Accepted Manuscript

Integration of Big Data analytics embedded smart city architecture with RESTful web of things for efficient service provision and energy management

Bhagya Nathali Silva, Murad Khan, Kijun Han

PII: S0167-739X(17)30517-4
DOI: http://dx.doi.org/10.1016/j.future.2017.06.024
Reference: FUTURE 3522

To appear in: *Future Generation Computer Systems*

Received date : 31 March 2017
Revised date : 25 May 2017
Accepted date : 20 June 2017

Please cite this article as: B.N. Silva, M. Khan, K. Han, Integration of Big Data analytics embedded smart city architecture with RESTful web of things for efficient service provision and energy management, *Future Generation Computer Systems* (2017), http://dx.doi.org/10.1016/j.future.2017.06.024

Integration of Big Data analytics embedded smart city architecture with RESTful web of things for efficient service provision and energy management

[1]Bhagya Nathali Silva, [2]Murad Khan, [3]Kijun Han

[1,3]School of Computer Science and Engineering, Kyungpook National University, Daegu, Korea

[2]Department of Computer Science, Sarhad University of Science and information Technology, Peshawar, Pakistan

[1]nathalis@netopia.knu.ac.kr, [2]murad.csit@suit.edu.pk, [3]kjhan@knu.ac.kr

Corresponding Author: Kijun Han

Abstract—Emergence of smart things has revolutionized the conventional internet into a connected network of things, maturing the concept of Internet of Things (IoT). With the evolution of IoT, many attempts were made to realize the notion of smart cities. However, demands for processing enormous amount of data and platform incompatibilities of connected smart things hindered the actual implementation of smart cities. Keeping it in view, we proposed a Big Data analytics embedded smart city architecture, which is further integrated with the web via a smart gateway. Integration with the web provides a universal communication platform to overcome the platform incompatibilities of smart things. We introduced Big Data analytics to enhance data processing speed. Further, we evaluated authentic data sets to determine the threshold values for intelligent decision-making and to present the performance improvement gained in data processing. Finally, we presented a representational state transfer (RESTful) web of things (WoT) integrated smart building architecture (smart home) to reveal the performance improvements of the proposed smart city architecture in terms of network performance and energy management of smart buildings.

Keywords: Smart city, Big Data analytics, Smart Home, Web of Things, RESTful Architecture

1 Introduction

The emergence of smart things has highly favored the notion of connecting everyday objects via the existing networks. The dramatic increase of smart devices led towards the realization of IoT [1]. The IoT is a continuously growing network that autonomously identifies and shares data among uniquely addressable devices [2]. A the concept matures, IoT notion is diversified into multiple interest areas pioneering numerous applications i.e. smart home, smart city, smart grid, and many more [3, 4, 5]. Initially, smart city concept was introduced to enhance the quality of life of urban citizens by effective utilization of public resources and services [6]. Smart city is built upon the IoT concept including smart community, smart transportation, smart energy, etc. [7]. The urban IoT optimizes transportation, surveillance, healthcare, and energy management with aid of autonomous data collection and data sharing. Accordingly, heterogeneous smart devices should be able to share collected smart city data at the application level despite of the platform variations. Consequently, it creates a significant demand for a common platform that facilitates successful cross platform communication among heterogeneous smart things. However, building a completely new universal platform is rigorous. So that, redefining an existing platform seemed to be the most feasible approach. Thus, academic and industrial experts identified web as a suitable candidate to offer cross platform communication for smart devices included in IoT applications i.e. smart home and smart city [8]. Thereupon, WoT notion was introduced to use web-associated technologies in IoT. In addition to the heterogeneity, smart city realization is further challenged by processing efficacy of enormous amount of data. In order to meet service requests, smart city architecture should allow real-time data processing. Therefore, embedding Big Data analytics to the IoT environment seemed to be an excellent fit that ensures flexible, reliable, and real-time data processing and decision-making [9].

In the recent past, multiple interest groups have focused on improving the usability of smart city architectures. Nevertheless, a majority of the attempts were designed to uplift individual aspects of smart cities i.e. water management, parking management, and so forth [10, 11]. Consequently, the practicability of smart city has been deteriorated due to lack of completeness. Therefore, a complete smart city architecture has become a crucial demand in the modern technological era. Further, heterogeneous smart devices in smart city architecture generates colossal amount of data. The rapid increase of data volumes exponentially degrades the performance of conventional data processing mechanisms. Thus, it is essential to utilize real-time processing mechanisms to make smart cities efficient and responsive [7]. In [6], an urban IoT was implemented to serve city administration tasks. It consists of a data collection system, street light monitoring system, and a gateway. SmartSantander testbed is used in [6] to determine the benefits of embedding Big Data analytics in smart cities. The emergence of wireless sensor networks (WSN) has improved smart city applications with aid of connected sensors. However, the improvements rely on effective communication protocols that consumes less power. Bluetooth, IEEE 802.15.4, IPv6 over low power wireless personal area network (6LoWPAN), and constrained application protocol (CoAP) are low power communication protocols, which are widely accepted for IoT applications [12]. Afore stated protocols serve smart thing integration at the network level. As a result, connecting heterogeneous smart devices at the application level of smart city architecture is hindered due to the platform incompatibilities. So that, integrating a cross-platform communicator into smart city architecture is much required to enable platform independent seamless communication among smart devices. The authors of [13] proposed a conceptual three tier pyramidal architecture with a wireless ubiquitous platform. Even though, above stated studies have addressed vital areas of smart city implementation, a complete smart city architecture capable of making intelligent decisions based on real-time data processing, while ensuring platform independency at the application level is still demanding. Addressing these challenges of smart city application will support to enhance the quality of service provision and efficient energy management of the network and smart buildings.

In this paper, we propose a complete smart city architecture embedded with Big Data analytics. Compared to traditional data analytics, Big Data analytics approach can extract information that is more intelligent, while maximizing data processing speed and improving reliability of decision-making. The urban IoT collects and shares data autonomously. So that, we familiarized the WoT concept into smart city architecture to make smart things accessible and manageable via open web standards. Smart city services are integrated with the web using a smart gateway to allow cross-platform communication at the application level. RESTful application program interfaces (API) expose services to the remote users. Smart home energy management is presented as an example scenario of controlling service operations using RESTful APIs. The proposed scheme achieved efficient home energy management by enabling remote web accessibility to smart objects integrated with WoT. Henceforth, the proposed smart city successfully improves data processing speed, energy management of smart city owing to embedded Big Data analytics and web integration via RESTful smart gateway. The rest of the paper is organized as follow. Section 2 elaborates on the state of the art on smart cities collaborating Big Data analytics and platform compatibility concerns for smart cities. The proposed is presented in Section 3. The results of the study and discussion is given in Section 4. Finally we concluded the paper in Section 5.

2 Related Work

A full-scale smart city architecture benefits researchers and industrial experts in various ways as it covers a variety of research approaches e.g. abstract concepts and complete set of services. In past few decades, many researchers attempted to define a generic architecture for IoT based smart city. Current state of the art presents various experimentation and test bed based smart city architectures that address challenges for a generic architecture.

Many studies proposed the deployment of IoT based smart city for one specific purpose i.e. waste management, noise pollution, and air quality monitoring [14, 15, 16]. However, these approaches do not reflect a standard solution. A test bed based architecture has been proposed in [17] to experiment a complete set of smart city services. The authors developed a large-scale infrastructure in Santander city to facilitate ease of experimenting. The SmartSantander city includes mobility support, security and surveillance, and scalability.

The experiment results reveal that the study covers several challenges in current literature. However, collected data are not tested for generalization. Therefore, the performance of the architecture may vary from one environment to another environment. Smart city collects Big Data from heterogeneous devices in the smart city. Thorough analysis of collected Big Data supports to design an efficient and generic smart city architecture. Big Data analytics uplifts generic smart cities to real-time controllable smart cities. Extensive analysis on Big Data becomes an asset for the smart city in future endeavors. Furthermore, carefully analyzed Big Data supports authorities to make critical decisions precisely, while timely made intelligent decisions increase the revenue of urban councils. City data and analytic platform (CiDAP) was developed to improve service designing and utilization using hadoop ecosystem [9]. CiDAP achieved a higher data processing throughput owing to its layered architecture. SCOPE [18] and FIRWARE [19] are similar projects that are developed based on Big Data analytics. These projects serve Big Data processing in real-time environment, even though they are not openly available to use in different contexts. Moreover, recent literature [15, 20] present several examples of using Big Data analytics in planning and developing smart cities. However, intelligent decision-making based on real-time processing of prodigious data is still a challenging task.

