2015 International Conference on Computational Science and Computational Intelligence

A Big-Data Centric Framework for Smart Systems in the World of Internet of Everything

Pedro A. Pena
Computer Science Department
Univ. of Miami, Coral Gables, USA
Email: p.pena3@umiami.edu

Dilip Sarkar
Computer Science Department
Univ. of Miami, Coral Gables, USA
Email: sarkar@cs.miami.edu

Parul Maheshwari
Computer Science Department
Univ. of Miami, Coral Gables, USA
Email: p.maheshwari@umiami.edu

***Abstract*—Current research and development efforts for integrating *Everything* — heterogeneous objects, devices, and people — in the Internet have advanced network technologies and protocols to develop the *Internet of Everything* (IoE). The emergence of smart systems such as Smart Cities and Smart Traffic Management Systems have been proposed to gather data from relevant sources via IoE. Recently, cloud computing has been introduced into the IoE paradigm to provide a multitude of services necessary to compose a smart system that utilizes big data. In this paper, we propose a big-data centric framework for smart systems that encompasses four phases: data collection, processing, management, and interpretation. The proposed framework facilitates a *modular* architecture for smart systems where *security* and *cognition* are interactive modules.**

***Keywords*-Cloud Computing; Cognition with Knowledge; Internet of Everything; Security; Smart Systems;**

I. Introduction

Currently, research on networking tools and protocols for creating *Internet of Everything* (IoE) have gained significant attention [1], [2], [3]. The growth of data from IoE is exponential, and it is putting a strain on IT infrastructure.

Architectures for smart cities have been proposed using IoTs in [4], [5]. These smart systems will add big data to today's 2.7 zetabytes of data[1]. As smart systems gather data from all sources, unauthorized access to data is a major concern and precautions are needed to secure data to preserve privacy of smart systems and individuals [6].

In this paper, we propose a generic framework for *smart systems* that includes cognition for data collection, processing, management, interpretation, and security. This is a generalization over the system proposed in [7] that has cognition for data management and interpretation only. Trust is a major concern because of all the security breaches in the recent decade, and experts predict an increasing trend. In our proposed framework, we include a module for security to increase safety and trust. We will discuss different methodologies used to protect privacy and sensitivity of data.

Cloud computing has been a favorite platform for IoEs [8], because it delivers i) *Software as a Service* (SaaS), ii) *Platform as a Service* (PaaS), iii) *Data Storage as a Service* (DSaaS) iv) *Infrastructure as a Service* (IaaS) [9]. Unlike past proposals, the building blocks of the proposed framework are cloud services. Utilizing the concept from [7], we propose to organize data into virtual objects for easy storage and management. We introduce a cognition module that gathers information from the overall system to create knowledge for intelligent services, such as detection and prevention of intrusions or malicious attacks. Our aim is to have a generic framework for all smart systems, including Smart City, Smart Hospital, Smart Education, etc. A system derived from the framework is expected to support horizontal integration with other smart systems to understand the world around us, and find interrelationships among them.

The rest of the paper is organized as follow: description of IoE enabling technologies is presented in Section II. An overview of a novel framework for smart systems is presented in III. The security framework is presented in Section IV, and then cognition framework is discussed in Section V. Finally we discuss the use of the framework for smart cities in Section VI.

II. IoE Enabling Technologies

In this section, we provide a very brief overview of the network and the cloud, two central components of IoE.

A. Network

Networking has received most attention on discussions of IoEs because they enable and provide communication among components, and supports interoperability of smart systems within IoEs. Since the network protocols handle the communication between heterogeneous objects, significant efforts have been dedicated to standardize these protocols. The following section lists the most important protocols by the type of service they provide. An extensive and comprehensive survey on this subject can found in [1].

1) Network Protocols:

- *Application Protocols*: *Data Distribution Service* (DDS), CoAP, AMQP, MQTT / MQTT-NS, XMPP, HTTP/*REpresentational State Transfer* (REST).
- *Service Discovery*: *Multicast DNS* (mDNS), DNS-SD.

