic Web Conference (ISWC-2011), pp. 23–38. CEUR Workshop Proceedings, vol. 839, Aachen (2011)
32. Dietze, S., Domingue, J.: Bridging between sensor measurements and symbolic ontologies through conceptual spaces. In: Proceedings of 1st International Workshop on the Semantic Sensor Web (SemSensWeb 2009), co-located with ESWC (European Semantic Web Conference) 2009, pp. 35–48. CEUR Workshop Proceedings, vol. 468, Aachen (2009)
33. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. The MIT Press, Cambridge, MA (2000)
34. Molina, M., Sanchez-Soriano, J.: Modeling sensor knowledge of a national hydrologic information system. In: Proceedings of SSW 2010 Workshop, co-located with 2nd International Joint IC3K Conference, pp. 23–31. INSTICC, Lisbon (2010)
35. Meinke, K., Tucker, J.V.: Many-Sorted Logic and its Applications. Wiley, Chichester (1993)
36. Brunner, J.-S., Goudou, J.-F., Gatellier, P., Beck, J., Laporte, C.-E.: SEMbySEM: a framework for sensors management. In: Proceedings of 1st International Workshop on the Semantic Sensor Web (SemSensWeb 2009), co-located with ESWC (European Semantic Web Conference) 2009, pp. 19–33. CEUR Workshop Proceedings, vol. 468, Aachen (2009)
37. Zarri, G.P., Sabri, L., Chibani, A., Amirat, Y.: Semantic-based industrial engineering: problems and solutions. In: Barolli, L., Xhafa, F., Vitabile, S., Hsu, H.-H. (eds.) Proceedings

of 2010 International Conference on Complex, Intelligent and Software Intensive Systems, pp. 1022–1027. IEEE Computer Society Press, Los Alamitos, CA (2010)
38. Pfisterer, D., Römer, K., Bimschas, D., Hasemann, H., Hauswirth, M., Karnstedt, M., Kleine, O., Kröller, A., Leggieri, M., Mietz, R., Pagel, M., Passant, A., Richardson, R., Truong, C.: SPITFIRE: towards a semantic web of things. IEEE Commun. Mag. **49**(11), 40–48 (2011)
39. Bassi, A., Bauer, M., Fiedler, M., Kramp, T., van Kranenburg, R., Lange, S., Meissner, S. (eds.): Enabling Things to Talk—Designing IoT Solutions with the IoT Architectural Reference Model. Springer, Heidelberg (2013)
40. Bauer, M., Bui, N., De Loof, J., Magerkurth, C., Nettsträter, A., Stefa, J., Walewski, J.W.: IoT reference model. In: Bassi, A., Bauer, M., Fiedler, M., Kramp, T., van Kranenburg, R., Lange, S., Meissner, S. (eds.) Enabling Things to Talk—Designing IoT Solutions with the IoT Architectural Reference Model, pp. 113–162. Springer, Heidelberg (2013)
41. De, S., Barnaghi, P., Bauer, M., Meissner, S.: Service modelling for the Internet of Things. In: Proceedings of 2011 Federated Conference on Computer Science and Information Systems—3rd Workshop on Software Services: Semantic-Based Software Services, pp. 949–956. IEEE Computer Society Press, Los Alamitos, CA (2011)
42. Kelaidonis, D., Somov, A., Foteinos, V., Poulios, G., Stavroulaki, V., Vlacheas, P., Demestichas, P., Baranov, A., Rahim Biswas, A., Giaffreda, R.: Virtualization and cognitive management of real world objects in the Internet of Things. In: Proceedings of 2012 IEEE International Conference on Green Computing and Communications, Conference on Internet of Things, and Conference on Cyber, Physical and Social Computing, pp. 187–194. IEEE Computer Society Press, Los Alamitos, CA (2012)
43. Trifa, V., Larizgoitia, I. (eds.): Design of the Object Virtualization Specification, COMPOSE Deliverable D2.1.1. IBM Research and COMPOSE Consortium, Haifa (2013)
44. Sottile, F., Franceschinis, M., Xiong, Z., Kasinathan, P., Smadja, P., Enjolras, P. Abreu, G., Severi, S., Oshiga, O., Vuppala, S., Poilinca, S., Ramakrishnan, A., Preuveneers, D., Hennebert, C., Tunaru, I., Denis, B., Suraty Filho, L.H., Saloranta, J., Macagnano, D., Destino, G., Salazar, M.-F., Monjas, M.-A., Melakessou, F., Andrushevich, A.: IoT Enabling Technologies and Future Developments—BUTLER Deliverable D2.5. Istituto Superiore Mario Boella (ISMB) and BUTLER Consortium, Torino (2014)
45. Perry, M., Herring, J. (eds.): GeoSPARQL—A Geographic Query Language for RDF Data, Version 1.0 2012 (Open Geospatial Consortium). http://www.opengis.net/doc/IS/geosparql/1.0. Accessed 20 May 2017
46. Preuveneers, D., Berbers, Y.: SAMURAI: A streaming multi-tenant context-management architecture for intelligent and scalable Internet of Things applications. In: Proceedings of 2014 International Conference on Intelligent Environments, pp. 226–233. IEEEXplore, New York (2014)
47. Pautasso, C., Wilde, E., Alarcon, R. (eds.): REST: Advanced Research Topics and Practical Applications. Springer, New York (2014)
48. Desai, P., Sheth, A., Anantharam, P.: Semantic gateway as a service architecture for IoT interoperability. In: Proceedings of 2015 IEEE International Conference on Mobile Services, pp. 313–319. EEEXplore, New York (2015)
49. Berners-Lee, T.: Linked Data (27 July 2006 2006, last changes 18 June 2009). http://www.w3.org/DesignIssues/LinkedData.html. Accessed 20 May 2017
50. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space, 1st edn. Morgan & Claypool, San Rafael, CA (2011)
51. Patni, H.K., Henson, C.A., Sheth, A.P.: Linked sensor data. In: Proceedings of 2010 International Symposium on Collaborative Technologies and Systems, pp. 362–370. IEEEXplore, New York (2010)
52. Barnaghi, P., Wang, W., Henson, C.A., Taylor, K.: Semantics for the Internet of Things: early progress and back to the future. Int. J. Semant. Web Inf. Syst. Spec. Iss. Sen. Netw. Internet Things Smart Devices **8**(1), 1–21 (2012)

53. Barnaghi, P., Presser, M., Moessner, K.: Publishing linked sensor data. In: Proceedings of 3rd International Workshop on Semantic Sensor Networks (SSN10), co-located with ISWC 2010, pp. 4–19. CEUR Workshop Proceedings, vol. 468, Aachen (2010)
54. Kolozali, S., Elsaleh, T., Barnaghi, P.: A validation tool for the W3C SSN ontology based sensory semantic knowledge. In: Proceedings of 2014 Workshop on Semantic Sensor Networks, co-located with International Semantic Web Conference 2014, pp. 83–88. CEUR Workshop Proceedings, vol. 1401, Aachen (2014)
55. Bernstein, A., Hendler, J., Noy, N.: A new look at the semantic web. Commun. ACM **59**(9), 35–37 (2016)
56. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F.: The Description Logic Handbook: Theory, Implementation, Applications. Cambridge University Press, Cambridge (2003)
57. Tarski, A., Mostovski, A., Robinson, R.: Undecidable Theories. Studies in Logic and the Foundation of Mathematics. North-Holland, Amsterdam (1953)
58. Davis, M.: Computability and Unsolvability. McGraw-Hill, New York (1958)
59. Zarri, G.P.: Advanced computational reasoning based on the NKRL conceptual model. Expert Syst. Appl. (ESWA) **40**, 2872–2888 (2013)
60. Trame, J., Kessler, C., Kuhn, W.: Linked data and time—modeling researcher life lines by events. In: Tenbrink, T., Stell, J., Galton, A., Wood, Z. (eds.) Proceedings of 11th International Conference on Spatial Information Theory, COSIT 2013. LNCS, vol. 8116, pp. 205–223. Springer, Heidelberg (2013)
61. Boussard, M., Meissner, S., Nettsträter, A., Olivereau, A., Salinas Segura, A., Thoma, M., Walewski, J.W.: A process for generating concrete architectures. In: Bassi, A., Bauer, M., Fiedler, M., Kramp, T., van Kranenburg, R., Lange, S., Meissner, S. (eds.) Enabling Things to Talk—Designing IoT Solutions with the IoT Architectural Reference Model, pp. 45–112. Springer, Heidelberg (2013)
62. Ibanez, F., Friess, P.: Putting the Internet of Things forward to the next level. In: Vermesan, O., Friess, P. (eds.) Internet of Things—From Research and Innovation to Market Deployment, pp. 3–6. River Publishers, Aalborg (2014)
63. Zarri, G.P.: Conceptual and content-based annotation of (multimedia) documents. Multimed. Tools Appl. **72**, 2359–2391 (2014)
64. Zarri, G.P.: Differentiating between "functional" and "semantic" roles in a high-level conceptual data modeling language. In: Murray, R.C., McCarthy, P.M. (eds.) Proceedings of 24th International Florida AI Research Society Conference, FLAIRS-24, pp. 75–80. AAAI Press, Menlo Park, CA (2011)
65. Zarri, G.P.: Representation and Management of Narrative Information, Theoretical Principles and Implementation. Springer, London (2009)
66. Zarri, G.P.: Integrating the two main inference modes of NKRL, transformations and hypotheses. J. Data Semant. (JoDS) **4**, 304–340 (2005)
67. Zarri, G.P.: Knowledge representation and inference techniques to improve the management of gas and oil facilities. Knowl. Based Syst. (KNOSYS) **24**, 989–1003 (2011)
68. Sabri, L.: Modèles sémantiques, raisonnements réactifs et narratifs pour la gestion du contexte en intelligence ambiante et en robotique ubiquitaire. Ph.D. Thesis, LiSSi Laboratory of the Paris-Est/Créteil (UPEC) University, Evry (2013)
69. Ayari, N.: Modélisation des connaissances et raisonnement à base d'ontologies spatio-temporelles. Application à la robotique ambiante d'assistance. Ph.D. Thesis, LiSSi Laboratory of the Paris-Est/Créteil (UPEC) University, Evry (2016)
70. Benazzouz, Y., Gurgen, L., Hennebert, C., Moreno Garcia, D., Munilla, C., Delsuc, J., Smadja, P., Atalla, S., Rizzo, F., Simone, A., Sottile, F., Nacabal, F., Castanier, F., Pascali, S., Vilei, A., Frà, C., Valla, M., Sancho, J., Shrestha, A.: Smart Object GW Platform Functional Specification (BUTLER Deliverable D4.3). Inno TSD and BUTLER Consortium, Sophia Antipolis (2013)

71. Medagliani, P., Leguay, J., Duda, A., Rousseau, F., Duquennay, S., Raza, S., Ferrari, G., Gonizzi, P., Cirani, S., Voltri, L., Monton, M., Domingo, M., Dohler, M., Vilajosana, I., Dupont, O.: Bringing IP to low-power smart objects: the smart parking case in the CALIPSO project. In: Vermesan, O, Friess, P. (eds.) Internet of Things—From Research and Innovation to Market Deployment, pp. 287–313. River Publishers, Aalborg (2014)
72. Ahlsén, M., Al-Akkad, A., Alcaraz, G., Asanin, S., Checcozzo, R., Franceschinis, M., Hreno, J., Jacobsen, M., Pastrone, C., Pramudianto, F., Spirito, M., Tomasi, R., Zimmermann, A.: Concepts and Technologies in Intelligent Service Structures 1 (EBBITS Deliverable D5.1.1). Fraunhofer FIT and EBBITS Consortium, Sankt Augustin (2010)
73. Fiedler, M., Meissner, S.: IoT in practice, examples: IoT in logistics and health. In: Bassi, A., Bauer, M., Fiedler, M., Kramp, T., van Kranenburg, R., Lange, S., Meissner, S. (eds.) Enabling Things to Talk—Designing IoT Solutions with the IoT Architectural Reference Model, pp. 27–36. Springer, Heidelberg (2013)
74. Spirito, M., Pastrone, C., Soldatos, J., Giaffreda, R., Doukas, C., Stavroulaki, V., Muñoz, L., Gutierrez Polidura, V., Gusmeroli, S., Sola, J., Agostinho, C.: Internet of Things applications —from research and innovation to market deployment. In: Vermesan, O, Friess, P. (eds.) Internet of Things—From Research and Innovation to Market Deployment, pp. 245–251. River Publishers, Aalborg (2014)
75. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human Pose recognition in parts from single depth images. In: Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1297–1304. IEEEXplore, New York (2011)
76. Chui, C.K., Chen, G.: Kalman Filtering with Real-Time Applications. Springer, Heidelberg (2009)
77. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. The MIT Press, Cambridge, MA (2005)
78. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 91–110 (2004)
79. Flitton, G., Breckon, T.P., Megherbi, N.: Object recognition using 3D SIFT in complex CT volumes. In: Proceedings of 2010 British Machine Vision Conference, pp. 11.1–11.12. BMVA Press, Durham (2010)
80. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Comput. Vis. Image Understand. Spec. Iss. Sim. Matching Comput. Vis. Multimed. **110**, 346–359 (2008)
81. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of 2005 IEEE computer society conference on computer vision and pattern recognition—CVPR 2005, pp. 886–893. IEEEXplore, New York (2005)
82. Saffiotti, A., Broxvall, M., Gritti, M., LeBlanc, K., Lundh, R., Rashid, J., Seo, B.S., Cho, Y.J.: The PEIS-ecology project: vision and results. In: Proceedings of 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-08), pp. 2329–2335. IEEEXplore, New York (2008)
83. Giurca, A., Gašević, D., Taveter, K. (eds.): Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches. Information Science Reference, Hershey, PA (2009)
84. Fugini, M.G., Teimourikia, M., Hadjichristofi, G.: A web-based cooperative tool for risk management with adaptive security. Future Gen. Comput. Syst. (Spec. Iss. Semant. Web Cooper.) **54**, 409–422 (2016)

Applications of IoT in Healthcare

Prabha Susy Mathew, Anitha S. Pillai and Vasile Palade

Abstract Internet of Things or IoT is an ecosystem of different physical objects provided with unique identifiers embedded with electronics, software, sensors and network connectivity which enables these objects to collect and exchange data without human intervention. The different technologies comprising IoT are Wireless sensor network, Cloud Computing, Micro-electromechanical systems (MEMS), Semantic technologies and future Internet. This concept makes it possible for the devices to be connected all the time everywhere, so it can also be referred to as Internet of Everything. Health care or Healthcare is the improvement of health in human beings by diagnosis, treatment and prevention of diseases, injury, and accidents, physical and mental impairments. It helps in the general physical, mental health and well being of people around the world. It comprises all the work done by Health professional in improving the primary care, secondary care and tertiary care of public. This chapter focuses on how the capabilities of Internet of things (IoT) can be leveraged in providing better Healthcare. In this chapter, various applications of IoT in healthcare as well as the challenges in the implementation are highlighted.

Keywords IoT · Predictive analytics · Telemetry · WSN

P. S. Mathew (✉) · A. S. Pillai
Hindustan University, Chennai, India
e-mail: prabha.susan102@gmail.com

A. S. Pillai
e-mail: anithasp@hindustanuniv.ac.in

V. Palade
Coventry University, Coventry, UK
e-mail: vasile.palade@conventry.ac.uk

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_11

1 Introduction

The demand for connected devices is found across multiple industries today. IoT has a plethora of applications in healthcare, like remote monitoring of patients health, tracking patients and equipments within the healthcare organization, smart beds to detect the occupancy, smart pill dispensers to monitor the patients intake of medicine and to send alert message to the care taker, etc. IoT can also provide early detection of certain health conditions of patients and provide rapid response to medical emergencies even so when the patient is on the move [1]. IoT technologies can help the healthcare organizations to reduce the cost by use of equipment tracking systems [2]. In addition, it can provide personalized care to patients, thereby improving the quality of healthcare services.

This chapter focuses on how the capabilities of the Internet of things (IoT) can be leveraged in providing better healthcare. It also discusses the key enabling technologies of the IoT (e.g., sensors and Wireless Sensor Networks (WSN)) , their characteristics and challenges.

The rest of the chapter is structured as follows. Section 2 discuss IOT in patient monitoring and Sect. 3 in Drugs and equipment monitoring. Sections 4 and 5 talks about IOT enabled maintenance of the medical equipment and cognitive computing. Section 6 presents a Case study on IOT enabled Health care, Various IoT implementation challenges pertaining to the healthcare sector is highlighted in Sect. 7, and, finally, Sect. 8 discusses future research directions.

2 Remote Physiological Monitoring

Remote Physiological Monitoring (RPM) (also known as Remote Patient Monitoring) uses digital technologies to track patients' vital signs and intervene if needed while the patient is at home. Technological advancements in sensors, easy availability of cellular technology, and the declining costs of embedded communication devices are opening up new avenues in the area of RPM. Organizations can improve their clinical outcomes and quality of service in a cost effective manner. It's a real boon to elderly as it helps in cutting down the travel to hospitals. Some of the essential physiological parameters like weight, blood pressure, heart rate, glucose level, etc. could be monitored remotely and appropriate medical advices can be given.

2.1 Technology Components

Figure 1 illustrates the key enabling technologies for patient monitoring systems. Recent advancement in technologies such as Wireless Sensor Network, RFID,

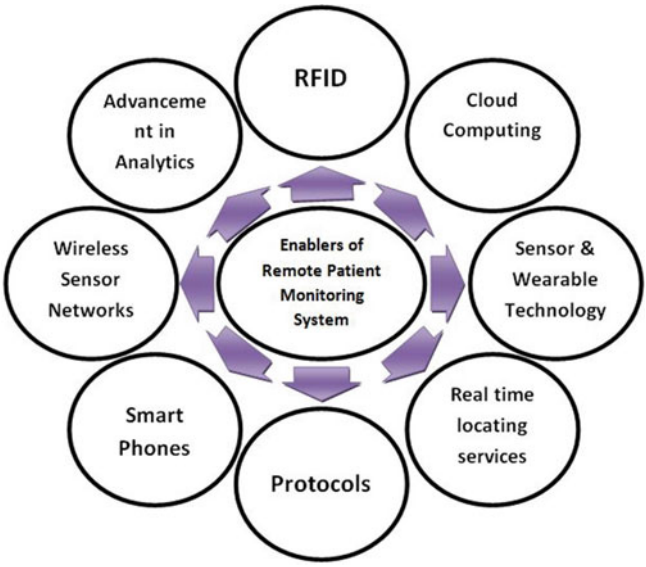


Fig. 1 Enablers of remote patient monitoring systems

embedded sensors, actuator nodes and micro-controllers, Cloud Computing; Smart phones have transformed the way in which traditionally the patients are now unobtrusively monitored.

2.2 Remote Monitoring Tools

The health monitoring tools monitor patient's vital signs, collect and forward the data electronically to the healthcare provider in a different location for assessment and medical assistance in the form of recommendations. Remote monitoring tools, such as glucose meter for measuring diabetes, pulse meter to check patients pulse rate, accelerometer (which is a movement monitoring sensor) are used to keep track of the patient's physical activity. Smart Phones, tablets, PDAs, computers are used to send the data to the central database where it can be viewed and analyzed by the healthcare provider.

2.3 Sensors

The sensors may be attached onto a patient's clothing, embedded in a watch, shoe, clothing, mattress, etc., or placed in a home in form of a motion sensor [3, 4]. The medical sensors can be classified into sensors that measure physiological

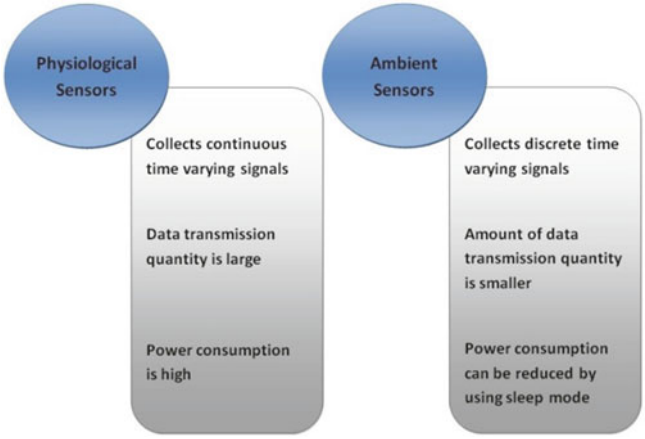


Fig. 2 Physiological and ambient sensors some of the most commonly used body sensors are [3, 5]

parameters and sensors that measures external environment, also known as ambient sensors. Figure 2 depicts the characteristic features of physiological and ambient sensors. The physiological sensors can be either wearable or implantable and are used to measure vital signs such as temperature, blood oxygen saturations, heart beat rate and pressure. The ambient sensors measures parameters such as the room temperature, light, sound and to detect falls.

Accelerometer: It is a device which measures the acceleration of a moving body. In healthcare applications it is used to observe and record body poster of the patient. It is also used to sense falls especially in cases where patients who are confined to bed.

Humidity and Temperature sensor: This type of sensor can be used to measure body temperature of the patient or that of the surrounding/external environment. The temperature measuring sensors can be either contact or contactless type sensors.

Sweat sensor: The biomarkers present in the sweat provide a wealth of information on sodium, chloride, potassium, glucose, amino acids, etc. It is very effective in the diagnose of diseases such as cystic fibrosis. Many instances of wearable sensors that are integrated to the textile have been used by athletes and patient to get information from their body fluids.

Respiration sensor: Respiration sensor can be an optical sensor used to monitor a patient during magnetic resonance image scanning. Respiratory rate monitoring is very effective for ambulatory measurements and is often used to monitor diseases such as sleep apnea or Chronic Obstructive Pulmonary Disease (COPD).

Blood glucose sensor: Glucose monitoring sensors are very crucial for diabetes patient to continuously monitor the glucose levels in interstitial fluids. These devices can be a bio-implant, which is implanted underneath the skin or a non-invasive device using infrared, optical sensors or ultrasound technologies.

Blood pressure sensor: High blood pressure leads to heart attacks, stroke etc. and blood pressure can change every minute, thereby requiring continuous monitoring. Wearable sensors that use pulse wave method give accurate reading without the need of any examiner as in the traditional way of examination [6].

Electrocardiogram sensor: ECG sensors measure the electrical impulse through heart muscles. ECG sensors use electrodes which must make contact with the skin.

Pulse oximetry sensor: Pulse oximetry is a sensor that uses non-invasive technique for measuring oxygenated hemoglobin in the blood. It is attached to fingertip of the patient from where the light wave is passed through the blood vessels. The variation of the light wave passing through gives the SpO2 measurement.

2.4 Ambulance Fitted with Sensors

In many cases of emergency, it is very difficult to perform diagnosis and treatment procedures during ambulance transport. This leads to a delay in the patient's diagnosis and treatment until arrival at the hospital. Patient often lose their lives in the ambulance due to the lack of necessary support systems during transportation while they are in the ambulance. Recent advancements in healthcare, reduced communication cost and a lot of research in the healthcare domain has led to improvement in quality of care provided to patients.

Ambulance telemetry is one such advancement where the automatic measurement and wireless transmission of vital data of patients inside an ambulance is made available to the doctors or medical centers for making critical treatment related decisions. Ambulance is fitted with sensors to capture the data; the data collected is passed on to the healthcare centers, thereby getting the necessary aid while the patient is still in the ambulance.

The ambulance telemetry uses several technologies to remotely provide patient the treatment and monitoring the vitals, while the critically ill patient is being transported to the hospital through ambulance [1, 7].

Polycam/Web Camera: One of the devices which are very useful for consultation from remote location is a polycam. A polycam can be connected to the network lines and to the TV at the medical facility/ambulance to monitor the patient vitals, such as heart rate etc. A web camera is often used as an alternative to polycam.

Internet: Once the healthcare organization's portal receives the vital parameters of the patient sent from the ambulance, they can be used for online consultation with doctors so that critically ill patient gets timely treatment while on the way to the hospital.

Wireless communication: Wireless technology has become an integral part of IoT based systems where devices communicate with each other from remote locations. Some of the communication devices used is Smartphone, GPS (Global Positioning System) units, Zigbee technology, Wi-Fi, Bluetooth etc. In a remote patient monitoring system the data collected from the sensor nodes are transmitted through wireless communication technologies, such as Zigbee or low-power

Bluetooth, to the concentrator or the IoT gateway where the data is further aggregated and transferred to the cloud for analysis [8]. Zigbee is designed for less power consumption and long battery life, making it the most desired device for remote sensing and monitoring applications.

2.5 Benefits of Using Remote Monitoring Devices

Remote Monitoring devices provide potential benefits to elderly patients, chronically ill as well as patients with mobility issues. It also helps healthcare professional to review patient's data from a remote location and intervene if needed.

Reduced Hospital Readmission: Remote monitoring devices can be both wearable and ambient sensors which are used to collect patient's health related data and alert the healthcare provider in case of a variation with respect to the threshold set for individual patients. This allows a timely intervention by the healthcare provider, thereby reducing the hospital admissions.

It helps patients manage their health data: Remote monitoring can help the patient to understand his/her own health condition and in certain cases, such as diabetes, self-administer the medicine when needed. Patient is always aware of the health condition, so can adopt healthier lifestyles.

Reduced travel time: Remote monitoring devices are of great help to patient in rural areas and those who have mobility issues. As the patients health parameters are continuously monitored by healthcare professionals from remote site, the need to be physically present at the medical facilities is not there anymore.

3 Sensor Enabled Drug and Equipment Tracking

Tracking drugs, patients and devices are becoming a very important aspect of healthcare industry as it promises to reduce cost and deliver quality treatment to patients. The ability to track the medical devices and the intake of medicine by patients will help care providers to manage their expenditure. The continuous streaming data coming from the tracking devices can be used to manage a number of chronic diseases effectively thereby reducing the healthcare cost.

3.1 Sensor Enabled Pills

Sensor enabled pills gives better insights on how deal with complex and chronic disease conditions. It enables the doctor to provide specific or tailor made treatment that suits the needs of individual patients [9, 10]. The ingestible pill on consumption

captures the status of vital health parameters but also sends the data to wearable device, which further sends it as a report to cloud where the healthcare providers can diagnose the disease and identify the effect of the drug on the vital organs.

Along with the regular medication the patient is prescribed sensor enabled pills that contains ingestible sensors. Once the sensor enabled pill is consumed by the patient it reaches the patient's stomach and sends signals on vital signs to the wearable device of the patient. The wearable records all the vital signs of the patient and relay it to both the patient and the healthcare professionals. Using this information, healthcare providers can monitor patients, track their activities and plan appropriate plan for the patients.

3.2 Smart Pill Bottles

Taking medicines at the right time would ensure well being in patient's health. Elderly patients often forget to take their pill and skip the dosage, which could lead to serious health conditions especially for a chronically ill patient.

A solution to this problem is use of sensor enabled smart pill bottles that recognizes when the patient has missed to take the medication and provides real time alert messages to caregivers and health care provides on non-adherence. Wireless and power efficient intelligent pill bottles with integrated cell phone technology enables patients' adherence to taking right dose of medicine at the right time. The pill bottle not just ensures patients take medication on time but also increases revenue for pharmaceutical companies, which otherwise it would result in lost sales for pharmaceutical companies. The Adhere Tech's smart pill dispenser solution monitors dosage in real-time, by keeping patients continually on track. It can alert patients when it's time to take their medicine, either through a phone call or text, or via blinking light directly in the bottle. It also helps by monitoring when the bottle was opened and how many pills were taken.

3.3 Medication Error Reduction Using RFID

Radio Frequency Identification technology is often used as a medical management solution for improving the functioning of operations within the pharmacy and for improving patient care and safety [11].

RFID is used for drug management, especially in monitoring the drugs stock level. There are several drugs that are rarely administered to the patient but have to be ready and available in the hospital such as ant venom, rabies etc. These medicines have a short shelf life, are slow-moving and are expensive too, resulting in a need to closely monitor their stock levels. This task, when manually done, takes more time and the error rates are high. RFID based drug management system is

used by pharmaceuticals to automate their restocking process and has been found to be a proven solution to reduce the time, error rates and increase the efficiency within the hospital pharmacy. Each of the medication trays/containers are attached with an RFID tag, which contains information such as the name of the drug, manufacturer's name, drug identification number, lot number and expiry date. The system periodically scans the tray and compares the details against the trays assigned quantities and drugs. It then generates a report about the drugs consumed, expired or about to expire and so on within seconds, so that appropriate actions can be initiated.

Healthcare is one such organization where accurate prediction of required supplies is difficult to plan ahead of time as the type of care or treatment required is not known in advance. As there is no definitive system to monitor inventory supplies, such as consumables, lab supplies and medications, they are often misplaced, lost or expired leading to inaccurate inventories and unnecessary last minute orders. Such unplanned rush orders increases the expenditure. RFID enable refrigerators and storage cabinets can continuously monitor the inventory levels and alerts the user if the inventory level reaches the minimum level, provides an alert for products reaching expiry, thereby allowing timely restocking, and it ensures critical supplies are available for patient care.

3.4 Equipment Tracking

Regularly scheduled maintenance of equipments, especially medical equipments are crucial for healthcare organizations. Yet in many hospitals, the majority of the time is spent on locating equipment, or it may be found that the equipment is in use and hence unable for servicing during the designated time. Even a slightest break down in power supply or malfunctioning of a life saving equipment functioning can bring about a catastrophic effect on the functioning of the hospital. Asset tracking system can alert the staff on the periodic repair or the replacement of the equipment.

Managing thousands of costly stationary and mobile equipments in a hospital is highly challenging. A mechanism to monitor and keep track of all these equipment could prove to be a useful solution. The right equipment when made available at the right time can prove to be extremely important to optimize supply of assets that can be life saving when dealing with critically ill patients [2].

In hospitals, often moveable assets, such as hospital beds, IV pumps, blood, stretchers, ECG machines, ventilators and other such assets are often misplaced or lost. It is common for hospitals to lose some percentage of its equipment annually as a hospital is a huge organization and many departments share equipments, making it difficult to track and locate assets. In places like hospitals, where life threatening emergencies can come any time, it becomes very important to keep a track of critical equipment, so that the efficiency of patient care is improved.