Smart city data collection predominantly relies on the connected smart objects. Generally, smart objects are controlled by battery operated sensors. Thus, energy efficient protocols are vital for IoT environments to enable energy aware communication. In [21], z-wave short range wireless technology is introduced for home automation purposes. Although it provides energy aware and cost effective communication, it does not provide interoperability with internet protocols. As a result, remote access for the smart objects via internet is restricted. In fact, a majority of smart things does not allow direct TCP/IP communication. Instead, they operate on dedicated protocols such as ZigBee and Bluetooth. WoT is the ideal candidate to expose smart things through the web to enable communication and remote accessibility. The WoT simplifies accessibility to smart things' functionalities and represent physical world entities as traditional web services [22]. Accordingly, REST architectural style [23] was introduced as a reference paradigm to integrate smart things into the web. Zeng *et al.* presented some research works that proposed approaches to implement WoT using REST principles [24]. Fielding *et al*. proposed a scheme to integrate smart objects with the web, which creates a significant impact on the state of the art. The authors suggested using REST architecture with hypertext transfer protocol (HTTP) [23]. HTTP is the de facto protocol used in RESTful applications. However, request/response structure and longer header formats consume more processing time and memory. Hence, using HTTP directly on resource-constrained objects is speculative and unrealistic. In addition to network resources, IoT environment should consider resource and energy efficiency of connected appliances. In [25], a resource aware smart home management system was proposed to manage home resources efficiently. Similarly, a smart controller scheduling system was proposed in [26]. The authors designed an algorithm based on dynamic programming to control smart home operations. However, the operation mechanism changes from one person to another, which increases the complexity of the system. Recent literature suggest integrating renewable energy sources to facilitate energy efficient smart home services. In [27], authors proposed a scheduling mechanism for thermal and electrical appliances. They compute energy from renewable energy sources and utilize that energy for operating certain appliances e.g. boiler and vacuum cleaner. However, implanting renewable energy sources in each home become less feasible due to various reasons e.g. weather conditions.

The above literature revealed important findings and existing challenges for a realistic, energy efficient, and intelligent smart city architecture. For example, designing a holistic smart city architecture, enable real-time processing of collected data, effective communication mode to connect smart objects and users, energy aware network for smart city energy management are the challenges that needs to be addressed. Thence, we propose a real-time data processing smart city architecture integrated with WoT to facilitate intelligent decision-making, seamless communication for smart objects, and energy efficient smart city environment.

3 Proposed Scheme

The proposed smart city architecture is embedded with Big Data analytics to serve real-time data processing and intelligent decision-making. Big Data embedded smart city is further integrated with RESTful WoT to allow accessibility and manageability of smart devices using web standards. Consequently, smart objects become controllable according to remote web requests. This feature has been used to manage energy utilization of smart appliances. A smart home connected to the smart home scenario is used as an example scenario to elaborate on energy management achieved as a result of RESTful web integration. Following sub sections provide detailed descriptions on Big Data analytics embedded smart city architecture and service representation over RESTful framework.

3.1 Big Data Analytics Embedded Smart City

The proposed smart city consists of four levels 1) data creation and collection level, 2) data processing and management level, 3) event and decision management level, and 4) application level as shown in Figure 1. The bottom layer of the architecture is liable for data generation and acquisition. The smart city framework aims to reduce unnecessary energy consumption and traffic congestion of the city, while maximizing the quality of health and other community services offered to urban citizens. Collection of data associated with daily operational tasks of the city plays a vital role in achieving afore stated goals. However, acquiring data in all operation aspects is tedious and challenging. In order to enhance the data acquisition of smart city, large number of sensors are deployed within the city area, in addition to other connected smart devices. Low cost and energy efficient sensors are widely used to sense phenomenon of interest from the real world. These sensors collect various types of data from the neighboring environment based on the deployed context e.g. smart home, transportation system, lighting system, etc. The proposed urban IoT is covered with all possible smart devices, sensors, and smart things, since the city becomes smarter with the expansion of connected devices [17]. Smart homes operate with the motivation to reduce household energy consumption. All home appliances are equipped with a ZigBee sensor to read real-time energy consumption. Similarly, ZigBee sensor is occupied in the RESTful web integration to control smart home appliances remotely over the web. Data collected at the bottom layer is transferred to data processing level to determine value added information. Vehicular and transportation system collects data to reduce city traffic congestion. Sensors implanted on roadsides collect number of vehicle entrance and departures between two points to calculate real-time traffic status at the data processing level. Weather forecast system obtains environmental parameters and weather parameters through deployed sensors. Collected data is transferred to the data processing level to derive intelligent forecasts. The smart city architecture incorporates several communication technologies i.e. ZigBee, Bluetooth, Wi-Fi, and data and cellular networks to communicate collected data to the data processing and management level.

As shown in Figure1, proposed smart city architecture utilizes multiple modalities i.e. HDFS, HBase, and HIVE to facilitate data storing and processing demands. HDFS acts as the storage medium of the smart city as it is the primary storage medium of Hadoop framework. Distributed nature of HDFS uplifts the performance of MapReduce by executing analysis on smaller subsets of a larger data set. Scalability demand of Big Data processing can be addressed with aid of HDFS. The proposed smart city occupies two nodes Hadoop cluster to obtain better performance. Autonomous real-time intelligent decision-making requires real-time read/write access throughout the cluster. So that, HBase is used to speed up the processing owing to its swift access capabilities and special characteristics i.e. real-time lookups, in-memory caching, and server side programming. Further, it supports fault tolerance, while augmenting usability. Querying and managing facility over the enormous amount of data on Hadoop is facilitated by HIVE. HiveQL queries data on Hadoop, since SQL querying is not possible on Hadoop. Upon storing respected data, the controls are transferred to event and decision management level.

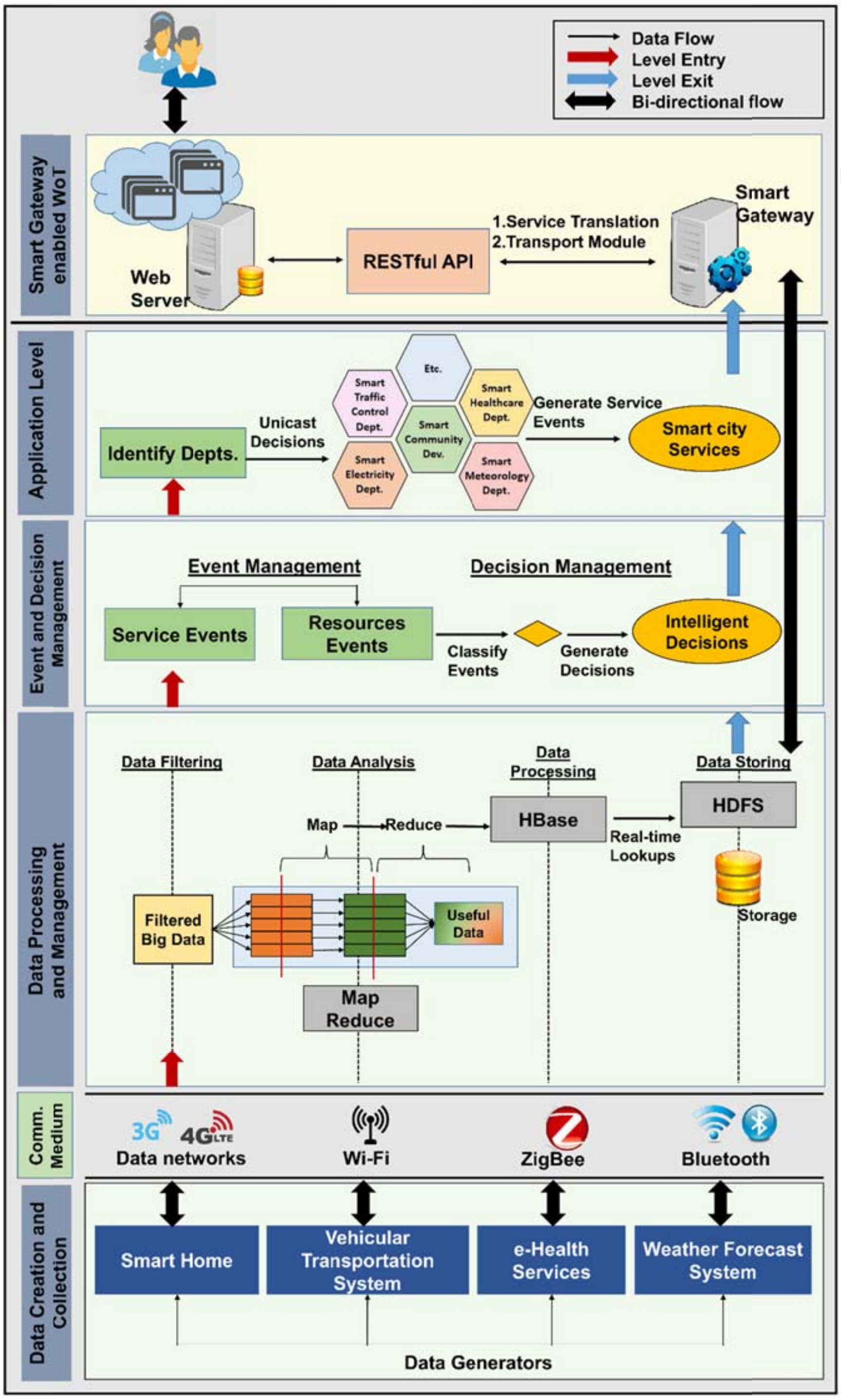

Figure 1 Working of the proposed Big Data embedded smart city integration with RESTful WoT

Event and decision management level performs two tasks 1) event management and 2) decision management. Aligning with the information received from the data processing and management level, event management classifies processed information into two events namely, service events and resource events. Service events control services offered by the smart city. For example, notifying urban citizens regarding street congestion is a service event. Similarly, resource events manage resources of the smart city e.g. water and electricity consumption. Decision management generates intelligent decision corresponding to a particular service event. Further, all decisions occupy a shared vocabulary (ontology) to describe intelligent decisions, in order to facilitate efficient processing at the application level. Then, the intelligent decisions are transferred to the application level.

