Applied Journal of Physical Science
Volume 2(1), pages 9-15, February 2020
Article Number: 3BBB61E12
ISSN: 2756-6684
https://doi.org/10.31248/AJPS2019.024
https://integrityresjournals.org/journal/AJPS

Review

Challenges of data capturing in smart cities

Mulikatu Ibrahim Yakubu[1*], Hafiz Uba Usman[1], Kamaluddeen Ibrahim Yarima[1], Oluwatosin Islamiyat Yusuf[1] and Kabiru Onotu Momoh[2]

[1]Department of Computer Science, Kano University of Science and Technology, Wudil, Nigeria.
[2]Department of Physics, Ahmadu Bello University, Zaria, Nigeria.

*Corresponding author. Email: muleekat246@gmail.com

Received 24th November, 2019; Accepted 14th January, 2020

ABSTRACT: Smart cities are becoming more ubiquitous. It covers all areas that makes the standard of living of the citizens easier. Sensors are implanted everywhere in the city to make the capturing of data much easier and in real time. These sensors communicate with each other using internet of things. There might be failure in the devices that are implanted in the cities which could make capturing the data at that moment more difficult because the data are in most cases critical and time sensitive. The data generated must keep pace with generation rates and must be used to get the desired results to make decision in an internet of things environment. The purpose of this paper is to discuss some processes of data collection to avoid data loss and the challenges being faced in capturing data in smart cities and the solution to minimize the challenges. This paper will serve as a starting point for explaining and discussing the challenges being faced in collecting real time data in the smart city environment.

Keywords: Internet of things, real time, sensors, smart city.

INTRODUCTION

A smart city is a future, better state of an existing city, where the use and exploitation of both tangible (e.g. transport infrastructures, energy distribution networks, and natural resources) and intangible assets (e.g. human capital, intellectual capital of companies and organizational capital in public administration bodies) are optimized (Popescul and Radu, 2016). It is a city in which its social, business, and technological aspects are supported by Information and Communication Technologies to enhance and improve the experience and quality of life of the citizens within the city. To achieve that, the city provides public and private services that operates in an integrated, affordable and in a sustainable way. In the smart city environment, sensors are implanted everywhere and enormous amount of data is being captured from the sensors (Felipe et al., 2015).

A few researchers have developed a vision for the construction of a smart city. The smart city construction in three layers; the first layer being the “perception layer”, in which different data are collected from various data sources such as cameras, GPS, sensor network; second layer, “network layer” is responsible for transmitting data, collected from layer 1 to data storage center, and layer 3 is dubbed, ”application layer” containing applications for analyzing and processing the massive data residing in the data storage center (Samih, 2019).

This research is focusing on the veracity of the data which is the truthfulness of the data that is being captured from the sensors. In the context of big data, one of the components of big data is the veracity which means “habitual fruitfulness”, so it is not just a matter of capturing the data but how it can make an impact or be of use in the decision making process. In capturing useless data, the amount of data both the true and untrue ones will be used to make decision which makes the decision obscure. Untrue data also occupies more space in terms of storage capacity. Some components of veracity include suspicious users, device malfunction and device failures. The introduction of Internet of things (IoT) makes people’s daily life easier, safer and more interesting. Business may no longer run out of stock or generate waste products, as involved parties would know which products are consumed and required. Traffic conditions can be achieved directly via smart phones or GPS, so that we can safely keep away from traffic jams or even accidents. All kinds of data are collected and analyzed to entertain people on the internet, for example, constellation interpretation or social hotspots (Zhang et al., 2013). According to Santos et al. (2017), a city's ability to produce and share relevant data that can be understood is critical that can be understood is critical and can be viewed as key indicator of a smart city. Furthermore, a city's ability to derive knowledge from this data and further, to use it to power innovation is an even better indication of a smart city. The important factors we ought to look out for is the Intelligent Internet of things (IoT) and the big data, process of data collection using sensors, challenges faced, and solutions.