3.5 *Availability Monitoring of Equipments*

The majority of the asset tracking solutions uses RFID technology to keep track of their equipment. RFID labels are attached to the equipment, such as patient's bed, stretchers, etc., which can then be traced through RFID readers in real time. The user can have the software downloaded either on mobile or on a system which can then enable the user to view the details and also to get location details of each of the equipments that are having an RFID tag. The user can track and locate equipments effortlessly and within a matter of few seconds. Such system reduces the search time of equipment due to real time location visibility of the equipment.

3.5.1 Tracking Equipment

In hospitals, which are huge organizations, with multiple floors, thousands of patients, staff and assets, tracking of equipment becomes more important than in other healthcare organizations. The asset tracking systems based on RFID uses all types of RFID technologies including active RFID, passive RFID, Real Time Location Systems (RTLS) and Wi-Fi [12, 13].

Different types of systems that are used for asset tracking are (Table 1).

Table 1 Asset tracking systems

Type	Range	Use case in healthcare
Passive RFID	5 m	Management of high cost Medical devices
Semi-passive RFID	Up to 20 m	Used where sensor data needs to be tracked with the asset
Active RFID	Up to 20–100 m	Asset Management, Drug Management, Track Vaccines
Real Time Location systems (RTLS)	Up to 100 m	Asset Management, Temperature and Blood Bank supply monitoring, Personnel locating, tracking both Employees and Equipment
WIFI	Read range is between 50 and 100 m, Social networking as mentioned in [14, 15] are	For security, personnel and patient tracking
NFC (Near Field Communication)	20 cm read range	Near Field Communication (NFC) enabled Mobile phones can be used to effectively administer medication at patient's bedside
GPS (Global Positioning System)	Longer than RFID Technology	GPS Tracking for Medical supplies and equipments no matter where they are located

3.6 Optimize Asset Usage

Asset tracking systems not just makes the location of the asset available but also gives the availability status, such as on/off, sterilized/unsterile, in-use/free, which proves to be very useful in managing and using the available resources effectively. IoT based asset management systems that uses Artificial Neural Network (ANN) and FL (Fuzzy Logic) approaches are used for demand forecasting of assets both for normal and abnormal conditions [16]. Systems as these can help optimize asset usage, as it collects data about the equipment utilization and movement for effective allocation of the equipment to the patient based on priority. Location based equipment utilizations details would be beneficial for an accurate demand forecasting of assets.

3.7 Benefits of Asset Tracking in Healthcare

- Reduces equipment search time
- Improves inventory management with reduction in errors and cost resulting due to last minute orders.
- As the staff can easily get information, such as availability and location of life-saving and critical care equipment, search time is saved and faster patient care is possible.
- Efficient staff and resource allocation.
- Enables visibility of equipments across the healthcare organization.
- Alerts for equipment would result in timely maintenance and effective utilization of equipments. Real time inventory management reduces the risk of out-of-stock situation.
- Ambulatory patients can be tracked while they move through the facility thereby making it easy to locate them for treatment.
- IT equipments can be tracked for improved security, as lost IT equipment would also mean loss of confidential healthcare data.

3.8 Smart Access

Access control is required to grant or deny access to restricted areas round the clock across the healthcare unit. Controlled access must be provided in areas such as consultation rooms, emergency room, store rooms, maternity area, pediatric area, operation theatre, intensive care units, pharmacy, parking garage and even server racks. Healthcare facilities are growing, they provide services round the clock and their facilities are publicly accessible security becomes a major concern. Access control devices, such as doors alarms, locks, biometric based entry, are evolving to

meet the new age threats [17]. Solutions that comply with the security requirements in healthcare, such as Health Insurance Portability and Accountability (HIPPA), are what the access control systems must provide. Healthcare officials need to ensure security and safety of confidential records of their patients, violence against their hospital staff and patients, prisoners/psychiatric care patient eloping from the facility, theft of equipments/documents, electrical fires, monitoring infants etc. Access control system from Kaba allows the hospital to trigger either full or partial shutdown in matter of seconds. It also provides other useful features such as scheduling management, IP and CCTV camera integration, visitor to the facility management, etc. Another similar system provides limited access and control to visitors and patients while giving free access to hospital staff. Some systems provide access control via Ethernet to form a fully integrated solution [18] to ensure safety and security in the healthcare organization, some of the commonly used components are: access control, CCTV, infant tagging, asset management, smart cards, intrusion detection systems, mass notification systems, intelligent electronic locks, alarm system, etc.

3.9 Intelligent Security Management

Intelligent and integrated security management systems can greatly benefit the healthcare organizations by providing them valuable insights to act upon in a timely manner. Power outages, electrical circuit overloads, clogged air filters, and theft of equipments and inadequately maintained ventilations could cost a lot not just to the hospital budget but also to the patient's health. Intelligent systems that could monitor and maintain humidity, ventilation, air pressure, electrical circuits for overload, tagged medical equipments, can prevent electrical fires, minimize the maintenance personnel for ventilation systems, reduces time taken to locate equipment and optimizes use of hospital resources. Intelligent electronics lock with access control provides real time access monitoring, ensuring greater security.

3.10 Card Readers

In recent times the demand for multi-technology card readers by the healthcare organizations has increased. Smart card readers provide staff the benefit of using one identification card to access multiple facilities within the healthcare organization. Healthcare professionals need to move from one facility to another, and during such times the patient data on the workstation can be compromised. Smart card readers ensure the security of patient data without compromising on the mobility of the healthcare professionals. The healthcare professionals need to authenticate themselves by keying in their pin number. High risk areas can be integrated with card reader and iris biometric reader for additional security.

4 Equipment Maintenance Using IoT

For healthcare organizations, the medical equipments are crucial in the prevention, diagnosis and treatment of illness. As these equipments directly affect the lives of humans, considerably care has to be taken for its right functioning. Traditional methods, such as fixing the maintenance schedules based on duration and having excessive inventory of parts to reduce the downtime, contributed to increase in operational cost and inefficient execution. Maintenance of equipments can be classified in two types:

Corrective maintenance: Corrective maintenance is done in the event of an equipment breakdown

Preventive maintenance: Preventive maintenance is a well planned and scheduled checking of the equipment to correct minor problems to avoid breakdown of the equipment. It aims to extend the life of the equipment and reduce the breakdown rates of a device.

The medical equipments are getting more and more sophisticated and specialized, making the maintenance of medical equipments a difficult task. These techniques and real field data of the equipment are very crucial to determine the condition of medical equipment which further facilitates in pr active maintenance.

4.1 Predictive and Preventive Maintenance Life Cycle

Predictive maintenance on the other hand is a forecasting technique to determine the rate of failure of equipment or its component. Once the rate of failure is known, the maintenance interval is then set so that components are replaced before the occurrence of a failure. Predictive maintenance ensures that the equipment is reliable and safe to use [19].

Harnessing the power of the Internet of Things (IoT) one can avoid unplanned downtime and dramatically maximize equipment uptime. With SAP (Systems, Applications, and Products) Predictive Maintenance and Service, large volumes of real time data can be analyzed and predictive insights can be applied—for gaining insights needed to increase availability of asset and satisfaction levels.

4.2 Predictive Medical Equipment Malfunctions

Predictive maintenance deals with using corrective maintenance to prevent unexpected downtime of equipment. Predictive maintenance uses the right information at the right context in the right time to avoid equipment failure. IoT based interconnected systems basically collect the details about the condition of the equipment such as uptime/downtime of the device, using a non-intrusive manner. Those data are then used to measure and compute the equipment performance trends and will be used to

accurately predict failures before actual failures take place. Predictive maintenance reduces the unplanned maintenance, thereby keeping equipment in usable conditions for a longer period of time and increasing the overall efficiency [20].

5 Design Challenges

Increasing interconnectivity and sharing of information can lead any system to vulnerability. Data security and privacy are still challenges in any IoT-based system. Although many security measures and privacy preserving algorithms are available, newer challenges require us to constantly devise new mechanisms to combat these issues. In Remote Patient Monitoring

5.1 Scalability

The devices of the IoT ecosystem in healthcare, especially the ones used for remote patient monitoring generate huge volume and velocity of data that is in the order of magnitude much larger than what is generated from the traditional systems. The system should be scalable to store, process and analyze large amount of data to get insight from the real time data that is remotely collected without latency. Technology, such as in-memory computing, provides solutions for integrating diverse data from multiple sources and performs analytics in near real time with almost zero latency [21].

5.2 Interoperability

A very crucial factor for an IoT based healthcare application lies in its ability to seamlessly integrate with multiple connected devices from different manufacturers. Gateways, IPV6 based solutions or standards such as IEEE 802.15.14 protocol designed for low data rate and for low followed by manufacturers, can be used to handle the issues pertaining to interoperability [22].

5.3 Connectivity and Reliability

As the numbers of devices connected to the internet are increasing, there is an inherent need to scale up the IP addresses as the IoT scales up. IPV6 plays a crucial role in IoT by providing 2138 IP addresses and features useful for deployment and operations of IoT [22]. IoT is enabled by devices that generally have low-performance properties due to their memory, energy and computation constraints. 6LoWPAN (IPV6 over Low Power Wireless Personal Area Network) has

been defined to extend internet to devices with constraint. It provides features such as scalability, end-to-end connectivity, mobility, flexibility, etc.

5.4 Privacy and Security

A major challenge of the healthcare data is to ensure security and privacy of the patient's data from being tampered, leaked or accessed by unauthorized people. The main security and privacy requirements are with respect to authentication, integrity, non-repudiation, authorization and confidentiality [23, 24].

5.5 In the Implementation of WSN

As described in [25], WSNs are composed of independent nodes having one or more sensors, whose wireless communication takes place over limited frequency and bandwidth. They are capable of collecting information of physical parameters, such as temperature, humidity, condition of equipment from locations that are difficult to reach or monitor and perform appropriate action. Every node in a wireless sensor network consists of:

i. Sensor
ii. Micro-controller
iii. Memory
iv. Transceiver
v. Power Source

WSN has limited communication range and the storage in each of the sensor node, so multi-hop transmission of information takes place between the source and the base station. The data collected by the various sensors in the wireless sensor network is then sent to the sink node for directed routing towards the centralized control also known as base station. Due to the multi-hop transmission, data uses different nodes to diversify the traffic load. The communication network formed by the use of wireless radio transceivers facilitates data transmission between nodes. Low power communication standard and selection of right propagation technique can be used to extend distance and reliability in the network [26, 27].

5.6 Design and Development of Sensors

Wireless sensor networks are often being used in real time applications, which are generally time bound and often critical. Designing of such system must ensure scalability, energy efficiency, fault tolerance and security of the data.

5.7 Data Fusion and Storage

In a wireless sensor network, the main task of the sensor nodes is to collect data periodically from the surrounding environment and aggregate and forward the data to a base station or sink. Data collected from multiple sources can contribute to storage of redundant data, as the same data may be available at different sources. Proper strategies must be devised to eliminate storage of such duplicate data during collection and transmission of data collected from sensors. WSN sensing data deals with many physical dimensions of a data, such as temperature, pressure, humidity, etc. IoT based intelligent systems require fusion processing of multidimensional heterogeneous data. New approaches and advancement in information processing is the key requirement for dealing with such multidimensional data in a WSN.

5.8 Energy Efficiency

Most of the applications of wireless sensor network use tiny sensors that collect information about environment without human interference. These tiny sensors have energy constraint as they are battery powered devices with small memory and low processing power. There are several energy efficient data collection strategies, such as data aggregation, routing protocols based on routing topologies, sleep-wake scheduling, transmission power adjustments, etc., that have been implemented to reduce the transmission energy expenditure [28]. TRAMA, T-MAC, and S-MAC are energy efficient MAC protocols which reduce the energy consumption. Another technique proposed by Rhee et al. [29] focuses on minimizing energy consumption at various levels of the system, such as the node design, the physical layer communications, MAC and network protocols and the system design. The technique uses i-Bean Network, which is an ultra-low power WSN. The reduction in power consumption is achieved by using techniques and schemes during node and network design, such as Star Mesh network topology, dual processor, Multihop Routing, use of progressive search.

6 Cognitive Computing in Healthcare Applications

Cognitive computing is simulation of human cognitive skills by using self learning systems such as natural language processing, computer vision, pattern recognition, data mining, deep learning, Machine learning algorithms, robotics, etc. to automate tasks which would otherwise require human intelligence. Cognitive computing is indispensible in our life, the human-machine interactions through IA (Intelligent Agents) is one such example. IAs that is built with advanced Cognitive Computing technologies are able to carry out more complex tasks. Apple's Siri, Alexa, Google

Assistant, Microsoft's Cortana, Amazon's Echo, Face book's M, Jarvis are some of the cognitive computing systems widely used for personal assistance [30, 31].

In a rapidly changing Health care system where better insight from the huge pool of data, quality of service and patient care at a reduced cost is the need of the time, cognitive technologies can be a game changer by automating tasks that usually requires human intelligence, reducing cost, improving performance and efficiency. Cognitive computing systems in healthcare can help the care providers take better decisions based on the insights. These insights are derived by analyzing and identifying patterns from the large and complex health care data. Cognitive systems that are used in Healthcare are often complex systems using multiple technologies such as Kafka, Spark, Hadoop, Smaza, Solr, and Cassandra, Hbase etc. to deal with huge and diverse data. Cognitive analytics will prove to be the best for data driven discoveries and decision making Cognitive IoT is the term that combine technologies such as cognitive computing and IoT. The diverse data collected from the connected devices in IoT such as sensor data, unstructured textual data, images etc. form the knowledge base for improved reasoning and decision making in a complex data driven environment [32, 33] (Table 2).

6.1 Tools and Techniques of Cognitive Data Science

Cognitive computing deals with Understanding, disambiguation, learning, reasoning and human-machine interaction so it relies on technologies such as voice recognition, Natural language processing, computer vision, neural networks, Bayesians statistics, a range of machine learning methods, support vector machine, voting algorithm, K nearest neighbor, random forest etc. and visualization [32, 34, 35].

Some of the tools and techniques used, and its application in cognitive data science is as discussed in Table 3.

6.2 Benefits of Cognitive Computing for Advances in Health Care

Cognitive technology in healthcare offers a wide range of benefits:

- Effective Communication: Improves communication between doctor and patient.
- Manage Content collection: Extracts medical knowledge from different sources (clinical trial reports, medical websites, etc.).
- Integrates diverse data: Cognitive technologies, integrates structured data with unstructured information about patients, treatments, drugs, etc. that is collected from multiple information centers.
- Personalized treatment—It help physicians make efficient diagnoses and treatment by providing access to relevant information

Table 2 Cognitive versus traditional computing

	Cognitive computing	Traditional computing
Data source	Can process structured, semi-structured and unstructured data	Not able to process unstructured data
Operations	Learning, Interference, etc.	Rule based
Learning	Self learning without being explicitly programmed	It has to be reprogrammed to make any change in rules
Output	Highly trusted and consistent results from huge and diverse data	Accuracy limited to the structured data provided

Table 3 Tools and techniques of cognitive computing

Tools and techniques	Purpose
Computer vision	The ability of the computer to identify objects activity in visual environments. It is used in Remote patient monitoring systems to monitor patient with the help of camera. Medical imaging technology using computer vision algorithms are used to analyze mammogram image and identify potential abnormalities that indicate breast cancer
Sensor processing	Sensors provide information about the environment, physical condition and position of the object it is attached to. In healthcare it monitors and provides real time health parameters of the patient to get better insight the condition of the patient and provide timely assistance in case of variation
Speech recognition	Helps in developing systems for customers self service by recognizing speech. Engage customers by conversing with them in natural language
Natural Language processing (NLP)	Ability of the system to process the natural language which is normally unstructured. Natural language processing helps in understanding how language works; from written to spoken, to understand its context and nuances and also to understand the sentiment. In healthcare it is used to understand the context/ sentiment from the unstructured clinical text, email, blogs, etc.
Inference engines	It is a computer program to gain inference by applying logical rules on the knowledge base. It is used in many decision support systems
Machine learning/Deep learning	To get insights from diverse and unstructured data thereby helps in improving diagnoses, predicting disease condition at an early stage and providing solutions for personalized patient care
Expert systems	It uses Artificial Intelligence to simulate human judgment and behavior of a human in medical field it is of great help in diagnoses and treatment of disease
Internet of Things (IoT)	Interconnectivity of physical devices in order to access, collect, analyzes and share information
Data Visualization	Used to visualize patterns in huge and complex medical data to improve clinical treatment decisions

6.3 Challenges in Cognitive Automation

Cognitive computing is known for automatically guiding and providing right inference to the user. Even though there are numerous advantages of cognitive computing there are certain challenges that cannot be ignored [36, 37]. Some of the challenges are:

Inference from the knowledge to guide automation: Efficiency of cognitive systems depends on the data that is being fed in to the system. Huge amount of structured and unstructured data has to be fed in to the system from which it can learn and make inferences to guide automation. Due to continuous advancements sometimes the data that is fed into the system may not be of use anymore. A solution must be identified to bridge the gap between the constant advances in knowledge and the time taken to codify it into the system for it to make right inferences. Data from multiple sources can be ambiguous so adequate care must be taken to remove any such ambiguities before the system could use the knowledge.

Lack of advanced computing skill: As this field is continuously evolving there is ever growing need for skilled resources in the market.

Full potential of big data has yet to be realized: Big data provides a knowledge base to the cognitive system which is collected from diverse locations and in different formats both structured and unstructured which acts as a foundation for making accurate and efficient decisions [23]. Associated with big data in healthcare domain, which is known for its rich, diverse and ever growing data, are some challenges.

Providing secure environment for integrating new data which is generating at a great speed for accurate insights

Effective means to handle noisy data, as it makes drawing inference from such data difficult and incorrect.

Embed cognitive computing into existing system: To gain useful insight and deeply engaging experience the cognitive services should be integrated with the existing system for complex discoveries and precise decision making. A lot of research is still needed to integrate various technologies in order to interpret massive amount of data.

7 Case Study Using Cognitive Computing in Healthcare

Case 1:

A case study reported from Japan in 2015 discusses the efficiency of cognitive Intelligence in healthcare. A patient who was diagnosed with acute myeloid leukemia, a type of blood cancer showed abnormally slow recovery post-chemotherapy.

The hospital then decided to use cloud-based, Artificial Intelligence powered IBM Watson to cross-check the woman's genetic data with its database and

millions of oncology based research papers from all over the world. Watson found out in just 10 min what the real cause of her illness was; while scientists would have taken nearly two weeks to identify the same. The patient was identified with a rare secondary leukemia caused by myelodysplastic syndromes. On the basis of Watson's Inference, the change in treatment improved patient's condition [38].

Case 2:

A solution developed by Aditi is based on the capabilities of IoT and Microsoft Azure Cloud platform: Cloud-based Hospital Hygiene System. Often, Hospital-Acquired Infections (HAI) can cause severe issues from both a clinical and economic perspective, contributing to increasing healthcare expenditures. The study reveals that nearly 20–40% of HAI gets transmitted from the hospital staff to patients, and it has been discovered that hospital staff involving themselves in proper hand hygiene only 55% of the time.

Hence the solution monitors the interaction of individual staff with a hand-sanitizer dispenser and records this information using Real-time Location System (RTLS) technology that are tagged to the employee badges and the hand sanitizer dispensers. This data accurately provides details of the compliance levels of hospital care providers. The data was also used to illustrate clinician-patient interactions, providing detailed data to help monitor and modify behavior, to improve the compliance levels further, thus saving the lives of patients and reducing the HAI.

Case 3:

The CHRISTUS Health System, St. Michael Hospital, Texarkana, invested in a remote patient monitoring system (RPMS). An Android Tablet along with Bluetooth enabled personal health devices such as weight scale, blood pressure monitor, and pulse oximeter were used to remotely monitor patients. The features provided by the system included customizable patient care plans for each patient, and simple and easy user interface for almost all patients to use. With added facilities for the patients to post queries, send vital parameters, view educational videos, interactive video conferencing with caregivers, etc. AT&T provided the Wireless connectivity, which is used to send data from personal health devices to a secure "cloud" where secure logging by authorized caregivers can be done using their browsers. The implementation of the system has resulted in decrease in hospital readmission and improved patient satisfaction along with 90% reduction in overall cost of care. The system promotes self care.

8 Future Research Directions

8.1 Privacy

Sensor networks deals with two main privacy concerns: data-oriented and context-oriented privacy. Data-oriented privacy focuses on securing the integrity of

data collected and transmitted from sensing system. Context-oriented privacy prevents the attacker form gaining access to contextual information, such as the time and location, from which the data originated. Privacy research in WSNs focuses on privacy attacks with respect to sensor data collected and the location of data sources. Some of the challenges identified, such as open environments and resource constraints, are common traits of the IoT based applications. A ring signature-based authentication proposed to preserve the privacy of a source node provides anonymity to the source, and the other members that are chosen from its neighborhood provide spatial anonymity. To balance the performance penalty due to the increased message size and the vulnerability to interaction attacks by an adversary is to reduce the number of signers used in the ring signature [39].

8.2 Security

WSNs are vulnerable to attacks due to their characteristics such as wireless communication, large scaled nature of the networks and limited capacity. Security attacks in the wireless sensor networks (WSNs) can be categorized under three categories: Goal-oriented, Performer-oriented and Layer-oriented. Public-key cryptosystems are too cumbersome for tiny sensor nodes which has resource constraints. The choice of right cryptographic technique is critical, as it may consume resources of the network, which in turn would affect latency; throughput and network lifetime. However studies have revealed that public key cryptography can be applied to sensor networks by using the right techniques that optimize the use of available resource and selecting appropriate algorithms. Such cryptographic approaches were introduced to remove the drawbacks of the sensor nodes and to improve performance. Both RSA and Diffie-Hellman based on the elliptic curve cryptography can be used on tiny sensor nodes, and the results shows that smaller keys can provide good results, leading to reduced computation time as well as the amount of data transmitted and stored [40]. These approaches prove beneficial for meeting security requirements in WSNs. Non-cryptographic mechanisms, such as random walk, phantom and randomized routing along with flooding have been used to hide the location of the source [39].

8.3 Layer Oriented Attacks

The layered architecture of WSNs make them more vulnerable to various kinds of attacks [40]. The physical layer of a wireless sensor network does the selection and generation of carrier frequency, modulation and demodulation, encryption and decryption, transmission and reception of data. Jamming technique is used to attack the physical layer of the WSN. In this type of Denial of Service (DoS) attack the

attacker continuously sends radio signals on the communication channel, thereby preventing communication, Node tampering: Here the node is tampered to extract sensitive information, Sinkhole, flooding and spoofing. The Link layer coordinates with neighboring nodes to access the shared wireless channel provides detection of data frames and error control. Few DoS attacks on this layer are: Collision: where two nodes at the same time transmits data on the same frequency channel resulting in small changes in the data causes a mismatch at the receiving end. Battery Exhaustion: It is a type of DoS attack causing unusually high traffic resulting in limited accessibility to other nodes in the network. Routing is the main purpose of Network layer. Spoofing, replaying and misdirection of routing information are some of the specific DoS attacks thereby disrupting the traffic in the network. High traffic in channels with high number of useless messages is caused by Hello flood attack. Homing: In this attack, the traffic is searched for cluster heads and key managers which can shut down the entire network. Selective forwarding: Here a compromised node sends data to a few selected nodes instead of all the nodes. Sybil: in such an attack, the attacker replicates a single node with multiple identities to the other nodes.

Data transmission reliability is provided by the transport layer of the WSN architecture. DoS attacks in this layer are: Flooding, De-synchronization where fake messages are created at endpoints requesting retransmissions for correction of non-existent error results in loss of energy at the end-points.

The application layer of WSN is responsible for traffic management. Different type of attacks can be carried out in this layer, such as repudiation, data corruption and malicious code. Such attack consumes network bandwidth and drains nodes energy.

8.4 Trustworthiness

Trustworthiness of the sensor data is very essential in controlling the business processes, especially when increasingly data is being shared across organizations. It is crucial to deal with trustworthy data for making decisions, as it directly affects the success or failure of the business process. Many commercial database management systems support semantic integrity constraints to ensure trustworthiness. However, such techniques are unable to solve complex problems of data trustworthiness. A cyclic framework for data trustworthiness proposed by Lim et al. in [41] for data streamed from sensor network uses a trust score which is assigned to each data item and sensor. Sensors with low trust score indicate malfunctioning or compromised sensors. In order to improve the performance of cyclic framework for data trustworthiness Razvani et al. combined techniques such as variance estimation for the initial trust score of sensors and the characterization of the statistical distributions of errors, which resulted in an approach that is very accurate in detecting colluding attacks [41].

8.5 Predictive Analytics

The main objective of organizations using predictive analytics is to get valuable insight from the data in order to make effective decisions. Basically, in predictive analytics, models are built to detect patterns and predict future outcomes. Healthcare organizations are extensively using predictive analytics techniques for understanding the historical data to predict future activities for supporting advanced capabilities, such as evidence-based medicine and clinical decision support systems. Data related to a patient such as vital health parameters, medications, allergies, treatment plan, etc., are used to identify patients at risk by implementing predictive analysis. These predicted data are used for personalized patient care, reducing long term healthcare cost and for improving the quality of patient care.

Some of the common challenges faced by most organizations employing predictive analytics are [42]:

- Massive Data: Healthcare data is huge, making the process of extracting predictions from it a difficult task. As data grows, the need for choosing the right massive scalable technologies, efficient analytics algorithms and recommendation engines become very crucial for getting the benefits from the predictions. Algorithms can be implemented using distributed file systems for processing jobs.
- Interoperability and transparency of predictive analytics in healthcare: In order to make use of data across practices, it is necessary to share healthcare information across multiple healthcare organizations as well as systems. Platforms that comply with open interoperable standards would scale wider. Transparency is another key issue, as the clinical decisions are to be made by patients, clinicians and other organizations that interact with them for business. At any given point, the clinicians should be able to look into the predictive model and understand how a certain prediction was arrived at [43].
- Data Inconsistency: There is a significant challenge in acquiring, managing and reconciling inconsistent data arriving from multiple data sources. For delivering actionable insights from such data analytics, we need to deal with choosing the most appropriate predictors form a large number of potential predictors. Data filtering, cleansing and normalizing strategies should help tackle such challenge and to provide actionable data for analytics.
- Regulatory requirements: Healthcare applications face strict security and privacy compliance requirements. The predictive analytics implementation should comply with HIPAA/Hi Tech standards.
- Poor Documentation: To leverage on big healthcare data, it is important to ensure quality of the data being used for analytics. There is a strong impact on quality of results obtained for actionable prediction if the data used is clean and accurate. Hence clean and accurate data needs to be ensured before running analytics. The accuracy of information would lead to more accurate predictions, contributing towards improved high quality service provided to the patients.

8.6 Smartphone in Conjunction with WSNs

The advances in sensor and wireless technology and the increasing use of smart phones have a major impact on connected healthcare. In remote patient monitoring systems, smart phones act as a gateway between sensors and a remote server. As smart phones are now coming with more storage and computational powers, it enables ubiquitous health monitoring. Integrated GPS tracking also allows one to monitor and locate the patient through the mobile devices in case of emergency or while in ambulance. Fall detection systems often are based on accelerometer sensors available in smart phones and GPS based methods [44]. In medical applications, communication protocols must adhere to standards. Bluetooth is standard for wireless communication between a body sensor node and a smart phone, because of its compatibility and its acceptance among standardization bodies. In the cases where data is processed locally, Bluetooth Low Energy (Bluetooth 4.0) can be used. ZigBee is also being increasingly used in healthcare monitoring applications. The lack of complementary transceivers on smart phones is however a major hindrance.

8.7 Social Sensors

The integration of social networks with sensor networks provides solutions that can sense the context and provide useful information to the user. Rapidly increasing real time data produced everyday through social networking platforms and smart phones can be made more useful using social sensing. Social sensing applications have numerous research challenges perspectives. Some of these challenges commonly faced in sensor-based social networking, as mentioned in [14, 15] are:

Privacy Issue: The social data collected through sensors may contain sensitive information (e.g. location data); therefore it becomes crucial to use privacy sensitive techniques before using the data for analysis. The most common method to conceal the actual data is by adding noise or aggregating the original data. PoolView is such a privacy-sensitive technique for collecting and using mobile sensor data. Most privacy-preservation schemes reduce the reliability of the data, whereas trust is based on high reliability of the data. While privacy-preserving data mining is a growing field, with considerable research already done, sensor data offers an additional challenge of dealing with interrelated multi-dimensional time-series data. Such correlations within such sensor data streams offer new opportunities for privacy attacks and hence the need to devise new approaches to combat it.

Trust issues: The volume of data collected from sensors and smart phones can be very huge. A patient or equipment tracking systems may track the location details of devices and users. Thus, design techniques which can compress and process huge amount of data is what is required to handle such huge data. Often the data that are collected through sensors are error prone, or they are inputs without any

verification, this leading to issues regarding the trustworthiness of the data collected.

Battery life: Wearable and mobile devices are operated using batteries, which have limited battery life. Sensors being a part of these devices can drain the battery life more quickly than others as they are involved in continuous data collection. Battery life is still a major constraint, thus the architectural design should understand the trade-offs and focus on increasing the efficiency without compromising the goals of the application.

9 Conclusions

Recent developments in sensor, cloud, WSN and big data technologies when used in healthcare for developing a connected system provide a platform for optimized use of medical equipments and personalized patient care. A connected healthcare system effectively monitors the medical equipments, personnel and patients remotely, thereby eliminating the need for personal monitoring and reduced cost.

From among the several challenges discussed in the chapter, privacy and security are of prime importance given the nature of the healthcare systems, as they deal with huge amount of sensitive data.

Cloud computing technology can be integrated with the WSN to boost reliability and availability in WSN. However security of the system should be given utmost importance while integrating technologies. Even cloud sensor framework can be employed to further improve interoperability, scalability and efficient utilization of WSN in healthcare applications.

Joint IoT systems which collaborate and exchange data with other healthcare organizations and medical labs can provide comprehensive detail an individual patient's course of treatment from beginning till end.