[1]http://wikibon.org/blog/big-data-statistics/

978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.62

- *Infrastructure Protocols*: 1) *Routing Protocol*: *Routing Protocols for Low Power and Lossy Networks* (RPL). 2) *Network Layer*: 6LowPAN, IPv4/IPv6. 3) *Link Layer*: The IEEE 802.15.4. 4) *Physical Layer*: EPC-global, *Long Term Evolution - Advanced* (LTE-A), *Bluetooth Low Energy* (BLE) or Bluetooth Smart, (Z-Wave).
- *Influential Protocols*: IEEE 1905.1 standard is considered most influential protocol for IoTs because it defines a network enabler for home networking supporting both wireless and wireline technologies.

As shown in Fig. 2, in the data collection phase sensors use the mentioned protocols to communicate with each other and the cluster head. The network protocols for the rest of the phases can also be seen in Fig. 2. Next, we introduce service-oriented architecture of clouds.

B. Cloud Architecture

It is generally believed that data for a smart system (such as Smart City, Smart Hospital, etc.) will be fed to the cloud to organize, understand, and store [8]. In this section, we discuss how services that clouds provide can be subscribed to construct a smart system.

1) IaaS: In Infrastructure as a Service (IaaS), a middleware is implemented for data processing to verify the data is pertinent to store in the cloud. The data is also post-processed to find patterns that can help us better understand the data we are receiving. The physical machines and virtual machines are stored in the IaaS, and the task of the engines in the IaaS is to mine the data. An example of an engine in IaaS is Hadoop or Mapreduce.

2) DSaaS: Data Storage as a Service (DSaaS) provides data storage and can be used for retrieval of information by a database manager. In DSaaS, security is essential to protect the data from malicious attackers. The data will be organized by historical, real time, and metadata.

3) PaaS: Platform as a Service (PaaS) provides the tools to work with the machines in the cloud. An example of a platform in the cloud is the LAMP stack, which is an acronym derived from initials of four original open-source components — Linux, Apache, MySQL, and PHP.

4) SaaS: Software as a Service (SaaS) is essential as it provides resources to the users for interpretation and visualization of data in the cloud. This component is expected to contain all the application software necessary to manage and operate smart systems for users and devices in the IoE. With all the data received from the city, government officials and citizens can visualize data at real time or historical data to understand trends in the city. For instance, SaaS may provide *smart applications* for monitoring traffic congestion, energy and water waste, and many others.

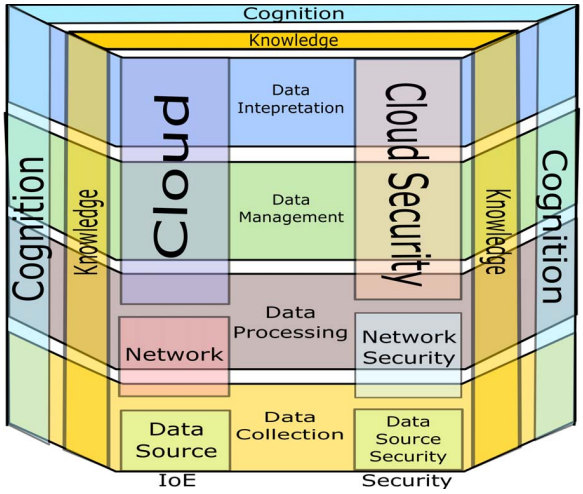

Figure 1. A big data centric framework for smart systems in the world of Internet of Everything. All functions are achieved through data collection, processing, management, and interpretation.

III. A Big-Data Centric Framework for Smart Systems

In this section we propose a ***big data*** centric generic framework for *smart systems*. Our objective is to embody Smart Cities, Smart Hospital, Smart Traffic Management Systems, etc. in a single framework. As shown in Fig. 1, we divide the system into four non-overlapping components: Data Collection, Data Processing, Data Management, and Data Interpretation. These four components can be organized and viewed from a data-flow sequence from both a manager and a user point of view. Before we describe each component of the system, an introduction of the system is provided next.