The IoT provides a platform for sensors and actuator devices to communicate seamlessly within the smart city environment and enables an increasingly convenient information sharing across platforms (Hashem et al., 2016). According to the report published by CISCO in 2008, things connected to internet surpassed the number of people living on earth. It is also added that these things would touch the limit of fifty billion by 2020, taking us to the world of digitization. These things interact and communicate with each other with the help of internet –it is called the Internet of Things (Ganesh, 2017).

Quintillion bytes of data are generated every day and it has become a tedious task to handle these volumes of data. Daily used devices like mobile phones, laptops, computers etc. make use of internet to connect to each other around the globe and in doing so these devices give birth to enormous amount of data. This data is collected at a storage unit called Data Center. The data is growing at an uncontrollable rate and is giving birth to Big data (Gehlot, 2016). Big data analytics can extract meaningful information from the oceans of data produced by sensor devices. Effective analysis and utilization of big data is a key factor for the success in many business and service domains, including the smart city application. The application of big data in a smart city has many benefits and challenges, including the availability of large computational and storage facilities to process streams of data produced within a smart city environment. One of the possible means to tap this benefit is a reliance on cloud computing services and IoT technologies (Hashem et al., 2016). The big data systems will store, process, and mine smart cities applications information in an efficient manner to produce information to enhance different smart city services. In addition, the big data will help decision-makers to plan for any expansion in either smart city services, resources, or areas (Al Nuaimi et al., 2015). The big data comprises of 6 V's which are, volume, velocity, variety, veracity, variability, and value.

- Volume can be defined as the amount of data that is generated from the sensors. In a smart city environment, quintillion amount of data is produced from the devices implanted in the city and also from millions of the citizens within the city.
- Velocity is the rate at which the data is generated from the sensors, the rate at which it is stored in the data centers and the rate at which it is being processed in real time using some big data tools.
- Variety consists of the types of data that is being generated from the sensors which includes, structured data, unstructured data and semi structured data. Data is captured from many data sources and from different vendors which makes the data to be in different formats.
- Veracity is how true is the data that is being generated. The data is generated from the sensors and the device can malfunction which will make the data not to be generated in real-time. Also, we can also have some situation due to device failure which makes the data not to be generated and decision cannot be made if data is not generated at that time.
- Variability means how the structure and meaning of data constantly changes especially when dealing with data generated from natural language analysis for example (Al Nuaimi et al., 2015).
- Value is the benefit that big data has brought to the smart city and how the tools benefits in the decision making process.

LITERATURE REVIEW

Spatial information of a smart city is generated from different types of sensors, controllers and computing terminals which are all maintained by computers and storage devices equipped in various departments and locations. Managing and coordinating these devices with different structures and wide-area distribution is not a trivial issue. Smart cities generate not only structured data such as temperature values, geographical coordinates but also a lot of unstructured data and semi structured data such as pictures, audios, and videos. Storing and managing these vast amounts of diverse data in several formats is a monumental task. In Samih (2019), the author stated that smart cities are responsible for thorough analysis of urban information, public affairs, decision support, real-time tasks and responding to users' requests on time. He mentioned that, Information on the source data such as the locations where it was collected, what streets were covered, what intersections were recorded, whether the data is from day or night, or sun or rain all needs to be recorded and associated with the data to aid in scene selection and to ensure the full portfolio of data requirements are being met. Scene selection is particularly important for supporting sensor fusion, where researchers combine data from different sensors and sensor types to use the combined information to perceive the environment more accurately. Information on the data's journey over time through various annotations, labeling needs, and training uses also must be tracked to maintain data integrity and usability. As local governments strive to deliver better services, cut costs and ensure their economies and communities are productive and vibrant, smarter decisions are vital. Data must drive these decisions, however, for most it will take a new way of thinking – becoming ‘data champions’ – to turn data into actions. The challenge is the ability to gather, manage, process and manipulate data safely, quickly and effectively to produce actionable insights. in Francis et al. (2018), the authors mentioned that, it is no easy task when many councils have complex sets of legacy IT systems and gaps in the specialist data science and technical skills needed to unlock the value from this unique asset. Effective analysis and utilization of big data is a key factor for success in many business and service domains, including the smart city domain. In Al Nuaimi et al. (2015), the authors reviewed the applications of big data to support smart cities and they discussed and compared different definitions of the smart city and big data and explores the opportunities, challenges and benefits of incorporating big data applications for smart cities. In addition, they stated that, it attempts to identify the requirements that supports the implementation of big data applications for smart city services. The review reveals that several opportunities are available for utilizing big data in smart cities; however, there are still many issues and challenges to be addressed to achieve better utilization of this technology. The use of big data in smart cities is characterized by a diverse set of dimensions, including data, data collection method, and value created with data. Classifying existing cases of big data use in smart cities could help make sense of this diversity by identifying categories that share a number of similar attributes. Furthermore, it may suggest reference models for data-based smart city transformation. In a business context, data may come from companies (e.g., business transaction, human re- source, and financial data) or customers (e.g., demographic, behavioral, and purchase history data). For example, data analysis may rely on business transaction data for business process management, or auto- mobile manufacturers may rely on customers' driving data. Big data use in smart cities can also be classified from these same perspectives, as indicated by the smart-city-related literature that discusses data sources and beneficiaries(Lim et al., 2018).