Healthcare organizations can introduce subscription plans for their patients by using smart monitoring devices so that patient can pay monthly fee for the use of a particular set of remote monitoring services. This will enable more participation from the patient's side thereby preventing complications in their health condition and thereby reducing their medical expenditure. The combination of diverse data sources, newer algorithms for mining patterns in data, and cognitive computing will continue to improve quality of patient care and also help physicians to make better and precise clinical decisions.

IoT is certainly going to change the way healthcare organizations function by facilitating new business models as a result of constant technology advancement and based on the massive real time data collected by the sensors.

References

1. Almadania, B., Bin-Yahyaa, M., Shakshukib, E.M.: E-AMBULANCE: real-time integration platform for heterogeneous medical telemetry system. Proc. Comput. Sci. **63**, 400–407 (2015)
2. Abd Rahman, N.A., Abidin, Z.M.Z., Vasudavan, H.: Malaysia Hospital Asset Tracking System using RFID Technology (MHATS). eHealth Symposium 2013
3. Acampora, G., et al.: A survey on ambient intelligence in health care. Proc. IEEE
4. Poncholi, H.: Mobile device for health care monitoring system using wireless body sensor network. Int. J. Electron. Commun. Comput. Eng. **3**(4) (2012). ISSN (Online): 2249-071X. ISSN (Print): 2278-4209
5. Abo-Zahhad, M., Ahmed, S.M., Elnahas, O.: A wireless emergency telemedicine system for patients monitoring and diagnosis. Int. J. Telemed. Appl. **2014**, Article ID 380787, 11 pp. (2014). http://dx.doi.org/10.1155/2014/380787
6. http://news.mit.edu/2009/blood-pressure-tt0408
7. Gotadki, S., Mohan, R., Attarwala, M., Gajare, M.P.: Intelligent ambulance, Int. J. Eng. Tech. Res. (IJETR) **2**(4) (2014). ISSN: 2321-0869
8. Hassanalieragh, M., et al.: Health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: opportunities and challenges. In: 2015 IEEE International Conference on Services Computing
9. Kuzela, C.: Smart Drugs: Where IoT Meets Healthcare, A Market Snapshot, 30 June 2015
10. Niewolny, D.: How the Internet of Things is revolutionizing healthcare. White Paper
11. Hsu, P.: Using RFID technology to reduce medication errors, 7 June 2015
12. Buchan, G.: Real Time Location Services and Asset Management GBS Solution Representative. IBM Corporation (2009)
13. http://www.oatsystems.com/content-library/library_content/5-Myths-RTLS-RFID-Healthcare_OATSystems.pdf
14. Aggarwal, C., Abdelzaher, T.: Social sensing. In: Managing and Mining Sensor Data. Springer (2013)
15. Moturu, S.T., et al.: Using social sensing to understand the links between sleep, mood, and sociability. In: Proceedings of IEEE International Conference on Social Computing (SocialCom 2011), Cambridge, MA, Sept 2011
16. Man, L.C.K., Na, C.M., Kit, N.C.: IoT-based asset management system for health-care-related industries. Int. J. Eng. Bus. Manage. (2015). https://doi.org/10.5772/61821
17. Nibbelink, S.: Hospitals meet security challenges with integrated security and facility solutions. White Paper, Jan 2012
18. Lorenzi, N.: Advanced access-control systems for health facilities. http://www.hfmmagazine.com/display/HFM-news. Accessed 01 June 2016
19. WHO Medical Device Technical Series, Medical Equipment Maintenance Programme Overview (2011)
20. Wollenhaupt, G.: IoT Slashes Downtime with Predictive Maintenance. http://www.ptc.com/product-lifecycle-report/iot-slashes-downtime-with-predictive-maintenance. Accessed 04 Mar 2016
21. Jara, A.J., Ladid, L., Skarmeta, A.: The internet of everything through IPv6: an analysis of challenges, solutions and opportunities. J. Wirel. Mobile Netw. Ubiquit. Comput. Dependable Appl. **4**(3), 97–118
22. Kulkar, A., et al.: Healthcare applications of the Internet of Things: a review. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **5**(5), 6229–6232 (2014)
23. Chen, Y., Argentinis, J.D.E., Weber, G.: IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. Sci. Direct Clin. Ther. 38(4), 688–701 (2016)
24. Xiao, X., Chen, C., Sangaiah, A.K., Hu, G., Ye, R., Jiang, Y.: CenLocShare: a centralized privacy-preserving location-sharing system for mobile online social networks. Future Gener. Comput. Syst. (2017). https://doi.org/10.1016/j.future.2017.01.035

25. Akyildiz, I.F., Georgia Inst. of Technol., Atlanta, GA, USA, Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. Commun. Mag. IEEE **40**(8), 102–114 (2002)
26. Guicheng, S., Liu, B.: The visions, technologies, applications and security issues of Internet of Things. In: 2011 International Conference on E-Business and E-Government (ICEE). IEEE (2011)
27. Qiu, T., Zhang, Y., Qiao, D., Zhang, X., Wymore, M.L., Sangaiah, A.K.: A robust time synchronization scheme for industrial Internet of Things. IEEE Trans. Ind. Inf. (2017). https://doi.org/10.1109/TII.2017.2738842
28. Rohankar, R., Katti, C.P., Kumar, S.: Comparison of energy efficient data collection techniques in wireless sensor network. In: 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015)
29. Rhee, S., Seetharam, D., Liu, S.: Techniques for Minimizing Power Consumption in Low Data-Rate Wireless Sensor Networks
30. Sharma, G.: Chatbots Mag. (2016)
31. Haziyev, S., Milovanov, Y.: Cognitive Computing: How to Transform Digital Systems to The Next Level of Intelligence, 11 Jan 2017
32. Schatsky, D., Ronanki, R., Petrov, P.: IBM Watson, cognitive technologies for health plans using artificial intelligence to meet new market demands, 27 Mar 2015
33. Matthews, S.: What is cognitive IoT? IBM big data and analytics Hub, 24 Mar 2016
34. Schatsky, D., Muraskin, C., Gurumurthy, R.: Cognitive technologies, the real business opportunities. Deliotte Review (2015)
35. Rick Francis, M.D.: Cognitive Computing in Healthcare: Current Capabilities and Future Directions, 14 Sept 2015
36. Samulowitz, H., Reddy, C., Sabharwal, A.: Cognitive Automation of Data Science (2014)
37. Davenport, T.H.: cognitive Computing in Healthcare: Early Adopters of IBM's Watson, 08 Dec 2015
38. Rohaidi, N.: Asian Sci. Mag. https://www.asianscientist.com/2016/08/topnews/ibm-watson-rare-leukemia-university-tokyo-artificial-intelligence/
39. Debnath, A., Singaravelu, P., Verma, S.: Privacy in wireless sensor networks using ring signature. In: 2014 Production and Hosting by Elsevier B.V. on behalf of King Saud University. http://dx.doi.org/10.1016/j.jksuci.2013.12.006
40. Chelli, K.: Security issues in wireless sensor networks: attacks and countermeasures. In: Proceedings of the World Congress on Engineering, vol. I (2015)
41. Garcia-Alfaro, J., et al.: Data Trustworthiness—Approaches and Research Challenges, pp. 17–25. Springer International Publishing Switzerland (2015). https://doi.org/10.1007/978-3-319-17016-9_2
42. Bhor, S.: Predictive Analytics in Healthcare. http://blog.harbinger-systems.com/2015/12/predictive-analytics-in-healthcare/. Accessed Dec 2015
43. Amarasingham, R., et al.: Implementing electronic health care predictive analytics: considerations and challenges. Health Aff. (2014). https://doi.org/10.1377/hlthaff.2014.0352
44. Patel, S., et al.: A review of wearable sensors and systems with application in rehabilitation. J. Neuro Eng. Rehabil. **9**, 21 (2012). https://doi.org/10.1186/1743-0003-9-21

Security Stipulations on IoT Networks

Sumod Sundar and S. Sumathy

Abstract Cyber physical systems consists a crowd of computing nodes and the material processes associated with it. The objects correlated with these embedded things may embrace of a central processor, sensor and actuator units annotated around various communication devices. These device capabilities with least human intercession are proficient to seize data from various environments that requires "smartness" and hence IoT can be briefed as Smart devices centre. As massive numbers of devices are coupled with IoT, there exists a colossal take up on the protocol standards wide with various communication capabilities. This is in regard of assuring the security standard diffusion between the objects of data transfer and the terminal it reaches to. The heterogeneity between these objects and application platforms is an encumbrance to the developers for implementing the architecture for particular services. Cloud platforms rescue the situation by storing, computing and visualizing data before transforming them into meaningful information. Botnets or Zombie army is a malware that takes control of a computer in which the attacker can squeeze into the network raising threat to the authenticity of devices and access to networks. Phishing and Spamming attacks are causing a severity to networks through insecure connections. The security facet of IoT has to be redefined in terms of confidentiality so that the end user is guaranteed with secure data and data integrity can be retained. Various technologies lay around Transport layer Security (TLS) that helps the network to maintain its privacy. This chapter first discusses the Constrained Application Protocol (COAP) associated with 6LoWPAN network. A 6LoWPAN network is a cluster of LoWPAN networks which comprises of low cost and low power devices. These networks are bearing passive and active attacks that affect the network's confidentiality causing its performance malfunction. In passive attacks, the attacker is abiding to spy on network and steal the confidential information. Denial of Service attacks make obscure scene to network causing

S. Sundar (✉)
School of Computer Science & Engineering, VIT Vellore, Vellore, India
e-mail: sumod.sundar@gmail.com

S. Sumathy
School of Information Technology & Engineering, VIT Vellore, Vellore, India
e-mail: ssumathy@vit.ac.in

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_12

performance degradation to the network; active attack is a label in the case. Next focus is to discuss on the protocol stack that congregate the standardized notations of the ISO/OSI and TCP/IP stacks. The stack is being dealt in industrial applications and then turn into de-facto standard that saturates the existing IoT growth on wireless nodes. Next we confer about the IPv6 Routing Protocol (RPL) for Low-Power and Lossy Networks. It consists of constrained nodes with low processing power which are typically unstable with low packet delivery rates. They are mainly battery controlled devices consuming less memory and energy; their traffic patterns are generally multipoint or multipoint-to-point and hence requires compromises with thousand of interconnected nodes. It integrates the method of multipoint-to-point traffic from devices inside the LLN towards a central control point and the point-to-multipoint traffic from the central control point to the devices inside the LLN. Following this, Time-based secure key generation approach that convolutes the local key generation at the both transmission ends is discussed. A time stamp is put up on the local transmitter. The validity of the secure keys is limited to a time interval and the reply attacks comprised on valid messages are removed. The key generation process is a procedure performed separately by both communication objects. Finally, the chapter perceive with Cognitive Security in IoT devices in which the arena uses authentication through well defined user properties and patterns. Cognitive solutions in wireless security become concrete, since conventional static security is meant with lack of privacy. The user is able to learn continuously from the network and machine learning approaches can be incessantly applied with the stipulated security problems. The mediators in the capillary network can monitor the parameters related with Cognitive security standard and raise the security with time based solutions.

1 Introduction

Internet of Things (IoT) has become a buzzword among people across the globe in present days. IoT sustainable connection among devices [1] incorporates from the body sensors to the recent cloud computing. It consists of wide cluster types of networks like distributed, grid, ubiquitous, and vehicular. These networks have become the prominent elements of the IoT era within a short period of time. The pivotal notion behind the IoT is sensors and they lead the IoT in applications ranging from smart parking to smart roads, air pollution control to early earthquake management, water leakage detection to river flood monitoring, stock calculation to smart grid management, patient fall detection to ultra violet radiation identification and item tracking to fleet management. Like any other network, IoT is vulnerable to threats and may affect the smooth communication in the network. The network may hold smart processing units, data centers (DCs), Bluetooth, Zigbee, smart phones, tablets, WiFi networks, household devices, RFID tags, wrist watches and so on. Certainly, the human intervention on the heterogeneous network at some level will tend to chances of threat and theft. IoT combines actual and virtual anywhere and

anytime, attracting the attention of both the maker and hacker. Therefore, efficient and effective defense mechanisms are of the utmost importance to ensure the reliability of the IoT [2, 3]. Protecting the devices is a real concern when multiple devices are connected. The U.S. Department of Energy (DOE) has identified attack resistance to be one of the seven major properties required for the operation of the smart grid [4]. IoT can bring effective utilization of existing data resources by intelligent management of environment that relies on it. The systems design should be more reliable, secure and independent. The intelligent smart systems and its wide environment can be enhanced with properties of hybrid interfaces in heterogeneous networks. It bridges the break between the existing systems to the current Internet world scenario. A framework is proposed to integrate smart homes into platform as a service cloud [5]. The regulatory feature of these systems is connected with the unique identification of the objects and processes related with it. Security policies like accountability, modification, ownership and classification contribute to the flow of information related to it. This can construct tenacious IoT networks by the combinatorial evolution of technology. Many cryptography-based methods [6–8] have been designed to ensure data integrity, confidentiality and access control for sensor networks. Several studies are done on key management in Wireless body area Sensor Networks (WBSN) [9]. Venkatasubramanian et al. [10] have introduced a physiological signal based key management scheme in WBSN. Law et al. [11] have analyzed the impact of light weight ciphers on wireless sensor networks.

The IoT can establish machine to machine communication that improves brain capabilities and enhances the machines to meet the computational power of human brain. The systems available in our world can be modeled as a connected network component. The network collection will carry a huge set of heterogeneous components and numerous spatio-temporal non linear interactions. The IoT environment can be made secured only with proper prior knowledge of the entire system.

Many techniques related to security in IoT such as Message queue telemetry transport (MQTT) [12], Advanced message queuing protocol (AMQP) [13], IEEE 802.15.4 [14], IEEE 802.15.4e [15], Data distribution service (DDS) [16], Internet Protocol version 6 (IPv6) [17], GS1 EPCglobal, and Constrained application protocol (CoAP) exist. Each protocol scheme has its own leads and snags, since choosing real-time protocol depends on the need of right IoT devices. This chapter will provide readers with an understanding about the security policies and mechanisms in complex IoT systems.

This Chapter discusses various attacks which are possible in the IoT connected network. The next chapter discusses the Constrained Application Protocol (COAP) associated with 6LoWPAN network. A 6LoWPAN network is a cluster of LoWPAN networks which comprises of low cost, low power devices. These networks bear Passive and Active attacks that affect the network's confidentiality causing performance malfunction. In passive attacks, the attacker abides to spy on the network and steal the confidential information. Denial of Service attacks are making obscure scene to network causing performance degradation to the network; active attack is a label in the case. Following this, the next chapter discusses on the protocol stack that congregate the standardized notations of the ISO/OSI and TCP/

IP stacks. The stack dealt in industrial applications then turn into de facto standard that saturates the existing IoT growth on wireless nodes.

Next section confers about the IPv6 Routing Protocol (RPL) for Low-Power and Lossy Networks. It consists of constrained nodes with low processing power which are typically unstable with low packet delivery rates. They are mainly battery controlled consuming less memory and energy. Its traffic patterns are generally multipoint or multipoint-to-point and hence need compromises with thousand of interconnected nodes. It integrates the method of multipoint-to-point traffic from devices inside Linked Local Network (LLN) towards a central control point and the point-to-multipoint traffic from the central control point to the devices inside the LLN.

Following this, the next section discusses on Time-based secure key generation approach that convolutes the local key generation at both the transmission ends. A time stamp is set on the local transmitter. The validity of the secure key is limited to a particular time interval and the reply attacks comprising valid messages are removed. The key generation process is a procedure performed separately by both communication objects.

Final section perceives with Cognitive Security in IoT devices in which the arena uses authentication through well defined user properties and patterns. Cognitive solutions in wireless security have become concrete since conventional static security is meant with lack of privacy. The user is able to learn continuously from the network along with the machine learning approaches and can be incessantly applied with the stipulated security problems. The mediators in the capillary network can monitor the parameters related with Cognitive security standard and raise the security with time based solutions.

2 Phase Attacks

An attack that removes nodes according to a measure of centrality will be referred to as centrality attack [18]. In [19–22] authors have done investigation about the efficacy of degree centrality attacks while removing the largest hub nodes with the intention to reduce the size of the largest module of the network. Node degree [23] is not the most effectual centrality measure for curtailing largest component size. A Sybil attack may affect and can cause damage to different layers of communication [24, 25].

2.1 Data Leakage or Breach

Data leakage can be internal or external, intentional or unintentional, authorized or malicious, involving hardware or software. Carrying the authorized data or credentials to another target in unauthorized means can be briefed as data leakage. It

can be consummated by simply remembering what was seen, by physical removal of tapes, disks and reports or by subtle means such as data hiding. Steganographic message transfers data secretly to its destination without any means of decryption keys. Data leakage is a severe threat to reliability. In cloud servers, data are habitually transferred between data stations and hence there is a great chance for leakage of data. The extent of data leakage can be minimized by the techniques of data leakage prevention (DLP).

2.2 Data Sovereignty

With minimal human intervention, the smartness of IoT applications like smart home, smart grid, and smart transportation can automatically detect the environment, collect data, communicate with each other, and perform corresponding actions [26]. Data sovereignty means that information stored in digital form is abiding the laws of the country. The IoT covers wide range of devices across the globe and is hence admirable with sovereignty. Reports on IoT attacks are described in [27].

2.3 Data Loss

Data loss is different from data leakage, where data leakage is a kind of revenge activity made on authority or employer. Data loss means drop of work accidentally due to hardware or software failure and natural disasters.

2.4 Data Authentication

Data is transmitting over different nodes in network. Hence, there is a chance for the data to be conceived by some intruders or external parties. The received data should be ensured from the trusted sender point and no invalid users interpret the data. The sender sources have to be verified and ensured on the possession of legitimate authority.

2.5 Attack on Availability

The availability of data is really a concern while supplying information to the target clients. The data mismatch or unexpected unbalanced form of data might be a result of obscured attack like Distributed Denial of Service (DDoS). Hence, the intended data do not reach the users due to unavailability of proper Data Centers (DCs).

2.6 Flooding by Attackers

DDoS is flooding of malicious or unsuited packets by attackers against the DCs. This type of overload threat can be easily spot by Matchboard Profiler. The firewall can detect similar user features and can be blocked at its gate.

2.7 Flooding by Legitimates (Flash Crowd)

A huge number of genuine users are prompted to get data at the DC resources simultaneously. This overloaded scene will deprecate the data centers. Improper buffering mechanisms will make the situation behave improperly with large number of requests over a time constraint.

2.8 Flooding by Spoofing Attackers

Impersonation is a means of giving acknowledgement for the incoming request and giving proper reply by preserving correct sequence number as a legitimate entity.

2.9 Flooding by Aggressive Legitimates

Aggressive legitimates are persons who are restive and randomly set off akin requests over a little span of time. This may lead to an overloaded condition and servers are flooded with requests of the genuine users and deteriorate the efficiency of data centers. These attacks are difficult to track since their legitimate characteristics will misrepresent the identification methods. Aggressive legitimates are regulated by observing the inter-arrival time between data packets and the back-off timers values.

2.10 Modification of Sensitive Data

After the transmission of data from sensors, it can be detained and customized, and will be transferred towards the destination node. Modification mainly occurs with different methods. Through content modification, some part of information will be changed without any reflection of data privacy. The sequence modification makes the extended delivery of data in disordered manner. This will swap the data to useless form. Time modification will deprecate the time stamps and often lead to replay attacks.

3 IoT Stack

Constrained application protocol (CoAP) [28] is a dedicated protocol for web transfer mainly for Low Power Lossy Networks (LLNs). It uses the User datagram Protocol (UDP) instead of Transfer Control Protocol (TCP). CoAP Uniform Resource Identifier (URI) is defined as coap[s]: <host> : <port>/<path> : <query>. It has identical request-response model of HTTP. The URI mapping is so significant to map multiple URIs with different protocol stacks. TCP connections are needed to get mapped towards the UDP segments (Figs. 1, 2 and 3).

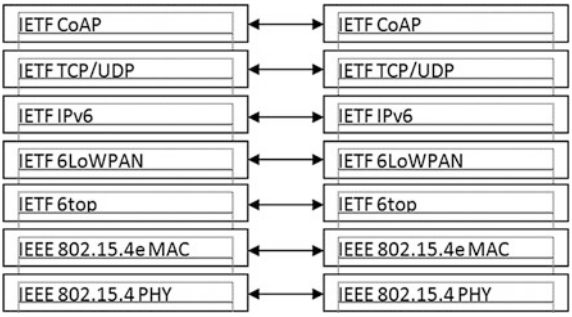


Fig. 1 IoT stack combining IETF and IEEE standards

```
{"id": "Jd93_jZ8Ls5V0qP",
 "ii": 1412941013,
 "is": "coap://tokenManager.um.es",
 "su": "aB4wSICIXC1pm2pkW9YMPQyFudc=CPhYdgOAQwcOYgURwP1q02WSv=",
 "de": "coap://smartObjectB.um.es",
 "si": "TqZaXuxZ5dmZU6k3PtiWwI3NrjH=7u5By5OHzl0Otq4TmkrZU2JPd=",
 "ar": [
  {"ac": "GET"
   "re": "position"
   "co": [{
     "t": 5,
     "u": trust,
     "v": 0.7}]}
 "nb": 1412941013,
 "na": 1412941456
} Legend: "id"-> identifier "ii"-> issued time "is"->issuer "su"->
subject "de"-> device "si"-> signature "ar"-> accessRights "ac"->
action "re"-> resource "co"-> condition "t"-> type "u"-> unit "v"->
value "nb"-> not before "na"-> not after
```

Fig. 2 Trust-aware capability token example

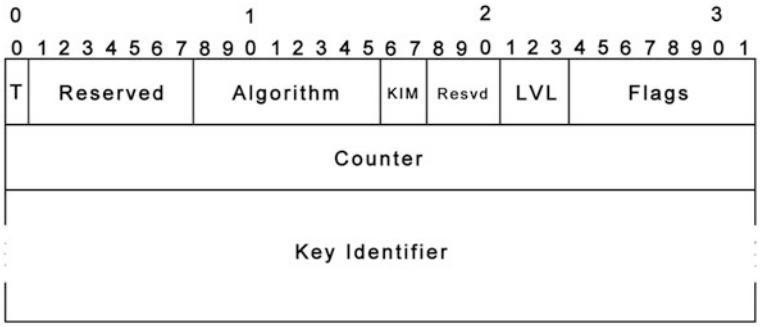


Fig. 3 Security fields

4 Capability-Based Access Control for IoT

Capability-based access control (CapBAC) has introduced as a viable method to be set up in IoT scenarios [29] yet in the existence of devices with tight resource constraints. It comprises a lightweight and flexible design that allows the authorization functionality to be embedded onto IoT devices. The main advantages needed for security constraints in distributed approach are scalability and interoperability. Capability [30] is a concept introduced such that a token or a key gives the permission to access the entity or object in the system of networks. This token is created with a set of privileges that are established to the entity which holds the token. Also, the token must be tamperproof and unequivocally identified in order to be considered in a real environment. So, it is necessary to consider suitable cryptographic mechanisms to be used even on resource-constrained devices, which enables an end-to-end secure access control mechanism. This concept is applied to IoT environments and extended by defining conditions that are locally verified on the constrained device. This feature improves the flexibility of DCapBAC because any parameter that is read by the smart object can be used in the authorization process. DCapBAC is based on JavaScript object notation (JSON) [31] as the representation format for the token. The use of emerging communication protocols like constrained application protocol (CoAP) [32] and 6LoWPAN, and a set of cryptographic optimizations for elliptic curve cryptography (ECC) are presented. DCapBAC along with a policy-based mechanism based on XACML [33] is used in SOCIOTAL, which is an access control mechanism to gather the access control privileges to be embedded into the capability token.

6lowPAN makes embedded nodes in a network to utilize the limited IPV6 address subset. The 6LoWPAN can be described as a hybrid gradient of IEEE 802.15.4 and IPv6 [34]. A 6LoWPAN network consists of one or more LoWPAN networks connected to the Internet. The CoAP can be defined as a software protocol designed for small, low-power sensors. Though the routers connected in between the networks regulate the incoming and outgoing control flow of LoWPAN, an adaptation layer is suggested by Internet Engineering task Force (IETF) 6LoWPAN

working group to optimize the packets of Ipv6. This is done in need to make the packets compatible with IEEE 802.15.4 link layer. The devices connected in the network will rely on the router to communicate rather than using the IPv6 header. Normally, the LoWPAN devices are distinguished by their effectiveness in various parameters like less power, less cost, low data rate, short radio range and so on. The implementation of LoWPAN topology will vary the connection of devices such as Reduced Function Devices (RFD) and Full Function Devices (FFD) to the Internet. With its structure, mesh topology enables the connection between the devices and proper communication among it. The star topology allows the coordinator node to regulate the communication of nodes in the network. The IETF-ROLL (Routing over Low-power and Lossy Network) working group proposed RPL (Routing Protocol for Low-power and lossy networks) as a solution for routing problem. 6LoWPAN networks will undergo numerous attacks on the security level that intend to cause direct damage to the network and will grab the confidential information from the network.

In an IoT network, various RPL instances could execute in parallel. The job of these instances will provide potential challenging criteria and constraints. With regard to the optimization objective like minimizing latency and minimizing energy, RPL separates the packet processing and forwarding. The LLN will show some variant properties at the router, since the router is evaluated to be used as head node. RPL expects an external mechanism to be triggered during the parent selection phase in order to verify link properties and neighbor reachability. RPL also expects an external mechanism to access and transport some control information, referred to as the RPL Packet Information, in data packets. Some RPL enable the association of a data packet with a RPL Instance and the validation of RPL routing states.

RPL has a procedure to distribute information on the dynamically formed network topology. The nodes are allowed to operate autonomously as this dissemination facilitates minimal configuration in the nodes. The router may have independent prefixes with its own. They are variables based on the origin of routers. Any router that is self prefixed is allowed to get broadcasted in the network. The RPL can also group hosts in a subnet network together and can label it with a common prefix. This intends that RPL can supply information directly to the grouped subnet network. The LLN links do not have the property of transitivity and hence the RPL broadcasts in the subnet do not get circulated in the network.

The RPL is able to circulate IPv6 Neighbor Discovery (ND) information like Prefix Information Option (PIO) and Route Information Option (RIO). Neighbor Discovery information maintains the real semantics for router to host and from router to router [35]. It guarantees the routing advertisements not to redistribute to other routing protocols. As a router, the RPL node may advertise the information from the options as required for the specific link like an ND Router Advertisement (RA) message.

RPL uses the following four values to identify and maintain a topology:

1. RPL Instance ID
2. DODAGID
3. DODAG Version Number
4. Rank

An RPL instance contains one or more DODAG roots. It may provide routes to certain destination prefixes, reachable via the DODAG roots or alternate paths within the DODAG. These roots may operate independently, or they may coordinate over a network that is not necessarily as constrained as an LLN.

An RPL instance may comprise:

- a single DODAG with a single root.
- multiple uncoordinated DODAG's with independent roots.
- a single DODAG with a virtual root that coordinates LLN sinks (with the same DODAGID) over a backbone network.
- a combination of the above as suited to some application scenario.

5 RPL Security

RPL ensures message confidentiality and integrity. It has its own mechanism for ensuring the properties [35]. It is deliberated such that link-layer mechanisms can be used when it is available and appropriate. RPL has three basic security modes. In unsecured mode, RPL control messages are sent without any additional security mechanisms. In this mode, it uses other security primitives like link-layer security to ensure application security requirements. In preinstalled mode, nodes connected with RPL instance uses preinstalled keys with them so that it is used to process and generate RPL messages with more security. In authenticated mode, the nodes may contain preinstalled keys like the previous mode; also used to link an RPL instance in form of a leaf.

The node should obtain the authentication details like key and certificate before establishing a successful RPL implementation. Every RPL message should possess a secure variant. The node is responsible to address the source of the authentication material for the implementation in authenticated mode. This secure variant is liable to ensure confidentiality and delay protection. Encryption is applied to retain the confidentiality of RPL ICMPv6 message. The variants enhance the replay protection and integrity and mainly security related details are kept between checksum and base.

In a secured packet, using any security algorithm or signature the cryptographic fields can be added. The cryptographic fields may contain the MAC or signature details and security transformation may provide an ICMPv6 secured RPL message.

Counter Time (T):

If the counter's Time flag is set, then the Counter field is a timestamp. If the flag is cleared, then the counter is an incrementing counter.

Reserved:

It is a 7-bit unused field. This field is initialized to zero by the sender and will be ignored by the receiver.

Security Algorithm:

The field denotes the encryption, MAC, and signature scheme the network uses.

Security Level (LVL):

It is a 3-bit field that indicates the provided packet protection. This value can be taken based on a per-packet basis. This ensures different levels of data authenticity and confidentiality.

KIM field:

It indicates the use of signatures and the meaning of the level field. The assigned values of security level are not well ordered. The higher value of LVL cannot guarantee the higher level of security.

5.1 Security Mechanisms

The security of RPL message can be observed by observing at the high-order bit of the RPL message. Apart from this to retain the security, secure versions of basic control messages like DIS, DIO, DAO, DAO-ACK are established with RPL, so that those messages that are found relevant with the network are only enabled with security features.

The dedicated features of RPL are used to a limited extent, since complexity in implementation and size matters in case of LLNs. Hence, more concise security provisions are physically difficult to implement. The implementation strategies should observe the basic security requirements and existing security mechanism that can avail while implementing in a network. The link-layer of the network is well utilized for achieving the security requirements without any RPL security. Any network implementation should utilize the integrity and confidentiality features and may be optional.

RPL consists of basically three security modes:

Unsecured: In this security mode, RPL uses basic DIS, DIO, DAO, and DAO-ACK messages that do not have any security sections. As a network that uses

any other security mechanisms like link-layer security, unsecured mode cannot entail if all messages are sent without any protection.

Preinstalled: In this mode, RPL apply secure messages. To join with an RPL Instance, a node should have a preinstalled key with it. The nodes in action use this key to supply message confidentiality, integrity, and authenticity. The key can be used by the node to join the RPL network as either a host or a router.