Application level is the upper bound of the smart city architecture that connects with the smart gateway enabled WoT to allow remote access to the smart city services. Exposing smart city services via the RESTful API will be described in the next section. All smart city services corresponding to service events and resource events are generated at the application level. This level is further divided into three layers i.e. departmental layer, services layer, and sub-services layer to empower precise event generation. Figure 2 presents the sub layers of the application level. Intelligent decision from the event and decision management level are unicasted to the bottom layer of the application level (departmental layer) i.e. smart community development department, smart traffic control department, smart healthcare department, and smart weather and meteorology department. As mentioned previously, intelligent decisions are described in accordance with a shared vocabulary to enforce unicasting of the decisions throughout the application level. The departmental layer ascertains high-priority decisions and low-priority decisions. At the departmental layer, respective departments store high-priority decisions and unicast further towards the upper layers (services and sub-services). However, the transmission of low-priority decisions terminates at the departmental layer. Services layer components accepts decisions forwarded from the respective department at the departmental layer e.g. smart home and waste management components accept messages from the smart community development department. The services layer transfers decisions to corresponding component at the sub-service layer. For example, smart home transmits decision messages to energy management and water management components. The sub services level generates respective smart city service to serve a specific citizen or a group of citizens.

Following example describes working of the proposed scheme for a real world scenario. Assume that sensors implanted on roadsides observe the number of vehicles arriving and departing between two defined locations (data creation and collection level). Collected data are transferred to the data management level via communication mediums i.e. data networks, Wi-Fi, ZigBee, and cellular networks. Consequently, these sensed data goes through MapReduce, HBase, and HDFS to complete processing of raw data (data processing and management level). Event management classifies received information and create a service event. Then, the decision management discover the corresponding intelligent decision and describe it according to the shared vocabulary e.g. street congestion (event and decision management level). In this regard, the decision is to notify potential travelers about the street congestion if sensed information surpass the pre-defined threshold. Sequentially the control transferred to the application level. Smart traffic control department at the bottom layer of the application level (departmental layer) receives the unicasted decision message. The message is forwarded to the corresponding service level component e.g. traffic congestion. Subsequently, alternative paths component of the sub-service layer receives the decision message. Finally, the notification component notifies travelers whom may enter the congested street. The smart traffic control department determines potential travelers by checking global positioning system (GPS) destination and current positioning of the vehicle.

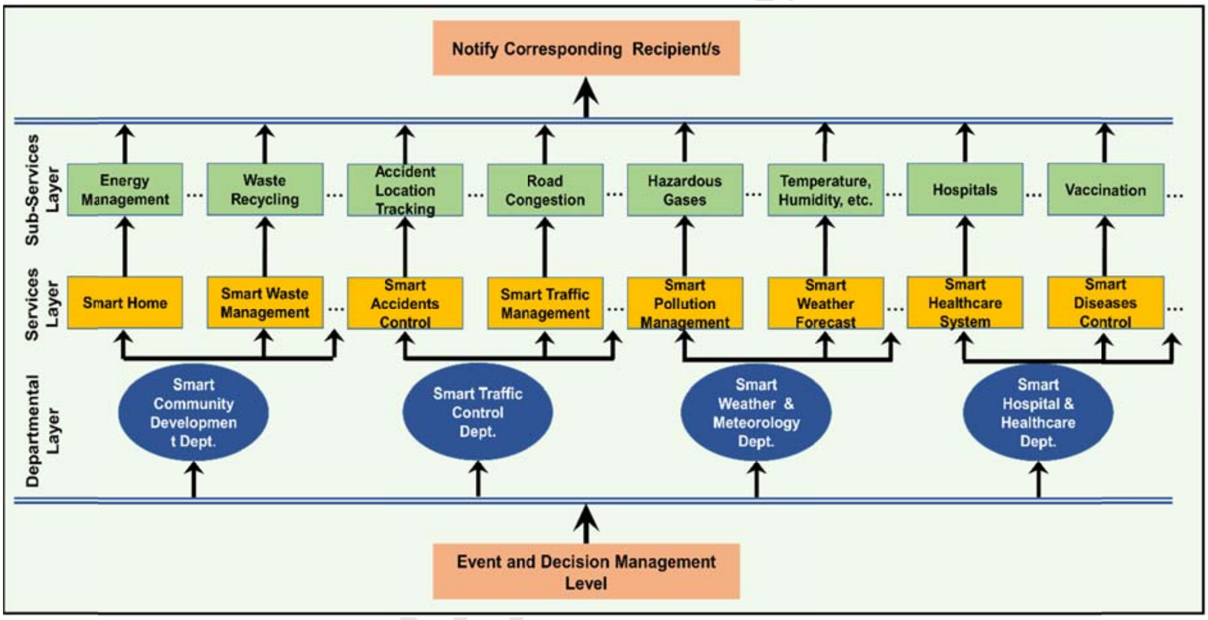

Figure 2 Sub layering at the application level for event generation and processing

3.2 Service Representation via RESTful Smart Gateway

The aim of introducing WoT is to design a stable and a generic architecture that provides ubiquitous connectivity between the smart objects and the web [14, 17]. In this section, we describe the integration of WoT to expose smart city services via a smart gateway. Henceforth, we introduce a smart home scenario to describe providing ubiquitous access to smart city services. Furthermore, we elaborates on energy management of smart buildings with aid of the said smart home example. We assumed that the appliances are fixed at specific locations within the smart home. A sensor that has a unique ID controls each appliance. Therefore, appliances are discovered by the corresponding sensor ID. Initially, each new device registers with a sensor in the management station (MS). Appliance registration reduces the performance deterioration cause by device discovery process. The proposed framework always communicates with the MS to expedite the performance.

The MS bridges household appliances with the web. RESTul API works on process management and service presentation. Multiple requests made on the same appliance are divided into multiple processes. Each process holds the ID of the sensor that controls the respective appliance. The processes are valid until the appliance is functioning properly. Whenever, the appliance is broken or removed from the smart home, all related processes will terminate autonomously. The users are allowed to access the smart home appliances remotely. They make requests over the web and the requests are consequently redirected to the smart home MS. For example, a user can request to view current energy consumption of the house. Worthy to note, that a smart home user can only access the services that are registered under his/her smart home. The smart home allows multi-user functionalities with aid of multi-processing environment, which handles multiple requests on the same appliance. Further, RESTful API enable the smart home users to access the services from the web via its presentation module. Smart home users can view and request for information using any web browser owing to RESTful web integration. Each appliance is represented as a hyperlink and users can explore the desired services by navigating across those hyperlinks. The proposed example smart home scenario is presented in Figure 3.

Significant increase of infrastructure energy consumption has highly influenced the current smart building architectures to introduce energy efficient mechanisms. Henceforth, we describe the unification of above stated smart home architecture with WoT to ensure efficient energy management of the smart buildings that are included in the proposed smart city architecture. Aforementioned smart home architecture is generalizable to all smart buildings with slight modifications according to the context. The smart home includes a device control unit (DCU) and MS. DCU operates in a semi-automated manner. Users manually switch on the appliances and the DCU switches off the appliances to save energy. Switch off controls are generated by MS at the absence of user/s. DCU is a collection of non-IP appliances and ZigBee sensors deployed in the smart home to gather environmental parameters and to access the assigned appliances. As stated, each appliance of the smart home is attached to a sensor to facilitate accessibility and manageability. The proposed architecture addresses energy efficiency of both household appliances and wireless sensor network (WSN). Each ZigBee sensor is connected to a ZigBee coordinator placed in each room and in kitchen. Coordinators are deployed to mitigate the interference from coexistence of multiple wireless communication technologies i.e. Wi-Fi and ZigBee and distance between sensors and MS. Interference reduction improves the energy efficiency of smart home WSN owing to lessen packet loss and packet retransmissions. Mindful coordinator placement shrinks the impact of distance between MS and sensors. In fact, a distant node consumes more energy to transfer a packet to the MS than a proximal node. However, careful deployment of coordinators in each room and the kitchen creates a comparatively similar proximity between sensors and coordinators. Sequentially, coordinators transfer data towards MS. Thus, improves energy efficiency of the smart home network by reducing unnecessary hops in the data transmission. Energy consumption of an appliance depends on its functioning time. Therefore, MS records switch on time for each appliance. DCU continuously checks on the user presence, since the proposed energy saving mechanism is based on user behavior. Whenever, the users are not available, DCU notifies MS. The MS notifies respective users and generate control events accordingly. Once the DCU receives the control message, it switches off the respective appliance to reduce energy wastage. Sequentially, the MS updates appliance status and switched off time.

