WILEY | Hindawi

Research Article

Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making

Bhagya Nathali Silva, Murad Khan, and Kijun Han

School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea

Correspondence should be addressed to Kijun Han; kjhan@knu.ac.kr

Academic Editor: Jaime Lloret

The concept of the smart city is widely favored, as it enhances the quality of life of urban citizens, involving multiple disciplines, that is, smart community, smart transportation, smart healthcare, smart parking, and many more. Continuous growth of the complex urban networks is significantly challenged by real-time data processing and intelligent decision-making capabilities. Therefore, in this paper, we propose a smart city framework based on Big Data analytics. The proposed framework operates on three levels: (1) data generation and acquisition level collecting heterogeneous data related to city operations, (2) data management and processing level filtering, analyzing, and storing data to make decisions and events autonomously, and (3) application level initiating execution of the events corresponding to the received decisions. In order to validate the proposed architecture, we analyze a few major types of dataset based on the proposed three-level architecture. Further, we tested authentic datasets on Hadoop ecosystem to determine the threshold and the analysis shows that the proposed architecture offers useful insights into the community development authorities to improve the existing smart city architecture.

1. Introduction

The novel concept of "connected everyday objects" over the existing network has been evolved with the emergence of the smart devices. The tremendous growth of the devices connected to the network has expanded the boundaries of conventional networks. This major breakthrough introduced Internet of Things (IoT) as the third wave of the web after static pages web (WWW) and social networking web [1, 2]. The IoT is an unceasingly growing network, capable of identifying and sharing data autonomously among heterogeneous devices, which are uniquely addressable. IoT has become the spotlight of attention among multiple interest groups due to the advancement of embedded device technology and rapid increase in the number of devices. The IoT concept has been matured with the attention of multiple interest groups and with the advancement of embedded device technology. This comes up with its productive applications like smart home, smart city, smart health, and so forth [3–6]. The

smart city notion is initially coined with the aim of utilizing public services and resources efficiently to increase the quality of services offered to the urban citizens [7]. In fact, the offered services, that is, transportation, parking, surveillance, electricity, healthcare, and so forth, are optimized with the autonomous data collection via the heterogeneous devices connected to the urban IoT. It is essential to process a large amount of data on a real-time basis in order to serve the service requests efficiently. Consequent to the immense increase in data volume, the general data processing and analytical mechanisms become impotent to satisfy the real-time data processing demand. Hence, the collaboration with Big Data analytics is considered to be the ideal first step towards a smarter city. It assures flexible and real-time data processing followed by intelligent decision procedures [8]. As a result of adopting Big Data analytics to the urban IoT, this enhances the quality of services provided by the smart city.

In addition, multiple efforts have been made by academic and industrial experts to realize the notion of the

smart city. However, many efforts on individual aspects of interest are seen in the literature [9–11] covering water management, garbage management, parking management, and so forth. Therefore, complete and resilient smart city architecture has become a crucial demand, as lack of integrity deteriorates the practicability. In addition, it has to facilitate autonomous behavior, real-time data processing, real-time decision-making, and smart energy consumption and customization. Thereupon, the processing and analyzing of the colossal amount of data become a necessity. Henceforth, the urban IoT integrates Big Data analytics for the realization of the smart city [12]. For example, a smart meter at a residential building collects the meter reading that is compared with a predefined electricity consumption threshold and, based on the result, the current energy demand is notified to the smart grid. Simultaneously, consumers are notified with the current level of energy consumption, allowing them to manage the energy utilization efficiently. Indeed, the preceding scenario generates a reasonable amount of data for a single house. Moreover, data processing and decision-making should be carried out in a timely manner. Nevertheless, thousands of residential and public infrastructures in the city generate a prodigious amount of data related to a single task as mentioned above. Thus, the unification of data sources and Big Data analytics is considered to be an expedient solution to facilitate real-time operation of the smart city.

Even though the smart city has become a buzzword in the modern technological era, the actual implementation is still in its infancy. In this regard, multiple efforts are made to implement a realistic smart city. An urban IoT, "Padova Smart City," was implemented to provide ICT solutions for the city administration [7]. The framework consists of a data collection system, street lighting monitoring system, and a gateway. By means of the collected environmental parameters, that is, temperature, humidity, and light, it assures the operation of streetlights. SmartSantander test bed in North Spain is used in [8] to determine the potential benefits of Big Data analytics for smart cities. The authors have analyzed temperature, traffic, season, and working days to define a network with many interacting parts, which behave according to individual rules. Smart city architecture from a data perspective is proposed in [13]. The architecture consists of six layers covering multiple aspects of a smart city. Moreover, three-tier pyramidal architecture is proposed in [14] to facilitate transactions among heterogeneous devices across a wireless ubiquitous platform. However, most of the proposed architecture types focus on specific area of interest such as lighting, traffic congestion, and water management. Thus, the claim is valid that there is a necessity of realistic smart city architecture competent enough to make real-time intelligent decisions to uplift the quality of urban IoT services. Figure 1 presents the overview of a conventional smart city that consists of smart community, smart transportation, smart grid, smart water management, and so forth.

In this paper, Big Data analytics are integrated with the smart city architecture to propose a realistic and feasible framework for the deployment of smart cities. The proposed architecture is capable of real-time intelligent decision-making, autonomous data collecting, and user-centric energy

customizing. However, the decision and control management is the most influential component for the realization of a smart city. Hence, the attainment of real-time and prompt decisions has become the utmost goal of the proposed scheme. Also, fusion techniques work to expedite the processing of the enormous amount of collected data in Big Data analytics. In this study, Hadoop is chosen as the storage and processing medium for the heterogeneous data. The Hadoop processing is followed by the generation of intelligent decisions related to the smart city operations. Finally, the actions or events corresponding to the decisions are executed.

The rest of the paper is organized as follows. Section 2 presents a detailed description of the recent literature and smart city management based on Big Data analytics. Section 3 gives a brief description of the proposed architecture. The results and analysis are presented in Section 4. Finally, the conclusion is outlined in Section 5.