Authenticated: Like in the preinstalled mode, RPL apply secure messages. To join with an RPL Instance, a node should have a preinstalled key with it. The nodes in action use this key to supply message confidentiality, integrity and authenticity. But, by using the key the node can join in the network as a host only. The node can join the network only after obtaining a second key from a key authority. RPL is compatible with symmetric algorithms only. This mode is well reached to the potential future cryptographic primitives. When a network utilizes the preinstalled mode or authenticated mode, it is signaled by the 'A' bit of the DAG configuration option. The unsecured mode uses RPL messages.

Joining a Secure Network: For connecting a node with a secured network, it should be pre-configured through a shared key to communicate with its neighbors. The node should listen to secure DIOs or trigger secure DIOs by sending a secure DIS.

Installing Keys: When a node is connected to the network, the keys should be installed dynamically. New keys can be supplied by the node to communicate with the authorization server.

Consistency Checks: The RPL nodes can be subjected under consistency check to make protection against replay attacks and synchronize counters.

Loop Avoidance and Detection: In RPL, loop is formed due to many factors like control packet loss. The mechanism for avoiding the loops is done with a view by reducing the total churns and states. RPL possess a reactive loop detection method that protects from network deterioration and it triggers repair of broken paths in it.

In IoT network and devices, phishing and spamming are becoming a severe issue. In January 2014, researchers at the security provider Proof point exposed an IoT cyber attack that happened on sophisticated devices. A network of connected home appliances such as home routers, televisions, and refrigerators are spammed and malicious e-mail spam are sent from routers. Privacy issues are raised from different corners while fixing it and became more complicated. In most cases the networks are bound with a single person authority and lack of privacy was observed.

In IoT scenarios, a number of technologies have been developed to achieve information privacy and security goals [36], such as transport layer security (TLS), which could also improve the confidentiality and integrity of the IoT. Onion routing encrypts and mixes Internet traffic from different sources, and encrypts data into multiple layers, using public keys on the transmission path. Finally, a recent in-depth review on the security aspects of IoT is provided [37].

6 Time-Based Secure Key Generation and Renewal

When the data is transferred in an insecure channel, the integrity of the data is a concern. An efficient time-based secure key generation method can be used for handling the proper keys in a secure connection. The local keys are synchronized and generated by using the symmetric encryption keys. This process has to be ensured at both the transmitter and receiver terminals. For enhancing the security level the generated keys should be changed after every transmission of data. The encrypted key is used by the sender to distribute the data. Similar to secure transmission methods the key has to be exchanged and synchronized between two communication terminals.

The principle of time-based secure key generation is schematically depicted in Fig. 4. In this approach, the key generation process is an operation performed independently by each communication party. In fact, unlike other key management algorithms, no additional messages are required to be exchanged to agree about a key, and the only requirement is that the key generation function should create the same keys for both communication parties based on the timestamp of the device. The validity of the secured key is restricted to a time interval, so that reply attacks based on valid messages sent using keys generated in past time intervals are discarded. Leveraging such features, we evince that, as a main advantage of the time-based secure key generation approach, there is no need for a server to manage

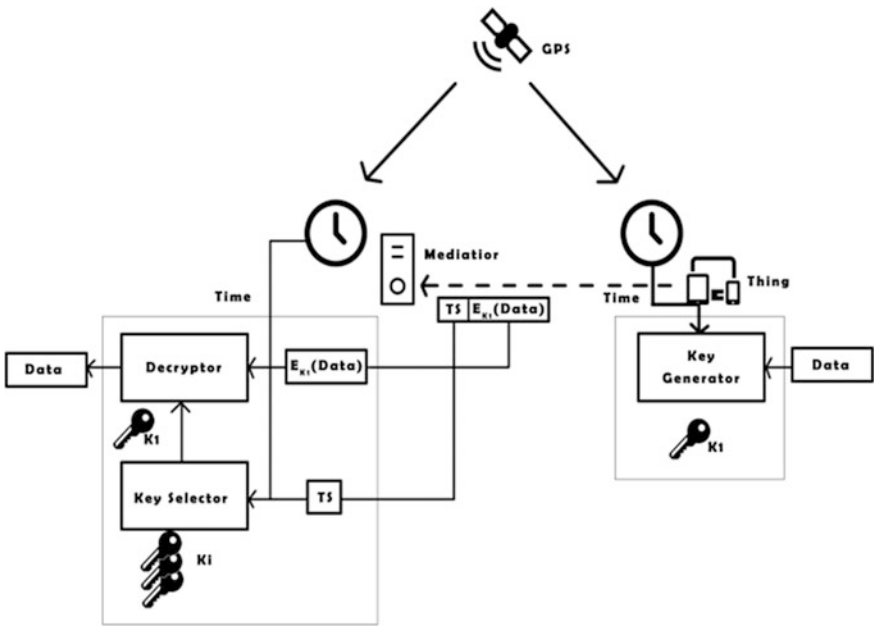


Fig. 4 Principle of time based key generation

secure keys. Moreover, the keys are generated locally on both sides of the communication link (i.e., transmitter and receiver) and are not shared along the connectivity link. It is assumed that clocks are locked to a global positioning system (GPS) timescale. This could be difficult to achieve for the IoT, since devices might not be able to receive GPS signals, or they might not be equipped with GPS. However, as shown in the following section, this principle can be extended to the considered heterogeneous IoT scenarios.

The security keys are bound with a time interval. Thus, the messages generated after the previous intervals are neglected. Comparing with other strategies, it doesn't need a dedicated server to maintain it. The replay attacks that focused with keys generated with past arena of time can be prevented and hence time based secure key generation is ensured. The constant advantage of this method is that it does not need to transmit the generated key through the transmission medium. The sender and receiver are synchronized with clock kept with global positioning systems scale of time. In the case of network connected with IoT devices there may be a chance of improper connection of IoT devices with GPS signals and the solution can be obtained.

The unidirectional IoT devices need to deal with synchronization of clock devices. A message sent by a transmitter without feedback need not to be received accurately. A generic non-IP unidirectional terminal performs following procedures to send data to the gateway/mediator:

1. It produces the encryption key locally with it, relatively with the time measured by a local clock.
2. It builds the message and encrypts it with the generated key.
3. Using the message text and generated key, it calculates the hash values.
4. The message is then sent to the gateway/mediator.

The message structure can be grouped into plain part and encrypted part. The plain part consists of time stamp that allows gateway/mediator to identify it locally; the hash can distinguish the transmitter identity; a security level parameter that maintains state data and sensor data. The encrypted part maintains a frame counter that decrements after every data transmit. When the gateway/mediator obtains the ciphered message, it will decipher it by producing exact decryption key starting from the timestamp attached. Based on the information supplied by the timestamp, the mediator can analyze and select the key to decrypt the given message. The temporal difference between the current time and the timestamp is calculated. If it exceeds the predefined threshold, the message will be discarded indefinitely. The cloud infrastructure service provides the processing power and hardware capacity needed for processing effective amount of data that is expected to handle situations such as frequent key generation in IoT devices.

7 Security Access Algorithms for Bidirectional Data Transmissions

The sender and receiver can transmit data packets in both directions in case of bidirectional data transmission. A mediator node in between the terminal nodes can observe the clock timings and broadcast the timing in a dedicated message. The identity of the message is kept in the plain part of the message [38]. Terminals can align their local clocks to the gateway/mediator terminal, and then generate the security keys. The propagation delays can be reduced since the devices are connected close with gateway. In the unidirectional scenario, the security keys have a time interval to keep with valid features. This will reduce unwanted transmissions by absorbing most retransmissions and the regeneration of new key can be done by time based generation algorithm. The terminal in the network is connected with different mediator gateways. This mediator is responsible for transmitting packets in both directions to recognize intermediate gateways with different clock times with minor variations.

If the terminal is not properly attaching mediator gateway identity in the transmitted message, the message will not be properly decrypted because of the gateway de-synchronization. In a network with IoT-gateway connections, the intermediate gateways transmit keys to the connected nodes. These keys are used to ensure the data identity and valid communication. The integrity of the broadcasted packet is to be ensured and this is a challenging task. Finally techniques are implemented by appending hash to the packet that is transmitted.

8 Cognitive Security

Traditional robust security mechanisms are inadequate mainly in wireless connections. They are deficient in preset infrastructure and needs isolation. Besides that, obliging wireless protocols are more susceptible. The dynamic network conditions do not allow genuine packets to be distinguished from anomalous ones. Due to unstable use of wireless technologies and the rapid evolution of mobile devices and applications, fully distributed control is a challenging question and may drop security management. Hence, mobile devices are severely investigated with more security trade-off.

Now-a-days, there exists need for a new approach for guaranteeing security since the adaptive security is insufficient. This new approach is termed cognitive security [38]. The word “cognitive” means conscious intellectual activity like knowing and perceiving. It is based on the prospect of being reduced to empirical factual data [39]. The security concerns are well solved by the techniques of effective knowledge representation. It handles authentication by using certain properties and trained patterns. The machine learning method can train the data with

effectual properties. The method has to be done with continuous process of analyzing dimensional features of packet data.

Figure 5 demonstrates a standard plan of cognitive security used for a capillary network. The cognitive engine gathers all the acknowledged data from the terminals in the capillary network through the mediator [4]. Various parameters are evaluated such as transmission–reception time difference of frames for each terminal, the transmission frequencies, the packet lengths and the queue lengths. In the unidirectional terminal scenario, the difference of timestamp connected with received frames supply information about the emission rate of the source. This can be compared with its target emission rate. In case of bidirectional terminals, their timestamp difference is evaluated at the mediator and should be cross checked with the base set value. With these evaluated parameters and on comparison with the historical data, a cognitive security-based algorithm should be properly used to ensure compatibility with security thresholds to neutralize all existing intruders or improper terminals that are not correctly working. The cognitive security engine can alter the back off time (BO) of the same terminals to amplify their possibility of accessing shared channel and transmitted frames. Whenever a traffic anomaly is detected at a certain terminal, the mediator will analyze the identity. The ID parameter at the terminal is analyzed and investigated to check as a potential disturber. If the disturber is found trusted and secure, the mediator alters the transmission parameters of terminals of the capillary network, to increase the bidirectional sent frames, and also notifies the ID disturber about the management entity of the capillary network. Evaluating at the opposite side, if the anomalous terminal is dependable, the mediator notifies the terminal ID to management entity that the terminal has been negotiated.

The mediator alters access parameters to a set of terminals. It depends on the information at the application level. Different possible access parameters that can be modified are:

1. Generation rate of the frames.
2. Reduction of the back-off time to repeat a new access to the channel.
3. Reducing the measured time to detect the presence of the transmission of another terminal.

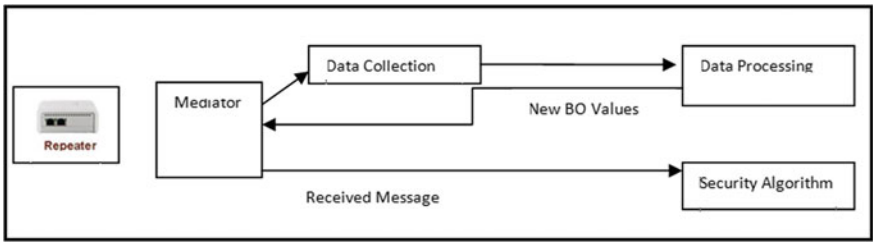


Fig. 5 Principle scheme of cognitive security work in the capillary network

9 Summary

Various security aspects and its counter measures were analyzed and discussed. Security has to be applied at the protocol level in order to cope up with the various attacks. Time based secured key generation helps the network abide to the replay attacks by properly time stamping the packets. Cognitive Security is distinguished by the ability to represent its features that guarantee adaptive security rather than the conventional security methods. A network of devices run on renewable energy sources can be modeled with IoT architecture and stipulations on IoT security can be evaluated. The scope of this work can be extended to study techniques to obtain robust IoT information and it's sharing over Internet clouds.

References

1. Naito, K.: A survey on the internet-of-things: standards, challenges and future prospects. J. Inf. Process. **25**, 23–31 (2017)
2. Chen, P.-Y., Chen, K.-C.: Optimal control of epidemic information dissemination in mobile ad hoc networks. In: IEEE Global Telecommunications Conference (GLOBECOM), pp. 1–5, Dec 2011
3. U.S. Department of Homeland Security: Strategic principles for securing the Internet of Things (IoT). Homeland Security, Nov (2015)
4. Eom, B., Lee, C., Yoon, C., Lee, H., Ryu, W.: A platform as a service for smart home. Int. J. Future Comput. Commun. **2**(3), 253–257 (2013)
5. Shen, J., Tan, H., Moh, S., Chung, I., Liu, Q., Sun, X.: Enhanced secure sensor association and key management in wireless body area networks. J. Commun. Netw. **17**(5) (2015)
6. Al Ameen, M., Liu, J., Kwak, K.: Security and privacy issues in wireless sensor networks for healthcare applications. J. Med. Syst. (2010)
7. Yu, F., Chang, C.-C., Shu, J., Ahmad, I., Zhang, J., de Fuentes, J.M.: Recent advances in security and privacy for wireless sensor networks. J. Sens. **2017** (2016)
8. Winkler, T., Rinner, B.: Security and privacy protection in visual sensor networks: a survey. ACM Comput. Surv. (CSUR) Surveys Homepage archive **47**(1) (2014)
9. Kausar, F.: Key management in wireless sensor networks: secure and efficient key generation, distribution and revoking in heterogeneous sensor networks (2012). ISBN:3659249955 9783659249952
10. Venkatasubramanian, K., Banerjee, A., Gupta, S.: PSKA: usable and secure key agreement scheme for body area networks. IEEE Trans. Inf. Technol. Biomed. **14**(1), 60–68 (2010)
11. Cazorla, M., Marquet, K., Minier, M.: Survey and benchmark of lightweight block ciphers for wireless sensor networks. In: International Conference on Security and Cryptography (SECRYPT) (2013)
12. IBM Corporation: Message queue telemetry transport (MQTT), June 2014
13. Vinoski, S.: Advanced message queuing protocol. IEEE Internet Comput. **10**(6), 87–89 (2006)
14. Vilajosana, X., Tuset-Peiro, P., Vazquez-Gallego, F., Alonso-Zarate, J., Alonso, L.: Standardized low-power wireless communication technologies for distributed sensing applications. Sensors (2014)
15. Palattella, M.R., Accettura, N., Vilajosana, X., Watteyne, T., Grieco, L.A., Boggia, G., Dohler, M.: Standardized protocol stack for the Internet of (important) Things. IEEE Commun. Surv. Tutor. **15**(3), 1389–1406 (2013)

16. Youssef, T.A., Elsayedm, A.T., Mohammed, O.A.: Data distribution service-based interoperability framework for smart grid test bed infrastructure. Energies (2016), Dec 2015
17. Sheng, Z., Yang, S., Yu, Y., Vasilakos, A., McCann, J., Leung, K.: A survey on the IETF protocol suite for the internet of things: standards, challenges, and opportunities. IEEE Wirel. Commun. **20**(6), 91–98 (2013)
18. Chen, P.-Y., Hero, A.O.: Assessing and safeguarding network resilience to nodal attacks. IEEE Commun. Mag. **52**(11), 138–143 (2014)
19. Iyer, S., Killingback, T., Sundaram, B., Wang, Z.: Attack robustness and centrality of complex networks, Apr 2013
20. Chen, P.-Y., Cheng, S.-M., Chen, K.-C.: Smart attacks in smart grid communication networks. IEEE Commun. Mag. **50**(8), 24–29 (2012)
21. Chen, P.-Y., Cheng, S.-M., Chen, K.-C.: Information fusion to defend intentional attack in internet of things. IEEE IoT J. **1**(4), 337–348 (2014)
22. Buscarino, A., Gambuzza, L.V., Porfiri, M., Fortuna, L., Frasca, M.: Robustness to noise in synchronization of complex networks. Nature (2013)
23. Chen, P.-Y., Hero, A.O.: Node removal vulnerability of the largest component of a network. In: Proceedings of IEEE GlobalSIP (2013)
24. Chen, P.-Y., Chen, K.-C.: Information epidemics in complex networks with opportunistic links and dynamic topology. In: Proceedings of IEEE Global Telecommunications Conference (GLOBECOM), pp. 1–6, Dec 2010
25. Chen, P.-Y., Lin, H.-F., Hsu, K.-H., Cheng, S.-M.: Modeling dynamics of malware with incubation period from the view of individual. In: 79th IEEE Vehicular Technology Conference (VTC Spring), pp. 1–5, May 2014
26. Sarkar, C., Nambi, S.N.A.U., Prasad, R.V., Rahim, A., Neisse, R., Baldini, G.: DIAT: a scalable distributed architecture for IoT. IEEE Internet Things J. **2**, 230–239 (2015)
27. Vijayalakshmi, A.V., Arockiam, L.: A study on security issues and challenges in IoT. Int. J. Eng. Sci. Manage. Res. (2016)
28. Shelby, Z., Hartke, K., Bormann, C., Frank, B.: Constrained application protocol (CoAP). In: IETF draft, January 2012
29. Hernández-Ramos, J.L., Jara, A.J., Marín, L., Skarmeta, A.F.: DCapBAC: embedding authorization logic into smart things through ECC optimizations. Int. J. Comput. Math. 1–22 (2014)
30. Dennis, J.B., Van Horn, E.C.: Programming semantics for multiprogrammed computations. Commun. ACM **9**(3), 143–155 (1966)
31. Crockford, D., RFC 7159: The JavaScript Object Notation (JSON) Data Interchange Format, IETF RFC 7159, March 2014. http://www.ietf.org/rfc/rfc7159.txt
32. Shelby, Z., Hartke, K., Bormann, C.: The constrained application protocol (COAP). IETF RFC **7252**, 10 (2014)
33. Rissanen, E.: Extensible access control markup language (XACML) version 3.0 oasis standard (2012)
34. Choi, S.I., Koh, S.-J.: Use of proxy mobile IPv6 for mobility management in CoAP-Based internet-of-things networks. IEEE Commun. Lett. (2016)
35. Winter, T., Thubert, P., Brandt, A., Hui, J., Kelsey, R., Levis, P., Pister, K., Struik, R., Vasseur, J.P., Alexander, R.: RPL: IPv6 routing protocol for low power and lossy networks. Request for Comments (RFC): 6550, March 2012
36. Weber, R.H.: Internet of things: new security and privacy challenges. Comput. Law Secur. Rev. **26**(1), 23–30 (2010)
37. Suo, H., Wan, J., Zou, C., Liu, J.: Security in the internet of things: a review. In: Proceedings of International Conference on Computer Science and Electronics Engineering (ICCSEE), vol. 3, pp. 648–651, March 2012
38. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements and future direction. Future Gener. Comput. Syst. **29**, 1645–1660 (2013)
39. Alabaa, F.A., Othmana, M., Hashema, I.A.T., Alotaibi, F.: internet of things security: a survey, vol. 88, pp. 10–28, June 2017

A Hyper Heuristic Localization Based Cloned Node Detection Technique Using GSA Based Simulated Annealing in Sensor Networks

D. Rajesh Kumar and A. Shanmugam

Abstract Due to inadequate energy resources, data aggregation from multiple sensors in Wireless Sensor Networks (WSN) is typically accomplished by clustering. But such data aggregation is recognized to be highly susceptible to clone attacks owing to the unattended nature of the network. Thus, ascertaining trustiness of the sensor nodes is crucial for WSN. Though numerous methods for cloned attack node isolation are provided in recent years, energy efficiency is the most significant issues to be handled. In this work, a Residual Energy and GSA based Simulated Annealing (RE-GSASA) for detecting and isolating the cloned attack node in WSN is given. Residual Energy-based Data Aggregation in WSN initially uses residual energy because the basis to perform aggregation technique with the sensor node possessing the maximum residual energy as the Cluster Head (CH). Next, Location-based Cloned attack on cluster nodes is given to enhance the clone detection probability rate. Here, the location and residual energy is taken into account to identify the presence of cloned attack nodes within the network. Finally, Gravitational Search Algorithm with global search ability is investigated to identify the cloned attack nodes and performs isolation through local optimal simulated annealing model. Simulation results demonstrate that RE-GSASA provides optimized energy consumption and improves cloned attack detection probability by minimizing the cloned attack detection time.

Keywords Wireless sensor networks · Data aggregation · Clone attack Gravitational search algorithm · Simulated annealing

D. Rajesh Kumar (✉)
Department of Information Technology, Bannari Amman Institute of Technology, Sathyamangalam 638401, India
e-mail: sangeraje@gmail.com

A. Shanmugam
Department of Electronics and Communication Engineering, SNS College of Technology, Coimbatore 641035, India
e-mail: dras_bit@yahoo.com

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*, Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_13

1 Introduction

A Wireless Sensor Network (WSN) is an infrastructure [1] that consists of sensing, computing and a communication process. WSN comprises a huge number of sensor nodes [2] which are placed in a finite area and it forms a network. These nodes sense the accessible data from the location and transmit the sensitive data to the base station. The Base Station (BS) verifies the data and stored for further needs. A WSN provides the potential of sensing various types of physical and environmental [3]. The WSNs includes the excessive capability in civil and military applications, such as smart home, smart grid, healthcare monitoring and intelligent transport.

The applications of WSN are employed [4] that monitors several physical values such as light, temperature, pressure, object detection, size, etc. The sensor includes the ability to broadcast and forward sensing data to the base station. The improvement of WSNs was introduced [5] in military applications such as battlefield surveillance. The WSNs are generally employed in technical settings, residential backgrounds and survival environments. Structural health monitoring, management applications, home mechanization and animal tracking are some of the models in WSNs applications.

A WSN is considered for the collection of sensor nodes in a cooperative network. Each and every node in the network has its own processing capability, memory, power sources, etc. The communication of data between nodes is operated in WSNs for better utilization of energy. Cluster based replication attack detection contains an identical copy of information during the transmission of aggregated data to the sink node. The mobile agents in WSNs identify the replicated nodes with the aid of replication attack approach. The data packets acquired from multiple sensor nodes are aggregated at an aggregator sensor node.

The aggregator sensor node forwards the packets to the BS only the aggregate values. During data aggregation, once a node is being captured by an adversary or malicious node and generates a clone of it and through it propagates into the entire network, causing serious threats to the entire network. Due to limited energy resources, data aggregation from multiple sensor nodes in WSN is habitually accomplished by uncomplicated methods such as clustering. However, such data aggregation is known to be extremely vulnerable to clone attacks due to the unattended nature of the network.

In order to detect and isolate the cloned attack nodes in wireless sensor network, a Residual Energy and GSA based Simulated Annealing (RE-GSASA) method is presented. Initially, Energy-based Data Aggregation is carried out using residual energy as the basic and the aggregation process is performed with sensor nodes that possessing the highest remaining energy as the cluster head. Then Location-based Cloned attack on cluster nodes is proposed for increasing the cloned attack detection probability rate and minimizing the cloned attack detection time. Here, the location and residual energy are considered to identify the presence of cloned attack

nodes in the network. Finally, Gravitational Search Algorithm (GSA) with global search ability is investigated to detect the cloned attack nodes and performs isolation through local optimal simulated annealing model.

2 Review on Clone Attack Detection and Isolation in WSN

In the node replication attack or clone node attack, the adversary forms the low-cost wireless sensor nodes termed clone nodes by [6]. An adversary captures a node after gathering all secret credentials. An adversary duplicates the sensor node and uses more than one clone into the network at all positions by reducing the network performance with internal attacks. The node replication attack is an essential one in sensor network security and plans many detection schemes against node replication attack. But, the clone attack detection probability rate is not higher.

An efficient clone detection scheme is designed [7] for sensor network where the selection is an essential one. The selection is used for clone detection schemes with device types, detection techniques, deployment plans and detection ranges. It is advantageous to use grid deployment knowledge in static sensor networks. However, the detection failed to remove the attack from the network.

A survey on Clone attack [8] addresses the existing problems in WSN. Position Verification Method (PVM) combines with Message Verification and Passing (MVP) for identifying, removing and avoiding the Clone node's entry. An optimization framework is designed [9] for selecting the parameters of the detection technique where the clone detection costs get reduced. The detection method is classified into four costs. They are: leaving undetected cloned nodes in the network, the cost of revoking nodes falsely recognized, communication cost and storage cost. A convex combination of cost describes the clone detection cost that reduced regarding the parameters of the detection method. However, the clone attack detection time is high.

The deterministic clone node detection is performed for anonymous RFID technique lacking priori tag IDentifiers (IDs). Three protocols, namely BASE, DeClone and DeClone+ are designed [10] for fast and deterministic clone node detection in RFID systems. The BASE controls observation where the clone tags, create tag cardinality higher than ID cardinality. The DeClone is used with new finding where the clone tags lead to the collisions that reconciled by means of re-arbitration. For DeClone, tree traversal verifies the unreconciled collisions. The DeClone+ integrates optimization techniques with fast clone detection when clone ratio is higher. Though the clone detection rate is high, the energy consumption for detecting the clone attack remained unaddressed.

A distributed Low-Storage Clone Detection protocol (LSCD) is designed [11] for WSNs. A detection route is identified in the perpendicular direction of the witness path with nodes in ring path. The detection route meets the witness path as the distance between two routes is comparatively lesser than witness path length. Clone node detection is carried out in a non-hotspot region where energy

consumption gets decreased with network lifetime enhancement. But, the communication cost is high in distributed LSCD.

Two new node clone detection protocols are designed [12] with trade-offs on network conditions. The first protocol is depending on a Distributed Hash Table (DHT) where decentralized, key-based caching and checking system identifies the cloned nodes efficiently. The second detection protocol termed randomly directed exploration has better communication results for dense sensor networks through probabilistic directed forwarding method with random direction and border determination. A distributed algorithm is designed [13] for detecting cloned nodes in WSNs. The drawbacks of leaving undetected cloned nodes in the network, communication cost and storage cost are addressed. An optimization framework is designed for selecting the clone node detection parameters based on costs and detection schemes. But, the energy consumption in an optimization framework while detecting the cloned attacks in the network is high.

An energy-efficient location-aware clone detection protocol is introduced in WSNs [14]. The designed protocol assures clone attack detection and increases the network lifetime. The location information about sensors is used and selects the witnesses to authenticate the legality of sensors and to report identified clone attacks. The ring structure forwards the data in energy efficient manner along the path to the witnesses and sinks. A new scheme designed [15] identifies the node clone attack in WSN through channel identification characteristic where the clone nodes are differentiated with channel responses between nodes. The detection scheme attained fast detection and reduced the data transmission cost with temporal and spatial uniqueness. But, the clone attack detection efficiency is less.

By considering the sparse feature of replicated nodes, a clone detection framework termed CSI is designed [16] depending on compressed sensing. CSI is based on the detection efficiency on compressed aggregation of sensor readings. CSI attained at the lowest communication cost and distributes network traffic uniformly over sensor nodes. They are attained with the sparse property of clones in the sensor network by clone attack.

3 Problem Definition

Wireless Sensor Networks (WSN) consist of a huge number of tiny and low-cost sensors [17–19] which are heavily organized around the target. They are more appropriate for applications such as battlefield monitoring, identifying the environmental pollutants, traffic monitoring, patient health monitoring in a hospital, etc. The WSN is frequently utilized in hostile environments [20, 21] and they are sensitive to attacks since, the resource limits nature of the sensor nodes. Clone node attack in WSN is one of the major issues developed [22] where the messages are monitored the received node is cloned.

Location-aware Energy-efficient Ring based Clone Detection protocol [23, 24] with clone detection protocol (ERCD) is introduced. The ERCD protocol ensures

the efficient clone attack detection. However, the energy consumption in ERCD protocol ring is high. Therefore, efficient clone attack detection scheme is required in WSNs. Hence, a Defense Mechanism for Clone Attacks in WSN [8] based on Gravitational Search Algorithm (GSA) where the witness nodes are employed to discover the clone attacks. The nodes in the channel are partitioned into witness node and the claimer node for efficiently detecting malicious node. The witness node checks the Node ID obtained from the claimer node along with the sequence and timestamp for efficient detection of clone attack. If a node contains a similar ID but different random sequences, it specifies the clone attack is occurring. The witness nodes continuously transmit request messages with a time stamp to the claimer nodes. In addition, to select the best witness node, GSA is used.

Gravitational Search Algorithm is applied to choosing the witness nodes in the network. GSA is a typical memory-less algorithm, but works proficiently like algorithm with memory. GSA provides the one of the finest optimal solution by mass and change in velocity of the object. Clone attack detection is executed by examining the behaviour of the neighbour nodes. After that isolation process is performed. Then, revocation procedure is activated by overflow the network with two incoherent response messages received by the witness node. The revocation process stops the clone attacks in the witness nodes. Defence mechanism based on GSA algorithm provides improved security to clone attacks with minimizing packet drop and enhancing packet delivery ratio. However, the harmful attacks in network remained unaddressed.

Dynamic Source Routing (DSR) protocol is designed for successfully detecting and isolating harmful attacks in the way of an improved DSR protocol. The DSR protocol depends on the throughput of the sensor network. When the throughput of the network, reduced to the particular threshold value, nodes in the network have monitored the status of nodes. The malicious node occurs sensor region is selected among source and destination. In case, if any, malicious node has existed in the sensory region, DSR protocol is used to detect the malicious node. The malicious node is the origin for producing the selective packet drop attack. After that detected node are isolated from the network. DSR routing protocol is used to detect the malicious node in the network, but the black hole node reduction rate of is less in the DSR routing protocol.

Therefore, an Improvised Hierarchical Black Hole Detection Algorithm is introduced [25, 26] for increasing the reduction rate of the black hole attack in WSNs. An intrusion detection system (IDS) is introduced in the hierarchical method for obtaining vitality efficient. IDS are used to protect the network nodes from black hole attacks. Hierarchical black hole detection approach is simple and depending on the transfer of control packets between sensor nodes and the base station. Hence, BS performs the part of the monitor node to detect any black hole attack. Hierarchical approach efficiently mitigates the effect of the black hole attack. However, hierarchical black hole detection approach consumes more energy to detect the black hole attacks.

