Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 83 (2016) 1232 – 1237

Procedia Computer Science

Second International Workshop on Mobile Cloud Computing systems, Management, and Security (MCSMS-2016)

Cloud Support Data Management Infrastructure for Upcoming Smart Cities

Muhannad Quwaider*, Mahmoud Al-Alyyoub and Yaser Jararweh

Jordan University of Science and Technology, P.O. Box 3030, Irbid 22110, Jordan

Abstract

This paper presents a novel large scale data management model for future smart cities. The system is exploiting the emerging cloud computing services availability over the globe by processing the collected data at the edge. To handle the large scale requirements in smart city, we introduce a Mobile Edge Computing (MEC) framework in order to increase the reliability of the deployed applications. MEC is a promising framework to reduce the cloud core utilization as well as to provide applications that require low latency to mobile end users. As near as possible to the location of the end user, MEC relates to the mobile network applications and data stream acceleration through caching and/or compressing of relevant data at the edge of the mobile network. Although MEC is not yet deployed in real-life systems, recent studies discussed some technical details and related concepts. This paper studies the definition of MEC and similar concepts with respect to typical application scenarios, like in smart cities, and provides scope and limitations that may be encountered when implementing and deploying MEC. Then, a MEC framework is used to explore the data management infrastructure for upcoming smart cities.

Peer-review under responsibility of the Conference Program Chairs

Keywords: Smart City; Data Management; Cloud Computing; Mobile Edge Computing (MEC).

1. Introduction

A city is called smart, if it uses information technology for data communication in order to improve the performance and quality of the civil services, decrease the consumption of the resources and involve interactive applications and

* Corresponding author. Tel.: +962-786-118-190; fax: +962-2-720-1077.
E-mail address: mqquwaider@just.edu.jo

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2016.04.257

active services with people. The areas/domains for which smart city technology are considered include water and waste, health care, government services, energy, traffic management and transportation. The goal of the developed applications in smart city is to improve the organization of city flows and to allow real time responses for any challenge. It is noticed that, a smart city application should be more prepared to response to the instantaneous challenge, rather than, for just a simple interaction relation between the city citizens. On the other hand, with the increasing importance of deploying smart city technologies and applications, which resulted in an increasing amount of data generated by these applications, the need for a holistic data management system increased due to the need of efficiently collecting and processing these data. Many recent examples show the importance of deploying smart city applications for surveillance and incident control. The availability of efficient data management system for smart city is crucial for detecting and preventing an emerging incidence.

Several terms for Mobile Cloud Computing (MCC) such as Mobile Edge Computing (MEC), fog computing, dew computing and mist computing have been introduced recently to promote the idea of pushing the computing resources from the core of the network to the "edge." The proposed platforms revolve around the basic idea that (mobile) devices with limited capabilities can get access to more powerful resources of nearby devices without incurring the cost of reaching the resources at the core. However, these platforms differ in certain things such as the notion of a network edge. The differences are discussed in detail in this paper in order to find the platform that fits well with the needs and constraints of future smart cities applications.

The goal of this paper is to develop cloud support data management infrastructure for upcoming smart cities using a multi-tier cloud system infrastructure. The core objective of the proposed system is to detect any instantaneous challenge in the collected data to reduce the consumption of the city resources and to improve the civilian life. The proposed system computes the end-to-end delay, from the application level to the global cloud of the city service provider in real-time manner. The proposed system is minimizing the data processing delay by choosing different cloud providers in different tiers so that the overall delay is minimized.

2. Related Work

Recently, with the development of city service technologies, people are requiring access to the services and the resources related to them at any time from any place. In this context, human subjects can take advantage of this fact, in terms of smart city by deploying Cloud-based solutions on the city services. On the other hand, the growth of city data, application service providers need a global city-services awareness system with advanced techniques and resources in terms of process, store, share and transfer capabilities.

Towards cloud based big data analytics for smart future cities has been introduced in[1]. The authors in this paper proposed prototype to demonstrate the effectiveness of the analytics service for big data analysis in order to provide citizen perception about crime and safety versus economy and employment. A grid web service in a smart city was proposed in[2]. Three tiers architecture of infra, middleware and portal where proposed to construct the proposed smart city. The authors in this paper showed that the proposed architecture can be serviced in android smartphone or tablet personal computers in the presence of big data. Towards a big data analytics framework for smart city applications has been proposed in[3]. The authors in this paper present a case study in the smart grid domain that illustrates the high level requirements towards such an analytical big data framework for proposed smart city. The authors in these proposed architectures did not consider the distribution of collected data in multi-tiers large scale settings or the time complexity of the processed data.

Scalability and cost are two important goals that need to be considered when we talk about processing of big data. Here we are talking about processing and communication cost for such big data. Authors in[4] presented replication, partitioning, caching and distributed control as four levels for multi-tier database application. These architectures can be seen in applications like in Google and Azure and with small modules bases. However, the proposed architectures do not serve in data processes as big data collection as in eHealth applications. The authors in[5] presented automatic virtual machine configuration for database workloads in order to share a common pool of data for running and management occurrences in the lower tiers without considering the scalability and the cost of the collected data.