PROCESS OF COLLECTING DATA IN SMART CITIES

Data is being captured by sensors implanted everywhere in the smart city. Smart traffic, smart mobility, smart fields, smart wells, smart waste management, smart energy, smart home, smart industry, smart factory, smart people, smart healthcare, smart infrastructure, smart police, smart grid, smart education, smart governance, smart safety and smart environment are the landscape of smart city and big data technologies. At present, a large amount of data is being generated from different data sources, such as smartphones, computers, sensors, cameras, global positioning systems, social networking sites, commercial transactions, and games. Given that the data generated in our present digitized world continuously grow, efficient data storage and processing facilities have posed challenges to the traditional data mining and analytics platforms. Big data analytics can extract meaningful information from the oceans of data produced by sensor devices. Effective analysis and utilization of big data is a key factor for the success in many business and service domains, including the smart city application. The application of big data in a smart city has many benefits and challenges, including the availability of large computational and storage facilities to process streams of data produced within a smart city environment. One of the possible means to tap this benefit is a reliance on cloud computing services and IoT technologies (Hashem et al., 2016). The real-time requirements of data storage and processing in the smart city are considered, the adoption of streaming architecture will guarantee the efficient and seamless communication between sensing devices within the smart city network. Such technology has been adopted recently with the introduction of many stream processing platforms, such as Apache S4, Storm, Spark streaming and Apache Kafka which can perform real time data streaming and capturing data even when there is device malfunction and device failure which also enables data storage and processing across various interconnected nodes (Hashem et al., 2016; Santos et al., 2017). Smart city data are collected from a variety of sources such as IoT devices, video surveillance systems, social networks, transport, government documents, or open data platforms, location-based services, and more. In addition, some data such as socio-economic data, contain sensitive personal related information such as social security number (SSN), name, age, home address, gender and health, etc. Therefore, smart city data have the characteristics of big data, including big volume, high velocity, veracity, value and variety. These pose a grand difficulty in dealing with the enormous data. In order to facilitate smart city data management, establishing a complete and flexible data management platform becomes very essential. This is a key step between data sources and the applications of using the data generated. Although some studies have been done to research on the big data platform of smart cities, most focus implementing a specific functional requirement and architectural design. At the same time, as there are many different tools and platforms with similar functionalities available in the big data community, we are often overwhelmed and confused by their features and capabilities that are needed in the smart city and IoT environments (Liu et al., 2017). The authors also stated that, some attempts have been made in managing the smart city data. Examples include the SCOPE, which is a cloud-based smart city open platform and ecosystem; CiDAP, which is a real-time smart city data platform; and FIWARE, which is a framework of providing intelligent application development in the Future Internet. They focus primarily on infrastructure development, data collection, test bench deployment or applications/services- specific development, but less emphasis on data sensitivity management.

There are three stages of collecting data in smart cities which are: The data are collected from data sources, then the data are transformed using the Big ETL tool, and finally the data is stored, published and retrieved. In the process of retrieving the data, sensitive data are filtered.