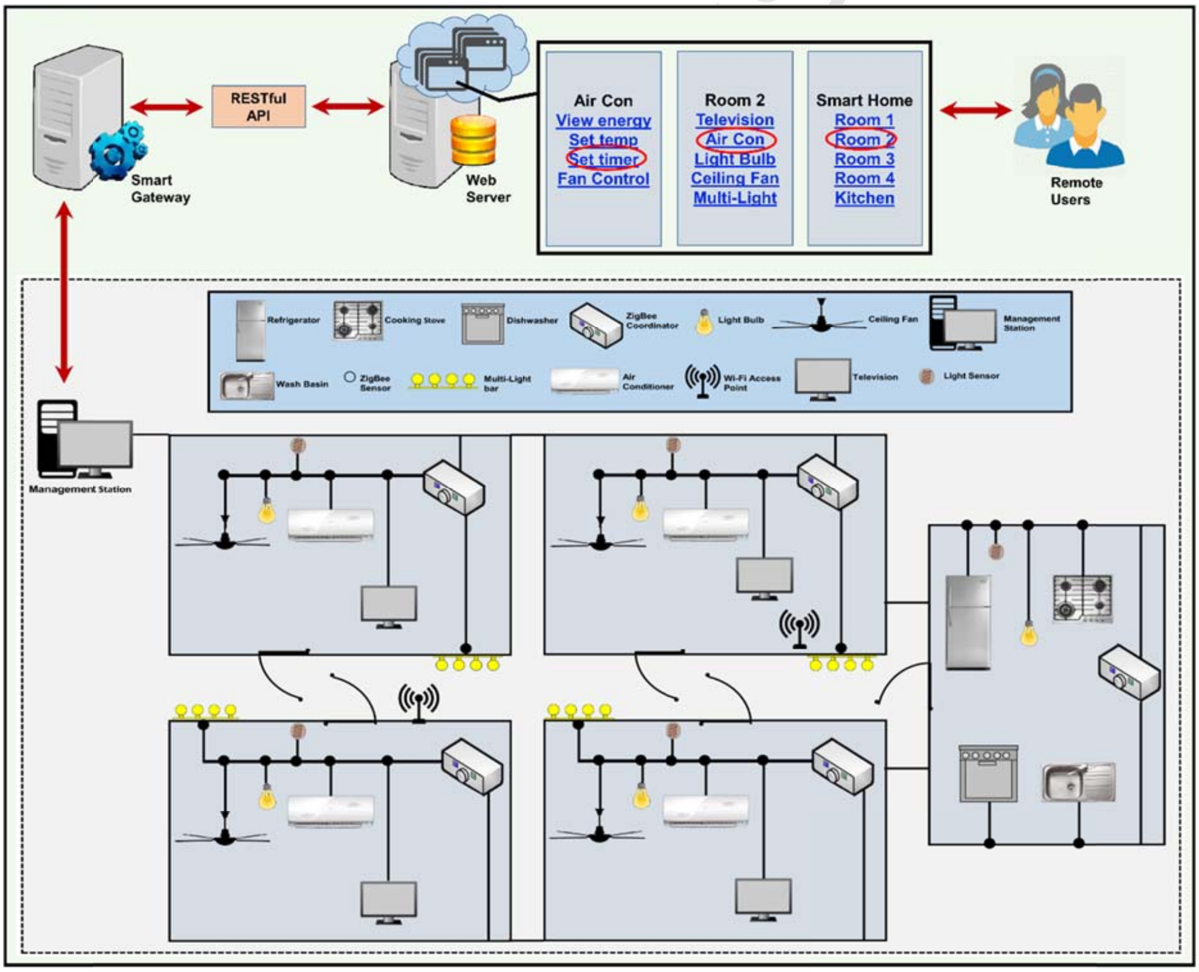

Figure 3 Proposed smart home architecture for service representation over the web

The smart city architecture is responsible for the data processing of smart buildings. However, processing all smart building data at the city level is a tedious task. Therefore, we introduced a MS to each smart building to control its operation. Data processing, decision-making, an event generation are embedded into the MS to avoid unnecessary data flow across the network and for a speedy processing. The MS incorporates an event management system (EMS) to enable communication between users and DCU. The MS stores all records corresponding to home energy consumption per month. At the end of the month, the MS sends a summary report to the data processing and management level of the smart city architecture and removes data from the local storage. Energy consumption records in the MS support smart home residents to conserve energy. Meanwhile, historic records transferred to the Hadoop storage are used for long-term energy management programs at the city level. Moreover, the MS is liable for the event generation of the smart building. Smart building even generation is in two types 1) locally driven events and 2) remotely driven events. Locally driven events are generated by the MS to minimize energy wastage of the smart home at the absence of the user. However, the MS does not switch off the appliances immediately. It notifies the respective user and opts to waits for β_T time to receive the user response as shown in figure 4. Users have the sole authority accept or reject the event suggested by MS. Thus, the users can adjust different tasks of the appliances e.g. control the cooling of refrigerator and air conditioner. If the MS does not receive any response from the user within β_T time, the MS autonomously generates a decision aligning with the human machine interaction (HMI) module. HMI is embedded into the MS to approve or reject the generated events considering previous user responses. Since HMI module makes future adjustments based on the past human interactions, it assures the reliability of the decisions. Moreover, user absence will not cause any adverse effects on the smart home operation thanks to the integration of HMI module. Remotely driven events are generated by remote smart home users The MS receives remotely generated events via the smart gateway.

The users can access smart home services through any web browser as a result of the service translation at RESTful framework. Remotely controlled events include turning on/off appliances, change settings, monitoring appliances, and monitoring energy consumption. Once the MS receives a remote user request, it immediately executes the request and β_T is not involved in remotely driven event generation. Initially, the MS determines the availability of information in storage, upon receiving the remote user request. If the information are available, it immediately responds to the user. For example, if a user request for the status of an appliance (on/off), the MS holds the latest information. Therefore, MS does not forward the request to the corresponding appliance. Instead, respond to the owner of the request promptly. On the contrary, users can also request for information that are not available in the MS. In that scenario, the MS forwards the service request to corresponding ZigBee coordinator and Zigbee sensor. For example, an user requests for the current room temperature of a particular area. Since the MS does not hold the latest information, it forwards the request to the respective sensor of that area via the coordinator. With aid of the received self-describing message, MS identifies coordinators and sensors to route the service request. The sensor sends the response message at the completion of service execution. Service response is wrapped as an HTTP response at the smart gateway and send further. While, routing the response, MS updates the status and energy consumption parameters of the respective appliance or sensor. Subsequently, the user can view requested service's response via a web browser.