An Optimized Weight-based Clustering Algorithm [27] with Security (OWCA) mechanism is used to reduce the energy consumption in WSNs. Initially, OWCA

algorithm partitions the network into different clusters. Clustering is a method in which sensor nodes does not require to transmit their data directly to the BS. The huge amount of energy loss has occurred on the directly transmitting the node to the base station. Therefore nodes in the cluster transmit the data to the CH. CH gather the data packets from all nodes and detect if any black hole node exists in the network. Then, the black hole node is cancelled from the sensor network. Hence the energy utilized in the network is reduced. However, during the transmitting process, OWCA mechanism consumes additional energy in the network. Therefore, the above-said methods and protocols are failing to improve the attack detection rate in WSNs. Hence, an efficient method is needed for detecting and isolating the clone and black hole attacks in WSNs.

4 Clone Attack Detection Overview

Wireless Sensor Networks (WSNs) is often deployed in competitive atmosphere and are susceptible to attacks because of the resource constrained nature of the sensors. Clone attack is one of the major issues in WSN where, the messages are overheard, the captured node is cloned and multiple nodes with same identity are produced by an attacker. Therefore, a Distributed Defence Mechanism for Clone Attacks based on Gravitational Search Algorithm (GSA) is designed [28] for overcoming the issues of the cloned attack in WSN. The nodes in the channels are separated into witness node and the claimer node for efficiently detecting the suspect nodes. The witness nodes are reliable for the suspect node detection, while the claimer nodes provide their identities in the detection process.

GSA is a new optimization algorithm that depends on the law of gravity. In GSA, agents are described as objects and their performance is predicted by their masses. All these objects are acquiring each other by the gravitational force and this force results in the global movement of the entire objects towards the entity which comprises heavier masses. Hence, masses are assisted by a direct form of communication based on gravitational force. In GSA, each agent mainly contains the four specifications that are positioned, inertial mass, active gravitational mass and passive gravitational mass. The position of the mass provides the result of the dilemma and its gravitational and ideal massed are constructed with the help of fitness function. Each mass signifies a result and the algorithm are managed by properly regulating the gravitational and inertia masses. The gravitational and inertia masses are assumed to concern by the heaviest mass and this mass is the optimal solution. The GSA requires to be treated as an isolated method of masses. In GSA algorithm, heavy masses equivalents to superior result as it moves slowly than lighter ones. Figure 1 shows the clone attack detection based on GSA algorithm.

The process of cloning attack detection is shown in Fig. 1. All nodes are separated into two main groups for effectively identifying the attacks in nodes. Those

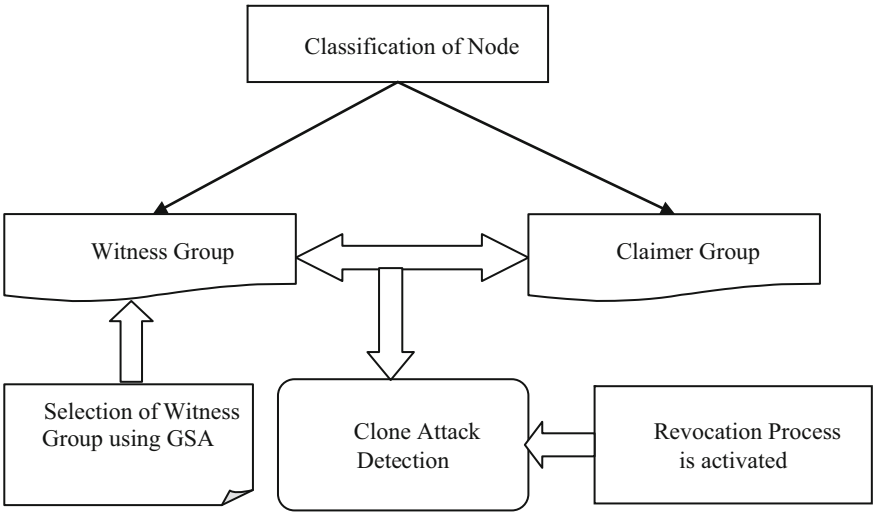


Fig. 1 Clone attack detection process

are witness group and claimer group. The nodes in the network channel periodically perform these two roles.

The witness node continuously broadcasts acknowledge messages to claimer node with time stamp. As a result, the entire neighbouring claimer nodes include the signal range of the witness nodes, receive the previous acknowledge messages. The claimer node broadcasts response messages to the witness nodes in time interval where response delay is less than the network channel coherence time.

The reply message includes the node ID and the pilot. The suspected nodes are discovered and they are deposited in the suspected group using node ID and the pilot. In order to verify the clone attacks, the suspect group is broadcasted to the claimer nodes. Hence, after collecting the broadcasted message, the claimer nodes forwards the reply message which consists of the node ID, sequence and time stamp. If the two nodes include the same ID and different random sequence, then the nodes are considered to be attacked. The process of selecting witness nodes in the network is performed by GSA that tends to discover the global optimum faster algorithms with higher convergence rate. Gravitational constant regulates the accuracy of the search, thereby decreasing with time.

The witness nodes are continuously broadcast request messages with a time stamp to the claimer nodes. The GSA is used to find best witness nodes set for witness node selection. After selecting the witness nodes, clone attack detection is carried out by examining the behaviour of the neighbour nodes. On detecting the clone attack, isolation process is activated to separate the clone attack in the witness nodes. Revocation process is performed through flooding the network with two incoherent reply messages obtained by the witness node. Revocation reduces the

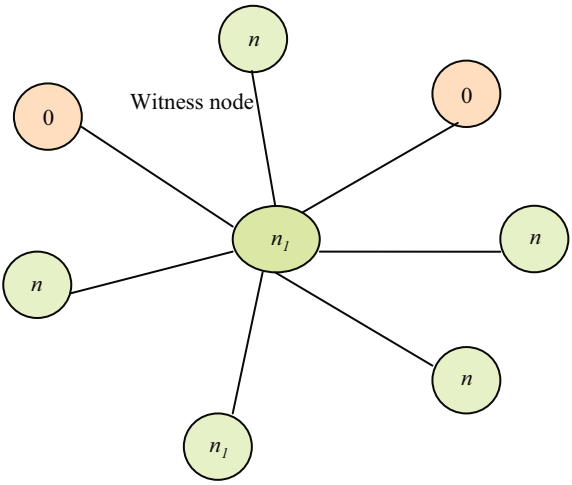


Fig. 2 Detection of malicious node

clone attacks in the witness nodes. GSA algorithm produces better defence to clone attacks with improving the packet delivery ratio and reducing packet drop.

Once, the witness node is selected, clone node attack detection starts identifying the behaviour of the neighbour node. Two or more malicious nodes placed a witness node region. Figure 2 shows the detection of malicious nodes. A witness node (n1) receives the record of their neighbour nodes (n). Witness node checks all the record and store the witness ID that consists of at least one same note ID with it into a same node ID. In addition, a random sequence is broadcasted to detect the suspect nodes. Hence, all witness nodes broadcasts randomly selected sequences to their neighbouring node and requests all their neighbour nodes to reply back the received sequence along with their ID. The broadcasted random sequences are different in various witness nodes.

5 Residual Energy and GSA Based Simulated Annealing

WSN consists of an unsecured environment and an adversary node is created due to an unsecured environment that can capture a sensor node, reconstruct it and produce a clone of the seized node. These clone node attacks pose serious threats because, with the single seized sensor node, the adversary node creates as many cloned nodes as he requires. A hierarchical distributed algorithm was introduced [29] for identifying node replication attacks by using a Bloom filter mechanism and a CH selection. But, optimizing the energy consumption remained unsolved.

Global Deterministic Linear propagation verification protocol (GDL) was introduced by [30] for detecting node replication attacks in WSN. GDL shares the node location information to numerous arbitrarily selected cells and then linear-multicasts the information for verification from the localized cells. However,

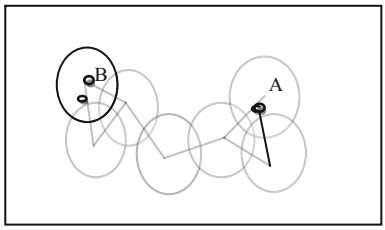


Fig. 3 Unit disk graph model using two sensor nodes

the detection performance of node replication attacks is not effective. In order to overcome the above said issue in WSN, Residual Energy and GSA based Simulated Annealing (RE-GSASA) method to efficiently detect and to isolate the cloned attack node in WSN. This RE-GSASA method improves the packet delivery ratio with multiple sensor nodes in the sensor network. The node energy optimization is performed well for efficient transmission and reduces the attack in the network. Therefore, Residual Energy and GSA based Simulated Annealing (RE-GSASA) method is proposed to efficiently detect and isolate the cloned attack node in WSN.

Let us consider a WSN which consists of a Base Station '*BS*', and many sensor nodes '$SN = SN_1, SN_2, \ldots, SN_n$' randomly deployed in a '1200×1200 m'. Let us further design the WSN in the form of an undirected graph '$G = (V, E)$', where 'V' and 'E' represent the set of sensor nodes and edges respectively. Prior to deployment, every sensor nodes in the network is assigned a key also called as the unique node ID 'SN_{ID}'.

A Unit Disk Graph model is used in which two sensor nodes are connected if their distance is below a fixed threshold '$\tau \leq 1$' and therefore '$p, q \in E$'. In the unit disk graph, the two sensor nodes only communicate within the network only the distance between them is at most R. Transmission radius (R) which is equivalent for all sensor nodes in the entire network. The radius of the disk of each node is R/2, and then the two nodes are communicated only if their corresponding disks intersect.

Figure 3 illustrates the Unit disk graph model using two sensor nodes A and B connected with radius R/2. All sensor nodes in the network are considered to be limited in communication and computation power and also the battery life. So, every sensor in the network directly communicates with 'n' other sensor nodes, also known as the neighbor nodes.

5.1 Residual Energy-Based Data Aggregation in WSN

WSN consists of a number of sensor nodes and a BS node. Here, a BS node is considered to be more secure and posses unlimited available energy than the other sensor nodes in the network. In WAN, sensors a node monitors the geographical

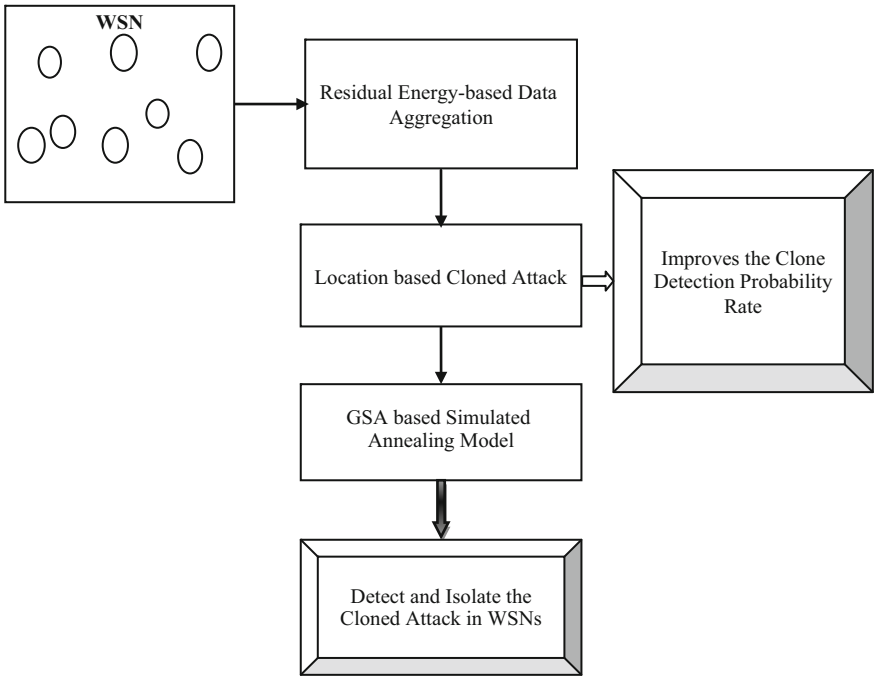


Fig. 4 Residual energy and GSA based simulated annealing method

area and collects the sensory information and communicate with the base station. The collected sensory information is aggregated at the intermediate sensor nodes to preserve energy by using an appropriate aggregation model. In this work, Residual Energy-based Data Aggregation (RE-DA) is used for collecting the sensed data from the diverse sensor nodes in WSN.

Figure 4 demonstrates that the Residual Energy and GSA based Simulated Annealing (RE-GSASA) method. The RE-GSASA method contains three processing steps for detecting and isolating the cloned attack in WSNs. At first, Residual Energy-based Data Aggregation collects data from multiple sensor nodes and provides that fused data to base station node according to the residual energy so as to avoid redundant transmission of data which in turn helps to minimize the energy consumption of sensor nodes.

In addition, the Location-based Cloned attack on cluster nodes is performed in RE-GSASA method thus improving the cloned detection probability rate. At last, Gravitational Search Algorithm based Simulated Annealing model with the global search ability to detect and isolate the cloned attack nodes.

A Data Aggregation and Authentication protocol (DAA) [31] combine the false data detection with data aggregation and privacy. For data aggregation with false

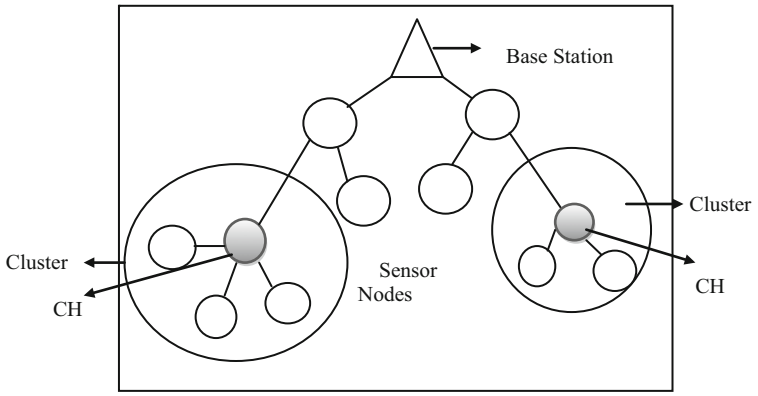


Fig. 5 Residual energy-based data aggregation

data detection, monitoring nodes of data aggregator perform the data aggregation operation. In addition, it also aggregates the small-size message code for data verification at their pair mates. For secret data transmission, the sensor nodes between successive data aggregators authenticate the data integrity of encrypted data than plain data. Hence, Residual Energy-based Data Aggregation model is designed in proposed RE-GSASA method for collecting the sensed data from the diverse sensor nodes in WSN.

Residual Energy-based Data Aggregation model collects the data from multiple sensor nodes and provides the combination of data to the base station according to the residual energy. Therefore, it avoids redundant transmission of data which in turns helps to reduce the energy consumption of sensor nodes. The main aim of Residual Energy-based Data Aggregation is to reduce the energy consumption during the clustering process by exchanging fewer packets between the nodes. Figure 5 shows the sensor node organization in clusters and communication between the nodes to the cluster head and then to the base station based on the residual energy.

As shown in Fig. 5, the proposed method selects its aggregators based on residual energy. It optimizes the total energy consumption of the aggregation process with the sensor nodes by selecting the cluster head that consists of highest residual energy. Similarly, other sensor nodes select their cluster head among the neighbours, according to the residual energy and the distance to the cluster head.

Many data aggregation schemes are depending on privacy homomorphism encryption on wireless sensor networks. In [32], it presents data aggregation schemes with enhanced security level as cluster heads aggregate the cipher texts without decryption. The aggregation functions are limited. The base station failed to

verify the data integrity and authenticity through linking the message digests or signatures to all sensing samples. The base station recovers all sensing data, termed as recoverable.

Let us consider the energy consumption '$E_{consump}$' for one iteration. The distance between the sensor node and the base station is considered for iteration and therefore packet transmission from a sensor node to the base station 'BS' is achieved. Hence, energy consumption is given be twice the distance between the sensor node and the base station. The mathematical formula for energy consumption is given in Eq. (1).

$$\mathrm{E_{consump}} = 2 \times (\mathrm{Dis_{SN \to BS}}) \tag{1}$$

For two iterations, the energy consumption is given in Eq. (2).

$$2 \times \mathrm{E_{consump}} = 2^2 \times (\mathrm{Dis_{SN \to BS}}) \tag{2}$$

Similarly, for 'n' iterations, the energy consumption is mathematically formulated as Eq. (3).

$$\mathrm{n} \times \mathrm{E_{consump}} = 2^n \times (\mathrm{Dis_{SN \to BS}}) \tag{3}$$

Hence, the residual energy for one sensor node after 'n' iterations is given in Eq. (4).

$$ResidualEnergy_n = E - n \times E_{consump} = E - 2^n \times (Dis_{SN \to BS}) \tag{4}$$

From the Eq. (4), 'E' denotes the initial energy of the sensor node, then the residual energy for all sensor nodes (i.e. total residual energy) in the network is formulated as Eq. (5).

$$ResidualEnergy_{total} = \sum_{i=1}^{n} E_i - 2^n \times Dis_i \tag{5}$$

From (5) E_i is the energy for all sensor nodes in the network. Obtained with the total residual energy, clustering of sensor nodes for data aggregation is performed based on residual energy. This in turn reduces the time duration for data transmission over the network, increasing the energy efficiency. Algorithm 1 shows the Residual Energy-based Data Aggregation algorithm.

Algorithm 1 Energy-based Data Aggregation Algorithm

```
// Residual Energy-based Data Aggregation Algorithm
Input: Base Station 'BS', Sensor Nodes 'SN = SN1, SN2, …, SNn', Base
Station 'BS', Iteration 'n'
Output: Optimized energy consumption
Step 1: Begin
Step 2:      For each Sensor Node 'SN' and 'n' iterations (to perform
             clustering)
Step 3:          Measure energy consumption using (3)
Step 4:          Measure residual energy using (4)
Step 5:          Measure residual energy for 'n' sensor nodes using
                 (5)
Step 6:      End for
Step 7: End
```

A distance weight taking account of the residual energy during clustering for data aggregation is presented in Algorithm 1. Through the algorithm, the energy consumption during the single iteration, two iterations and 'n' iterations are measured. With this the residual energy for one sensor node, all the sensor nodes in the network is obtained. With the aid of obtained residual energy of all sensor nodes, Residual Energy-based Data Aggregation algorithm efficiently performs the data aggregation, which in turn helps to reduce the energy consumption during the data aggregation in an effective manner.

5.2 Location-Based Cloned Attack on Cluster Nodes

Once, the sensor nodes are clustered based on the similar residual energy model from Energy-based Data Aggregation algorithm, the cloned attack on clustered nodes are identified. Cloned attacks occur in different cases. The proposed RE-GSASA method analyses the occurrence of a cloned attack on clustered nodes based on similar residual energy.

Randomized Multicast is NDFD that address the needs with high communication overhead. For reducing the communication overhead, two NDFD protocols, namely random walk and Table-assisted random walk are designed [33]. The Random walk based approach is developed to detect the cloned attack in WSN. But, the security level is not at the required level. The proposed method analysis the occurrence of a cloned attack on clustered nodes based on similar residual energy and provides better security on the data aggregation process.

During the data aggregation process every sensor node 'SN' in the network sends its ID and location information '$(SN_{ID}, SN_{r,c})$' to the base station 'BS' through other sensor nodes (i.e. neighboring nodes). Here, 'SN_{ID}' symbolizes the

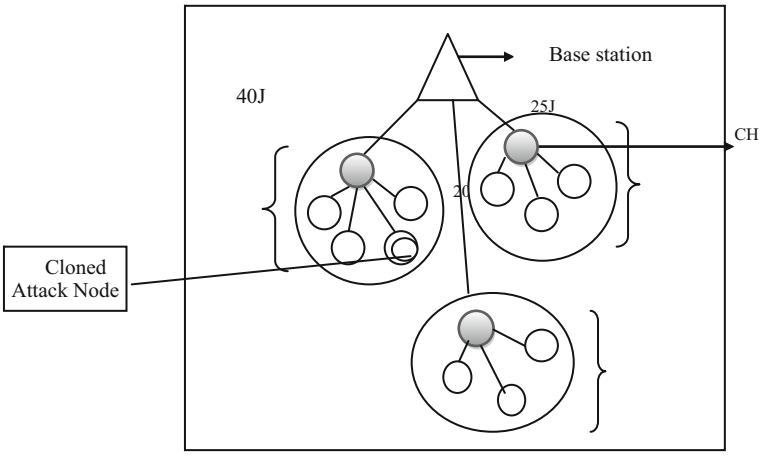


Fig. 6 Structure of cloned attack on cluster nodes based on similar residual energy

sensor nodes ID whereas '$SN_{r,c}$' symbolizes the sensor node positioning or location represented in the row and column 'r, c' respectively. Figure 6 shows the occurrence of a cloned attack on cluster nodes based on similar residual energy.

Figure 6 shows the structure of the cloned attack on cluster nodes based on similar residual energy. The base station checks the node IDs upon receiving the location claims from all the sensor nodes. When the base station receives two different locations named as '$SN_{ID1}, SN_{r,c}$' and '$SN_{ID2}, SN_{r,c}$', they contains similar node ID '$SN_{ID1} = SN_{ID1}$' and similar residual energy 'Residual EnergySN_{ID1} = Residual EnergySN_{ID1}'. But when there is a different location, then the base station concludes that at least one of the sensor nodes is a cloned attack node, or else the base station realizes that no attack node is present in the network and proceeds with the data transfer.

A distributed detection protocol is presented [30] to counteract node replication attacks. The presented scheme sends the node location information to many selected cells and then linear-multicasts information for verification from localized cells [34]. However, the adversary nodes are not correctly identified by the designed scheme. Therefore, residual energy designs an algorithm of cloned attacks to detect and isolate the adversary nodes. Algorithm 2 shows the algorithmic representation of Cloned attacks on Clustered Nodes.

Algorithm 2 Cloned Attacks on Clustered Nodes

Input: Base Station 'BS', Sensor Nodes '$SN = SN_1, SN_2, \ldots, SN_n$', Base Station 'BS', Sensor Node ID '$SN_{ID} = SN_{ID1}, SN_{ID2}, \ldots, SN_{IDn}$', Location information '$(SN_{r,c})$', Residual energy 'Residual EnergySN_{ID}'
Output: Improves Clone attack detection probability
Step 1: Begin
Step 2: For each iterations and 'n' sensor nodes
Step 3: Sensor node sends it ID 'SN_{ID}' and location information '$SN_{r,c}$' to thebase station 'BS'
Step 4: If $(SN_{IDi} = SN_{IDj})$ && (Residual EnergySN_{IDi} = Residual EnergySN_{IDj}) && $(SN_{IDi\ r,c} <> SN_{IDjr,c})$
Step 5: Occurrence of cloned attack node
Step 6: Else
Step 7: Occurrence of normal node
Step 8: End if
Step 9: End for
Step 10: End

Algorithm 2 gives the algorithmic description of the cloned attack on clustered nodes. The cloned attack on clustered nodes is said to occur only if two sensor nodes possess a similar node ID and residual energy based on different locations [35]. Here, any one of the sensor nodes is considered as the cloned attack node by the base station.

5.3 GSA Based Simulated Annealing

Finally, a GSA based Simulated Annealing model is applied to select the witness node that detects the cloned attack node and isolate the cloned attack node in WSN. From [36], it is observed that Gravitational Search Algorithm has global search ability. However, the algorithm lacks a local search mechanism. On the other hand, Annealing Algorithm, though not found to be good in global searching, can be applied to obtain the local optimal solution. The structural diagram of the GSA based Simulated Annealing model is illustrated in Fig. 7.

Therefore, the proposed RE-GSASA method applies an integrated Gravitational Search Algorithm (to detect cloned attack node) based Simulated Annealing (isolate the cloned attack node) to render both the global and local search capabilities. To start with the cloned attack node detection using Gravitational Search Algorithm is presented followed by the optimal annealing model.

Let us consider a network with 'n' sensor nodes (masses). The position of the 'ith' node is defined in (6):

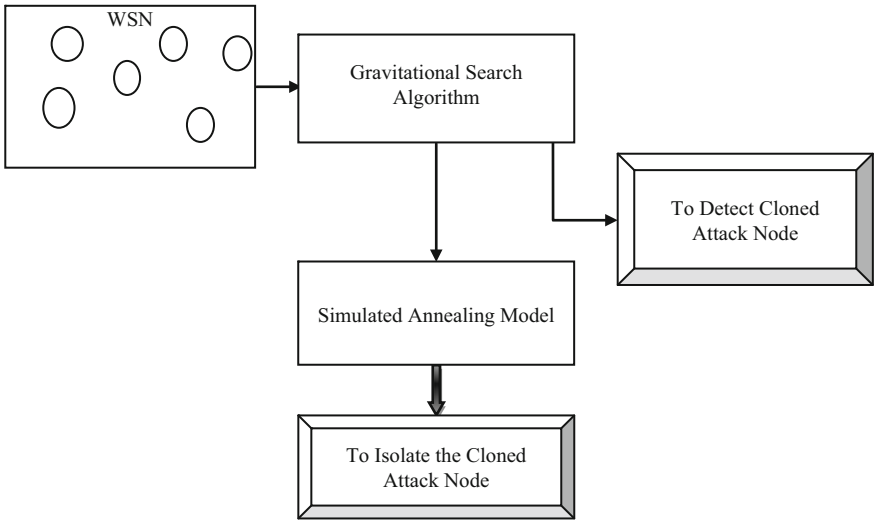


Fig. 7 Structural diagram of the GSA based simulated annealing model

$$X_i = X_i^1, X_i^2, \ldots, X_i^k, \ldots, X_i^n \tag{6}$$

From Eq. (6), 'X_i^k', represent the position of the 'ith' sensor node in 'kth' dimension. At a particular time 't', the force acting on mass 'i' from mass 'j' is given in (7):

$$F_{ij}^k = G(t) \times \frac{M_{pi}(t) \times M_{aj}(t)}{Dis_{ij}(t) + \varepsilon} \tag{7}$$

In Eq. (7), 'M_{pi}' and 'M_{aj}' represents the passive and active relational mass related to the mass of sensor nodes 'i' and 'j' respectively. The gravitational constant is represented by '$G(t)$' with a small constant 'ε'. The distance between the sensor nodes 'i' and 'j' at time 't' are represented by '$Dis_{ij}(t)$'. To further obtain stochastic characteristics, it is presumed that the total force that acts on sensor node 'i' in a dimension 'e' to be an arbitrarily weighted sum of 'eth' component of the forces applied from other sensor nodes is as given in the Eq. (8).

$$F_i^e = \text{ran}\ F_{ij}^k(t),\ j = 1,\ j \neq i \tag{8}$$

From Eq. (8), 'ran' represents the random number in the interval '0, 1'. Hence, according to the law of motion, the acceleration of the sensor node 'i' at a time 't' in 'kth' dimension is given in Eq. (9).

$$A_i^e(t) = \frac{F_{ij}^k}{M_{ii}(t)} \tag{9}$$

The acceleration of sensor node represents in Eq. (9), '$M_{ii}(t)$', represent the inertial mass of the sensor node 'i' at time 't'.

As an individual position update plan in Gravitational Search Algorithm (GSA) causes damage to the individual position and the local search ability of GSA is weak, an enhanced algorithm is designed [37, 38]. The algorithm designed Simulated Annealing into GSA with Metropolis-principle-based individual position for improving the particle movement. Therefore, in order to measure the fitness value to determine the node which is the witness node, gravitational and inertia masses are computed. A sensor node possessing heavier mass represents a more efficient node. By assuming the equality of gravitational and inertia mass, the values of the masses are evaluated with the aid of map of fitness.

$$m_i(t) = \frac{Fit_i(t) - Worst(t)}{Best(t) - Worst(t)} \tag{10}$$

$$M_i(t) = \frac{m_i(t)}{\sum_{r=1}^{n} m_r(t)} \tag{11}$$

From Eqs. (10) and (11), the gravitational and inertial mass is attained using '$Fit_i(t)$' and it denotes the fitness value for 'i' sensor node at time 't' and the 'Best(t)' and 'Worst(t)' is obtained as given below. For a minimization problem, the 'Best(t)' and 'Worst(t)' is formalized as Eqs. (12) and (13).

$$Best(t) = \min Fit_n(t) \tag{12}$$

$$Worst(t) = \max Fit_n(t) \tag{13}$$

For a maximization problem, the 'Best(t)' and 'Worst(t)' is formalized as Eqs. (14) and (15)

$$Best(t) = \max Fit_n(t) \tag{14}$$

$$Worst(t) = \min Fit_n(t) \tag{15}$$

Therefore, the sensor nodes with the right fitness value are assigned as the witness node that detects the presence of cloned attack on cluster nodes with similar residual energy. Once the cloned attack nodes are detected by the witness node that possesses the best fitness value, they (sensor node) are isolated using simulated annealing that operates the annealing of optimal individuals (sensor nodes).

At each step, the simulated annealing considers certain neighbour sensor nodes '$SN^{'}$' of the current sensor node 'SN'. Probabilistically node is chosen between moving the system to sensor node '$SN^{'}$' or staying at 'SN' with the objective of

isolating the cloned attack node. It is repeated until the detected cloned attack node is isolated from the network.

Algorithm 3 presents the simulated annealing process to isolate the detected cloned node in WSN. Algorithm 3 shows the algorithm using the optimal annealing to operate the optimal individual to improve the capability of local optimization.