CHALLENGES OF CAPTURING DATA IN SMART CITIES

The problems and challenges being faced in data capturing in smart cities are as follows: privacy, communication, heterogeneity, security, safety, city models, lack of testbeds, platform maintenance, data management, energy management, scalability, technological challenges, data quality, data and information sharing, smart city population, cost, and data sources and characteristics. These challenges are further explained below.

Privacy

Privacy is the most cited challenge to implementing a smart city platform (Hassani and Silva, 2018). Protecting data collected from all the areas of smart city and also from citizens within the city and enterprises is important. The data collected in the smart city environment may include confidential information related to the government and the citizens, so the data needs to be highly secured to prevent it from unauthorized use and malicious attacks. A solution to privacy includes restricting the data to be captured or to be used in making decisions. What should be done again is to request for license to use the data from the citizens and the government in such a way that once the data is captured, it should be made aware to the citizens so as to avoid being meddling into people’s privacy. Cryptography and anonymization should also be used to tackle the problem of privacy.

Communication

The city enables communication among heterogeneous devices which makes the detecting of errors from a certain sensor device difficult. The devices in the cities consists of sensors, global positioning services (GPS) and also the sensors are different. Communication from different devices makes the capture of data difficult. Since the smart cities of the future will incorporate a massive number of devices, enabling communication among these devices will be a challenge. To tackle this problem, we need to limit the number of devices used in the cities and also making the devices to be multi-processed so as to avoid plenty devices because involving massive number of devices makes handling of these devices difficult.

Heterogeneity

Ensuring the interoperability of devices and applications, a smart city platform has to define standards across heterogeneous devices, systems, and domains. This is a challenge because of the differences between the devices in a smart city, and the difficulty of relating data from different sources (Felipe et al., 2015). A solution is to use a city unified model.

Security

The main challenge most likely to be encountered in security management lies in the availability of voice telecommunication IoT supported device. The telecommunication model is crucial in realizing the role of noise sensors in the environment. By a different token, the privacy and security of transmitted data in an IoT network is crucial in the development of a smart city. As a result, the use of Integrated Radio Frequency Identification sensors adequately addresses security issues in a cellular IoT system and the utilization of a cellular network for the IoT devices addresses scalability and reliability risk (Alharbi and Soh, 2019). Protecting the city data, services, and infrastructure using access tokens, devices, and cryptography. Unauthorized users accessing city services without permission may cause a lot of harm. Infrastructure and technologies used at different levels of data processing are designed and implemented by different suppliers, without the possibility of standardization and proper protection.

The problem is that they are sensitive data, often gathered without our explicit consent. For example, messages, personal pictures, appointments, bank account information, contacts and others are stored in our smart phones in full awareness, with more or less security measures put in place. But an average smart phone comes with various sensors like gesture sensor, proximity sensor, Red Green Blue (RGB) light sensor, gyro sensor, accelerometer, geomagnetic sensor, barometer, and hall sensor. Such sensors can capture location, movements, time stamps, even private conversations and background noises. The use of these sensors by different applications, the quantity and the purpose of collected data are not fully understood and controlled by their owners (Popescul and Radu, 2016).

A solution to data security in smart cities is an onion model. In order to adequately protect a smart city, a lot of measures provided by various actors are needed. An overall view of these solutions is presented in the Figure 1.