2. Related Work

The rapid development of the smart city system diverts the focus of many researchers and architects towards an efficient communication and standard architectural design. Standardizing the smart city models can provide various benefits to the researchers and engineers in different contexts, naming standalone communication paradigm, detailed layering architecture, processing of information in real time, and so forth. In addition, the smart city architecture covers a variety of research approaches ranging from abstract concepts to a complete set of services. Recently, the researchers are working on deriving various solutions to present generic architecture of IoT-based smart city. Similarly, various schemes have been proposed in the current literature that follows thorough experimentation and test bed based simulations to overcome the challenges. A scheme based on experimenting a complete set of smart city services on various test bed modules has been proposed in [15]. The authors in [15] developed the physical implementation of a large-scale IoT infrastructure in a Santander city. The experimental facility is designed to be so user-friendly so that the experimenter can test the facility in different urban environments and smart city planning. A variety of new mechanisms were developed following the Santander city requirements. These mechanisms include mobility support, security and surveillance systems, large-scale support, scalability, and heterogeneity in a smart city environment. The test bed results show that the proposed architecture covers several challenges in the current literature. However, the data collected from various sensors is not tested for future urban planning and designing. Therefore, the architecture can guarantee better services in one environment but may show poor performance in another environment. Similarly, the demands of the user in an IoT-based smart environment rapidly change. Hence, it decreases the chances of understanding the context and dynamicity of the IoT-based smart user. On the other hand, the IoT is not yet matured to deploy it as generic standard for designing smart services such as smart homes and smart cities because of the following two major reasons: (1) the current IoT-based solutions are limited to specific application domain
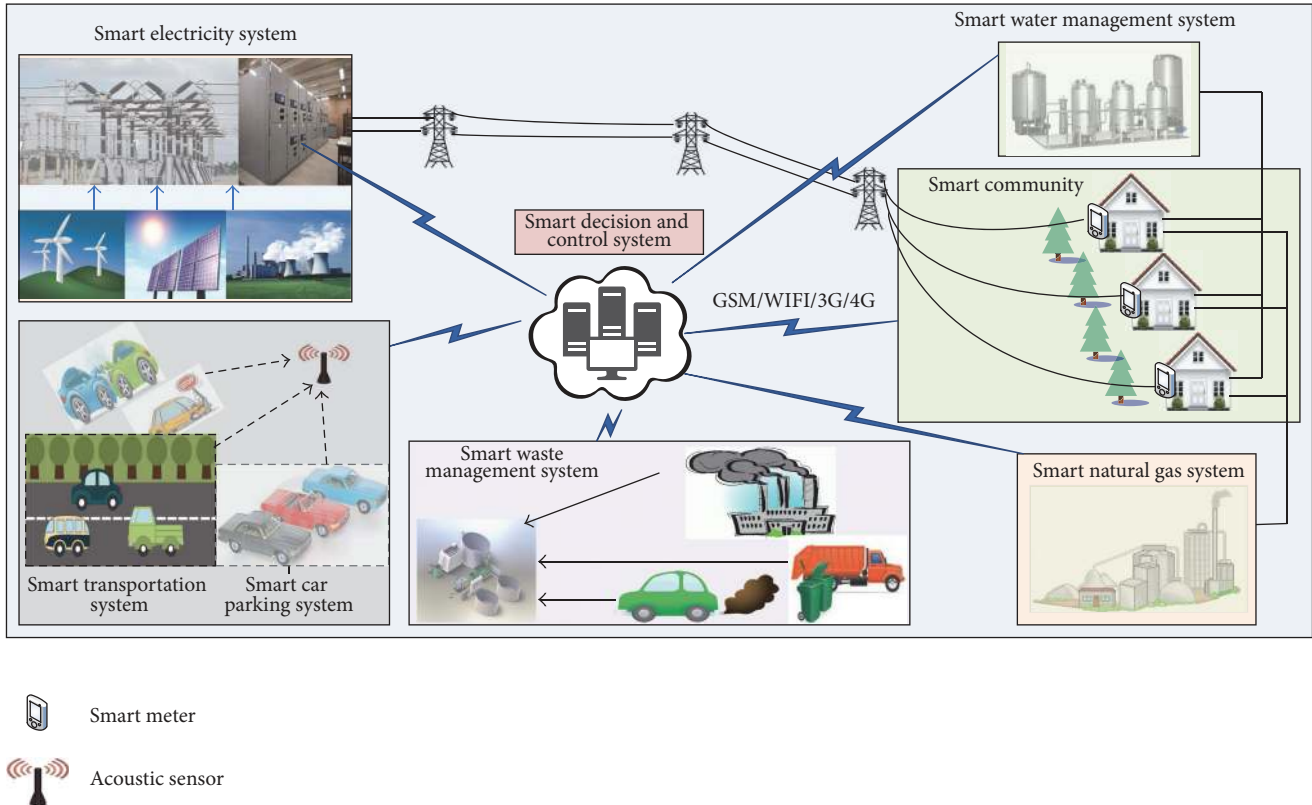
Figure 1: Typical smart city architecture.

and (2) new technologies and optimization techniques are good in one area but may be not in another. For example, wireless sensor networks (WSN) suffered high packet loss in a heterogeneous wireless environment. In addition, the deployment of IoT for one particular purpose such as waste management, air quality, noise pollution, and so forth does not reflect a standard solution [16–18]. Similarly, wireless local area network can provide low-cost services but it provides a narrow coverage compared to other technologies. Therefore, the researchers have come up with several solutions which ultimately lead to a generic communication model covering a wide set of services [19–22]. Moreover, a generic communication model can be achieved by integrating the WSN with the existing infrastructure and, thus, helps in achieving a real IoT environment with multifaceted architecture [23].

In order to design efficient and generic smart city architecture, the Big Data that is obtained from the existing smart city should be carefully examined and analyzed. The process of collection of data can be done by placing sensors in various locations in a smart home or smart city environment. Offline processing of Big Data can help in designing and planning of the urban city environment. However, it does not help in performing real-time decisions. Various techniques based on Hadoop ecosystem are developed to analyze the data for better usage and designing of the services for a smart city. For example, architecture called City Data and Analytics Platform (CiDAP) has been proposed in [8]. The authors developed layered architecture of data processing between the data sources and applications. The entire architecture consists of

different parts such as data collection unit (IoT broker) and IoT agent (a repository to store data), a Big Data processing module, and a city model communication server providing the communication facilities with an external object. The data from different applications is collected and is passed to the city model server. The city model server processes the data and passes it to the IoT broker. The IoT broker separates the data based on the sensors' IDs and assigns an index number to the data. Finally, the IoT broker sends data to IoT agent for further processing. The proposed scheme achieves a higher throughput in processing of the data. Similarly, various other projects are developed based on Big Data analytics such as SCOPE [24] and FIRWARE [25]. These projects help in various aspects and provide different mechanisms to deal with Big Data in the real-time environment. However, they are not openly available to the researchers and engineers for use in different environments.

The wireless-based technologies such as wireless sensor network, wireless LAN, 3G/4G, and LTE play a vital role in providing *always best-connected* services in the smart city environment [26]. These technologies are employed in various fields and sectors of the smart city such as health care, transportation, schools, universities, and marketing. Moreover, these technologies enable a real-time communication with the smart cities devices. Thus, the data generated by the smart city sensors can be efficiently processed to take real-time decisions. However, real-time decisions require fast and efficient data processing tools. For example, Hadoop presents a solution to process the big amount of data in possible time.

Table 1: The amount of data collected in one year.

Collection frequency	1/day	1/hour	1/30 min	1/15 min
Records collected	365 m	8.75 b	17.52 b	35.04 b
Terabytes collected	1.82 tb	730 tb	1460 tb	2920 tb

m: million, b: billion, and tb: terabyte.