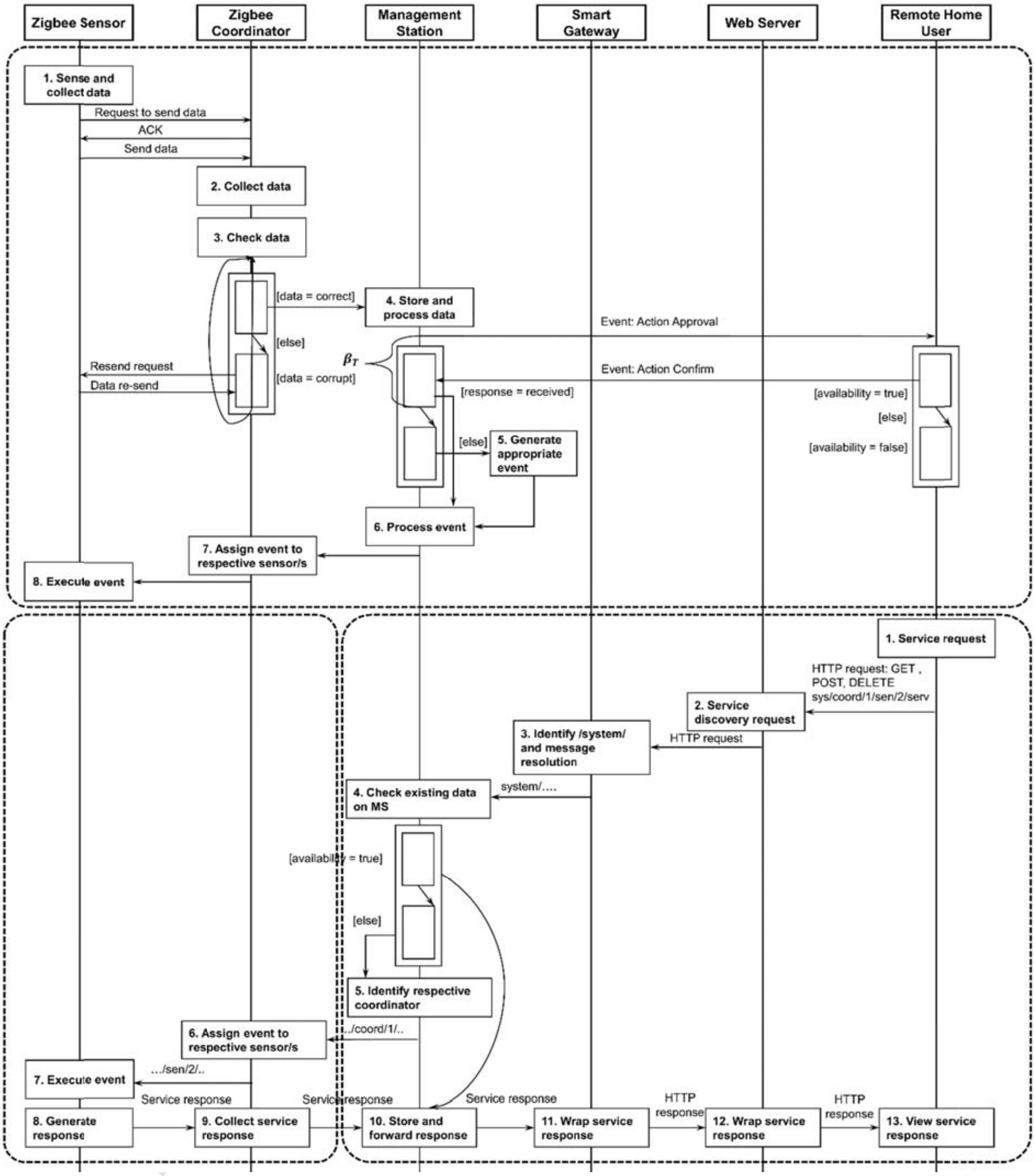

Figure 4 Event management and communication flow for in-house controls and remote user controls

The proposed smart home architecture occupied the web as the communication medium between remote users and smart objects. Herein, we propose a smart gateway architecture to compensate the HTTP and TCP/IP requirements of non-IP devices, such as ZigBee sensor, ZigBee coordinator, Bluetooth sensor, etc. The proposed smart gateway facilitates transparent web process, while hiding the internal details of the network. The smart gateway consists of two layers i.e. 1) transport module (TM) and 2) device service module (DSM). These two layers collaboratively resolves services requests and service responses to facilitate seamless communication. The proposed smart building utilizes ZigBee as the communication protocol. ZigBee is a proprietary protocol that was developed based on IEEE 802.15.4. The DSM resolves service protocols of the non-IP appliances. In addition, DSM maps the device functionalities to the RESTful APIs. The RESTful API architecture binds uniform resource identifier (URI) to the functionalities of smart things. So that, the remote smart home users can access the smart home services. Further, URI binding supports in generating self-describing service requests to access physical devices via the MS. The MS is liable to transport the resolved user requests to the MS. Similarly, DSM receive service responses and transform proprietary response messages to HTTP format to forward to the web server to allow access via any web browser.

4 Results and Data Analysis

In this study, we proposed a smart city architecture that can overcome the existing drawbacks of a smart city in terms of performance efficiency and energy efficiency. Performance improvement of the smart city highly depends on the processing and analysis of data from previous studies that belongs to various fields e.g. transportation, healthcare, community development, etc. Herein, we processed authentic data obtained from various sources to confirm the performance improvement gained through embedded Big Data analytics. Dataset information are mentioned below in Table 1. Initially, the datasets consist with fuzzy and raw data entries. We applied MapReduce on raw data to expedite the processing of smart city data. Notably, it improves the performance efficiency of Hadoop. Thus, Big Data analytics supports in boosting real-time data processing. In this section, we present dataset information, simulation setup to integrate smart city components with WoT, and results on real-time data processing and energy management of smart buildings.

4.1 Dataset Information

We analyzed datasets obtained from authentic and reliable sources. These datasets include 1) water consumption data collected from meter readings of 61263 houses in Surrey, Canada [28], 2) city traffic, 3) parking lot data in Aarhus City, Denmark [29], and a dataset on toxic gases and materials that has been collected in Aarhus City, Denmark [30]. The hazardous gases dataset was analyzed to determine ozone, carbon monoxide, sulphar dioxide, and nitrogen dioxide gases.

Table 1 Dataset Information

Sources	Dataset	Size
Surrey City, Canada [28]	Water Consumption	4 MB
Aarhus City, Denmark [29]	Traffic Data	3.04 GB
Aarhus City, Denmark [30]	Pollution Data	77.25 MB
Aarhus City, Denmark [29]	Parking lots	0.20 MB

4.2 Simulation Setup for Data Processing and Smart Home Integration with WoT

Collected data were analyzed using a two node Hadoop cluster on Ubuntu 16.04 LTS having core I5 processor and 8GB RAM. Considering the analyzed data, thresholds were defined on the output data. The thresholds are specific to the dataset size used for the analysis. The threshold values are shown in Table 3. The smart home architecture along with RESTful framework, MS, smart gateway, and a web server is simulated as shown in Figure 3. The MS and the smart gateway are installed on Core I3 CPU 3.60 GHz computer. The RESTful framework and web server were hosted in a computer with windows server 2008 operating system. We simulated and tested the entire smart home scenario using C# programming language. Each household appliance is attached with a ZigBee sensor. A sensor is a programming function, which follows the IEEE 802.15.4 protocol architecture for communication. In this context, we did not consider the energy consumption of sensors. Importantly, the RESTful framework is a separate program and it is not separately installed on each sensor. Therefore, the sensor nodes require energy only for sensing and transmitting information to the coordinator available at the shortest distance.

Table 2 Threshold Values and Event Generation Analysis

Dataset	Size	Threshold	T_β
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

4.3 Dataset Analysis and Processing Performance

The data were analyzed extensively to determine the expected threshold values for each dataset. These threshold values helped us to validate the reliability of event generation of the smart city. Accordingly, we defined a threshold value for event generation (T_β), which includes time taken to process data, generate an event, and to notify respective user/s through the corresponding departments that are shown in Figure 2. Experimented dataset analysis revealed that the data processing time significantly increases with larger datasets. However, in real-world scenario data that requires real-time processing are available in streaming form. Therefore, the processing time degradation with the data size will not have any adverse effects on the system performance. Nevertheless, it is crucial to facilitate immediate data processing on high-speed data. The simulation setup analyzed the datasets at the data processing and management level. In this study, we used authentic data sets to represent data creation and collection level. Initially, we used these analyzed data to determine threshold levels for each dataset. Secondly, we used the same data sets to validate the reliability and performance of proposed event and decision management and application levels. The analysis results were used to generate various events i.e. traffic intensity warnings, parking availability notifications, hazardous gas intensity warning, etc., in order to enhance the quality of life of urban citizens. Analysis results of the datasets are shown in Figure 5.