In our proposed framework we have separated IoE from its security. As can be seen in Fig. 1, the left column shows IoE enabling technologies and their roles in each phase. There is a parallel security column to the right. Behind the IoE and security columns is the knowledge layer. The knowledge layer gathers and maintains both past and present knowledge of the system. Behind the knowledge layer sits the cognitive layer. This layer utilizes knowledge in the layer above it to make IoE intelligent from user and system perspectives.

One could think each component as a phase of operation of a smart system. In the sequel, we use the terms component and phase of smart systems interchangeably. Henceforth, in the following sections, we will explain the system by each phase in the data flow. This will help the reader understand the transition of the system as well as understand how the system is organized. For each phase, we will briefly discuss how the framework functions, and then, we will discuss how the security and cognition fits in the respective phase.

A. Data Collection

As shown in Fig. 2, data for a smart system will come from sensors, social media, crowd, as well as, offline devices

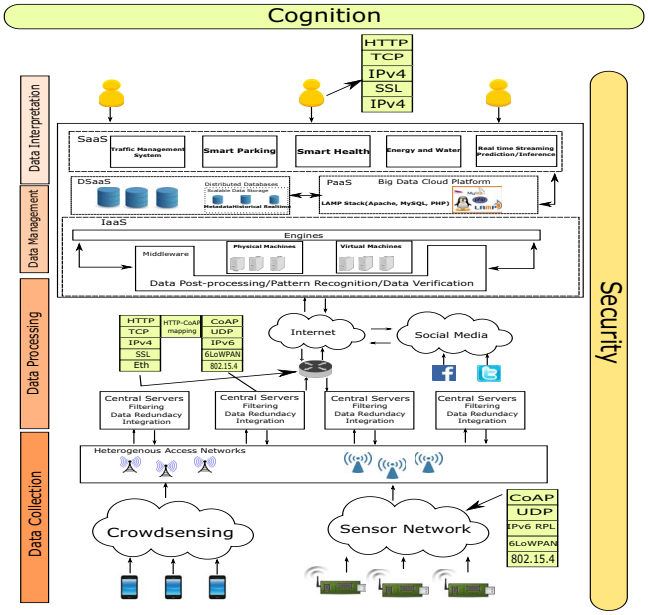

Figure 2. Block diagram of a cloud based architecture for Internet of Everything (IoE)

and legacy sources. For intelligent operation, both present and past data are very critical. A brief role of each data source is provided next.

Sensors and Sensor Networks: The wired and *Wireless Sensor Networks* (WSNs) consist of cameras, sensors, *Road Side Units* (RSUs), and many other sensors. These sensors, for example, collect data for temperature, pressure, light intensity, traffic conditions, etc. The heterogeneous data from a set of inexpensive and unreliable sensors should be preprocessed before communicating them to the cloud for storage. For this purpose, a set of sensor nodes form a cluster and communicate their readings using 3G, GSM, UTMS, WiFi, WiMax, BLE, and ZigBee networks to a node with higher computing power than the sensors. This node is known as a cluster head.

Crowdsensing: The use of mobile phones to collect data in an environment for a smart application is called *crowdsensing*. The burden of deployment and operation of a sensor network, at least partially, is reduced from crowdsensing. The cellphone user that provides data must have incentives to perform these tasks. The crowdsensing platform needs a front end application as well as a crowdsensing manager. Another important aspect for crowdsensing is security and privacy of the data providers. The crowdsensing manager assigns tasks to the mobile users and also collects data from the mobile phones.

Social Media: Social media can serve as an indirect source of data to smart systems. Moreover, information from social media can be used to connect the data that comes from sensors and crowdsensing. For example, social media data can be mined to find causal effects of crises to better understand problems that a city faces. While data from social media is free, it could be very subjective and may need very powerful and cognitive software for extracting information before using the data for any smart system.