In addition, employing any existing tool to process Big Data depends on three properties of Big Data, that is, velocity, variety, and volume. However, processing a huge amount of data in the minimum possible time and performing real-time decision are a challenging task. Therefore, the recent research presents several models to process the data in the offline form. Thus, the outcomes can be used for management of urban planning. In order to elaborate the idea of urban planning based on Big Data analytics, we present a few example scenarios. The energy consumption recorded by smart meters in a time span of one year is shown in Table 1 [27]. The information clearly illustrates the exponential growth of data generation. The amount of data collected was calculated assuming 5 kilobytes per record [27].

The table shows that the amount of data collected by 1 million meters per 15 mins in one year is equal to 2920 TB. Thus, this high amount of data cannot be processed at once. Therefore, sophisticated tools and techniques are required to process the data and come up with proper planning and management. Similarly, processing the parking data from various parking garages in a smart city can help in designing smart parking systems. The vehicular data from various roads of a city can be used to design a smart transportation system. Moreover, this data can be used in the development of roads and bridges in various places in the smart city. Similarly, several examples of using Big Data analytics in planning and developing of smart cities services are presented in recent literature [17, 28]. However, real-time decision-making and processing on such a large amount of data are still a challenging job. In addition, an efficient smart city can be built by considering the following two points: (1) generic communication model and (2) real-time Big Data analytics.

The above literature reveals some important challenges that need to be addressed, for example, designing a generic communication model, real-time Big Data analytics, and acquisition of data from sensors in a smart city. Therefore, in this paper, we identify the need for an efficient and generic communication model for future smart cities based on Big Data analytics and integration of WSN.

3. Proposed Scheme

The proposed smart city architecture comprises three levels: (1) data generation and acquisition level, (2) data management and processing level, and (3) application level. A brief overview of the proposed smart city architecture is provided in the next subsection followed by detailed description of three levels of the proposed framework.

3.1. Overview. The layering architecture and working flow of the proposed smart city architecture are illustrated in Figure 2. Both layering and workflow are presented in a top-down manner starting from data generation and acquisition level to data management and processing level to application level. The proposed city architecture encompasses smart community development department, smart traffic control department, smart weather forecast department, and smart hospital and health department. The aforementioned components are liable for the collection of heterogeneous data within the city suburbs, thus acting as the bottom level of the proposed framework. These components are further connected with the smart decision and control system via heterogeneous access technologies such as GSM, Wi-Fi, 3G, and 4G. The autonomous decision-making uplifts the reliability as well as the practicability of the proposed scheme. Upon receiving the collected data, intelligent decisions are carried out by the smart decision and control system, situated in the middle level of the smart city framework. Moreover, the middle level regulates the events conforming to the made decisions. The event generation is taken place at the top level (application level), upon the reception of autonomous decisions.

The utmost goal of this study was to exploit realistic smart city architecture to enhance the data processing efficacy to enable real-time decision-making. In this paper, we proposed smart city architecture that incorporates Big Data analytics. In fact, there are previous studies, which integrated Big Data analytics into the smart city architecture. However, the proposed scheme is not a conventional Big Data embedded smart city as it performs explicit data filtration using Kalman filter (KF) prior to the Big Data processing. Data filtration is performed to further expedite the data processing. The KF applies threshold based filtration to distinguish between valuable and noisy data. Thus, it reduces the load that requires further processing. Similarly, we occupied a Hadoop two nodes' cluster for the Big Data processing. As shown in the Results and Data Analysis, the unification of data filtration and system architecture has enhanced the throughput of the smart city, while reducing the processing time. Thus, the proposed scheme was able to fulfill the demand for smart city architecture capable of processing data and making decision in real time.

3.2. Data Generation and Acquisition Level. A realistic smart city not only includes a prodigious amount of data but also includes complex and comprehensive computation and multiple application domains. The realization of the smart city implementation relies on all forms of data and computation due to their indispensability [13]. The smart city notion aims to optimize residential resources, to reduce traffic congestion, to provide efficient healthcare services, and to perform the water management. The acquisition of data associated with the daily operational activities become vital in terms of achieving the preceding aims. However, the data acquisition has become tedious and challenging due to the massive amount of data created by people and other connected devices. For the sake of further processing, the phenomena of interest from the real world are sensed and identified. Consequently, conversion into digital data employs various mechanisms. Low-cost and energy efficient
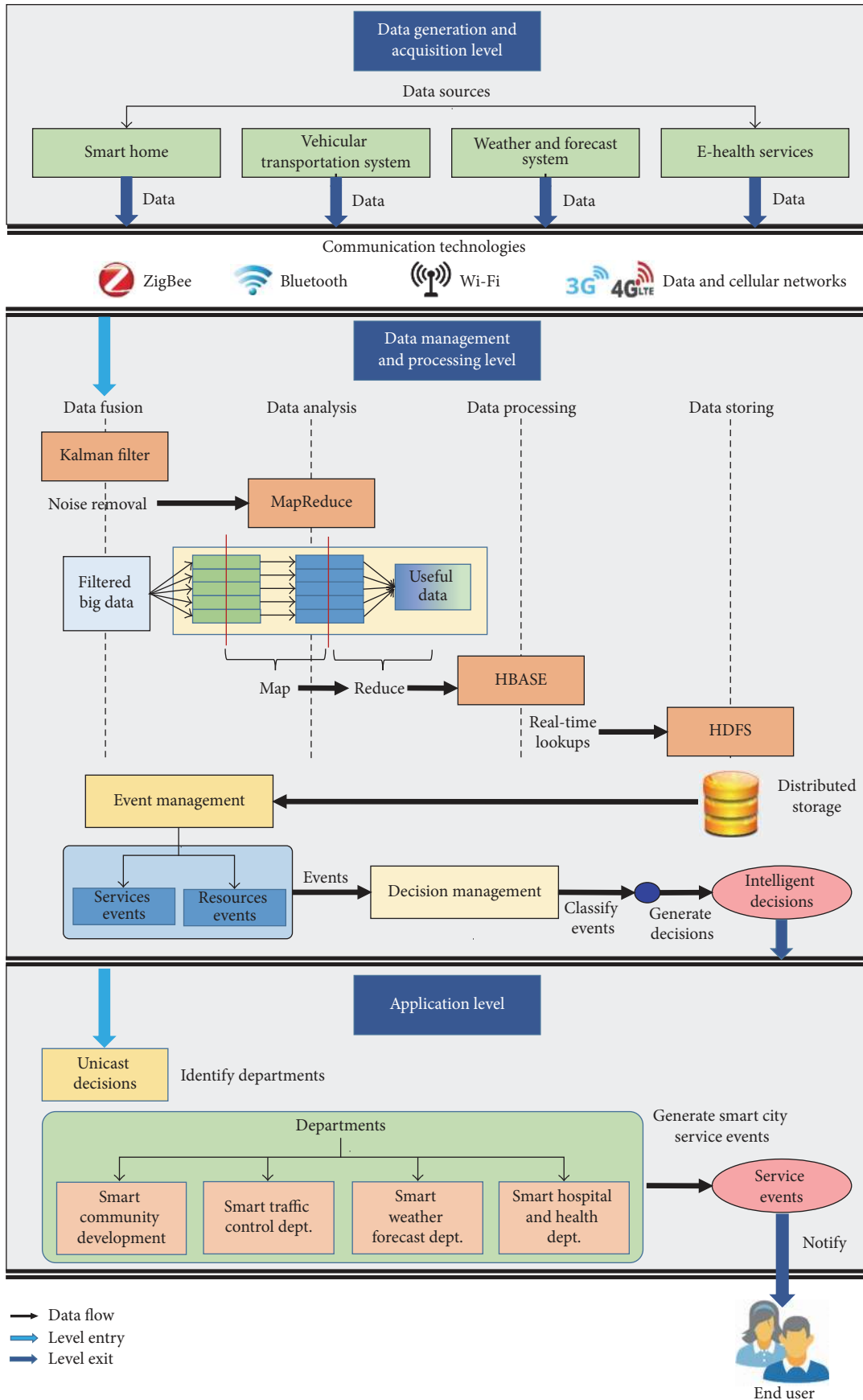
FIGURE 2: Working of the proposed architecture.