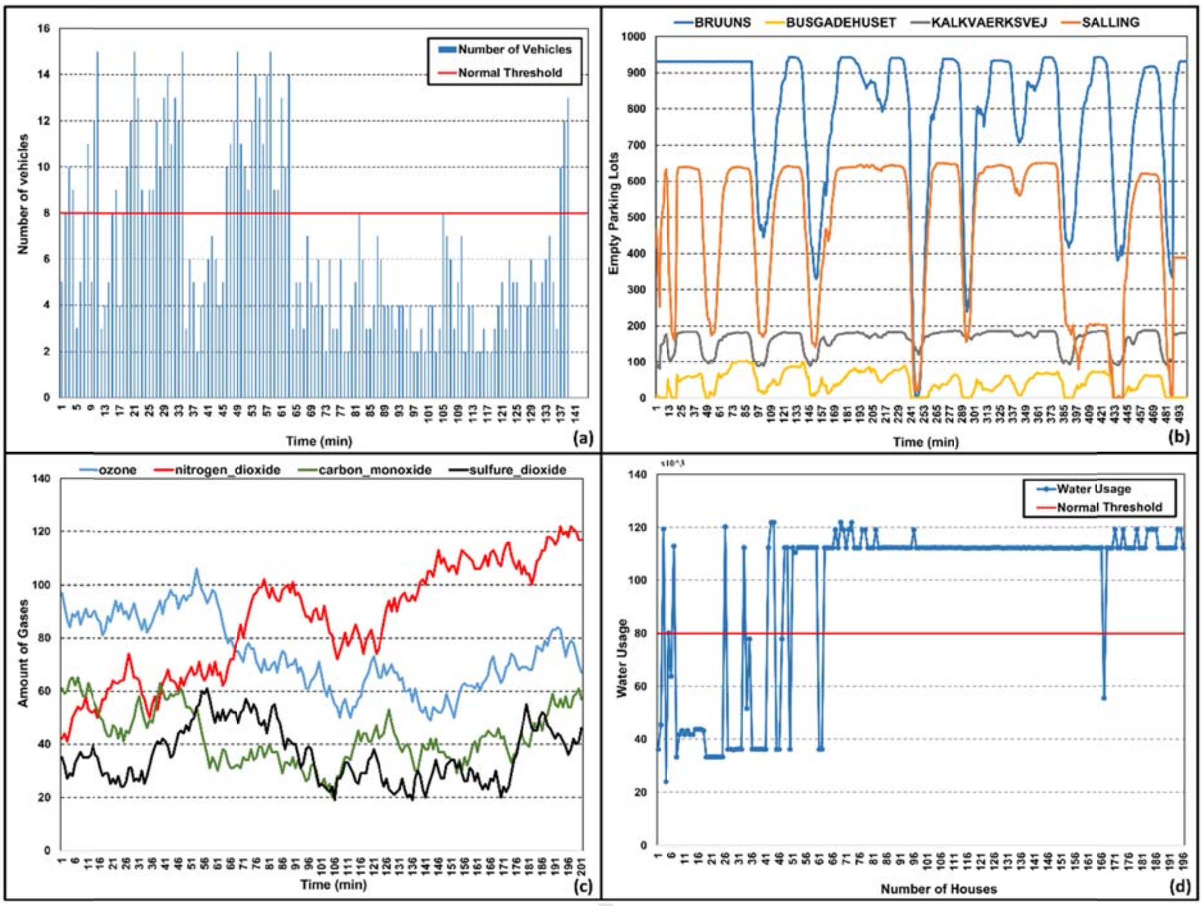

Figure 5 Data analysis results for the datasets. (a) Number of vehicle at Aarhus City, Denmark (b) Empty parking lots in various places in Aarhus, Denmark (c) Amount of pollution at different time of day in Aarhus, Denmark (d) Water usage at different houses of Surrey, Canada

Traffic intensity warnings are sent to the travelers if the number of vehicles in a specific portion of the road surpasses the threshold value. According to the data analysis, we defined eight as the threshold number of vehicles as shown in Figure 5(a). Whenever the actual number of vehicles exceeds the threshold, data processing level filters that as a valuable data forward it towards respective departments along with the decisions made at the event and decision management level. This process consume T_β time with an offline dataset, since it analyze the complete dataset. However, it is capable of generating events instantly with streaming data. Similarly, the proposed architecture serves citizens with real-time parking data. The sensors deployed in the parking garages updates availability of empty parking lots as shown in Figure 5(b). Citizens can find available parking lots by checking the database instead of physical checking. Indeed, it makes the parking process a lot more convenient. Moreover, we found that that the parking lot availability highly depends on the number of vehicles in the city as well as on the population. Thus, having more parking lot details can improve the reliability and availability of service provision. In addition, the proposed smart city architecture enables a citizen to reserve a parking lot prior reaching the destination. Further, deployed sensors immediately updates the availability upon departure of a vehicle that has occupied a particular parking lot.

Increasing volume of the toxic gases is another major problem that arise with increasing number of vehicles, waste production, and pollution, which causes serious health conditions of urban citizens. Hence, we analyzed the pollutant density of Aarhus City, Denmark. The air quality sensors can obtain information regarding the concentration levels of pollutants and hazardous gases. Figure 5(c) presents toxic gas concentration of Aarhus City. The data processing level utilized existing data and evaluated the event generation efficiency of the proposed smart city architecture. Pollutant dataset T_β demotes the time taken for the event generation after processing the complete data set. These actions are transmitted to users via the weather and meteorology department. Accordingly, users are warned to be careful, while visiting polluted areas. In long term, these precaution warnings can reduce adverse effects on health problems that are caused by toxic gases and pollutants. Further, the proposed smart city concern on the water consumption, since excessive water usage can become a critical problem in near future. Figure 5(d) illustrates the water consumption information of Surrey, Canada. According to the analysis, each house consumes an average of 80000 liters of water per month. In order to reduce excessive water consumption, we generate warning messages at the threshold. Once the water usage reaches 80000, event and decision management level of the proposed architecture generates warning messages and notifies the users via water management department.

The performance improvement of the proposed scheme was compared with a single node Hadoop system and a JQuery based system. Introducing data filtration before data processing has significantly influences the data processing time. As shown in Figure 6(a), processing of the proposed scheme has reduced remarkably compared to other two schemes. Figure 6(b) illustrates the performance enhancement gained in terms of throughput. With a smaller data amount, the proposed scheme produced a slightly increased throughput. However, single node Hadoop system and JQuery based system followed an extremely slow rise in throughput with larger data amount. Contrastingly, the proposed two node Hadoop cluster gained a rapid and steady growth with increasing data amount. The results revealed that the proposed scheme has improved the data processing of the smart city. Consequently, the performance improvement has supported the smart city to provide reliable services in various contexts in a timely manner.

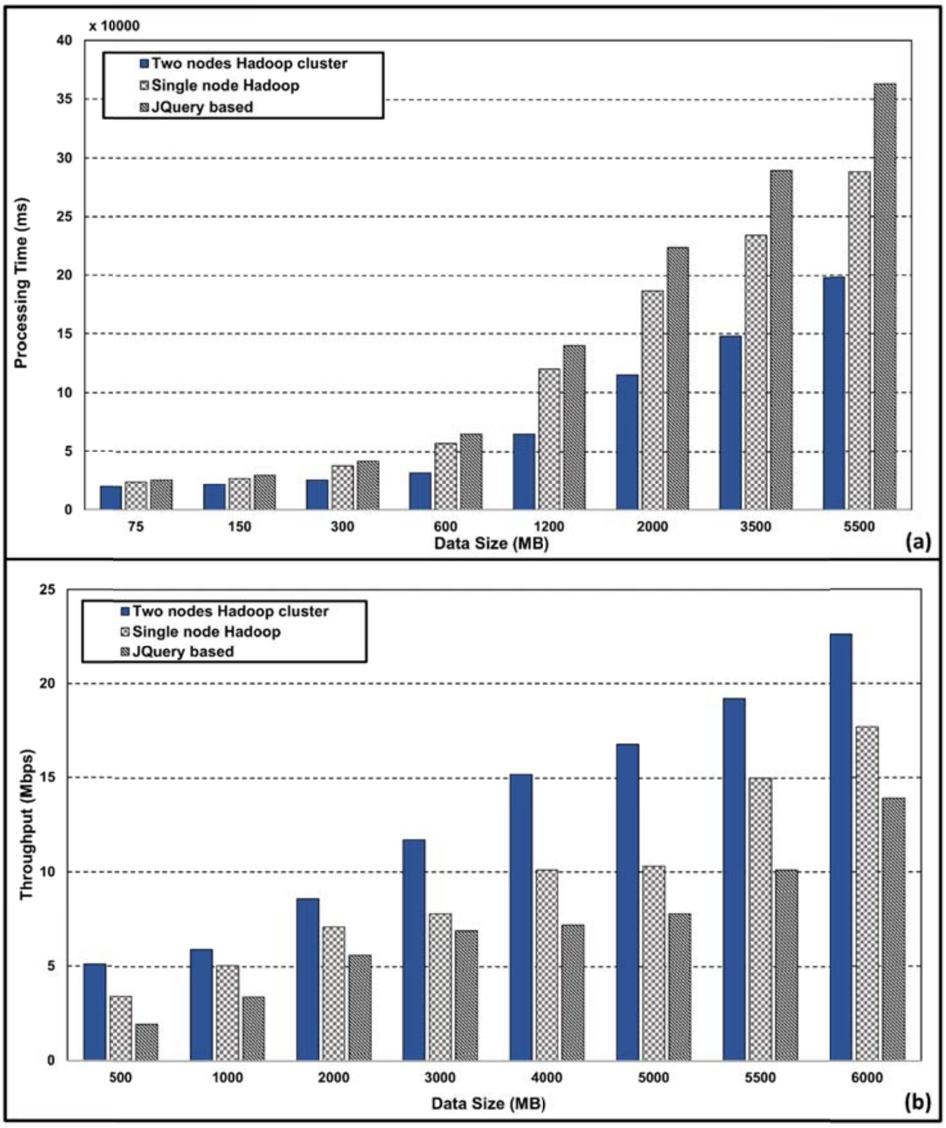

Figure 6 Proposed system performance analysis. (a) System processing time analysis (b) System throughput analysis

4.4 Energy Management of Smart Buildings

Energy management of smart buildings is simulated with the smart home architecture integrated with the WoT. Users of the smart home switch on the appliances, but the proposed web integrated energy management scheme take care of the switching off the appliances. The proposed scheme is tested against a pure WSN architecture and a relay based architecture. In a pure WSN, sensors deployed in the smart home directly control the appliances. Thus, the energy consumption of the network is highly influenced by the distance between sensors and MS. In fact, pure WSN based smart home is capable of switching off appliances when the users are not available. Although it switch off the appliances, user preferences for a particular action is not considered. Moreover, the pure WSN architecture does not favor the user context. Hence, pure WSN controlling is limited to a few appliances i.e. light, heater, and air conditioner. Therefore, pure WSN services are not adequate to cater real-time energy management scenarios, where smart operations highly depend user preferences and behaviors. In the case of relay based WSN, three relay sensors are used to forward data to the next available hop. Unfortunately, these systems do not record the functioning time of the appliances. Therefore, home users are unable to adjust the household appliance for better usage concerning the previous history of the household appliances. In other words, these systems does not provide the functionality of smartness and thus not suitable to use in the future WoT-based smart buildings. On the contrary, the proposed ZigBee coordinator based smart home management system enables prior records based appliance management. Moreover, it generates appropriate decisions, if the home user is not available or offline.