Algorithm 3 Optimal Annealing Algorithm

Input: Sensor Node '$SN = SN_1, SN_2, .., SN_i, \ldots, SN_n$'
Output: Improves cloned attack detection time
Step1: Begin
Step 2: Let $SN = SN_i$, where $i = 1$ to n
Step 3: For $n = 0$ through SN_n
Step 4: Pick a random neighbor $SN_{new} \leftarrow$ Neighbor (SN_i)
Step 5: If Prob $\left(\begin{matrix}\text{Energy}_{\text{Residual}}(SN_i),\\ \text{Energy\_Residual}(SN_{new})\end{matrix}\right) > Ran(0,1)$
Step 6: Move to the next sensor node $SN_i \leftarrow SN_{new}$
Step 7: Cloned attack node does not exist and outputs sensor node SN_i and terminate test
Step 8: End if
Step 9: If Prob $\left(\begin{matrix}\text{Energy}_{\text{Residual}}(SN_i),\\ \text{Energy\_Residual}(SN_{new})\end{matrix}\right) < Ran(0,1)$
Step 10: Move to the next sensor node $SN_{new} \leftarrow SN_i$
Step 11: Cloned attack node exists and outputs sensor node SN_{new} and terminates test
Step 12: End if
Step 13: End for
Step 14: End

As shown in Algorithm 3, Optimal Annealing Algorithm starts from a sensor node 'SN_i' and continues to process till a maximum of 'n' steps are arrived at. During the process, a randomly chosen neighbor of a given sensor node 'SN' is generated. The annealing schedule is defined by the 'Energy_Residual' function, which should yield the residual energy to use, given the presence of the sensor nodes in the network. The application of GSA based Simulated Annealing reduces the cloned attack detection time.

6 Experimental Evaluation

A Residual Energy and GSA based Simulated Annealing (RE-GSASA) method is implemented in the NS2 simulator using DSR protocol. The sensor network consists of 500 nodes, placed in a random manner in the WSN that generates traffic for

Table 1 Simulation parameters

Parameters	Values
Network simulator	NS 2.34
Network area	1200 m × 1200 m
Protocol	DSR
Number of sensor nodes	50, 100, 150, 200, 250, 300, 350, 400, 450, 500
Number of data packets i.e., size of data block	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Range of communication (m)	30
Speed of node (m/s)	0–10
Simulation time (ms)	500
Number of runs	10

every 10 m/s. The base station node collects the data from different sensor nodes with a range of 10–100 and forwards the data to the base station node. The size of each data packet size varies from 100 to 512 KB. The simulation time varies from 500 simulation seconds to 1500 simulation seconds and the following metrics like energy consumption during data aggregation, clone attack detection probability rate and packet delivery ratio in WSN is measured. Table 1 lists the set of an input parameter and evaluates the performance of the RE-GSASA method.

In proposed RE-GSASA method, the performance is carried out by comparing with two existing methods. The proposed RE-GSASA method compared against the existing, Hierarchical Distributed Algorithm [29] and Global Deterministic Linear propagation verification protocol (GDL) in WSN [30]. For performance evaluation, the comparison is made from existing works, whereas the simulation parameter values are provided in Table 1 with the positions of nodes. To illustrate the simulation results for RE-GSASA method, experiments were conducted and the average values are taken. The performance of proposed Residual Energy and GSA based Simulated Annealing (RE-GSASA) method is measured in terms of following.

(i) Energy Consumption
(ii) Cloned Attack Detection Probability Rate
(iii) Cloned Attack Detection Time
(iv) Packet Delivery Ratio

7 Performance Analysis of Re-GSASA Method

The proposed Residual Energy and GSA based Simulated Annealing method is compared with two existing methods. They are namely, Hierarchical Distributed Algorithm [29] and Global Deterministic Linear propagation verification protocol (GDL) in WSN [30]. To evaluate the proposed method, the following metrics are used.

7.1 Performance Analysis of Energy Consumption

Energy consumption for data aggregation at the base station node is the product of energy consumed by a single node and the total nodes in WSN.

$$EC = Energy_{SN} \times Total_{SN} \tag{16}$$

From Eq. (16), 'EC' is the energy consumption for data aggregation at the base station node whereas 'SN' represents the sensor nodes. The consumption of energy is measured in terms of Joules.

Table 2 shows the experimental values of energy consumption obtained during data aggregation with respect to different number of sensor nodes. For experimental purposes, sensor nodes range from 50 to 500 nodes. The comparison is made with two existing methods namely, Hierarchical Distributed Algorithm [29] and Global Deterministic Linear propagation verification protocol (GDL) [30].

From the above table, it is illustrative that the energy consumption of proposed RE-GSASA method attains lower energy consumption when two other existing state-of-methods.

Table 2 Tabulation for energy consumption

Number of nodes	Energy consumption (J)		
	Hierarchical distributed algorithm	GDL	Proposed RE-GSASA
50	112	93	76
100	134	117	87
150	153	127	105
200	174	149	124
250	182	164	148
300	211	186	163
350	234	204	187
400	256	237	211
450	287	261	235
500	306	279	257

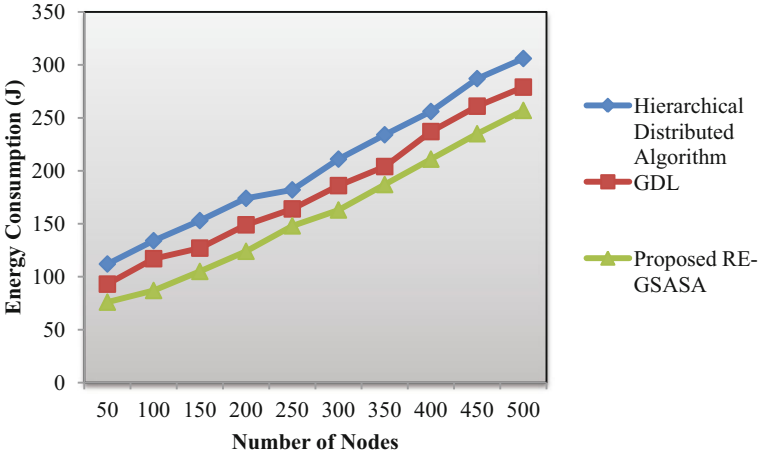


Fig. 8 Measure of energy consumption

Figure 8 shows the measure of energy consumption using proposed RE-GSASA method with existing Hierarchical Distributed Algorithm [29] and Global Deterministic Linear propagation verification protocol (GDL) [30]. As shown in Fig. 8, the proposed method consumes less energy for data aggregation in the sensor nodes. Based on the residual energy in the RE-GSASA method, every forwarding sensor node chooses its aggregators and the sensor node having the highest residual energy is selected as the cluster head. Thus the clustering of sensor nodes is performed based on residual energy. The transmission of these clustered nodes data packet consumes much battery power resulting in a more energy drain. RE-GSASA then provides the fused data to the base station that further does not require any additional process to aggregate so very less energy is discharged. In the RE-GSASA method [39], during data aggregation the sensor nodes exhaust less energy whereas in Hierarchical Distributed Algorithm and GDL more energy is exhausted. Therefore, a RE-GSASA method minimizes the energy consumption due to efficient clustering based on residual energy. Thus, energy consumption is minimized by 33% when compared to Hierarchical Distributed Algorithm and 16% when compared to GDL respectively.

7.2 Performance Analysis of Cloned Attack Detection Probability Rate

Cloned attack detection probability rate is defined as the ratio of difference between the overall sensor nodes and cloned attack node to the overall sensor nodes in WSN. The clone attack detection probability rate is measured in terms of percentage (%).

Table 3 Tabulation for cloned attack detection probability rate

Number of nodes	Cloned attack detection probability rate (%)		
	Hierarchical distributed algorithm	GDL	Proposed RE-GSASA
50	66.23	69.78	76.21
100	67.23	71.26	78.22
150	69.21	73.54	79.84
200	71.58	74.89	81.42
250	72.4	75.69	83.11
300	73.69	77.82	84.78
350	75.46	79.36	85.96
400	75.89	81.42	87.32
450	77.24	82.76	89.12
500	79.98	84.33	91.55

$$\mathrm{CADP} = \frac{(\mathrm{SN_i} - \mathrm{SN_{CN}})}{\mathrm{SN_i}} \times 100 \tag{17}$$

From Eq. (17) the cloned attack detection probability rate 'CADP' is obtained using the overall sensor nodes 'SN_i' and the cloned attack sensor nodes 'SN_{CN}' during data aggregation in WSN.

Table 3 shows the comparative value of the cloned attack probability rate for various numbers of sensor nodes ranges from 50 to 500 nodes. Whenever sensor nodes in the network send data, due to the presence of replica or cloned attack nodes in WSN, packet drop has occurred [40]. Hence, the possibilities of packet drops occur during data packet transmission create a clone for a specific sensor node and introduce the false data through the sensor nodes and send to the base station. The base station collects data from the sensor node. As a result, clone attack node gets aggregated at the base station. The targeting results of clone attack detection probability rate using the RE-GSASA method provides better results when compared with other state-of-the-art methods.

Figure 9 represents the simulation results of the cloned attack detection probability rate for different methods, when there is only one clone node in the network. The clone attack detection probability method must have a high probability of node replication detection. The node detection probability is designed with independent sensor networks and different number of sensor nodes. From the above figure, it observes the networks with any node density, detection probability of methods are acceptable. Therefore, the detection probability of the RE-GSASA method provides comparatively better results by increasing the size of the networks. By applying Location-based Cloned attack on cluster nodes, the detection probability rate is improved. They compare the sensor node ID and location along with the information regarding to the residual energy for improving the cloned attack detection probability rate. As the result, the cloned attack detection probability rate is

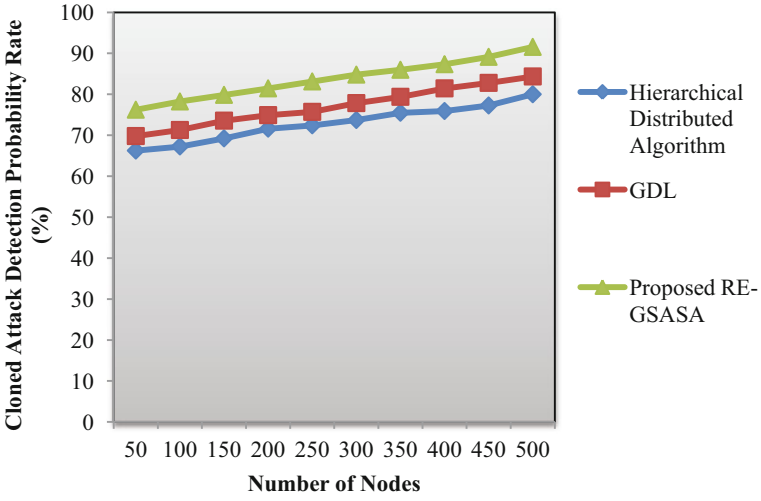


Fig. 9 Measure of cloned attack detection probability rate

improved in the RE-GSASA method by 12% compared to Hierarchical Distributed Algorithm [29] and 8% compared GDL [30].

7.3 Performance Analysis of Cloned Attack Detection Time

The cloned attack detection time is defined as the measures of time taken to identify the cloned nodes in WSN. It is specified by the product of the total number of nodes on the network to the time taken for identifying the cloned attack nodes. The detection time is measured in terms of milliseconds (ms).

$$\text{CADT} = \text{Number of Sensor Nodes} \times \text{Time (Identifying the Cloned Nodes)} \quad (18)$$

The cloned attack detection time 'CADT' is measures using the Eq. (18) based on the total number of sensor nodes. While the cloned attack detection time is lower, the method is said to be more efficient.

Table 4 shows the comparison value of cloned attack detection time for different number of sensor nodes using proposed RE-GSASA method with existing Hierarchical Distributed Algorithm [41] and GDL. Here, a number of sensor nodes range from 50 to 500. The comparison is made with two existing methods namely, Hierarchical Distributed Algorithm [29] and Global Deterministic Linear propagation verification protocol (GDL) [30] in WSN. From the above table, it is illustrative that the cloned attack detection time of proposed RE-GSASA method attains lower when to other existing state-of-methods.

Table 4 Tabulation for cloned attack detection time

Number of nodes	Cloned attack detection time (ms)		
	Hierarchical distributed algorithm	GDL	Proposed RE-GSASA
50	22.37	18.98	16.23
100	23.48	19.67	17.28
150	25.61	22.32	19.63
200	27.48	22.74	20.56
250	28.93	24.21	21.85
300	30.45	25.86	22.56
350	31.26	27.69	24.85
400	32.78	29.15	26.12
450	33.82	31.21	27.86
500	35.12	32.87	29.35

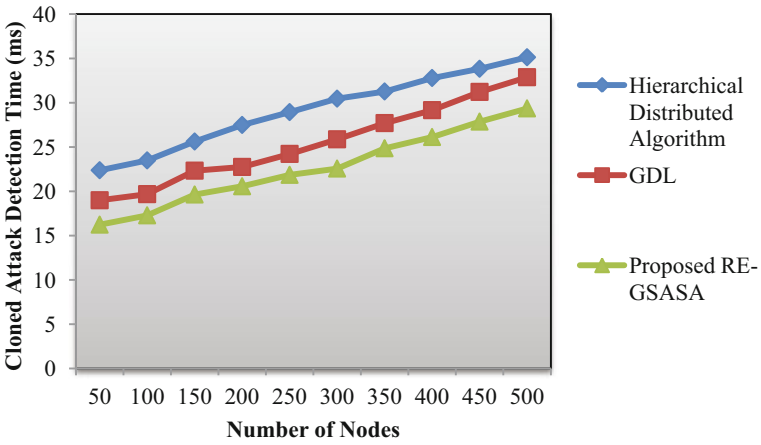


Fig. 10 Measure of cloned attack detection time

Figure 10 shows the experimental analysis of cloned attack detection time for proposed RE-GSASA method with existing methods. In the RE-GSASA method, with an increase in the number of nodes, the routing time also increases. As shown in Fig. 10, the cloned attack detection time is reduced using the proposed RE-GSASA method.

With the construction of an optimal annealing algorithm, the cloned attack detection time is reduced using the proposed RE-GSASA method. While increasing the number of sensor nodes, the cloned attack detection time is also getting increased using all the three methods. But, comparatively the cloned attack detection time is lower using proposed RE-GSASA model.

The simulated annealing presents the process to isolate the detected cloned node in WSN and with the application of GSA based Simulated Annealing reduces the

cloned attack detection time. Therefore, the RE-GSASA method reduces the cloned attack detection time by 30% when compared to Hierarchical Distributed Algorithm [29] and 13% compared to GDL [30].

7.4 Performance Analysis of Packet Delivery Ratio

The average packet delivery ratio is the ratio of the number of data packets received successfully at the base station and the total number of packets transmitted by the sensor nodes. It is measured in terms of percentage (%).

$$\mathrm{PDR} = \frac{P_r}{P_s} \times 100 \tag{19}$$

From the Eq. (19), the packet delivery ratio 'PDR' is obtained. Here, the number of packets received at the base station node is denoted as 'P_r' and the number of packets sent by the sensor node is given by 'P_s' respectively.

Table 5 shows the comparison value of the packet delivery ratio for different number of packets using proposed RE-GSASA method that ranges from 10 to 100 packets. It is compared elaborately with existing methods namely, Hierarchical Distributed Algorithm and GDL. From the table values, proposed RE-GSASA method increases gradually though not linear for different number of packets when compared to the other methods.

From Fig. 11, the measure of the packet delivery ratio is presented for three methods according to the different number of packets and data packets sent. As illustrated in Fig. 11, the RE-GSASA method performs relatively well when compared to two other methods Hierarchical Distributed Algorithm [29] and GDL [30] This is because of the GSA based optimal Simulated Annealing to detect the

Table 5 Tabulation for packet delivery ratio

Number of packets	Packet delivery ratio (%)		
	Hierarchical distributed algorithm	GDL	Proposed RE-GSASA
10	71.23	75.69	81.23
20	73.44	77.65	82.69
30	74.89	79.23	84.72
40	75.69	81.23	86.31
50	77.32	82.9	87.98
60	79.11	84.78	89.63
70	80.3	86.21	91.57
80	81.87	86.87	92.74
90	83.11	88.12	94.21
100	84.89	89.57	95.68

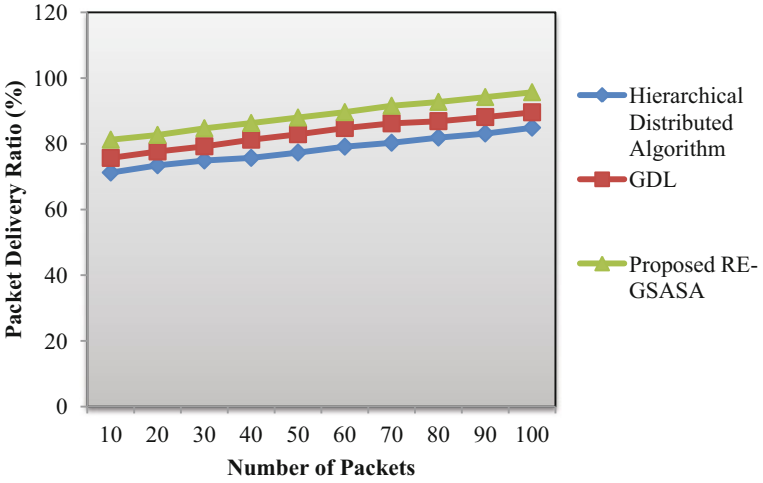


Fig. 11 Measure of packet delivery ratio

presence of cloned attack nodes based on the similar residual energy and isolates the node in an optimal manner that simulates the annealing pattern to operate optimal individual gravitational and inertial mass on the basis of node density. Furthermore, the RE-GSASA method simulates the annealing model based on the residual energy in addition to the sensor node ID and location [42]. Therefore, the RE-GSASA method improves the packet delivery ratio by 12% when compared to Hierarchical Distributed Algorithm and 6% compared to GDL.

8 Summary

A novel Residual Energy and GSA based Simulated Annealing (RE-GSASA) method is introduced to reduce the energy consumption during data aggregation and improve the packet delivery ratio. Here, residual energy is based on the clone attack detection probability rate with varying node density and packets. The main objective is achieved by performing data aggregation using residual energy-based clustering. It selects the optimal cluster head node and transmitting the data packets to the base station. Initially, a Location-based Cloned attack is designed on cluster nodes to measure the occurrence of a cloned attack on clustered nodes based on similar residual. It performs energy saving and obtains optimum attack detection probability route for different sensor nodes with different packets in WSN. Finally, GSA-based optimal Simulated Annealing model is investigated to obtain the witness node that detects and isolates the cloned attack nodes from the network. Thus, it ensures maximum packet delivery ratio.

Future, GSA based Simulated Annealing Black hole attack Detection model is developed to identify and isolate the black hole attack nodes in WSNs. This model is employed with a Location Based Black Hole Attack Possibility for attaining the black hole attack detection.

References

1. Singh, S.K., Singh, M.P., Singh, D.K.: Routing protocols in wireless sensor networks—a survey. Int. J. Comput. Sci. Eng. Surv. (IJCSES) **1**(2), 63–83 (2010)
2. Islam, M.M., Hassan, M.M., Lee, G.-W., Huh, E.-N.: A survey on virtualization of wireless sensor networks. Sensors (Open Access Journal), 2176–2207
3. Sonule, M.G., Nikam, S.: Role of sensor virtualization in wireless sensor networks. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **5**(1), 699–703 (2014)
4. Khera, S., Mehla, N., Kaur, N.: Applications and challenges in wireless sensor networks. Int. J. Adv. Res. Comput. Commun. Eng. **5**(6), 448–451 (2016)
5. Alsheikh, M.A., Lin, S., Niyato, D., Tan, H.-P.: Machine learning in wireless sensor networks: algorithms, strategies, and applications. IEEE Commun. Surv. Tutor. **16**(4), 1–24 (2012)
6. Khan, W.Z., Aalsalem, M.Y., Saad, N.M.: Distributed clone detection in static wireless sensor networks: random walk with network division. PLoS ONE **10**(5), 1–22 (2015)
7. Cho, K., Lee, D.H.: Low-priced and efficient replica detection framework for wireless sensor networks. Peer-to-Peer Netw. Appl. **7**(4), 743–749 (2014). Springer
8. Maheswari, P.U., Kumar, P.G.: Dynamic detection and prevention of clone attack in wireless sensor networks. Wirel. Pers. Commun. 1–12 (2016). Springer
9. Bonaci, T., Lee, P., Bushnell, L., Poovendran, R.: A convex optimization approach for clone detection in wireless sensor networks. Pervasive Mob. Comput. **9**, 528–545 (2013). Elsevier
10. Bu, K., Xu, M., Liu, X., Luo, J., Zhang, S., Weng, M.: Deterministic detection of cloning attacks for anonymous RFID systems. IEEE Trans. Ind. Inform. **11**(6), 1255–1266 (2015)
11. Dong, M., Ota, K., Yang, L.T., Liu, A., Guo, M.: LSCD: a low-storage clone detection protocol for cyber-physical systems. IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst. **35** (5), 712–723
12. Li, Z., Gong, G.: On the node clone detection in wireless sensor networks. IEEE/ACM Trans. Netw. **21**(6), 1799–1811 (2013)
13. Bonaci, T., Lee, P., Bushnell, L., Poovendran, R.: Distributed clone detection in wireless sensor networks: an optimization approach. In: IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, pp. 1–6 (2011)
14. Zheng, Z., Liu, A., Cai, L.X., Chen, Z., Shen, X.S.: Energy and memory efficient clone detection in wireless sensor networks. IEEE Trans. Mob. Comput. **15**(5), 1130–1143 (2016)
15. Wen, H., Luo, J., Zhou, L.: Lightweight and effective detection scheme for node clone attack in wireless sensor networks. IET Wirel. Sens. Syst. **1**(3), 137–143 (2011)
16. Yu, C.-M., Lu, C.-S., Kuo, S.-Y.: Compressed sensing-based clone identification in sensor networks. IEEE Trans. Wireless Commun. **15**(4), 3071–3084 (2016)
17. Mishra, A.K., Turuk, A.K.: A zone-based node replica detection scheme for wireless sensor networks. Wirel. Pers. Commun. **69**(2), 601–621 (2013). Springer
18. Mishra, A.K., Turuk, A.K.: Residual energy-based replica detection scheme for mobile wireless sensor networks. J. Secur. Commun. Netw. **8**(4), 637–648 (2015)
19. Mishra, A.K., Turuk, A.K.: Node coloring based replica detection technique in wireless sensor networks. Wirel. Netw. **20**(8), 2419–2435 (2014). Springer

20. Zhu, B., Setia, S., Jajodia, S., Roy, S., Wang, L.: Localized multicast: efficient and distributed replica detection in large-scale sensor networks. IEEE Trans. Mob. Comput. **9**(7), 913–926 (2010)
21. Chaudhari, H.C., Kadam, L.U.: Wireless sensor networks: security, attacks and challenges. Int. J. Netw. **1**(1), 4–16 (2011)
22. Deore, M.R., Patil, R.V.: A review: detection of clones in wireless sensor network. Int. J. Comput. Sci. **6**(2), 316–323 (2015)
23. Zheng, Z., Liu, A., Cai, L.X., Chen, Z., Shen, X.S.: ERCD: an energy-efficient clone detection protocol in WSNs, In: Proceedings of IEEE INFOCOM (2013)
24. Qiu, T., Zhang, Y., Qiao, D., Zhang, X., Wymore, M.L., Sangaiah, A.K.: A robust time synchronization scheme for industrial internet of things. IEEE Trans. Industr. Inf. (2017). https://doi.org/10.1109/TII.2017.2738842
25. Kumar, A., Khoslay, A., Sainiz, J.S., Singh, S.: Meta-Heuristic range based node localization algorithm for wireless sensor networks, pp. 1–7. IEEE Conference Publications (2012)
26. Karuppiah, A.B., Dalfiah, J., Yuvashri, K., Rajaram, S.: An improvised hierarchical black hole detection algorithm in wireless sensor networks. In: International Conference on Innovation Information in Computing Technologies (ICIICT), pp. 1–7 (2015)
27. Lal, C., Shrivastava, A.: An energy preserving detection mechanism for blackhole attack in wireless sensor networks. Int. J. Comput. Appl. **115**(16), 32–37 (2015)
28. Maheswari, P.U., Thenmozhi, L., Ganeshkumar, P.: Distributed detection of clone attacks in wireless sensor networks using RED-ANT algorithm. J. Comput. Appl. (JCA) **6**(3), 43–46 (2013)
29. Znaidi, W., Minier, M., Ubéda, S.: Hierarchical node replication attacks detection in wireless sensor networks. Int. J. Distrib. Sens. Netw. **2013**, 1–12 (2013). Hindawi Publishing Corporation
30. Zhou, Y., Huang, Z., Wang, J., Huang, R., Yu, D.: An energy-efficient random verification protocol for the detection of node clone attacks in wireless sensor networks. EURASIP J. Wirel. Commun. Netw. **163**, 1–12 (2014). Springer
31. Ozdemir, S., Çam, H.: Integration of false data detection with data aggregation and confidential transmission in wireless sensor networks. IEEE/ACM Trans. Netw. **18**(3), 1–14 (2010)
32. Chen, C.-M., Lin, Y.-H., Lin, Y.-C., Sun, H.-M.: RCDA: recoverable concealed data aggregation for data integrity in wireless sensor networks. IEEE Trans. Parallel Distrib. Syst. **23**(4), 727–734 (2012)
33. Zeng, Y., Cao, J., Zhang, S., Guo, S., Xie, L.: Random-walk based approach to detect clone attacks in wireless sensor networks. IEEE J. Sel. Areas Commun. **28**(5), 677–691 (2010)
34. Hidoussi, F., Toral-Cruz, H., Boubiche, D.E., Lakhtaria, K., Mihovska, A., Voznak, M.: Centralized IDS based on misuse detection for cluster-based wireless sensors networks. Wirel. Pers. Commun. **85**(1), 207–224 (2015). Springer
35. Shaukat, H.R., Hashim, F., Sali, A., Abdul Rasid, M.F.: Node replication attacks in mobile wireless sensor network: a survey. Int. J. Distrib. Sens. Netw. **8**, 1–15 (2014). Hindawi Publishing Corporation
36. Rashedi, E., Nezamabadi-pour, H., Saryazdi, S.: GSA: a gravitational search algorithm. Inf. Sci. **179**(13), 2232–2248. Elsevier
37. Tong, C.: Gravitational search algorithm based on simulated annealing. J. Convergence Inf. Technol. (JCIT) **9**(2), 231–237 (2014)
38. Ho, J.-W., Wright, M., Das, S.K.: Fast detection of mobile replica node attacks in wireless sensor networks using sequential hypothesis testing. IEEE Trans. Mob. Comput. **10**(6), 767–782 (2011)
39. Dhanaraj, R.K., Shanmugam, A., Palanisamy, C., Natarajan, A.M.: Optimal clone attack detection model using an energy-efficient GSA based simulated annealing in wireless sensor networks. Asian J. Res. Soc. Sci. Hum. 6(11), 201 (2016)
40. Huang, D.-J., Teng, W.-C.: A defense against clock skew replication attacks in wireless sensor networks. J. Netw. Comput. Appl. **39**, 26–37 (2014). Elsevier

41. Conti, M., Di Pietro, R., Spognardi, A.: Clone wars: distributed detection of clone attacks in mobile WSNs. J. Comput. Syst. Sci. **80**(3), 654–669 (2014). Elsevier
42. Taylor, V.F., Fokum, D.T.: Mitigating black hole attacks in wireless sensor networks using node-resident expert systems. In: Wireless Telecommunications Symposium (WTS), pp. 1–7 (2014)
43. Vasserman, E.Y., Hopper, N.: Vampire attacks: draining life from wireless ad-hoc sensor networks. IEEE Trans. Mob. Comput. **12**(2), 318–332 (2013)

Review on Analysis of the Application Areas and Algorithms used in Data Wrangling in Big Data

Chiranjivi Bashya, Malka N. Halgamuge and Azeem Mohammad

Abstract This study performed a content analysis of data retrieved from 30 peer-reviewed scientific publications (1996–2016) that describe the applied algorithm models for data wrangling in Big Data. This analysis method explores and evaluates applied algorithm models of data applications in the area of data wrangling methods in Big Data. Data wrangling unifies messy and complex data by a procedure of planning, which involves, clustering, and grouping of untidy and intricate sets of for easy access for the purposes of trending themes useful for business or company planning. This application of data wrangling is not only for business use, but also for the convenience of individuals, business users that consume data directly in reports, or schemes that further process data by streaming it into targets such as data warehouses, called data lakes. This method sets- up easy access and analysis of all untidy data. Data streaming procedure are exceptionally useful for planning, small and big businesses, all around the world who use data non-stop and constantly to produce emerging trends, structure and schemes that inadvertently makes a difference when sustaining and customising business by simply streaming data it into warehouses, or in other words data storage pools. This study analyzed and found that commonly used statistical figures and algorithms are used by major data application, however the information technology area certainly faces security challenges. However, Data wrangling algorithms used in different data applications such as medical data, textual data, financial data, topological data, governmental data, educational science, galaxy data, etc. could use clustering methods as it is much effective than others. This study has analyzed and found significant comparisons and contrasts between algorithms along with data applications and evaluated them to identify certain superior methods over others.

C. Bashya · M. N. Halgamuge (✉) · A. Mohammad
School of Computing and Mathematics, Charles Sturt University, Melbourne, VIC 3000, Australia
e-mail: MHalgamuge@studygroup.com; malka.nisha@unimelb.edu.au

C. Bashya
e-mail: chiranjivibashyal7@gmail.com

A. Mohammad
e-mail: AMohammad@studygroup.com

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*, Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_14

Moreover, it shows that there is a significant use of medical data in the big data research area. Our results show that data wrangling when clustering algorithm can solve medical data storage issues by clustering algorithms. Similarly, clustering algorithms are frequently used for clustering data sets to analyze information from raw data. Fifty percent of the literature found that clustering algorithms for Data wrangling method is beneficial for algorithms used in different data applications to thoroughly analyze and evaluate their importance. After the analysis of Clustering algorithm, suggestions are made for applications used by medical data for the data wrangling purposes.

Graphic Abstract A pictorial representation of the abstract of this research is shown in Fig. 1.