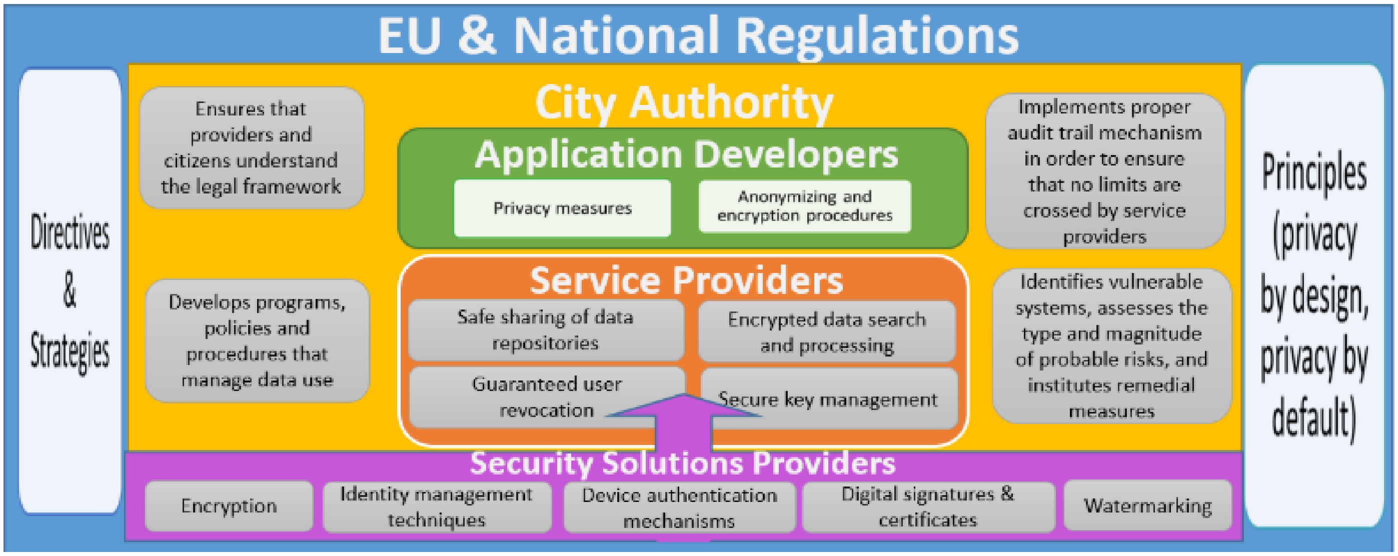

Figure 1. Security measures for a smart city – an onion model (Popescul and Radu, 2016).

Safety

Although emergency services have drastically improved, cities continue to lose lives due to delays in report and response time. Therefore, it is important to not only focus on crime and risk prevention but also mitigation strategies. The police or the firefighters are sometimes informed of accidents or crimes only after occurrences. This delay can be eliminated by installing gunshot, accident and noise sensors around the city. Also, connecting smoke sensors to a central repository, which is based on an artificial intelligence system, can facilitate timely response by police, ambulance and firefighters (Alharbi and Soh, 2019).

City models

To make intelligent decisions in real time, a very effective, quality and efficient model needs to be created. Creating a model in smart cities tends to be difficult because understanding the city itself is difficult which makes creating the model more difficult. A unified model of the city is required, so that the huge amount of heterogeneous data generated can be shared among applications and services. Also, semantic web and technologies should be used.

Lack of testbeds

There are no sufficient testbeds to experiment smart city solutions. Without testbeds, it is difficult to perform tests and experimentation to discover the real challenges that deploying a smart city platform will present. Testbeds should be created so as to perform tests and experiments before processing data that is captured. Simulators should be introduced to create testbeds.

Platform maintenance

Maintaining a middleware to manage and process millions or billions of devices connected to the platform is difficult. City systems and infrastructures should always be maintained by monitoring using an alert tool

Data management

Storing data and processing data in smart cities is a challenge because the amount of data that is being captured is very large and also some algorithms are used to make decisions. The bigger the data, the more difficult it is to process. To extort useful knowledge from the data tends to be difficult because of quintillion amount of data being captured. One solution to data management is to use some big data tools like Hadoop, Apache Spark, Apache Kafka, Apache Flume, Apache Hive, Cassandra, Zookeeper, Impala, Not only SQL (No SQL) and some relational databases and processing tools.

Energy management

The electricity used by deployed devices in the city needs to be well managed. To manage the energy in the smart city environment, there is need to ensure that applications and services in the cities do not fail due to power outages.

Scalability

The growth of devices and users connected to the smart city platforms should be used. As the day goes by, the number of users in the city tends to increase in population and also devices are needed to balance the equation. So, peer to peer algorithms and decision models are needed to solve the problem of growth of population in the city. Distributed tools and algorithms can also be used.