Offline and Legacy Data Sources: While most of the IoT and IoE architecture omits offline and legacy data, because *directly* they are not part of IoT or IoE, we believe any smart system must consider these data sources. For instance, legacy data that has been gathered over many years from operations of a city is very valuable for building a smart system.

1) Security: Security at the *data collection phase* will mainly focus on encrypting the data when sent to the network. At this level of the system, a malicious attacker can target an area of sensors and affect the data packets in the network. Therefore, we have included an *Intrusion Detection System* that can detect any intruders in the network.

2) Cognition: The cognitive system will mainly monitor the data at this phase to find any emergency event that needs to be sent to the application layer to alert user. Also, the cognitive system will help security by monitoring the network for any new attacks designed by a malicious user. The system mainly *learns* to find these attacks. This knowledge is given to the security system for further actions.

B. Data Processing

Data processing phase spans over the network and cloud (see Fig. 1). We discussed earlier that at the data collection phase some preprocessing increases accuracy and security, but the major data processing is done at the cloud before archiving. Data preprocessing may include filtering the data using some filters, such as Kalman filter. This part of the system utilizes edge computing at edge cloud or cloudlets. If involved, local servers also communicate with the cloud manager to avoid sending redundant data. When the preprocessed data reaches the cloud, the data is then further processed and integrated with data packages from other local servers. An illustration of the data processing phase can be seen in Fig. 2.

1) Security: At this phase, the security system's goal is to have the data intact as possible. This will be done by *Secure Data Aggregation*. This will verify that the data has not been manipulated by colluders in the network. The data will also be encrypted for secure communication through the network.

2) Cognition: The cognitive system mainly provides knowledge to the local servers to discover emergent events through the preprocessed data that comes from all local servers. This is crucial to immediately notify users of any event that is time sensitive. In the cloud, it is expected to use machine learning algorithms for maintaining integrity of data. It should be noted that this phase utilizes DSaaS and SaaS of cloud.

C. Data Management

At the *data management phase*, the processed data needs to be stored in the database. The distributed databases store the metadata, historical data and realtime data. The data is managed by PaaS that provides tools to access the data when user queries. This phase is vulnerable to attacks to access sensitive data. The DSaaS and PaaS can be seen in Fig. 2.

1) Security: In this phase, the security is focused on protecting the data from being accessed by an authorized user. Thus, the security system may use techniques such as T-Coloring to fragment the data and store pieces of the data at different parts of the database. Also, as a protection layer, there is an Intrusion Detection System to find any manipulated network packets or other known attacks.

2) Cognition: The cognitive system also provides knowledge to the Intrusion Detection System in the security system if there is an unknown attack in the network. The cognition system also helps with the management of data. Semantic knowledge is extracted from the data and organized into virtual objects and composite virtual objects for application specific knowledge.

D. Data Interpretation

At the *data interpretation phase*, users have access to applications and can request specific data from the system. There will also be applications that will monitor the system and wait for alerts from the system to notify the user of an event. Requests from users will be processed through the system and queried through the database. If data is not available through the cloud, an alert will be sent to the data sources to fetch data needed by users. Examples of smart applications is depicted in Fig. 2 in the data interpretation phase.

1) Security: The security system assigns a virtual machine and a secure connection is established with the user and the virtual machine. This allows the Intrusion Detection System to find a common ground with the user and from there find any anomalies in the network. Also a *Secure Socket Layer* (SSL) is used to create a secure connection.

2) Cognition: The cognition system in the data interpretation phase mainly provides knowledge to the applications through a *Smart System Knowledge Base* (SSKB) that can create inferences and forecast models from all the data received in the data collection phase. This knowledge base then feeds knowledge to the smart systems.