The energy consumption records of four different appliances i.e. fan, television, air conditioner, and refrigerator are presented in Figure 7(a), 7(b), 7(c), and 7(d) respectively. The simulation results revealed that the proposed scheme has obtained an average of 15% - 20% energy efficiency compared to existing pure WSN and relay based smart home systems. Similarly, we analyzed the energy consumption data of all appliances for a duration of six hours. The simulation generated random user activities. In this scenario, the system opts to wait for β_T time to receive user response. If the user does not react within the defined time, previous user contexts are utilized to execute necessary actions. Figure 8(a) illustrates the energy consumption of all household appliances. Compared to relay based WSN and pure WSN, proposed ZigBee coordinator based WSN achieved an average of 15% energy efficiency.

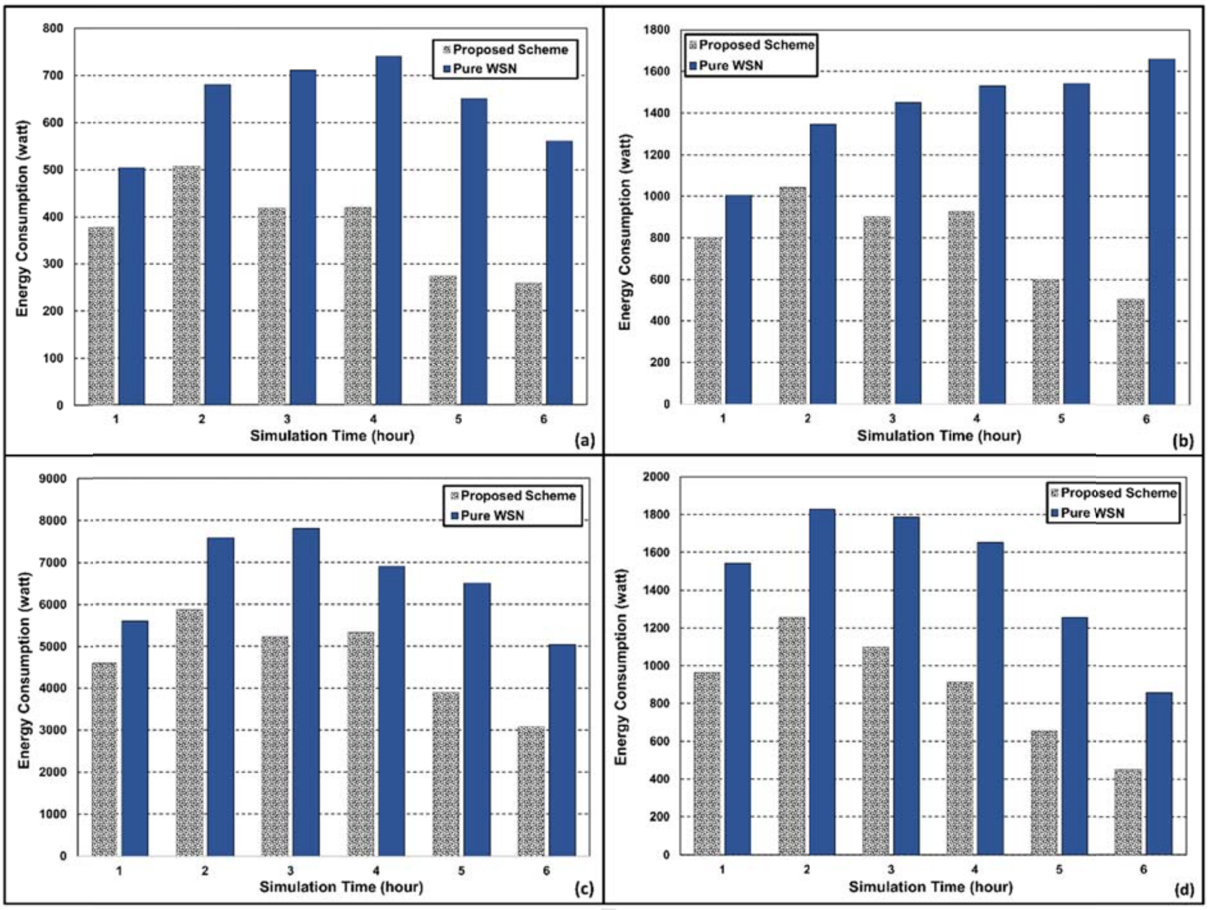

Figure 7 Energy consumption of various household appliances. (a) Energy consumption of fan (b) Energy consumption of television (c) Energy consumption of air conditioner (d) Energy consumption of refrigerator

We simulated a large number of service requests over the web, in order to evaluate the performance of proposed WoT architecture under a heavy load. In general, the number of appliances of the smart home is more than the number of smart home users. As a result, the response time for various requests is negligible. However, we evaluated the system with less number of appliances to assure reliability and efficiency of the proposed WoT based smart home architecture. The duration of the experiment was set to 10 minutes and tested for different performance parameters. In the first setup, web pages are hosted on one computer, while MS and other sensors are hosted on a separate computer. However, all three components belonged to the same network. Figure 8(b) depicts the response time with 3, 5, and 7 sensors that are attached to household appliances. The results revealed response time is increased with three appliances. However, the response time is decreased with increasing number of appliances. Effects of interference is another major concern of the proposed architecture. Whenever, the sensors reside closer to Wi-Fi access point that causes coexistence interference, which lead to transmission failures. We divided all sensor nodes in the smart home architecture into four groups, in order to simulate the fail transmission scenario. Sensors that are 3, 6, 9, and 12 meters apart from the Wi-Fi access point belongs to group A, B, C, and D respectively. High interference within 3 meters range results a higher number of failed transmissions in group A. According to the experimented results, 10% of all attempts were failed when the request arrival frequency is 120 requests per minute. Although group A is highly affected from interference, rest of the groups presented a significant decrease with the increasing physical distance. Figure 8(c) illustrates fail transmission percentage of group A operates with 3, 5, and 7 appliances that serve multiple requests. Worthy to note that we presented the worst conditions of the proposed architecture. Thus, embedding the proposed scheme in real-world scenario benefits the existing energy management approaches significantly. Discovery time variation with the number of sensors is presented in Figure 8(d). Initially, all sensors are registered in the web server via MS and ZigBee coordinator. The discovery time of the sensors increase with number of sensors that are turned on simultaneously. Thereupon, we experimented the time consumption to discover 30 sensors deployed in the smart home. According to the results obtained from the experiment, proposed scheme discovers 20 sensors immediately after turning on at once. The time consumed to discover 20 sensors is 19 seconds. However, the time consumption for sensor discovery increases rapidly if 20 sensors are turned on at once.