sensors have become a promising mechanism to acquire heterogeneous data from the urban IoT. The city becomes smarter, along with the expansion of the number of connected devices [15]. Hence, the realization of the proposed smart city architecture begins with the extensive deployment of heterogeneous sensors within the city suburbs. These sensors are liable for the collection of real-time data from the neighboring environment. The deployed context determines the type of collected data, that is, smart home, vehicular transportation system, healthcare management system, and meteorology system.

The bottom layer of the proposed scheme consists of multiple components. The key concern of the smart home is to enhance the energy utilization of the residential buildings. The home appliances are equipped with a sensor, which determines the real-time energy consumption and moves to the middle layer afterward. The data processing layer defines a threshold value for particular household's energy consumption. A data filtration process is performed by the fusion techniques to determine the values exceeding the threshold and thus optimizes further processing. Consequently, the decisions made at the middle level send to the smart community development in application level, which notifies energy consumption of a particular household to the respective residents. Meanwhile, it empowers the energy usage customization of residential buildings. The prime objective of the vehicular transportation system is to reduce the city traffic congestion. The data processing level defines the mean time that is taken to travel between two stated points. The sensors implanted on the roadsides collect vehicle entrance and departure between two points. The embedded fusion techniques determine the roads with congestion by analyzing the current travel time of stated locations, which exceeds the defined mean time. Thence, vehicular transportation system autonomously generates alternative paths and notifies the travelers via the application level. The utmost goal of the meteorology department is to ascertain the weather conditions and other environmental parameters. For example, the sensors implanted in certain locations determine the carbon monoxide (CO) concentration of the city. These sensors convey the acquired data to the middle level for filtering and processing accordingly to facilitate decision-making and event generation.

The proposed city architecture occupies multiple communication technologies; ZigBee, Bluetooth, Wi-Fi, and data and cellular networks to transmit sensed data to the data management and processing level.

3.3. Data Management and Processing Level. The data management and processing level acts as the mediator between the data acquisition and application levels. Since the crucial processes such as filtering valuable data, analyzing, processing, storing, decision-making, and generating events are carried out in this layer, this layer is considered as the brain of the proposed framework. In order to perform the aforementioned tasks, multiple modalities are embedded into this layer. Initially, the enormous amount of sensed data is filtered by fusion mechanisms to obtain valuable real-time and offline data. The MapReduce paradigm is used for the

data analysis, while manipulation and storing are performed by Hadoop distributed file system (HDFS), HBASE, and HIVE.

The fusion techniques enhance the data processing efficiency by applying data filtration. Kalman filter (KF) is used to perform data filtration in the proposed framework [29]. The KF is an optimal estimator, which removes noise from the sensed data [30, 31]. The working mechanism of KF in different steps for sensor data filtration is shown as follows.

Working of KF for Sensor Data Filtration

(1) Initialization

T_k: state transition model (applied to the previous state f_{k-1})
O_k: observation model
Q_k: covariance of the process noise
R_k: covariance of the observation noise
C_k: control input model (applied to the control vector v_k)
$w_k \sim \mathcal{N}(0, Q_k)$

(2) Computing the new state f_k using the previous state f_{k-1}

$$f_k = T_k f_{k-1} + C_k v_k + w_k$$
$$h_k = O_k f_k + u_k \quad u_k \sim \mathcal{N}(0, R_k) \tag{1}$$

(3) Current state estimation from the previous state

Predicted state

$$\widehat{f}_{k|k-1} = T_k \widehat{f}_{k-1|k-1} + C_k v_k \tag{2}$$

Predicted covariance

$$G_{k|k-1} = T_k G_{k-1|k-1} T_k^T + Q_k \tag{3}$$

(4) Combining current prediction with the current observation

Current observation

$$\widetilde{x}_k = h_k - O_k \widehat{f}_{k|k-1} \tag{4}$$

Observation covariance

$$S_k = O_k G_{k|k-1} O_k^T + R_k \tag{5}$$

Optimal gain

$$K_k = G_{k|k-1} O_k^T S_k^{-1} \tag{6}$$

Update state (prediction and observation)

$$\widehat{f}_{k|k} = \widehat{f}_{k|k-1} + K_k \widetilde{x}_k \tag{7}$$

(5) Update covariance (prediction and observation)

$$G_{k|k} = (I - K_k O_k) G_{k|k-1} \tag{8}$$

It initially assumes the current state f_k is evolved from the previous state f_{k-1}. The current state observation is denoted by h_k. $\widehat{f}_{k|k-1}$ represents the estimation of f at time k, while the estimation accuracy is denoted by $G_{k|k-1}$. It deduces valuable data from a large set of indirect and uncertain data. Since the KF works recursively, it processes data on arrival. Thus, it assures the real-time operation of the smart city. Moreover, it facilitates immediate processing with a minimal memory consumption. As KF removes noise from data, the data processing level utilizes its capability to infer the best estimate from a larger set of real-time data. Thereupon, the KF is manipulated to determine valuable data corresponding to the predefined threshold values. For example, the roadside sensors of the streets and roads generate a massive amount of city traffic data. Nevertheless, further processing of uncongested street data is a superfluous task. Thence, the KF determines best fitting sensed data in accordance with the predefined thresholds. Ultimately, it reduces the amount of futile data resulting in a swift analysis.