IV. Security Framework

An overview of the security system can be found in Fig 3. As can be seen from this figure and Fig. 1, in the proposed security framework corresponding to every IoE component, there is a parallel security component. We have maintained this to decouple security of one phase from another for ease of implementation and operation. Also for modeling purposes, security of overall system can be derived

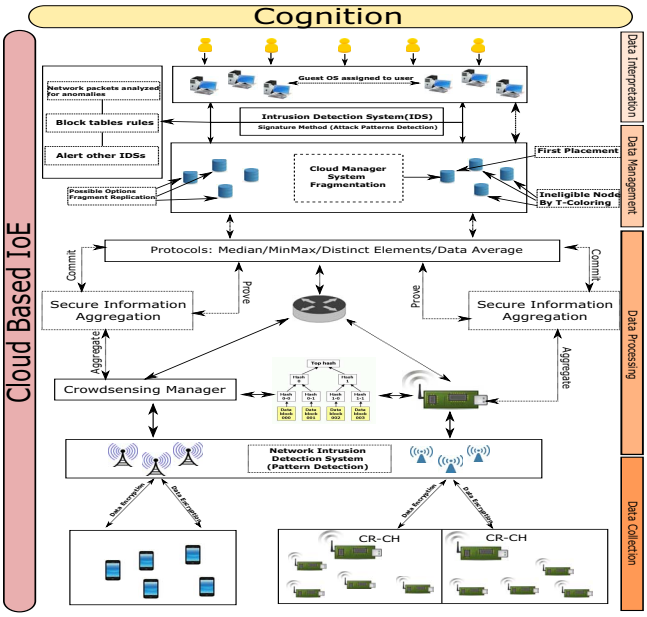

Figure 3. Block diagram of a security architecture for IoE

if security of each component is known. We briefly describe intended functions of each component next.

A. Data Collection

In the data collection stage, the sensors will be collecting data for the smart systems [3]. As shown in Fig. 3, a Network Intrusion Detection system implemented at this level to detect any malicious attacks is necessary. As proposed in [10], a network-based IDS that captures the traffic from the network and analyzes the traffic to detect possible intrusions is a possible solution. We will also provide more improved models in the future. Figure 3 illustrates a possible architecture for security of IoE.

B. Data Processing

To protect this part of the system, one may use a commit-aggregate-prove method [11], or a similar method. The secure information aggregation method is illustrated in Fig. 3.

1) Aggregation: The aggregator collects the data from sensors and locally computes the aggregation result. Each sensor shares a key with the aggregator, such that the aggregator can verify the authenticity of each sensor reading to prevent sensor impersonation, but not flawed data from a corrupt sensor.

2) Commit: The aggregator commits the collected data, which ensures that the aggregator uses the data provided by the sensors. The statement can be verified by the home server, or middleware to verify the correctness of computed results.

3) Prove: In the prove part, the aggregator and the home server utilize a protocol in which the aggregator communicates the aggregation result and the commitment to the server and proves to the server that the reported results are correct using interactive part of the protocol. The middleware will be able to detect if a sensor is compromised.

C. Data Management

The data management of the system will consist of storing the data in the data storage of the cloud. This data can be sensitive to attacks because of the privacy of the data. The data can be protected by using fragmentation of the data and storing the fragments in different parts of the cloud storage. The fragments will be stored far way from each other. Hence, this will prevent attackers from obtaining relevant information from the data by attacking one node. Moreover, the attacker will need to attack many nodes to find relevant information. The probability of this happening is very small. By separating the fragments of the data at different places, the data will be safe from attackers. As mentioned earlier, T-Coloring or a similar method can be deployed [12]. T-Coloring can also be shown in Fig. 3.

D. Data Interpretation

In the data interpretation layer, the security will only consist of user verification. Therefore, it will require a secure connection with the user and the virtual machine. SSL can be used to establish a secure link. As shown in Fig. 3 in the data interpretation phase, there will also be an IDS to detect any anomalies in the network. Henceforth, the network packets will be analyzed for anomalies with a block tables rule. If any suspicious activity is found, it will be reported to a higher layer that will be in charge of alerting other IDSs in the system. Moreover, security protocols will be executed to secure communication in the network. More information see [10].