Keywords Data wrangling · Big data · Algorithms · Clustering Decision tree · Data application · Medical data · Financial data

Abbreviations

IT	Information Technology
t-SNE	t-Distributed Stochastic Neighbor Embedding
SOM	Self-organizing Map
NCD	Normalized Compression Distance
GHSOM	Growing Hierarchical Self-organizing Map
GCS	Growing Cell Structure
IGG	Incremental Grid Growing

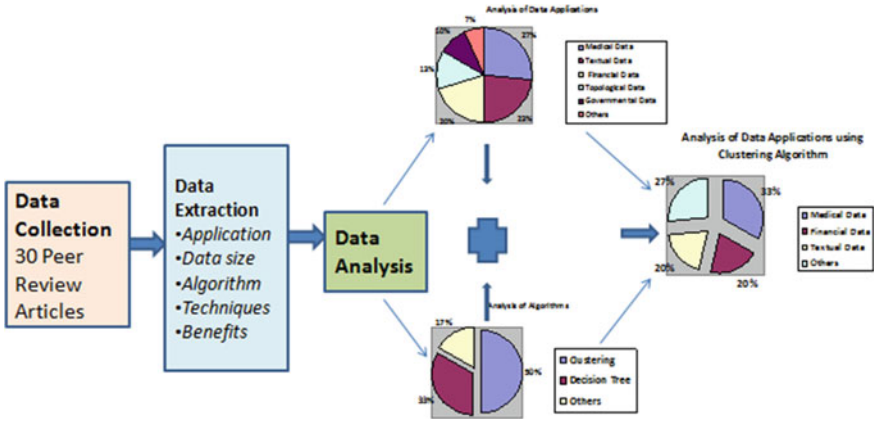


Fig. 1 Graphical abstract of the research: The data was collected from thirty peer review articles that features areas such as, data size, algorithms, techniques, and their benefits are studied and resulted and this present study will analyze these results to look at trends of these studies. The data application areas and algorithms are tools that are used to analyze statistical results to gather data. From the results, we further analyzed the percentage of the data application areas which relates to the most frequently used algorithm, i.e., clustering algorithm

1 Introduction

The world has totally entered the data performing age called "Wrangler" this method has introduced a new way to unify messy data and the word "wrangling" means to wrangle through direct interaction with data presented in a visual interface. The broad utilization of IT advancements has made a difference in data sources that have been expanding at a phenomenal rate while making the structures and types of information progressively complex due to confidentiality issues. Detailed analysis and use of enormous information will take a vital part in futuristic planning that advances the financial rise of nations and improve the expansion of organizations [1]. Conducting research on normal issues of enormous data, particularly on achievements of innovations will empower organizations and eliminate the drawbacks of incomplete data and their interconnection. This helps to remove the vulnerabilities brought by the repetition or potential deficiency of data. Data wrangling helps with the extraction of knowledge from information for data science, arithmetic, topology, medicine, and financial matters. It contains a lot of societal data that consequently can be seen as a system mapped to society. Investigating huge data and discovering pieces of information can help us see the current situations, and track the seasonal patterns [2].

A vast majority of researchers spend time on data planning (gathering, cleaning and sorting out) before they can start further data analysis. There are few significant data jobs like data investigation, data representation, and so forth. However, the most critical and tedious task is, data wrangling as data storage administrator is works with large amounts of data which is extremely arduous. In today computerized company data compiles easily in areas where huge projects with large amounts of complex computer-generated data occur each day if not second.

Different organizations collect data the wrong way around with the use of multiple sources, such as some companies, clean the raw data to anticipate the data and then take required information from the data in order to take decisions to support the emerging trends [3]. Usually, the general approaches and practices for data wrangling suitability are assessed before to produce the anticipated effects, as analysts interactively discover new ways to manipulate the data and immediately see results. Besides, Wrangler follows the user's data alterations and mechanically generates an encryption or a script that applies it on datasets. Some popular methods can be used for the analysis purposes according to the requirement of the organization [4, 5]. This research has found that, the benefits of algorithms used in different data applications to analyze their importance for certain professions are extremely useful. This research seeks to explore data wrangling techniques that are used in modern data mining systems and how large organizations can benefit from extracting useful information. Thus, we can understand how efficient an algorithm can be in corresponding data applications. Finally, this research compares the different algorithms and data applications that help us in making accurate decisions while using data wrangling and algorithms in different data applications.

2 Materials and Methods

Data wrangling, as a process, commonly takes place after an arrangement of raw data which starts with separating the information in a set from the information source [6].

2.1 Collection of Raw Data

The raw data collected for this study is from 30 peer-reviewed scientific publications (1996–2016) these publications all have different data applications that have been studied and analyzed in this study. Among the review articles, most of the data was found to be medical science while financial data was also found to be used significantly in big data. The attributes compared were data applications, data size, algorithms and their benefits. These attributes were found to be the most important contents in the published articles to be studied or analyzed.

2.2 Data Inclusion Criteria

To assess data inclusion criteria a comparison table was drawn to include attributes such as Author, Applications, Data size, Algorithms, and Benefits. The different data applications are used for those research articles that are clustered with respect to the algorithms used. The Most of the data applications are used to cluster algorithm for data wrangling purposes. Nevertheless, in our analysis, we excluded studies when no whole parameters are revealed and the publication is not published in peer-reviewed scientific publications.

2.3 Analysis of Raw Data

Furthermore, the review articles were clustered according to the data applications criteria such as medicine, finance, text, government, topology, galaxy, and education for the analysis purpose. On the other hand, those articles were again separated according to the algorithms used. The algorithms used in those review articles mostly found that clustering algorithm and decision tree are a better solution. Later on, the analyzed outputs allowed finding out the percentage of the use of clustering algorithm in the recent data application areas in Big data.

The total number of data application areas and algorithms used to review the chosen from were categorized into different clusters. Then, the percentage of data

application areas are then used for data wrangling purposes and calculated. Similarly, the percentage of the algorithms used in the review articles for data wrangling purposes was calculated. These calculations are further used for further analysis of the most frequently used algorithm by major data application areas.

2.4 Comparison of Research Studies

See Table 1.

3 Results

The data gathered from different 30 peer-reviewed scientific publications (1996–2016) (Table 1) were analyzed with respect to the data areas and algorithms. The graphical representations of the analysis of the data applications and approaches, and algorithms are shown in the Figs. 2, 3 and 4.

3.1 Analysis of Data Application Used for Data Wrangling

The data applications are clustered according to the major categories such as medicine, finance, text, government, topology, galaxy and education. Then, the calculation for the analysis purposes was also carried out using the following formula:

*Percentage of the Data Application areas = (Number of studies that used the Data Application area * 100/Total studies have been used for this research) %,* for an example: Percentage of Medical Data = (8 * 100/30) % = 26.67%.

Similarly, the percentages for other data applications were calculated and the results are shown in the Fig. 2.

The analysis of the data applications areas shows that most of the research is done with the use of medical data, i.e., 26.67% of the review articles used for this analysis are medical data. Secondly, textual data with 23.33% of the review articles are also used for conducting the research. Financial data with 20% is used after medical and textual data. Topological data and governmental data were then used 13.33% and 10% respectively. Other data applications only cover 6.67% of the total applications for this analysis. Some educational data and Galaxy data were used for some minor contribution.

Table 1 Description of algorithms: algorithms used in published studies

	Author	Applications	Data size	Algorithms	Techniques	Benefits
1	Gulwani [7]	Textual data	Sufficient sets	Search algorithm	(1) Search algorithm (2) Dealing with ambiguity (3) Inductive specification (4) Domain specific language	(1) Enable by-example interaction for any relevant feature in any software (2) Define the next generation of programming experience
2	Heer et al. [8]	Textual data	Big data	Decision tree	(1) Guide predictive methods, and decide on a desired result (2) Visualization, interaction, prediction, disambiguate, and compile	Significant performance benefits (at least 2x faster on median) over traditional approaches
3	Terrizzano et al. [9]	Financial data	75 data sets	Schema mapping	(1) Procuring data a. Vetting data b. Obtaining data c. Describing data (2) Grooming data (3) Provisioning data (4) Preserving data	(1) Gain prominence in an enterprise's core operational business processes (2) Automatic data interpretation
4	Endel and Piringer [10]	Textual data	Big data	Clustering	(1) Raw data are arranged and cleaned (2) Database stores the information for analytical purposes (3) Known data are organized for visualization (4) Extra information from other sources is merged to exported file	(1) Data quality (2) Merging and linking (3) Reproducibility and documentation (4) Error tolerance (5) Transformation and editing
5	Savinov [11]	Financial data	Large class of users	Decision tree	(1) Multiple data sources (2) No dimension table (3) No measure attributes (4) Importing data (5) Defining links	(1) Efficient in the visualization (2) Recommended for schema mapping

(continued)

Table 1 (continued)

	Author	Applications	Data size	Algorithms	Techniques	Benefits
					(6) Defining a new table (7) Defining new columns	
6	Parisot et al. [12]	Governmental data	Lot of data about products and users retrieved from EBay or Amazon	Decision tree	(1) Data retrieving (2) Inconsistent data and sampling (3) Features selection (4) Clustering (5) Computation of predictive models (6) Visual analytics approach	Efficient and useful decision support
7	Blankenberg et al. [13]	Galaxy data	Galaxy's built-in data and additional downloaded data	Clustering	(1) User's choices and parameters are recorded (2) Data manager instruments evacuate the specialized weights of guaranteeing the reproducibility and provenance of implicit reference information (3) Inspect the after-effects of individual data manager executions and view the present condition of Galaxy's implicit information registries (4) Data manager system parses the yield for new information table sections and values. These qualities are empowered progressively	(1) Data manager system gives a way to deal with guaranteeing reproducibility and provenance following of reference information (2) Efficient utilization of data manager devices over fluctuating server situations and on the cloud
8	Ceusters et al. [14]	Medical data	390 people	Ontological realism	(1) Cross-checking the study set (2) Annotating the datasets (3) Building an executable template (4) Selecting from appropriate realism-based ontology (5) Implementing an algorithm (6) Generating statistics for results	75% accuracy

(continued)

Table 1 (continued)

	Author	Applications	Data size	Algorithms	Techniques	Benefits
9	Kandel et al. [15]	Financial data	35 people	Decision tree	(1) Discovery of the data (2) Field definitions (3) Ingesting semi-structured data (4) Data integration (5) Advanced aggregation and filtering (6) Profiling	As the data is diverse and large, the quality of analysis and speed of the wrangling process is better.
10	Grimes et al. [16]	Medical data	41 data sets	Clustering	(1) Calculation of distance matrices (2) Spearman and euclidean distance (3) t-SNE (4) Clusters (5) Visualization	(1) Increases the resolution of clusters (2) Clusters with meaning
11	Kandel et al. [17]	Textual data	Big data	Visualization	(1) Identify potential problems (2) Create columns and rows (3) Merge data in visualization tool (4) Make transformations and finalize data	(1) Visualization helped in detecting potential problems (2) Transformation helped in obtaining clean and final data
12	Kandel et al. [18]	Governmental data	Big data	Sorting algorithm	(1) Mapping transformations of input and output (2) Incorporate data from outer tables (3) Manipulate table structures and schemas (4) Generate values	(1) Data cleaning is faster than excel (2) Rapid navigation to the desired transformation
13	Zengin et al. [19]	Educational data	531 people	Decision tree	(1) A five-point likert type material (2) Network dependencies, analysis of variance, clusters, t-test and decision tree	(1) The data with high quality can be categorized and predicted (2) Visual presentation of data and the findings can be easily understood

(continued)

Table 1 (continued)

	Author	Applications	Data size	Algorithms	Techniques	Benefits
14	Guo et al. [20]	Governmental data	Big data	Clustering	(1) Transform supports the cleaning and reformatting task (2) Wrangler generates executable code using Python and JavaScript (3) Exploration and learning (4) Reduced sets of formally defined operator	Helpful when embedded with reactive suggestion
15	Espejo et al. [21]	Medical data	Big data	Decision tree	(1) Extracting decision trees (2) Learning rule-based systems (3) Learning discriminant functions (4) Other representations	The quality of classification is enhanced
16	Wu et al. [22]	Textual data	Big data	Decision tree	(1) Lattice concepts	Decision can be explained
17	Tasdemir and Merenyi, [23]	Topological data	Big data	Clustering	(1) Connecting matrix and preserving topology	Helps to make decision if data can be done or SOM is needed
18	Oehmen and Nieploch [24]	Medical data	Big data	Decision tree	(1) Global arrays are used in the database (2) The processes are grouped (3) The results are reported	Helps in solving scalability problems as well as memory limitations
19	Datta et al. [25]	Topological data	Big data	Decision tree	(1) Exact local algorithms (2) Majority voting (3) Frequent-item-set mining (4) Monitoring a K -means clustering (5) Approximate local algorithms	Good performance in accuracy and communication cost
20	Cilibrasi and Vitanyi [26]	Textual data	Big data	Clustering	(1) Finding similarity distance (2) Normal compression (3) Background in Kolmogorov complexity (4) Finding compression distance	NCD minimizes the similarity distances with respect to the reference compressor

(continued)

Table 1 (continued)

	Author	Applications	Data size	Algorithms	Techniques	Benefits
					(5) Clustering using the quartet method	
21	Saraiya et al. [27]	Medical data	Big data	Clustering	(1) Tree view (2) Time searcher (3) Hierarchical clustering explorer (4) Functional genomics	Produces a more practical and effective evaluation
22	Au et al. [28]	Medical data	Big data	Clustering	(1) Initialization (2) Assignment of attributes (3) Computation of mode (4) Termination	Helps in finding good configurations and selecting significant gene and getting good classification results
23	Figueiredo et al. [29]	Financial data	Big data	Decision tree	(1) Data pre-processing (2) Definition of number of classes (3) Consumer characterization	It is able to handle large datasets, robust and a practical application
24	Jiang et al. [30]	Medical data	10,000 datasets	Clustering	(1) Gene clustering (2) Class validation	The result of clustering is biologically meaningful
25	Pedrycz and Bargiela [31]	Financial data	Big data	Clustering	(1) Design (2) Interpretation and validation of granular clustering	Emphasizes the transparency of results obtained
26	Seo and Shneiderman [32]	Medical data	6,000 genes	Clustering	(1) Overview in a limited screen space (2) Dynamic query controls (3) Coordinated displays (4) Cluster comparisons	Big achievement in the understanding of molecular biology
27	Rauber et al. [33]	Topological data	Big data	Clustering	(1) Use of growth process (2) Analysis of GHSOM characteristics	Training time is decreased and uncovers hierarchical structure of data Easier to overview different clusters
28	Alahakoon et al. [34]	Topological data	Big data	Clustering	(1) Use of GCS (2) Use of IGG	Used on data sets whose information is not available to find cluster presence

(continued)

Table 1 (continued)

	Author	Applications	Data size	Algorithms	Techniques	Benefits
29	Karypis et al. [35]	Textual data	Big data	Clustering	(1) Modelling the data (2) Modelling cluster similarity (3) Relative interconnectivity and closeness (4) Results and comparison	Effective in capturing relative closeness and interconnectivity
30	Keim and Kreigel [36]	Financial data	10,000–100,000 data sets	Clustering	(1) The methods use pixels, geometry, icons, hierarchy and graphs	Data mining tasks, grouping similar data, retrieving similarities is easier using this technique

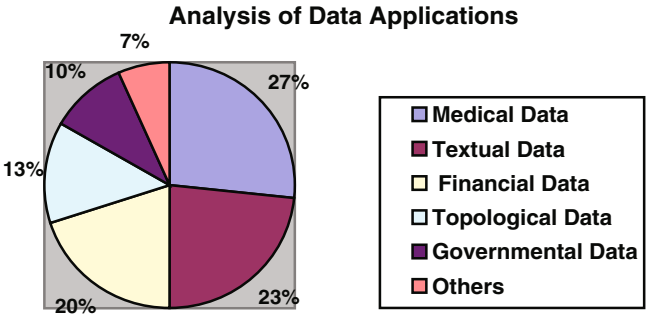


Fig. 2 The analysis of data applications used for data wrangling and their percentage

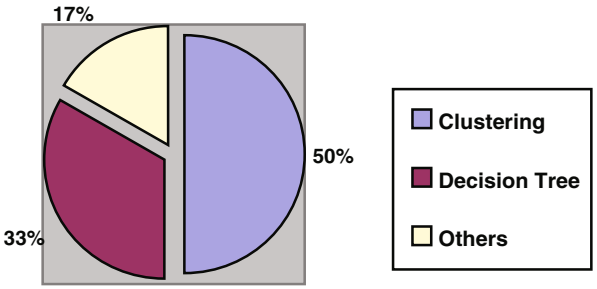


Fig. 3 Shows the analysis of algorithms that have been used for data wrangling

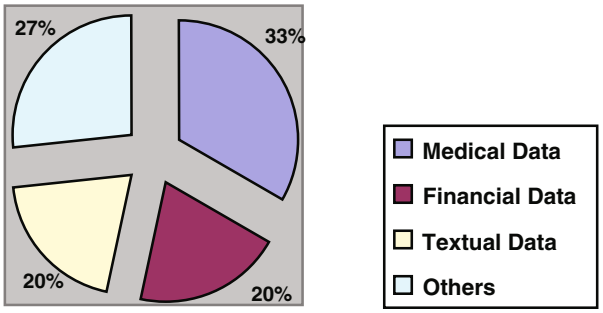


Fig. 4 The analysis of data applications using clustering algorithms

3.2 Analysis of Algorithms Used for Data Wrangling

The algorithms used in the research articles were clustered according to the similarity of their methods. The calculation for the analysis purpose was carried out by

using the following formula: *Percentage of the Algorithm used = (Number of* articles *using algorithm*100/Total articles have been used for the research) %,* for an example: Percentage of the Clustering Algorithm = (15 * 100/30) % = 50%.

Similarly, the percentages for other algorithms are calculated and the results are also shown in the Fig. 3.

After the analysis of algorithms used by the review articles, this study has found that, the clustering algorithm used by the articles is 50%. This indicates that most frequently used algorithm in the data wrangling is clustering algorithm as 50% is not a small number. On the other hand, 33.33% of the research articles used Decision Tree algorithm approach. Some other algorithms such as Ontological realism, Sorting, Search, Visualization and Schema mapping algorithms were also used in some of the review articles.

3.3 Analysis of the Mostly used Algorithm by the Mostly used Data Applications

Firstly, the number of review articles that used clustering algorithms were separated according to the data application areas. Most of the review articles have used medical data implemented the clustering algorithm for data wrangling purposes in most of their review articles, i.e., five medical data applications out of eight is some medical data applications that has used clustering algorithm whereas the remaining three review articles have used a decision tree algorithm. The highest frequency of the data applications have used research articles that were calculated according to the number of the clustering algorithms used in the review articles are for data wrangling purposes. The calculation for the analysis purpose was carried out by using the following formula:

*The percentage of mostly used data application in clustering algorithm = (Maximum of data application used * 100/Total number of clustering algorithm used).* For example: Percentage of Medical data has used Clustering algorithm = (5 * 100/15) % = 33.33%.

Similarly, the percentages of the mostly used data application for mostly used algorithm was calculated and the results were shown in Fig. 4.

Five out of fifteen published articles have used clustering algorithms. These articles are mainly medical articles that have implemented data in order to cluster algorithms (33.33%) and the data applications have used clustering algorithm. This shows that, medical data has used clustering algorithm most frequently out of the rest. It can also be concluded that while conducting the medical data analysis, clustering algorithm can be the first choice for the data mining purposes.

4 Discussion

When the data sets were clustered according to the data applications of the data wrangling processes, it was found that clustering algorithm is the most commonly used algorithm wrangling processes. This research concludes on a note, stating that, the medical data is the most applicable method used for data wrangling processes, as there is great scope in data wrangling in medicinal science. The related trends of the data applications identified the patterns of their appearances in groups and they are recognized accordingly. In order to find the different aspects of the research articles, the patterns of the data were used. The data gathered from articles were then classified into different clusters of data applications as well as algorithms which led the research forward.

After clustering, the articles were analyzed in a group to find the similar and dissimilar properties of the applications and algorithms. The analysis formed the representation of similar and different trends of the data sets. The interdependency between the data sets was then recognized. When one of the data applications occurred, we analyzed whether a specific data application is used frequently in an algorithm or not. This helped us to find out the conclusion about the relationship between the data applications and algorithms used in the research articles.

This study observed that medical data application for clustering algorithm is one of the best options. There are some studies which use clustering algorithm for medical data wrangling [16, 27, 28, 30, 32]. On the other hand, some studies [21, 24] used Decision Tree algorithms instead of clustering algorithms for medical data applications. Furthermore, exploring different techniques for Big Data databases [37, 38], security [39, 40] and prediction and pattern analysis [41] could be an interesting path to explore in the future.

5 Conclusion

This study, performed a content analysis which comprised of data extracted from 30 peer-reviewed scientific publications (1996–2016) describing the applied algorithm models for data wrangling in Big Data area. The Clustering algorithm which is sensitive data application area was the most applicable area for using data wrangling. Fifty percent of the research articles have used clustering algorithms such as clustering algorithm divides the datasets into different clusters with much more similarities and dissimilarities between the clusters. The concluded analysis shows that the data application areas and medical data are found to be used in most of the research articles. Furthermore, it has been found that the medical sector, being a sensitive data application area, was the most applicable area for using data wrangling. It has been concluded that clustering algorithm is suggested to be used by medical data applications for the data wrangling purpose. Since all of the small sample sizes, as there could be less accuracy for the results obtained. This research

can be extended by collecting more data in order to get the highest accuracy in the results obtained.

Author Contribution C. Bashyal and M.N. Halgamuge conceived the study idea and developed the analysis plan. C. Bashyal analyzed the data and wrote the initial paper. M.N. Halgamuge helped to prepare the figures and tables, and finalizing the manuscript. All authors read the manuscript.

References

1. Vlahogianni, E.I., Karlaftis, M.G., Stathopoulos, A.: An extreme value based neural clustering approach for identifying traffic states. Intell. Transp. Syst., 320–325 (2005)
2. Jin, X., Wah, B., Cheng, X., Wang, Y.: Significance and challenges of big data research. Big Data Res. **2**(2), 59–64 (2015)
3. Sarikaya, A., Correli, M., Dinis, J., O'Connor, D., Gleicher, M.: Visualizing co-occurrence of events in populations of viral genome sequences. Comput. Graph. Forum **35**(3), 151–160 (2016)
4. Meena, K., Lawrance, R.: Semantic similarity based assessment of descriptive type answers. In: International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE), pp. 1–7 (2016)
5. Medhane, D.V., Sangaiah, A.K.: ESCAPE: effective scalable clustering ap-proach for parallel execution of continuous position-based queries in position monitoring applications. IEEE Trans. Sustain. Comput. (2017). https://doi.org/10.1109/TSUSC.2017.2690378
6. Padua, L., Schulze, H., Matković, K., Delrieux, C.: Interactive exploration of parameter space in data mining: Comprehending the predictive quality of large decision tree collections. Comput. Graphics **41**, 99–113 (2014)
7. Gulwani, S.: Programming by Examples (and its applications in Data Wrangling) (2016)
8. Heer, J., Hellerstein, J.M., Kandel, S.: Predictive interaction for data transformation (2015)
9. Terrizzano, I., Schwarz, P., Roth, M., Colino, J.E.: Data wrangling: the challenging journey from the wild to the lake (2015)
10. Endel, F., Piringer, H.: Data wrangling: making data useful again. IFAC-PapersOnLine **48**(1), 111–112 (2015)
11. Savinov, A.: ConceptMix—self-service analytical data integration based on the concept-oriented model. In: Proceedings of 3rd International Conference on Data Management Technologies and Applications (2014)
12. Parisot, O., Vierke, G., Tamisier, T., Didry, Y., Rieder, H.: Visual analytics for supporting manufacturers and distributors in online sales (2014)
13. Blankenberg, D., Johnson, J., Taylor, J., Nekrutenko, A.: Wrangling galaxy's reference data. Bioinformatics **30**(13), 1917–1919 (2014)
14. Ceusters, W., Hsu, C.Y., Smith, B.: Clinical data wrangling using ontological realism and referent tracking (2014)
15. Kandel, S., Paepcke, A., Hellerstein, J., Heer, J.: Enterprise data analysis and visualization: an interview study. IEEE Trans. Vis. Comput. Graphics **18**(12), 2917–2926 (2012)
16. Grimes, M., Lee, W., van der Maaten, L., Shannon, P.: Wrangling phosphoproteomic data to elucidate cancer signaling pathways. PLoS ONE **8**(1), e52884 (2013)
17. Kandel, S., Heer, J., Plaisant, C., Kennedy, J., van Ham, F., Riche, N., Weaver, C., Lee, B., Brodbeck, D., Buono, P.: Research directions in data wrangling: Visualizations and transformations for usable and credible data. Inf. Vis. **10**(4), 271–288 (2011)
18. Kandel, S., Paepcke, A., Hellerstein, J., Heer, J.: Wrangler: interactive visual specification of data transformation scripts (2011)

19. Zengin, K., Esgi, N., Erginer, E., Aksoy, M.: A sample study on applying data mining research techniques in educational science: Developing a more meaning of data. Proc. Soc. Behav. Sci. **15**, 4028–4032 (2011)
20. Guo, P.J., Kandel, S., Hellerstein, J.M., Heer, J.: Proactive wrangling: mixed-initiative end-user programming of data transformation scripts (2011)
21. Espejo, P.G., Ventura, S., Herrera, F.: A survey on the application of genetic programming to classification (2010)
22. Wu, W., Leung, Y., Mi, J.: Granular computing and knowledge reduction in formal contexts. IEEE Trans. Knowl. Data Eng. **21**(10), 1461–1474 (2009)
23. Tasdemir, K., Merenyi, E.: Exploiting data topology in visualization and clustering of self-organizing maps. IEEE Trans. Neural Netw. **20**(4), 549–562 (2009)
24. Oehmen, C., Nieplocha, J.: ScalaBLAST: a scalable implementation of BLAST for high-performance data-intensive bioinformatics analysis. IEEE Trans. Parallel Distrib. Syst. **17**(8), 740–749 (2006)
25. Datta, S., Bhaduri, K., Giannella, C., Wolff, R., Kargupta, H.: Distributed data mining in peer-to-peer networks. IEEE Int. Comput. **10**(4), 18–26 (2006)
26. Cilibrasi, R., Vitanyi, P.: Clustering by compression. IEEE Trans. Inf. Theor. **51**(4), 1523–1545 (2005)
27. Saraiya, P., North, C., Duca, K.: An insight-based methodology for evaluating bioinformatics visualizations. IEEE Trans. Vis. Comput. Graphics **11**(4), 443–456 (2005)
28. Au, W., Chan, K., Wong, A., Wang, Y.: Attribute clustering for grouping, selection, and classification of gene expression data. IEEE/ACM Trans. Comput. Biol. Bioinform. **2**(2), 83–101 (2005)
29. Figueiredo, V., Rodrigues, F., Vale, Z., Gouveia, J.: An electric energy consumer characterization framework based on data mining techniques. IEEE Trans. Power Syst. **20** (2), 596–602 (2005)
30. Jiang, D., Tang, C., Zhang, A.: Cluster analysis for gene expression data: a survey. IEEE Trans. Knowl. Data Eng. **16**(11), 1370–1386 (2004)
31. Pedrycz, W., Bargiela, A.: Granular clustering: a granular signature of data. IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) **32**(2), 212–224 (2002)
32. Seo, J., Shneiderman, B.: Interactively exploring hierarchical clustering results [gene identification]. Computer **35**(7), 80–86 (2002)
33. Rauber, A., Merkl, D., Dittenbach, M.: The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. IEEE Trans. Neural Netw. **13**(6), 1331–1341 (2002)
34. Alahakoon, D., Halgamuge, S., Srinivasan, B.: Dynamic self-organizing maps with controlled growth for knowledge discovery. IEEE Trans. Neural Netw. **11**(3), 601–614 (2000)
35. Karypis, G., Han, E., Kumar, V.: Chameleon: hierarchical clustering using dynamic modelling. Computer **32**(8), 68–75 (1999)
36. Keim, D., Kriegel, H.: Visualization techniques for mining large databases: a comparison. IEEE Trans. Knowl. Data Eng. **8**(6), 923–938 (1996)
37. Vargas, V., Syed, A., Mohammad, A., Halgamuge, M.N.: Pentaho and Jaspersoft: a comparative study of business intelligence open source tools processing big data to evaluate performances. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **7**(10), 20–29 (2016)
38. Kalid, S., Syed, A., Mohammad, A., Halgamuge, M. N.: Big-Data NoSQL databases: comparison and analysis of "Big-Table", "DynamoDB", and "Cassandra". In: IEEE 2nd International Conference on Big Data Analysis (ICBDA 2017), pp 89–93, Beijing, China, 10–12 March (2017)
39. Kaur, K., Syed, A., Mohammad, A., Halgamuge, M. N.: Review: an evaluation of major threats in cloud computing associated with big data. In: IEEE 2nd International Conference on Big Data Analysis (ICBDA 2017), pp. 368–372, Beijing, China, 10–12 March (2017)
40. Pham, D.V., Syed, A., Mohammad, A., Halgamuge, M.N.: Threat analysis of portable hack tools from usb storage devices and protection solutions. In: International Conference on

Information and Emerging Technologies (ICIET 2010), pp. 1–5, Karachi, Pakistan, 14–16 June (2010)
41. Gupta, A., Mohammad, A., Syed, A., Halgamuge, M.N.: A comparative study of classification algorithms using data mining: crime and accidents in denver city the USA. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **7**(7), 374–381 (2016)

An Innovation Model for Smart Traffic Management System Using Internet of Things (IoT)

Amardeep Das, Prasant Dash and Brojo Kishore Mishra

Abstract Traffic management is the focus area for most urban dwellers and planners. Congestion is the most important major obstacle that has been seen in many countries including India. Countries To avoid this obstacle means how to manage the traffic smoothly. Traffic congestion mainly focuses on the signals failure, reduced law enforcement and improper traffic management. Existing foundation can't be extended increasingly and subsequently the main choice accessible is to enhance the administration of the traffic. Traffic congestion is not a good sign for our country as well as it creates a negative impact starting from economy to the leaving standard. Consequently the opportunity has already come and gone to viably deal with the traffic congestion. Many methods are designed to manage the traffic and minimize the congestion. Out of all the techniques, infrared sensor, inductive loop detection, video data analysis, wireless sensor network, etc. are used to somewhat solve the congestion in the traffic and to manage the traffic smartly. But in the above said methods having some demerits like much time to take for installation, maintenance cost is very high. Actually, our objective is to develop a new technology or method; that will solve the above problems and produce better result within a stipulated time. To overcome the challenges, a new method arises called as Radio Frequency Identification (RFID). By this innovation, it will require less time for establishment with lesser expenses when contrasted with different strategies for traffic blockage administration. Utilization of this new innovation will prompt lessened traffic jam. It refers to small electronic devices that consist of a small chip and an antenna. It plays a vital role in intelligent traffic management system technologies to sense the presence and movement of tagged objects; the traffic will be monitored and managed automatically using this system.