The proposed scheme stores and processes data in Hadoop framework. Thus, MapReduce has been selected as the mechanism for analyzing filtered data. MapReduce works in two steps. First is the mapping process where the set of filtered data is converted into another set of data. Next is the Reduce process which combines the data created in mapping process and results in a set of values that are reduced in amount. Data storing and processing play a major role in the realization of a smart city. As shown in Figure 2, the proposed framework utilizes multiple techniques, that is, HDFS, HBASE, HIVE, and so forth, to facilitate the above requirements. The storage demand of the proposed smart city is facilitated by HDFS, which is the primary storage of Hadoop. Since the storage of HDFS is distributed, it augments the MapReduce execution on smaller subsets of larger data cluster. In addition, HDFS assists the scalability demand of the Big Data processing. In order to favor the autonomous decision-making, the real-time read/write facility over the complete cluster is essential. Hence, HBASE is used to enhance the processing speed on Hadoop as it offers real-time lookups, in-memory caching, and server side programming. Further, it enhances the usability and the fault tolerance. HIVE provides querying and managing facility over the large amount of data that resides on the Hadoop cluster. Since SQL cannot be used to query on HIVE, we have used HiveQL to query the data on Hadoop cluster. Finally, the derived intelligent decisions are transferred to the application level of the smart city framework.

3.4. Application Level. Application level resides on top of the proposed framework. Thus, it is liable for the generation of actions corresponding to the conveyed autonomous intelligent decision. The application level is the mediator between data management level and the end user. Figure 3 presents the extended layering structure of the application level that is proposed for performance improvement of service generation. The application level is subdivided into three layers, that is, departmental layer, services layer, and subservices layer. Department layer is the boundary at the data management and processing level. Subservices layer acts as the boundary for end users. The autonomous decisions from the data processing level are unicasted to the specific departmental service, that is, smart community development department, smart traffic control department, smart weather forecast department, and smart hospital and healthcare department. The intelligent decisions of the data processing level describe the decision according to a shared vocabulary (ontology). The ontology is used to unicast the events throughout the application level. The respective departments distinguish the high-level events and the low-level events. The high-level events are stored at the departmental level and are forwarded in unicast to the recipients, whereas the low-level events are not moved further. Sequentially, the corresponding service event layer's component receives the unicast event from the departmental events. For example, the service events, smart home and waste management, are readily available to receive the departmental events from the smart community development department. Similarly, the service events are further categorized into subservice events, that is, water management and energy management under the smart home services events. The subservices events layer generates the respective event and transmits to the embedded notification component. Finally, the notification component determines the specific recipient with respect to generated event. Accordingly, it notifies the user with the generated event for the event execution.

Assume the sensors implanted on a particular city observe a street congestion. The congestion level is analyzed at the data processing level. Subsequently, the data processing level generates the appropriate intelligent decision. At the same time, the decision is communicated to the application level. The ontology determines the respective departmental event according to the decision message, that is, street congestion. Accordingly, the event is unicasted to the smart traffic control department at the application level. The departmental level determines service event component as traffic congestion. Sequentially, the generated event is forwarded to the subservice level of alternative paths. Finally, the alternative path event is notified to the respective recipient via the notification component of the application level. Moreover, the paths are notified to the potential travelers, who may enter the congested street. The smart traffic control department determines the fact by the GPS destination check and the current positioning of the vehicle.

4. Results and Data Analysis

The designing of a smart city free from existing issues entirely depends on the processing and analysis of the previous data that is obtained from various sources, that is, transportation, community department, health care, and so forth. We obtain such data from various authentic sources as is given in "*Working of KF for Sensor Data Filtration*" part. Initially, the data is fuzzy and consists of raw data entries. Therefore, on top of Hadoop system, we filter the data through KF according to our requirements that result in significant optimization of the processing time and performance efficiency of Hadoop. Moreover, the filtration process helps in processing the real-time data with less amount of time.
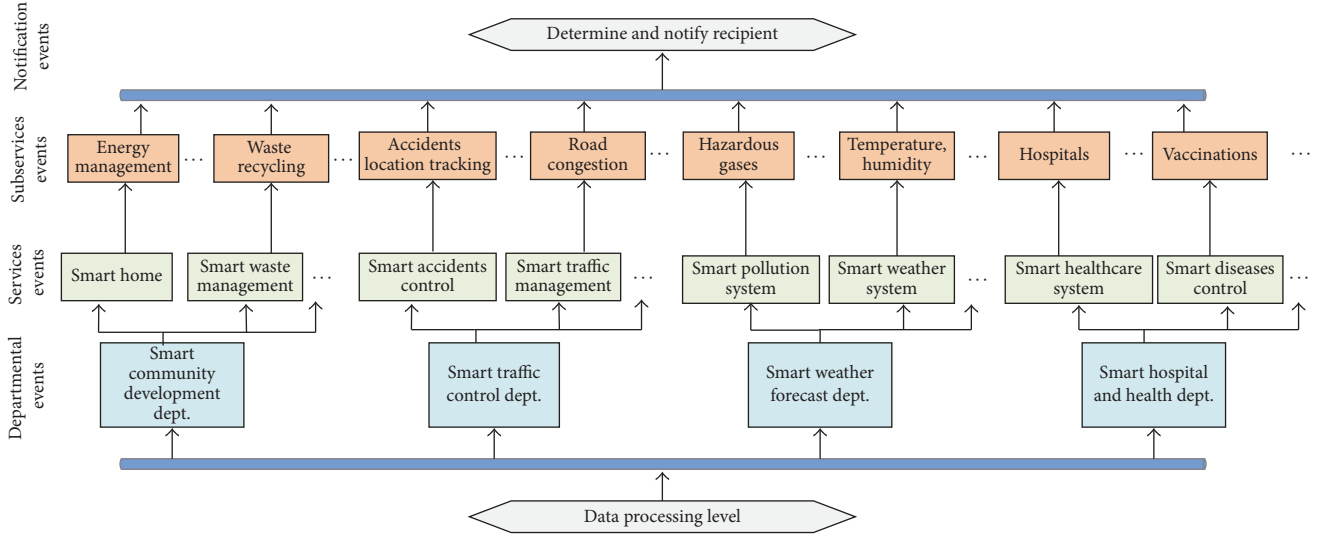
Figure 3: Event generation and processing at application level.

Table 2: Datasets information.