V. Cognition Framework

We believe that every smart system that is integrated into the world of IoE must have its own knowledge-base as well as a cognitive framework not only for safe and secure operations, but also for timely delivery of services to its users and devices. For instance, all operations must be grouped into classes based on delivery time requirements. To provide the fastest route to an emergency vehicle, some data collection operation may be delayed. However, this delay in transportation of collected data must be recognized by the transport protocol at the network level; the network connection should not be changed because of late delivery of expected data.

At this time, very scant literature exist on cognition of IoT/IoE. We will provide some outline and leave the details as our future work. We hope to provide a much elaborate description in the near future. An overview of the cognition system can be found in Fig. 4.

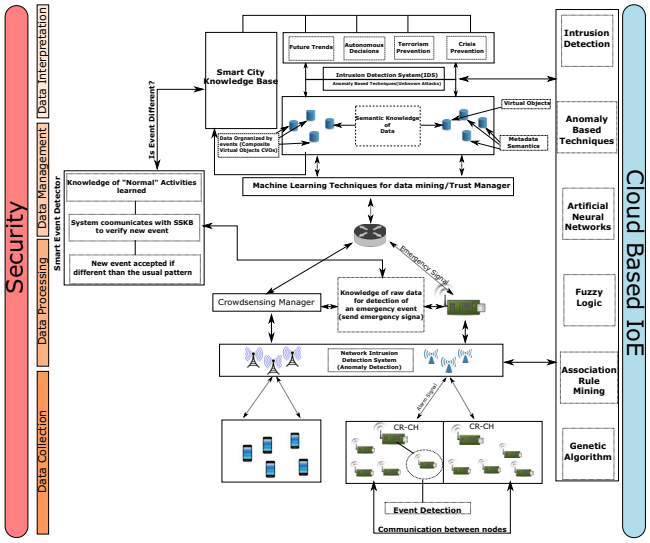

Figure 4. Cognition architecture

A. Data Collection

In [13] it was proposed that an event detection system sends an alarm to the cluster head to report an emergency signal to the cloud. As shown in Fig. 4, alarm signal can be evaluated at the data processing phase by the cognitive system to direct required services, if necessary.

B. Data Processing

A *Smart Event Detector* (SED) can be implemented to learn the frequent patterns from everyday sensing and this information can be gathered in the *Smart System Knowledge Base* (SSKB). If a new event is detected, SED verifies with SSKB, which in turn sends a signal to SED to allow data integration (see Fig. 4). If the event goes through, the data is integrated and data mined to extract knowledge. This information is integrated in the SSKB and stored in databases.

C. Data Management

When the data is processed and transported to the cloud, the data becomes *Virtual Objects* (VOs) [7]. A group of VOs with similar semantics can be combined into a *Composite VO* (CVO) for integration into the knowledge-base, and in the future, it can be utilized for intelligent operations of the system. This can be seen in Fig. 4 in the data management phase.

D. Data Interpretation

In this layer, the data will be ready to visualize. As can be seen in Fig. 4, the *Smart System Knowledge Base* (SSKB) is located in the data interpretation phase. The SSKB gathers all the data from the VOs and CVOs and constructs future models that can be used to avoid future events or it can be used to lead to a wanted event. Thus, SSKB can bring

a plethora of advantages to the city by providing a way to plan future events. This can help, for example, with avoiding stock-market bubbles, inflation, terrorism, crisis, etc. The SSKB will provide all knowledge needed for a smart system to thrive.

VI. Applications for Smart Systems: Smart City

In this section we briefly discuss how an architecture of a smart city can be derived from our proposed framework for modular implementation and deployment.