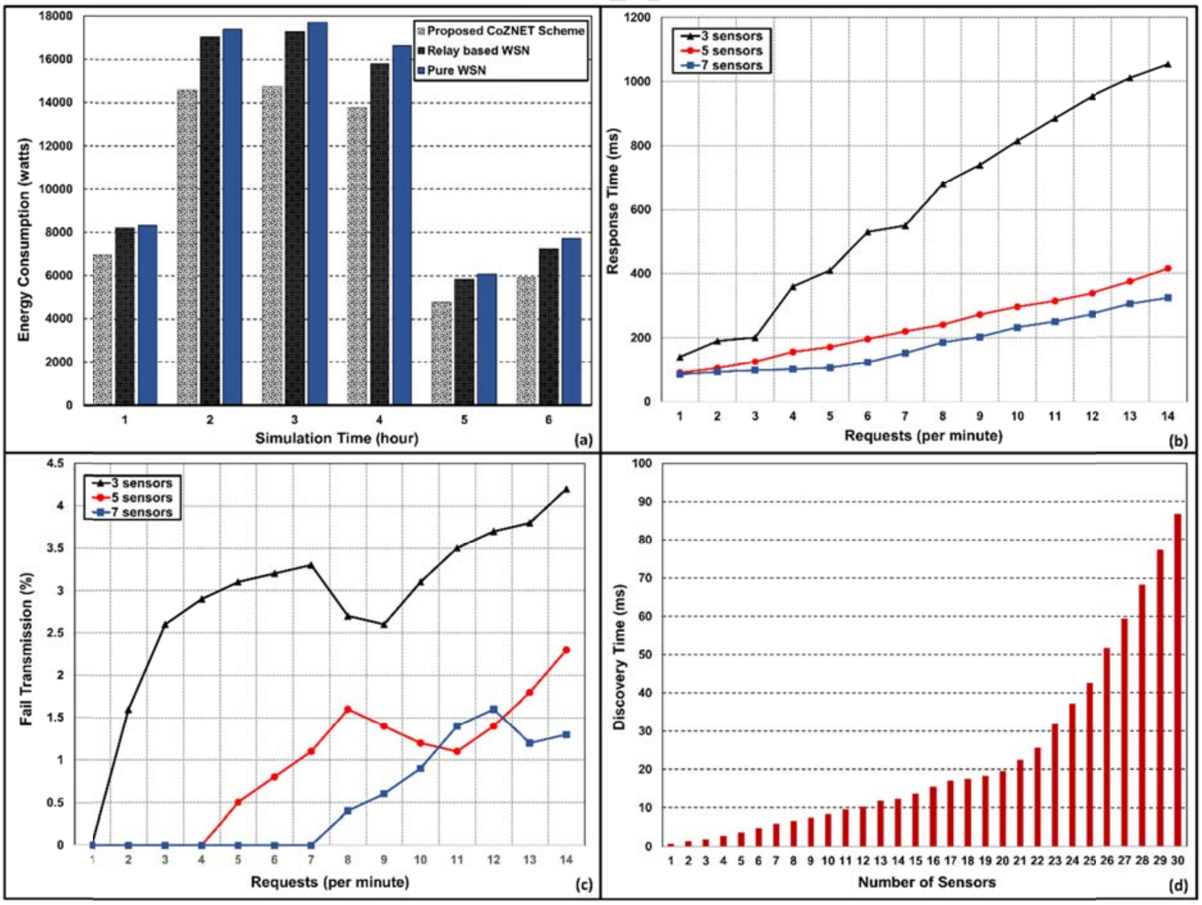

Figure 8 The effects of proposed WoT architecture. (a) Energy consumption analysis for all household appliances (b) Response time analysis (c) Fail transmission analysis (d) Sensor discovery time analysis

5 Conclusion

Transforming a conventional city into a smart city requires compatibility of connected devices, ubiquitous access to urban services and resources, and ability to process voluminous data. Therefore, we proposed a smart city architecture that is capable of processing colossal amount of data with aid of data filtration followed by Big Data analytics using a Hadoop ecosystem. The smart city architecture was further extended with WoT integration to facilitate platform compatibility and ubiquitous access to urban services.

The proposed system analyzed various types of data to present the data processing efficiency gained through the embedded Big Data analytics. Processing time and throughput were significantly improved with the proposed two-node Hadoop cluster at the data processing and management level of the smart city. Moreover, WoT integration has greatly supported the energy management of smart buildings. The beauty of the proposed scheme is it coordinates operations to minimize energy consumption according to user preferences. We simulated a smart home scenario to evaluate the energy consumption efficacy and the impact of WoT integration on network performance.

Acknowledgments

- This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B03933566).
- This study was supported by the BK21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering. Kyungpook National University, Korea (21A20131600005).

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 & Tutorials*, vol. 17, no. 4, pp. 2347 - 2376, 2015.

[2] M. 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, a. 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] 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.

[7] S. P. Mohanty, U. Choppali and E. Kougianos, "Everything you wanted to know about smart cities: The Internet of things is the backbone," *IEEE Consumer Electronics Magazine* , vol. 5, no. 3, pp. 60-70, 2016.

[8] D. Guinard, "A Web of Things Application Architecture - Integrating the Real-World into the Web (Doctorate Dissertation)," 2011. [Online]. Available: https://s3.amazonaws.com/guinard.org/webofthings-thesis-domguinard.pdf. [Accessed June 2016].

[9] 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 *2015 IEEE International Congress on Big Data (BigData Congress),* New York, 2015.

[10] T. h. Kim, W. C. Fang, C. Ramos, S. Mohammed, O. Gervasi and A. Stoica, "Ubiquitous Sensor Networks and Its Application," *International Journal of Distributed Sensor Networks,* vol. 2012, 2012.

[11] R. E. Barone, T. Giuffre, 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.

[12] Z. Shelby, K. Hartke, C. Bormann and B. Frank, "Constrained Application Protocol (CoAP) draft-ietf-core-coap-18," 1 June 2014. [Online]. Available: https://datatracker.ietf.org/doc/rfc7252/.

[13] S. V. Nandury and B. Begum, "Smart WSN-based ubiquitous architecture for smart cities," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI),* Kochi, 2015.

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

[15] 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.

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

[17] L. Sanchez, L. Muñoz, J. Galache, P. Sotres, J. Santana, V. Gutierrez, R. Ramdhany, A. Gluhak, S. Krco, E. Theodoridis and D. Pfisterer, "SmartSantander: IoT experimentation over a smart city testbed," *Computer Networks,* vol. 61, no. 2014, pp. 217-238, 2013.

[18] "SCOPE: A Smart-city Cloud-based Open Platform and Ecosystem," [Online]. Available: http://www.bu.edu/hic/research/scope/.

[19] "FIWARE Open Source Platform," [Online]. Available: “FIWARE Open Source Platform.

[20] I. J. Lee, "Big data processing framework of road traffic collision using distributed CEP," in *APNOMS,* Hsinchu, 2014.

[21] D. Guinard, V. Trifa, F. Mattern and E. Wilde, "Architecting the Internet of Things," in *From the Internet of Things to the Web of Things Resource-oriented Architecture and Best Practices*, Berlin, Springer, 2011, pp. 97 - 129.

[22] E. Wilde, E. C. Kansa and R. Yee, "Web services for recovery.gov," School of Information, University of California, Berkeley, CA, USA, 2009.

[23] R. T. Fielding and R. N. Taylor, "Principled design of the modern web architrcture," *ACM Transactions on Internet Technology (TOIT)* , vol. 2, no. 2, pp. 115-150, 2002.

[24] D. Zeng, S. Guo and Z. Cheng, "The Web of Things: A Survey," *Journal of Communications,* vol. 6, no. 6, pp. 424-438, 2011.

[25] J. Y. Son, J. H. Park, K. D. Moon and Y. H. Lee, "Resource-aware smart home management system by constructing resource relation graph," *IEEE Transactions on Consumer Electronics,* vol. 57, no. 3, pp. 1112-1119, 2011.

[26] A. N. Alvi, S. H. Bouk, S. H. Ahmed, M. A. Yaqub, M. Sarkar and H. Song, "BEST-MAC: Bitmap-Assisted Efficient and Scalable TDMA-Based WSN MAC Protocol for Smart Cities," *IEEE Access,* vol. 4, no. PP, pp. 312-322, 2016.

[27] L. Liu, Y. Liu, L. Wang, A. Zomaya and S. Hu, "Economical and Balanced Energy Usage in the Smart Home Infrastructure: A Tutorial and New Results," *IEEE Transactions on Emerging Topics in Computing,* vol. 3, no. 4, pp. 556 - 570, 2015.

[28] "Dataset Water meters," [Online]. Available: http://data.surrey.ca/dataset/water-meters. [Accessed 12

January 2016].
[29] T. Dataset, "Dataset Collection," [Online]. Available: http://iot.ee.surrey.ac.uk:8080/datasets.html#traffic. [Accessed 21 2 2016].
[30] P. Dataset, "Dataset Collection," [Online]. Available: http://iot.ee.surrey.ac.uk:8080/datasets.html#pollution. [Accessed 21 2 2016].

Bhagya Nathali Silva received the B.S. and M.S. degree in Information Technology from Sri Lanka Institute of Information Technology, Colombo, in 2011. She is currently a candidate of School of Computer Science and Engineering in Kyungpook National University, Daegu, Korea. Her area of expertise includes architecture designing for Internet of Things, Machine-to Machine Communication, Cyber Physical Systems, and Communication Protocols, etc.

nathalis@netopia.knu.ac.kr

Murad Khan Received the B.S. degree in computer science from university of Peshawar Pakistan in 2008. He completed his Ph.D. degree in computer science and engineering from School of Computer Science and Engineering in Kyungpook National University, Daegu, Korea. Dr. Khan published over 40 International conference and Journal papers along with two books chapters in Springer and CRC press. He also served as a TPC member in world reputed conferences and as a reviewer in numerous journals such as Future Generation Systems (Elsevier), IEEE Access, etc. In 2016, he was awarded with Qualcomm innovation award at Kyungpook National University for designing a Smart Home Control System. He was also awarded with Bronze Medal in ACM SAC 2015, Salamanca, Spain, on his distinguished work in Multi-criteria based Handover Techniques. He is a member of various communities such as ACM and IEEE, CRC press, etc. His area of expertise includes ad-hoc and wireless networks, architecture designing for Internet of Things, and Communication Protocols designing for smart cities and homes, Big Data Analytics, etc.

murad.csit@suit.edu.pk

Kijun Han received the B.S. degree in electrical engineering from Seoul National University, Korea, in 1979 and the M.S. degree in electrical engineering from the KAIST, Korea, in 1981 and the M.S. and Ph.D. degrees in computer engineering from the University of Arizona, in 1985 and 1987, respectively.
He has been a professor of School of Computer Science and Engineering at the Kyungpook National University, Korea since 1988

Highlights

- A Big Data analytics embedded smart city architecture is proposed
- The proposed system integrates web with the smart control system using a smart gateway syste
- The system is evaluated based on authentic data sets from various reliable data sources

ACCEPTED MANUSCRIPT