Sources	Dataset	Size
Surrey city, Canada [32]	Water consumption	4 MB
Aarhus city, Denmark [33]	Traffic data	3.04 GB
Aarhus city, Denmark [34]	Pollution data	77.25 MB
Aarhus city, Denmark [33]	Parking lots	0.20 MB

4.1. Dataset Information. The datasets are obtained from various authentic and reliable sources. These datasets include the following: (1) the energy and water consumption data of smart homes of survey, Canada, is obtained from the meter readings of around 61263 houses [32] and (2) the transportation and vehicular data used in the proposed analysis is obtained from number of vehicles on various roads in Aarhus city, Spain [33]. The datasets used for the analysis are openly available and authenticated. Water consumption data are covered by Open Government License of City of Surrey, Canada. Traffic data, parking lot data, and pollution data are semantically annotated datasets for the CityPulse EU FP7 project and the data is licensed under Creative Commons Attribution 4.0 International License. The dataset contains very useful information, for example, the number of vehicles and their average speed between two points on a road. (3) The parking lot dataset contain the information of various parking lots of Aarhus city, Denmark. The dataset is generated from various parking garages from May 2014 to Nov 2014 and (4) the pollution dataset consists of the information of various toxic gases such as ozone, carbon, sulfur and nitrogen dioxide, and so forth [34]. The dataset consists of the detail of some other hazardous materials but we filter out the entries of those materials for performing real-time decisions. Information corresponding to each dataset is mentioned in Table 2.

4.2. Results, Analysis, and Event Generations. The data collected from various sources is analyzed with two-node

Hadoop cluster on Ubuntu 16.04 LTS having Core I5 processor and 8 GB RAM. The rationale behind this analysis is to determine normal threshold values for the actual implementation as well as to evaluate the performance of the proposed scheme in terms of processing time and data throughput. Moreover, various thresholds are defined on the output from Hadoop system. The threshold is specific to the dataset size used for the analysis. The threshold values are shown in Table 3.

Whenever the amount of data in a particular time exceeds the normal threshold, an event is generated to the respective department. In order to validate the proposed event generation system, the time taken to process the data, generate an event, and send it to the respective user (T_α) via the respective department is shown in Table 3. The experiment reveals that as the dataset size increases the time required to process it significantly increases. However, in the case of real-time processing the data is always available in streaming form and, therefore, the size of the data does not affect the system. But it is essential to make a system that directly processes the data with high speed. Therefore, the proposed filtering helps in minimizing the processing time of the Hadoop system.

The proposed system efficiently processes the data and generates various events such as traffic intensity warnings when the number of vehicles increases on a particular road compared to the normal threshold. Figure 4 illustrates number of vehicles between two defined locations for a period. The threshold value obtained from the analysis is used to determine valuable data and generate events accordingly. The threshold value was eight vehicles between two defined locations on a certain path. We set the normal threshold to 8 vehicles on a specific portion of the road. The proposed system generates the warning events in real time and sends them to the respective department. The department then broadcast the message to the vehicles headed towards that particular road. A smart parking lot system helps the citizens to get the information of empty parking lots in the surrounding

TABLE 3: Threshold and event generation time analysis.

Dataset	Size	Threshold	T_α
Water consumption	4 MB	80 cubic liters	11.23 s
Traffic data	3.04 GB	8 vehicles	212.88 s
Pollution data	77.25 MB	80%	16.97 s
Parking lots	0.20 MB	<10/parking garage	3.67 s
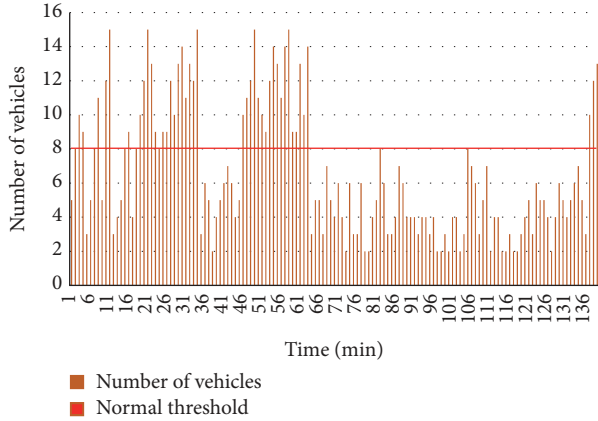
FIGURE 4: A number of vehicles at different roads in Aarhus, Denmark.

FIGURE 5: Empty parking lots in various places in Aarhus, Denmark.

area. The parking dataset is carefully analyzed and the empty parking lots information is sent to the respective department. The department maintains a database of empty parking lots. The citizens can check the database before parking the vehicle in the surrounding parking lots. Thus, it helps the citizens to find an empty parking lot without physical checking of the entire parking garage. Moreover, the entry for a parking lot is deleted when a citizen occupies it. Figure 5 shows various empty parking lots during a different time of the day. Initially, the Bruun city has more parking lots during daytime. However, the parking lots are rapidly filled with the passage of time. The parking lots also depend on the population and number of vehicles in a city. Thus, using the data obtain from different parking lots can be used to fulfill the parking lots requirements of the city. The shopping malls, departmental stores, and offices parking lot's data can be analyzed for better management to serve the customers with the information related to the empty parking lots. Moreover, a citizen can reserve a parking lot prior to reaching the destination.

The excessive water usage can become a critical problem in near future. Therefore, we analyzed the water usage to come up with an appropriate solution for water management. The dataset contains the water consumption information of Surrey, Canada. Figure 6 shows that each house consumed more than 80000 to 90000 liters of water each month. The normal threshold for household water consumption was obtained from the data analysis performed on the water consumption dataset. This amount of water consumption is very high and therefore in future it can become a serious problem. However, the proposed decision mechanism generates various events to the water management department to take necessary actions
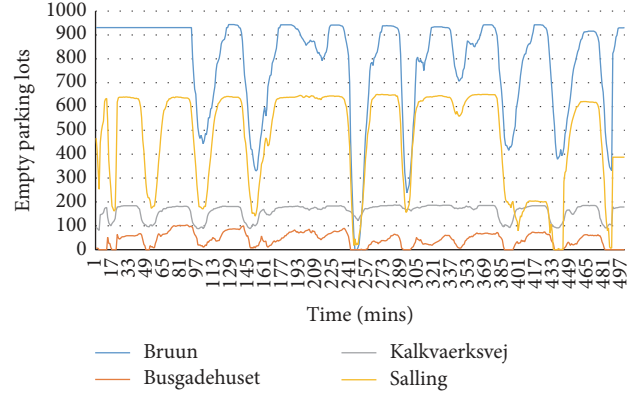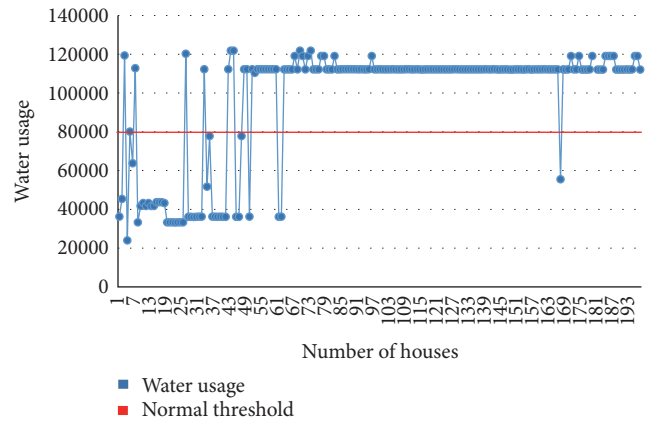

FIGURE 6: Water usage at different houses of Surrey, Canada.

to control the water consumption. The warning event is generated when the water consumption exceeds 80000 liters. Moreover, new water generations methods can be developed to fulfill the requirements of the citizens.