1) Data Collection: For gathering data, sensors, such as *Road Side Units* (RSUs), cameras, and induction loops, will be installed around the city. Crowdsensing and social media are also very important. Thus, Smart Citizens will be a big part of Smart City, where they will deliver data for many components of the system.

2) Data Processing: As presented in our proposed framework, when there is an emergency event in a city, such as a big traffic accident blocking intersections, emergency vehicles can be directed without delay, while collected data will be held at the source or in a local storage.

3) Data Management: Smart cities will generate heterogeneous data coming from sensors, mobile phones, old archives, etc. Hence, in the data management phase, the data needs to be integrated with relevant data that will help the SSKB detect useful information for Smart City applications.

4) Data Interpretation: Using Smart Parking, Smart Light, Traffic Management Systems, and the rest of smart features will help smart cities manage all aspects of the city. Inferences and forecast models will be simulated for knowledge of *future* events. With a *Traffic Management System* (TMS) [14] incorporated in the the application layer of the framework, government officials will know where traffic is more denser, and citizens or emergency vehicles can reroute their trips accordingly.

VII. Conclusion

The networking technologies of IoEs has been addressed extensively [2], [3], [1]. Much efforts have been devoted to standardization of network protocols, but a framework that can handle the flow of big heterogeneous data from IoEs remains mostly unexplored. Resources from cloud computing have been shown useful for managing and interpreting data. However, data collection and data processing phases of smart systems have not been discussed in an integrated framework yet. In this paper, we addressed these issues by proposing a framework that is generic enough for any smart system in the domain of IoEs. The framework allows design of modular architectures for reliable and intelligent systems. However, details of the knowledge gathering processes for the cognitive layer and the operations of the cognitive layer are yet to be addressed.

References

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols and applications," *IEEE Communications Surveys and Tutorials*, 2015.

[2] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through internet of things," *IEEE Internet of Things Journal*, 2014.

[3] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, 2014.

[4] S. Madakam and R. Ramaswamy, "100 new smart cities," *IEEE*, 2015.

[5] T. Clohessy, T. Acton, and L. Morgan, "Scaas-a future roadmap for e-government smart city cloud computing initiatives," *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, 2014.

[6] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: Privacy and data mining," *IEEE*, 2014.

[7] P. Vlacheas, R. Giaffreda, Stavroulaki, D. Kelaidonis, V. Foteinos, G. Poulios, P. Demestichas, A. Somov, A. Biswas, and K. Moessner, "Enabling smart cities through a cognitive management framework for the internet of things," *IEEE Communications Magazine*, 2013.

[8] S. Sowe, T. Kimata, M. Dong, and K. Zettsu, "Managing heterogeneous sensor data on a big data platform: Iot services for data-intensive science," *2014 IEEE 38th Annual International Computers, Software and Applications Conference Workshops*, 2014.

[9] P. Mell, "What's special about cloud security?" *IEEE Computer Society*, 2012.

[10] S. Kene and D. Theng, "A review on intrusion detection techniques for cloud computing and security challenges," *IEEE Sponsored 2nd International Conference on Electronics and Communication Systems(ICECS '2015)*, 2015.

[11] B. Przydatek, D. Song, and A. Perrig, "Sia: Secure information aggregation in sensor networks," *SenSys '03*, 2003.

[12] M. Ali, K. Bilal, S. Khan, B. Veeravalli, K. Li, and A. Zomaya, "Drops: Division and replication of data in cloud for optimal performance and security," *IEEE Transactions on Cloud Computing*, 2015.

[13] H. Salameh, M. Dhainat, A. Al-hajji, R. Aqeli, and M. Fathi, "A two-level cluster-based cognitive radio sensor network: System architecture, hardware design, and distributed protocols," *2015 IEEE International Conference on Cloud Engineering*, 2015.

[14] S. Djahel, R. Doolan, G. Muntean, and J. Murphy, "A communications-oriented perspective on traffic management systems for smart cities: Challenges and innovative approaches," *IEEE Communications Surveys and Tutorials*, 2015.