Data and information sharing

Sharing data and information among different smart city departments is another challenge. Each government and city agency typically has its own warehouse or silo of confidential or public information. Most of which are often reluctant to share what might be considered proprietary or privacy data. In addition, some data may be governed by certain privacy conditions that make them hard to share across different entities (Nuaimi et al., 2015). The users should be given a tutorial on what their data is used for and the benefits it will render to the society.

Traffic control and parking

Cities lose millions of dollars because of traffic congestion. Owing to temporary immobility, quality time and fuel are lost on roads. Besides, these congestions diminish the quality of life within cities. Moreover, it is evident that the cause of the congestion is the difficulty in finding parking spaces. Additionally, transport management is crucial in an era of electric cars; as a result, the adoption of smart metering tools is efficient in addressing the challenge being faced (Alharbi and Soh, 2019).

Data quality

Sensor data collected through a third party without a centralized control could have been produced by sensors that are faulty, wrongly calibrated, or beyond their lifetime. What should be done is to always make sure that the devices are intact and are well equipped. The devices should be very strong in order to avoid device malfunction and also the device should have a tracking unit, that will check in real time if a device is working properly.

Smart city population

The amount of data created in a smart city due to the population and also due to the areas that comprises the smart city application tends to be too large. The larger the citizens in the city, the larger the data and also the more the congestion of traffic in the city. Ultimately, the goal is to develop and deploy smart city applications that are smart enough to evolve and intelligently handle the rapid growth of big data to generate better result.

CONCLUSION

In this paper, some processes of collecting smart city data were discussed, some tools used in collecting the data were mentioned, challenges being faced in capturing data in smart cities and solutions to minimize the challenges were listed. Some related works on the processes of capturing data in smart cities were reviewed explaining the tools being used to avoid data loss when there is device failure.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6, Article Number 25.

Alharbi, N., & Soh, B. (2019, August). Roles and challenges of network sensors in smart cities. In *IOP Conference Series: Earth and Environmental Science* (Vol. 322, No. 1, p. 012002). IOP Publishing.

Felipe, E., Santana, Z., Chaves, A. P., Gerosa, M. A., Kon, F., & Milojicic, D. S. (2018). Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture. *ACM Computing Surveys*, 50(6), 1-37.

Francis, P., Piers, H. S., & Toni, J. (2018). *The smart city data challenge: How local governments can unlock data to help their citizens*. KPMG publication. Retrieved from https://assets.kpmg/content/dam/kpmg/au/pdf/2018/smart-city-data-challenge.pdf.

Ganesh, E. N. (2017). Development of SMART CITY Using IOT and BIG Data. *International Journal of Computer Techniques,* 4(1), 36-41.

Gehlot, R. (2016). Storage and retrieval of data for smart city using hadoop, SSRG International Journal of Computer Science and Engineering, 3(5), 85-89.

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, 36(5), 748-758.

Hassani, H., & Silva, E. S. (2018). Big Data: a big opportunity for the petroleum and petrochemical industry. *OPEC Energy Review*, 42(1), 74-89.

Lim, C., Kim, K., & Maglio, P. P. (2018). Smart cities with big data : Reference models , challenges , and considerations. *Cities*, 82, 86-99.

Liu, X., Nielsen, P. S., Heller, A., & Gianniou, P. (2017, August). SciCloud: A scientific cloud and management platform for smart city data. In *2017 28th International Workshop on Database and Expert Systems Applications (DEXA)* (pp. 27-31). IEEE.

Popescul, D., & Genete, L. D. (2016). Data security in smart cities: Challenges and solutions. *Informatica Economică*, 20(1), 29-38.

Samih, H. (2019). Smart cities and internet of things. *Journal of Information Technology Case and Application Research*, 21(1), 3-12.

Santos, H., Furtado, V., Pinheiro, P., & McGuinness, D. L. (2017). Contextual data collection for smart cities. *arXiv preprint arXiv:1704.01802*. Retrieved from https://arxiv.org/abs/1704.01802.

Zhang, Q., Huang, T., Zhu, Y., & Qiu, M. (2013). A case study of sensor data collection and analysis in smart city: Provenance in smart food supply chain. *International Journal of Distributed Sensor Networks*, 9(11), 382132.