As the number of factories and vehicles is increasing, the waste production and pollution rise dramatically. Thus, controlling waste management and pollution is becoming a critical issue. In order to design a solution to handle these consequences, we analyzed the pollution data of Aarhus city at the various times of day. As shown in Figure 7, the quantity of the Ozone (O_3) is particularly high at the different times of the day. The decision system generates various events to the weather and forecast and health department to circulate a message among citizens to take great care while visiting the
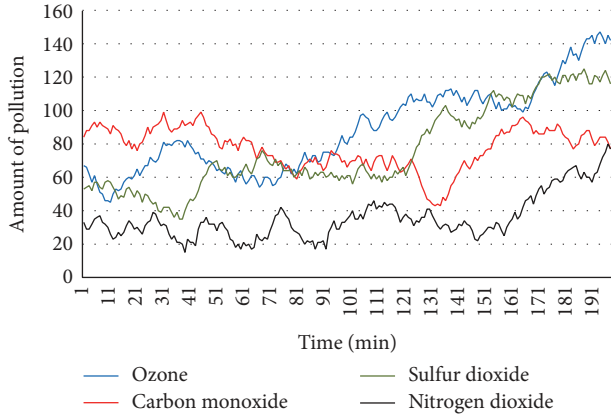
FIGURE 7: Amount of pollution at different time of day in Aarhus, Denmark.
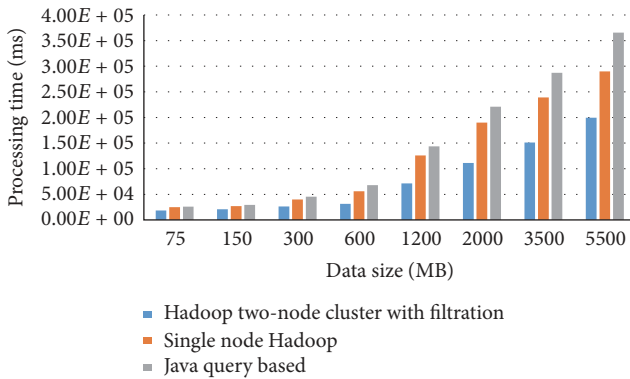

FIGURE 8: Processing time analysis.

FIGURE 9: Efficiency of the system in context of throughput.

polluted areas. Moreover, the environment control department can take necessary actions against those firms which are generating high pollution. Thus, using the pollution data can help in the future planning of the smart city as well.

The processing time of the proposed system is compared with the single node Hadoop and Java query system. The filtration of data and removing irrelevant entries from the data significantly reduce the processing time. As shown in Figure 8, as the dataset size increases, the single node Hadoop and Java query based systems required high processing time. Thus, implanting such systems in the real-time data processing environment affects the proposed decision and event generations schemes. Moreover, the efficiency of the proposed system in the context of throughput is measured as shown in Figure 9. Initially, the single Hadoop node and Java query based system process the data with similar speed. But, with the increase in dataset size, the speed of processing is highly decreased. However, the proposed scheme efficiency is significantly high compared to the single node Hadoop and Java query based system.

The analysis also shows that a city can be made smarter by analyzing data obtained from various departments. Moreover, the living style of the citizens can be improved and the comfort level can be increased by informing the citizens with
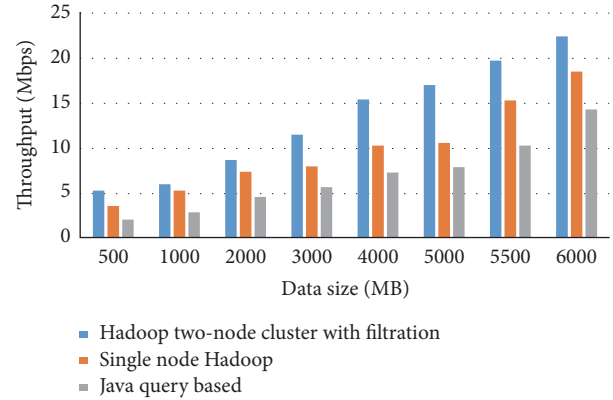
the usage of various services such as electricity and water consumption, traffic intensity on a road.

5. Conclusion and Future Work

The extensive expansion of IoT has encouraged the urban networks to be smarter, coining the notion of "smart cities." However, the realization of the smart city is still emerging, since the transformation of the conventional city operations requires novelty, networking, and processing ability of voluminous data. Therefore, the researchers and industrial experts are keen on shaping baseline architecture for a realistic smart city. In this paper, we proposed architecture of a smart city based on Big Data analytics. The key concern of this study is to ensure intelligent decision management and control center, which mediates the data acquisition sources and applications. Based on testing various types of datasets, we showed how Big Data can be used for future smart cities development and planning based on the existing data from various sources. However, our system is designed for specific goals and does not reflect a solution in general to every system present in a smart city. Moreover, a scalability option is provided in order to extend the current work in future.

As is already mentioned this work targets specific issues of a smart city to facilitate a more advanced environment for testing data in real time as well as offline. The data fusion functionality is used to reduce processing of Hadoop ecosystem on irrelevant and inappropriate data. Multiple technologies are used on top of the Hadoop storage to facilitate analysis and decision-making processes. Finally, real-world dataset of Surrey (Canada) and Aarhus (Denmark) cities are analyzed to derive the threshold values. In this study, we conceptually proposed the threefold smart city architecture for real-time decision-making. In future endeavors, we plan to carry out a simulated experiment to confirm the accuracy and efficiency of the proposed framework. Moreover, we plan to evaluate the generalizability of the proposed model, in order to standardize this smart city architecture.

Competing Interests

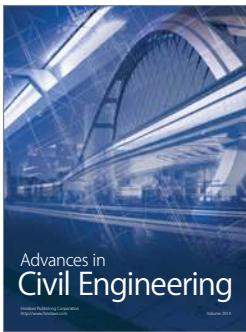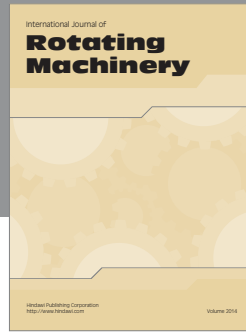The authors declare that they have no competing interests.

Acknowledgments

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*, vol. 17, no. 4, pp. 2347–2376, 2015.

[2] M. A. Razzaque, M. Milojevic-Jevric, A. Palade, and S. Clarke, "Middleware for internet of things: a survey," *IEEE Internet of Things Journal*, vol. 3, no. 1, pp. 70–95, 2016.