A. Das (✉) · P. Dash · B. K. Mishra
C. V. Raman College of Engineering, Bidyanagar, Janla, Khordha,
Bhubaneswar 752054, Odisha, India
e-mail: amardeepcvrp@gmail.com

P. Dash
e-mail: prasant.oitburla@gmail.com

B. K. Mishra
e-mail: brojokishoremishra@gmail.com

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7_15

The data collected from this system will be sent to a centralized server for further analysis. Moreover, the traffic signal lights at crossing points are based on traffic density of roads intersecting at that point. This chapter discusses about an architecture which integrates Internet of things and other moving components like data management techniques to create a model for traffic management and monitoring. The model comprises of a single platform where this platform will communicate with the large number of decentralized heterogeneous components.

Keywords Traffic management · RFID · GSM · Congestion
IOT

1 Introduction

The smart traffic management system, a speculative informational, clever, effective and mingled new transport framework that works to expand the productivity of transport foundation by new age data innovation, data communication transfer technology, electronic control technology and computer processing technology. The creative energy of smart city is inconceivable without utilizing smart traffic management system. The smooth and fault free activity development is the key of smart city [1]. Traffic congestion is the result of poor infrastructures, low speed and violation of traffic rules.

Although many traffic management and control strategies utilized on interurban systems are legitimate, with some traffic administration designs close key conurbations coordinating the interface to urban systems, [2] urban traffic administration mainly includes traffic signal administration and co-ordination, need and enhancements to open transport and a more exhaustive portability administration approach, given specifically that a considerably more noteworthy extent of journeys in urban ranges are normal trips (e.g. driving). These days, the quantity of vehicles has expanded exponentially, however the bedrock limits of streets and transportation frameworks have not created in a comparable approach to productively adapt to the quantity of vehicles going on them. Because of this, street sticking and activity connected contamination have expanded with the related unfriendly societal and budgetary impact on various markets around the world [3]. A static control framework may square crisis vehicles because of roads turned parking system. WSNs have increased expanding consideration in movement location and maintaining a strategic distance from street blockage. WSNs are exceptionally popular because of their quicker exchange of data, simple establishment, less support, smallness and for being more affordable contrasted with other system choices. There has been huge research on Traffic Management Systems utilizing WSNs to maintain a strategic distance from clog, guarantee need for crisis vehicles and cut the Average Waiting Time (AWT) of vehicles at crossing points [4].

The purpose of Smart Traffic Management is mainly improvised for looking after the Set off data of a region to manage the Traffic along that area and implement

various useful technologies which are been required by various persons like vehicle owners, pedestrians, police officers etc. Mainly the purpose of Smart traffic management system is to give the details which can be used and they can be implemented in their daily life. The problems which have been occurred at their presence can be solved by this Smart Traffic.

1.1 Background and Motivation

Traffic management is the arranging, observing and control or impacting of activity. It expects to: boost the adequacy of the utilization of existing foundation; to guarantee dependable and safe operation of transport; to address ecological objectives; and guarantee reasonable assignment of framework space (street space, rail openings, and so forth.) among contending clients. Traffic monitoring system developed so far are primarily focused on structured traffic that is not the case in a country like ours. Development of overhead structures can't be considered as a viable option since it increases the cost substantially, the same goes for under the road construction. Necessity to analyze traffic pattern, near real time reporting and simultaneous conduction of smooth traffic flow. Our country (India) is the second largest population of world; according to that vehicles are increased day to day life. Here, the questions arise! how to avoid the congestion in the road; that means traffic management.

Traffic management has since quite a while ago existed in some frame, from the beginning of railroad flagging or movement lights on city lanes, yet the improvement and execution of modern coordinated applications in light of Intelligent Transport Systems (ITS) has developed apace lately, because of effective research and technological advances. This has been pushed by acknowledgment of the need to oversee transport organizes all the more adequately keeping in mind the end goal to boost the utilization of existing framework, give a solid support of the end client and increment security, while lessening negative natural impacts. To manage the traffic in the road; we are doing very good job by manually that means one or more traffic police are appointed there to serve the people from the congestion in road. But, it is difficult to monitor and manage. That's why, the new concept Internet of Things (IOT) is implemented to control the traffic.

1.2 Internet of Things (IOT)

In Today's Scenario it is certainty that, number of vehicles is expanding exponentially, however infrastructure for transportation isn't adequate to fulfil their necessities. Because of this, valuable time is being lost each day. This additionally prompts enormous financial issues [5]. Principle issue happens when this traffic

clog costs life of somebody. It ought not astonish that activity blockage influences all crisis vehicles, which can be excessively unsafe to influenced individuals.

1.2.1 What Is IOT?

The gigantic system of gadgets associated with the Internet, including advanced mobile phones and tablets and practically anything with a sensor on it—cars, machines in production plants, jet engines, oil drills, wearable devices, and more. These "things" collect and exchange data. All the non living organs are connected with each other for sharing the information in a global network.

IoT and the machine-to-machine (M2M) innovation behind it—are bringing a sort of "super perceivability" to almost every industry. Imagine utilities and telcos that can predict and prevent service outages, airlines that can remotely monitor and optimise plane performance, and healthcare organisations that can base treatment on real-time genome analysis. The business possibilities are endless. From the Fig. 1, it is observed that how things are connected and sharing the data among themselves. The numbers of connections are increased rapidly from time to time; shown in Table 1.

The quick growth of information technology (IT) has presented a hyper connected society where objects are linked to mobile devices and the Internet and communicate with each other [1]. In this century, we would like to be connected with anything anytime and anywhere, which is already occurs in various places through over the world. The core part of this hyper connected society is IoT, that is

Fig. 1 The diagram of internet of things

Table 1 Number of connected devices from time to time

Year	Number of connected devices
1990	0.3 million
1999	90.0 million
2010	5.0 billion
2013	9.0 billion
2025	1.0 trillion

also treated as Machine to Machine (M2 M) communication or Internet of Everything (IoE).

$$\text{IOT} = \text{Physical Object} + \text{Controller, Sensor, Actuators} + \text{Internet}$$

1.2.2 Why IOT Require?

Internet of Things (IoT) makes our reality as conceivable as associated together. These days we practically have web foundation wherever and we can utilize it at whatever point. Implanted figuring gadgets would be presented to web impact. Normal cases for inserted processing gadgets are MP3 players, MRI, traffic lights, microwave stoves, washing machines and dishwashers; GPS even heart observing inserts or biochip and so on.

IoT tries to set up cutting edge network (with the guide of web) among these said gadget or frameworks or administrations to little by little make robotization in all ares. Picture that everything is associated with accumulate and all data would be connected to each other over standard and diverse convention space and applications. More or less IoT needs to associate every single potential question connect each other on the web to give secure, comfort life for human [6].

Recent investigates appear by 2020 we have more than 20 billion gadgets which utilizes IoT. IoT does that as a result of controlling on gadget and lower cost on radio. Yet, these huge fields make difficulties, for example, lacking IP address, creating perfect and valuable convention and condition. The accompanying reasons are to require IOT. They are:

(a) **Get more out of your current IT resources**:
Begin with your current IT resources and expand upon them. Include a couple of new gadgets, associate them to the cloud, and empower them to converse with each other, to the representatives and to clients. Change your business by using the information those gadgets create with business knowledge apparatuses to have further understanding into what the clients and workers need.

(b) **Become more proficient**:
Associating gadgets and frameworks can enable you to shave minutes from a client's login procedure, hours from restocking stock, or days from routine framework updates and improvements. At the point when information streams flawlessly amongst gadgets and through the cloud, you can access and utilize it

more effectively than any time in recent memory. That implies investing less energy pulling reports, and additional time making new services and products based on your new insights.

(c) **Find better approaches to enchant your clients**:
From the slightest utilized fitting room in the store to the keyboard that drive the most grounded coupon deals, each bit of information is a sign to the items and encounters your clients are looking for. Picture rising examples and foresee conduct to expect patterns and give your clients what they need, before they even know they need it.

(d) **Increase agility**:
Information bits of knowledge can enable you to react all the more rapidly to rivalry, inventory network changes, client request and changing economic situations. Gathering and breaking down information gives you speedy knowledge into patterns, so you can change your creation movement, tweak your upkeep timetable or find more affordable materials.

1.2.3 Applications of IOT

Though, IOT gives the solution easily for the society; and also it is very easy to implement. The most and famous application area where IOT is used. They are:

- **Environmental monitoring**
This utilization of the IoT regularly utilize sensors to aid ecological assurance by observing environment or soil quality and can incorporate zones like checking the development of untamed life and their living spaces. Advancement of asset obliged gadgets associated with the Internet additionally implies that different areas like earthquake or tsunami early-detection frameworks can likewise be utilized by crisis administrations to give more viable guide. IoT gadgets in this application ordinarily traverse an extensive geographic zone and can likewise be portable.
- **Infrastructure management**
The key application of the IoT is railway tracks, bridges, on-and offshore-wind farms are the checking and controlling operations of urban and rural infrastructure (building, bridges and dam). IoT system can be utilized for checking any events or changes in basic conditions that would negotiation be able to safety and increment risk. It is used meant for fix and maintenance actions in an efficient manner, by distributing jobs among service providers and users of these services. These devices controls required infrastructure like bridges to give access to ships. Handling of IoT devices used for monitoring and operating infrastructure is possible to improve event management and emergency reply management, quality of service, up-times and decrease costs of process in each infrastructure related areas. Areas such as waste management can profit from computerization and optimization that could be brought in by the IoT (Fig. 2).

Fig. 2 Infrastructure management

- **Energy management**
  Coordination of detecting and activation frameworks, associated with the Internet, is probably going to improve the total energy utilization. This is normal for IoT gadgets to incorporated with all types of energy expending gadgets (switches, electrical plugs, bulbs, TVs, etc.) and have the capacity to speak with the utility supply organization keeping in mind the end goal is to adjust power generation and its use. Such gadgets would likewise offer the open door for clients to remotely control their gadgets, or midway oversee them by means of a cloud based interface, and empower advanced capacities like planning (e.g., remotely powering on or off heating systems, controlling ovens, changing lighting conditions etc.). Other than locally established power management, the IoT is particularly significant to the Smart Grid since it gives frameworks to accumulate and follow up on power and power-related data in a mechanized manner with the objective to enhance the effectiveness, reliability, financial matters, and supportability of the creation and appropriation of power. Using advanced metering infrastructure (AMI) gadgets associated with the Internet backbone, electric utilities can gather information from end-client associations, as well as oversee other distribution automatic gadgets like transformers and reclosers.
- Industrial Applications
  Engineering Applications look at the idea of thing remembering the ultimate objective to constant enhancing to have an average publicizing, for instance, who are most interested to which thing and how this thing can find exhibiting with which minor changes; appeared in Fig. 3.

Fig. 3 Industrial application

- **Medical and Healthcare Systems**
  Medicinal services Systems shows signs of improvement understanding state by checking and controlling their heart rate or pulse or even for their diet.. Smart tablet shows the number of dosages and its ingredient needed for a particular patient.

  Now a day's some hospitals implements new bed allocation system that can sense at what time the bed are occupied and when a patient is trying to get up. It can also regulate itself to guarantee suitable pressure and support is applied to the patient without the physical interaction of nurses shown in Fig. 4.
- **Building and Home Automation**
  In all types of home appliances that have the potential to monitor and remote control such as, ventilation, security lock lightening, heating, air condition, telephone system, television to make a comfort, secure, with low energy consumption (Fig. 5).
- **Transport Systems**
  Transportation Systems helps in automatic configuration in traffic lights, smart parking, and traffic camera to detect which road has heavy traffic and offer automatically less crowd road [7–9], or smart camera which fine driver in high speed. Figure 6 shows the smart parking system.

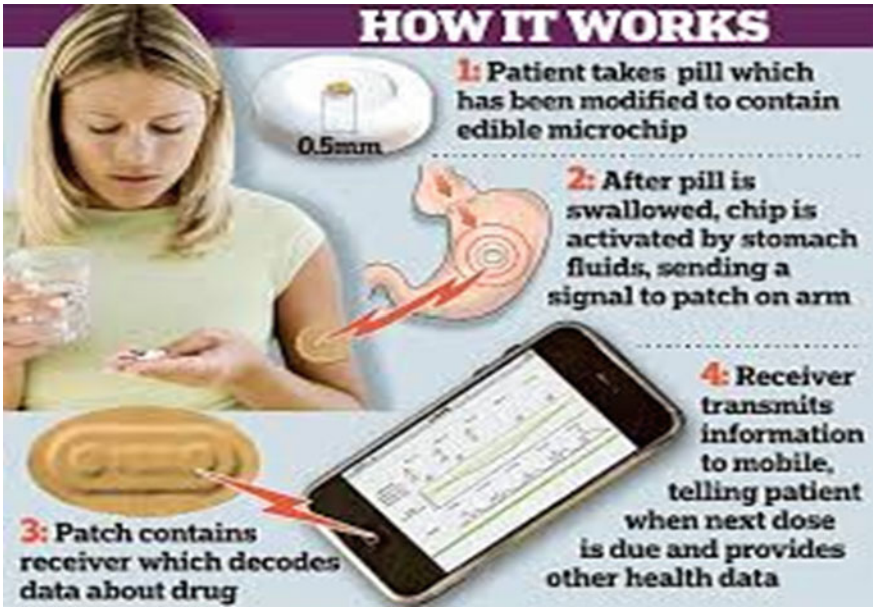


Fig. 4 Medical and healthcare systems

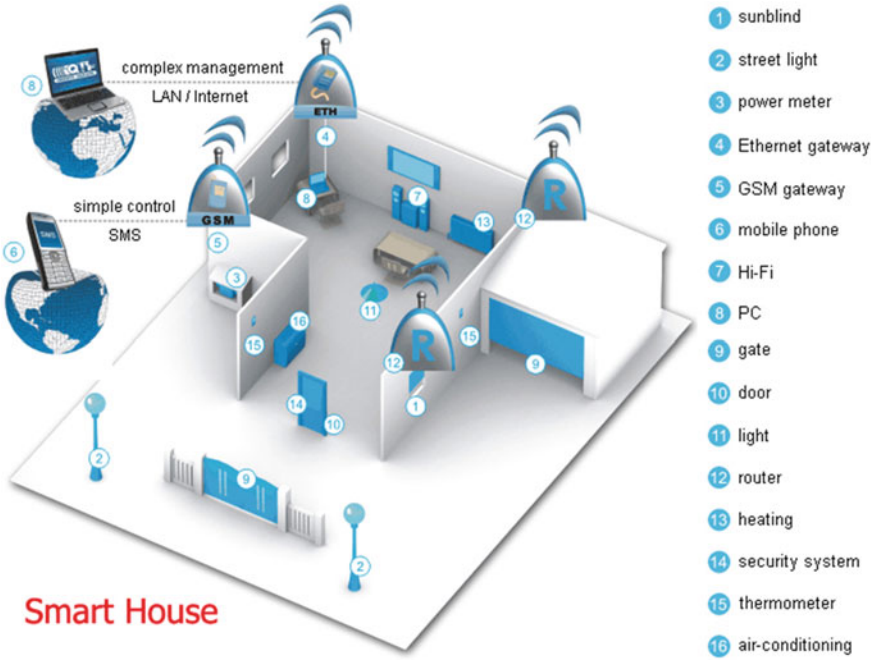


Fig. 5 Building and home automation

Fig. 6 Transport system

1.2.4 Advantages Towards New Era

In 2010 as a result of rapid proliferation of smart phones and tablets the number of connected devices exceeded the number of people (the number of connected objects increased to 12.5 billion, while the entire population was 6.8 billion people). It introduced the new generation of IoT and removes the concept of internet-of-people.

At present moment we are observing fast development of the Internet of things. Now more and more devices are being connected to network, as we have a strong desire to do everything around us "smart", up to road signs. According to scientific forecasts, by 2020 the number of connected devices will have been 60 billion! Thus, the Internet of things is a trend of our time.

2 Traffic Management System

The Traffic Management System is a key component of Intelligent Transportation System (ITS) domain. The Traffic Management System view is a top-down management perception that integrates technology primarily to improve the flow of vehicle traffic [10] and improve safety.

Our intelligent traffic specialist solution for road traffic control System offers the ability to acquire real-time traffic information. Traffic expert enables operators to perform real-time data analysis on the information gathered. Traffic management measures are aimed at improving the safety and flow of traffic utilizing traffic capacity more effectively [2, 11].

The major goals of the ITS domain is:

- Increase transportation system efficiency
- Mobility Enhancement
- Safety Improvement
- Reduce fuel consumption and environmental cost
- Increase economic productivity
- Create an environment for an ITS market.

Our objective is all the devices are connected and sharing the information smoothly over globally. Many methods are designed to manage the traffic and minimize the congestion. Out of all the techniques infrared sensor, inductive loop detection video data analysis, wireless sensor network, etc. are used to solve the congestion in the traffic and manage the traffic smartly.

2.1 Inductive Loop Detection

Induction loops are utilized for transmission and gathering of correspondence signals, or for recognition of metal objects in metal detectors or vehicle indicators. Multiple rounds of insulated cable are situated in a low cut out in the street, a lead inside the cable passes through the street-side pull box towards the controller and the electronic element positioned in the controller cabinet [11, 12]. The induction of the wire has changed depending upon the number of objects passes through the loop or stops. Change of induction is directly proportional to frequency. Due to the frequency change a electronic signal is forwarded to the control unit; this signal indicates the presence of the vehicle. Inductive loop detection helps to know the vehicle existence, its movement and also counts the number of vehicles passing through an exact location.

2.2 Video Analysis

An intelligent camera which has a unit to process data, different sensors to sense the stimuli and a unit for communication. The traffic always checks using an intelligent camera [11]. The video which is captured by the camera after that compressed to shrink the transmission bandwidth. The summary of video analysis depends on the raw video data and then calculates the traffic statistics. This statistic has the information about vehicles frequency, its average speed and path occupancy. The problems linked with video analysis are—(i) high relatively cost (ii) affect of deep smog or rains (iii) at the evening surveillance requires proper street lighting.

2.3 Infrared Sensors

The main objective of infrared sensor is used to sense or emit the infrared radiation. The sensors [11] are also capable of measuring the heat being emitted by an object and detecting motion; Infrared Radiation. It is used to detect the energy coming from the various types of objects like vehicle, road surface etc. The data can be captured the object by this electronic device is focused on to an infrared reactive objects using an optical method which then converts the energy into an electrical signals.

The captured energy of the infrared sensors is determined against infrared sensitive objects with an optical scheme which subsequently transforms the energy into the electric signals. These stimuli are used to check the traffic. It is used for signal management, recognition of pedestrians during crosswalks and communication of traffic information [13]. The major drawback of infrared sensors is that the usefulness of the system can be affected due to fog; also it has a complex installation and maintenance process [2, 6].

2.4 Wireless Sensor Network

By the help of the various technologies related with wireless sensor networks (WSNs) have been used to detect the traffic and avoids road congestion. WSNs are very trendy due to their faster transfer of information, easy installation, less maintenance, compactness and for being cheaper compared to other network options [12, 14, 15]. There has been significant research on Traffic Management Systems using WSNs to avoid congestion, ensure priority for emergency vehicles and cut the Average Waiting Time (AWT) of vehicles at intersections. In recent decades, researchers have started to monitor real-time traffic using WSNs, RFIDs, ZigBee, VANETs, Bluetooth devices, cameras and infrared signals.

3 Proposed Traffic Management System

3.1 Introduction and Objective

According to a new concept RFID, any vehicles deployed along with a RFID tag. The tag maintains the entire information about the vehicles. The tag identifying all vehicles exclusively and alerts the driver for getting various traffic messages. The RFID controller can be fixed with open signaling system. As per the Fig. 7, every signal have knows the data about all the vehicle passes through it. Here, each vehicle considered as an object and when it crosses through a signal, signal can repeatedly keep the data that means count of the vehicles passing by it. That traffic

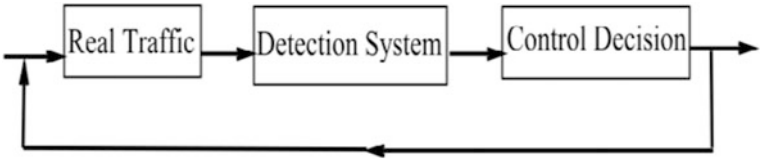


Fig. 7 Intelligent transportation system (ITS)

congestion should be maintained by variable counts. Every signal has a threshold value i.e. red and green. The timer can be dynamically controlled by the vehicles frequency crossing the signal per second.

Calculate the lowest frequency of vehicle passes per second and kept it in the controller. When this lowest frequency is arrived, the controller must give an instruction to the signal for turning red. Hence the signal is managed dynamically. Let's assume the maximum threshold value of green and red signal are 30 and 20 s respectively. The lowest frequency of vehicles passing by is 5 and that value is stored in the controller. Initially, the green signal will come (by default), when the timer begins through a peak value of 20, the frequency of the vehicles crossing the signal per second is 10, then after 10 s, that decreases to 5, and at that point of time automatically the RFID controller gives an instruction to the signal to turn red. As a result the red signal appears and it's subsequently signal in that connection turns green. This method continues in a cycle.

Due to the vibrant control, the signal helps in decreasing the waste of time and measuring the traffic congestion as a priority basis on a known vehicular traffic street. The proposed system checks traffic congestion in the street. No of vehicles passing the signal per second reached more than the maximum threshold time, then the congestion will arise at that location. When the congestion is detected, a notification command or message can send to its next former signal's controller as a temporary basis to stop the traffic through the RFID. The signal is turn red when the message has reached at its next point and according to that RFID message the traffic will work. After releasing the congestion at the crossing point, the particular signal's controller will forword a latest message to its previous controller signifying to start the traffic flow another time in that direction. After getting this message from the controller of the past signal place the red light OFF and turns the signal as green and restart the signal cycle as before. We have studied the flow of the proposed traffic management system from the Fig. 8.

3.2 RFID

The RFID is an electronic device which has small chip and antenna. The information is collected through antenna and stores it into the chip. The major component of RFID are RFID controller and RFID tag.

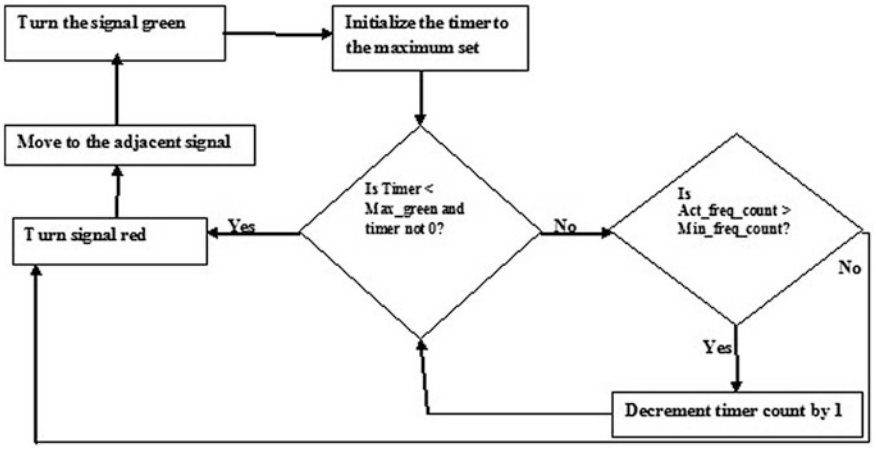


Fig. 8 System architecture

RFID Controller:

The RFID interrogator is currently available in the RFID controller and it is used to communicate with the RFID tag. The interrogator receives the signal or data as an input and passed it to the controller. Inter process communication sends commands and data messages from the controller components. Controller core is situated in the RFID controller. The controller core listens to the interrogators and based on the configuration; it can execute read/write operations upon the RFID tag or can do both listening and performing operations. The RFID controller can have serial interface during which external GSM/GPRS devices can be interfaced with it to make a dual radio device.

RFID Tag:

It is a mobile device which utilized the radio frequency electromagnetic fields towards move data and helps to identify and track the objects. Normally, two types of devices are used; one is active and other is passive. In case of active RFID, the battery is installed inside, whereas in passive it is not installed. Passive RFID based on outside source for its work. Tags data is situated in a non-volatile memory. Tag comprise of a Radio Frequency transmitter and receiver. Each tag has a unique serial number.

3.3 Steps of Proposed Work

Table 2, describes the variables which are used for the proposed model. Here, we considered two signals depending upon the congestion on traffic as a green or red.

Table 2 Variables are used as an input for the proposed work

Max_{th}_Red	Keeps the highest value where signal is red
Max_{th}_Green	Keeps the highest value where signal is green
Min_{th}_count	Calculate the lowest frequency of vehicle passes per second and kept it in the controller
Act_count	Calculate the real frequency (F) by the following formula: $F = \sum$ vehicles/second
Clock rate or timer	Keeps the real clock count

Thus, we require minimum and maximum value for the two signals within a particular time period and also count the number of frequencies.

From the above Fig. 7, we clearly observe the flow of data of our proposed work. Two steps are normally used to mange; they are as follows:

Step 1: Initially, the light should be Green.
Timer or Clock rate compares with Maximum threshold of Green signal and Timer is not Zero:
Then if (actual frequency of the vehicle (Act_count) > minimum threshold of vehicle (MIN_{th}_count))

Remain the signal Green and reduce the count rate by 1.

Else if (actual frequency of the vehicle (Act_count) <= minimum threshold of vehicle (MIN_{th}_count))

Go to Step 2.

Step 2: Create the signal Red and turns the adjacent signal as Green.
Go to Step 1.

4 Conclusion and Future Work

To overcome the drawbacks and shortcomings of existing work on traffic management system; the proposed work is designed and developed. The implementation of existing system cost is very high and it also depends upon the environmental condition. An efficient algorithm should fulfill the criteria like minimum cost, easy installation procedure and effectively manage the congestion at traffic. Our proposed system fulfils all the criteria for measuring, controlling and avoiding the traffic. The procedure is gainful than the present system. Here the survey shows about the problem which arises at metropolitan location throughout the globe caused by congestions along with the linked sources. Mostly, metropolitan location is most horrible with this situation. Congestions comprise a harmful effects on the

monetary condition of a nation, on the surroundings and so the in general quality of life. The proposed method can be improved by using powerful communication network other than GSM. This proposed model is used intelligent and plan for the future using transport scheduling tools to classify the packages of measures, which can best meet objectives such as, eliminating road accidents, minimizing emissions, improved accessibility and growing the economy of the all over the world.

References

1. Awasthi, P.K.: Smart traffic management system: the back bone of smart city. SSRG Int. J. Civil Eng. (SSRG-IJCE) **3**(7) (2016)
2. Das, S., Roychowdhury, P.: Smart urban traffic management system. https://doi.org/10.13140/RG.2.1.3414.6324
3. Sivasankar, P., Brindhavathy, B.: IOT based traffic monitoring using raspberry Pi. Int. J. Res. Eng. Sci. Technol. (IJRESTs) **1**(7) (2016)
4. Uddin, A.: Traffic congestion in Indian cities: challenges of a rising power. Kyoto of the Cities, Naples, 26–28 Mar 2009
5. AIOTI WG01—IERC: Internet of Things Applications. 15 Oct 2015
6. Vyas, D.A., Bhatt, D., Jha, D.: IoT: trends, challenges and future scope. IJCSC **7**(1), 186–197. Sept 2015–March 2016
7. Elavarasi, T., Kuppusamy, P.: The smart transportation using IoT and intelligent transport system in GPS localization. Int. J. Innovative Res. Comput. Commun. Eng. (An ISO 3297: 2007 Certified Organization) **4**(6) (2016)
8. Sherly, J., Somasundareswari, D.: Internet of things based smart transportation systems. Int. Res. J. Eng. Technol. (IRJET), **02**(07)| (2015)
9. Traffic Detector Handbook: Third Edition—Volume I Publication no. FHWA-HRT-06-108. Oct 2006
10. Kamoji, S., Nambiar, A., Khot, K., Bajpai, R.: Dynamic vehicle traffic management system. IJRET Int. J. Res. Eng. Technol.
11. Lanke, N., Koul, S.: Smart traffic management system. Int. J. Comput. Appl. **75**(7), 0975—8887 (2013)
12. Kafi, M.A., Challal, Y., Djenouri, D., Doudou, M., Bouabdallah, A., Khelladi, L., Badache, N.: A study of wireless sensor network architectures and projects for traffic light monitoring. ANT (2012)
13. Wu, B.F., Kao, C.C., Juang, J.H., Huang, Y.S.: A new approach to video-based traffic surveillance using fuzzy hybrid information inference mechanism. IEEE Trans. Intell. Trans. Syst. **14**(1) (2013)
14. Kafi, M.A., Challal, Y., Djenouri, D., Doudou, M., Bouabdallah, A., Badache, N.: A study of wireless sensor networks for urban traffic monitoring: applications and architectures. In: The 4th International Conference on Ambient Systems, Networks and Technologies (ANT 2013), pp. 617–626. Elsevier, Procedia Computer Science 19 (2013)
15. Lee, U., Gerla, M.: A survey of urban vehicular sensing platforms. Comput. Netw. **54**, 527–544 (2010)

Index

A. K. Sangaiah et al. (eds.), *Cognitive Computing for Big Data Systems Over IoT*,
Lecture Notes on Data Engineering and Communications Technologies 14,
https://doi.org/10.1007/978-3-319-70688-7