[3] 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*, vol. 1, no. 2, pp. 112–121, 2014.

[4] J. Pan, R. Jain, S. Paul, T. Vu, A. Saifullah, and M. Sha, "An Internet of Things framework for smart energy in buildings: designs, prototype, and experiments," *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 527–537, 2015.

[5] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: a comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.

[6] M. Khan, B. N. Silva, and K. Han, "Internet of things based energy aware smart home control system," *IEEE Access*, vol. 4, pp. 7556–7566, 2016.

[7] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.

[8] B. Cheng, S. Longo, F. Cirillo, M. Bauer, and E. Kovacs, "Building a big data platform for smart cities: experience and lessons from santander," in *Proceedings of the 4th IEEE International Congress on Big Data (BigData Congress '15)*, pp. 592–599, New York, NY, USA, July 2015.

[9] T.-H. Kim, R. C. Wai-Chi Fang, S. Mohammed, O. Gervasi, and A. Stoica, "Ubiquitous sensor networks and its application," *International Journal of Distributed Sensor Networks*, vol. 8, no. 7, pp. 1–3, 2012.

[10] R. E. Barone, T. Giuffrè, S. M. Siniscalchi, M. A. Morgano, and G. Tesoriere, "Architecture for parking management in smart cities," *IET Intelligent Transport Systems*, vol. 8, no. 5, pp. 445–452, 2014.

[11] M. Khan, S. Din, S. Jabbar, M. Gohar, H. Ghayvat, and S. C. Mukhopadhyay, "Context-aware low power intelligent SmartHome based on the Internet of things," *Computers & Electrical Engineering*, vol. 52, pp. 208–222, 2016.

[12] A. J. Jara, D. Genoud, and Y. Bocchi, "Big data in smart cities: from poisson to human dynamics," in *Proceedings of the 28th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA '14)*, pp. 785–790, IEEE, Victoria, BC, Canada, May 2014.

[13] W. Rong, Z. Xiong, D. Cooper, C. Li, and H. Sheng, "Smart city architecture: a technology guide for implementation and design challenges," *China Communications*, vol. 11, no. 3, pp. 56–69, 2014.

[14] S. V. Nandury and B. A. Begum, "Smart WSN-based ubiquitous architecture for smart cities," in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI '15)*, pp. 2366–2373, Kochi, India, August 2015.

[15] L. Sanchez, L. Muñoz, J. A. Galache et al., "SmartSantander: IoT experimentation over a smart city testbed," *Computer Networks*, vol. 61, pp. 217–238, 2014.

[16] N. Maisonneuve, M. Stevens, M. E. Niessen, P. Hanappe, and L. Steels, "Citizen noise pollution monitoring," in *Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government*, Puebla, Mexico, May 2009.

[17] X. Li, W. Shu, M. Li, H.-Y. Huang, P.-E. Luo, and M.-Y. Wu, "Performance evaluation of vehicle-based mobile sensor networks for traffic monitoring," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1647–1653, 2009.

[18] T. Nuortio, J. Kytöjoki, H. Niska, and O. Bräysy, "Improved route planning and scheduling of waste collection and transport," *Expert Systems with Applications*, vol. 30, no. 2, pp. 223–232, 2006.

[19] G. Santucci, "Internet of the future and internet of things what is at stake and how are we getting prepared for them," in *Proceedings of the eMatch Conference*, pp. 1–20, Oslo, Norway, September 2009.

[20] T. Ojala, "Open urban testbed for ubiquitous computing," in *Proceedings of the International Conference on Communications and Mobile Computing (CMC '10)*, pp. 442–447, IEEE, Shenzhen, China, April 2010.

[21] S. Jabbar, A. A. Minhas, M. Imran, S. Khalid, and K. Saleem, "Energy efficient strategy for throughput improvement in wireless sensor networks," *Sensors (Switzerland)*, vol. 15, no. 2, pp. 2473–2495, 2015.

[22] S. Kolozali, M. Bermudez-Edo, D. Puschmann, F. Ganz, and P. Barnaghi, "A knowledge-based approach for real-time iot data stream annotation and processing," in *Proceedings of the 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom), IEEE and Cyber, Physical and Social Computing(CPSCom), IEEE Internet of Things (iThings)*, Taipei, Taiwan, 2014.

[23] M. Zorzi, A. Gluhak, S. Lange, and A. Bassi, "From today's INTRAnet of things to a future INTERnet of things: a wireless- and mobility-related view," *IEEE Wireless Communications*, vol. 17, no. 6, pp. 44–51, 2010.

[24] SCOPE A Smart-city Cloud-based Open Platform and Ecosystem, https://www.bu.edu/hic/research/highlighted-sponsored-projects/scope/.

[25] FIWARE Open Source Platform, https://www.fiware.org/.

[26] M. Gohar, J.-G. Choi, S.-J. Koh, K. Naseer, and S. Jabbar, "Distributed mobility management in 6LoWPAN-based wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 620240, 12 pages, 2015.

[27] T. Dunne, *Big Data, Analytics, and Energy Consumption*, Lavastorm Agle Analytics, Boston, Mass, USA, 2012.

[28] I. J. Lee, "Big data processing framework of road traffic collision using distributed CEP," in *Proceedings of the 16th Asia-Pacific Network Operations and Management Symposium (APNOMS '14)*, Hsinchu, Taiwan, September 2014.

[29] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control Theory and Applications*, vol. 4, no. 8, pp. 1303–1318, 2010.

[30] L. Kleeman, Understanding and Applying, http://biorobotics.ri .cmu.edu/papers/sbp_papers/integrated3/kleeman_kalman_basics .pdf.

[31] D. Li, S. Kar, J. M. F. Moura, H. V. Poor, and S. Cui, "Distributed Kalman filtering over massive data sets: analysis through large deviations of random Riccati equations," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1351–1372, 2015.

[32] Dataset Water meters, http://data.surrey.ca/dataset/water-meters.

[33] T. Dataset, Dataset Collection, http://iot.ee.surrey.ac.uk:8080/ datasets.html#traffic.

[34] P. Dataset, Dataset Collection, http://iot.ee.surrey.ac.uk:8080/ datasets.html#pollution.

Journal of
Engineering

The Scientific
World Journal

International Journal of
Rotating
Machinery

Journal of
Sensors

International Journal of
Distributed
Sensor Networks

Advances in
Civil Engineering

Journal of
Control Science
and Engineering

Journal of
Robotics

Journal of
Electrical and Computer
Engineering

Advances in
OptoElectronics

VLSI Design

International Journal of
Navigation and
Observation

Modelling &
Simulation
in Engineering

International Journal of
Aerospace
Engineering

International Journal of
Chemical Engineering

International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components

Shock and Vibration

Advances in
Acoustics and Vibration

Hindawi

Submit your manuscripts at
https://www.hindawi.com