The authors in[6] presented a framework that analyzed the processing of big data in a parallel manner. Such big data is collected through wireless sensor networks of several deployments (the authors concentrated on environmental monitoring systems). Because of the special nature of the huge amounts of data, big data models and MapReduce solutions have been adopted. The MapReduce solution model presented involved four stages: acquisition, aggregate, range, and spatio-temporal analysis. The authors in[7] presented a survey of MapReduce. They highlighted different advantages and disadvantages of using MapReduce in processing data of huge nature. The authors discussed some of these advantages such as parallel execution of massive data in a timely manner, simplicity of the MapReduce framework, its scalability and its fault tolerance. The authors also mentioned that the MapReduce framework should not be considered as a replacement of the traditional DBMS systems. In fact, the integration of both the MapReduce and the DBMS should be the concentration of new frameworks and/or paradigms so that the advantages of both can be realized. The authors in[8] introduced a new direction of using the MapReduce framework that is different from the traditional usage of MapReduce to handle and process big data. In their paper, they extended the MapReduce to be used in handling and processing several small, almost real time and interactive jobs. This involves a data of heterogeneous nature and small jobs of huge number and interactive. This in fact shifted the MapReduce to be applied to many systems with query processing systems.

Intercloud, or what is called cloud of clouds, is an interconnected global network of networks (an extension of the Internet)[9]. Intercloud is used for federation of Cloud computing environments in utility oriented base, and it should support scalable applications in different tiers[10]. The first published work on Intercloud was few years ago after the spread of cloud computing infrastructure in many aspects[10-13]. In this context we should differentiate between two terms, MultiCloud and Federation. MultiCloud indicates when a client uses independent and multiple clouds without implying volunteer sharing and interconnection of providers' infrastructures. While a Federation indicates when there are a set of interconnected volunteer cloud providers sharing of resources among each other[14]. None of the above works integrated data processing, storing, sharing and transmitting with Intercloud as part of global health awareness system, and as a coupled with of MapReduce model in multi-tiers cloud provider.

In this paper we present a novel multi-tier cloud provider data management infrastructure for upcoming smart cities. The proposed system optimizes data processing time in multiple tiers in the presence of scalable data collection. The proposed system integrates the end data in large scale monitoring system in order to detect any instantaneous challenge in the collected data to reduce the consumption of the city resources and to improve the civilian life. To the best of our knowledge, the novelty of the proposed work has never been addressed in any previous studies.

3. Upcoming Smart City Infrastructure

The growth in exploiting the available computing and communication resources to handle different large scale applications, such as distance learning, space monitoring and climate change monitoring, open the doors for new applications to benefit from these enormous resources. Locally controlled city monitoring systems are becoming an urgent need with the increase in civil services that are not contained within some regions of a city. These civil services include water and waste, health care, government services, energy, traffic management and transportation, are based on large scale monitoring for the developed city with different sizes.

Developing such smart city systems is a challenge for many reasons. First, effective systems should cover large geographical areas with different terrains and population in order to provide the required services. This will create a challenge in deploying the systems and maintaining their operations. Second, achieving efficient smart city system on a large scale should be accompanied with reliable computing infrastructures that facilitate data sharing and storage in a timely manner. The availability of the computing resources cannot be guaranteed, many target resources are suffering from a shortage in advanced computing and communication systems. Third, the computing infrastructure should span the entire process starting from the data collection system in one location all the way to the global decision making system that covers the entire city. This will require collective efforts from different parties such as the governments, health organizations, and computing and communication service providers. In this paper we are focusing on the latter one.

Figure 1 shows the top level overview of our proposed smart city system. The system is composed of three cores. These cores from the bottom to the top are *Monitored Area, MEC* and *Cloud*.

- *Monitored Area (MA)*: Monitored area is representing the area at which the end data need to be collected in order to provide certain civil services. This data can represent human subject's data, traffic rate, pollution, water and waste usage, etc. Data collection sensor nodes and communication capabilities devices are deployed in the MA. Such devices have the ability to communicate with *MEC* service provider using different communication technologies, such as *Bluetooth*, *Wi-Fi* or 3G technologies, as shown in Figure 1. The number of MAs can range from few MAs to hundreds of MAs, depending on the applications and the size, the density and the sub-regions of the city. The size and location of the MA depends on the number of services and applications that need to be covered by *MEC*[15, 16].

- *Mobile Edge Computing (MEC)*: It is composed of a local cloud system, which represents inexpensive, resource efficient, easy to deploy and moveable computing system with communication capabilities connecting it with *MAs* end data collection and other computing facilities on the *Cloud*. A *MEC* consumes less amount of power for processing and communication with an acceptable quality of service compared with *Global Cloud Providers*[15]. The number of *MEC* in the city depends on the size of the monitored area and the number of services and applications. Each *MEC* can handle data from tens or may be hundreds of *MAs* within the covered area. The covered area of *MEC* is determined by the transmission range of the *Bluetooth* and *Wi-Fi* technologies[17]. At the same time, *MEC* has the capability of running MapReduce operations[18] to extract the abnormal data in the covered area to take any urgent action.
- *Cloud and Services*: It is composed of a *MEC* cloud system, which represents cloud resources that are specific to a zone, or a region, which can only be accessed by *MAs* sensor nodes. This core is the central processing and coordinator system that aims to process the collected data from different *MEC* providers in large scale to generate useful facts and observations. Then, these observations are used to find abnormal phenomena within the collected data, or to detect or predict any loss, waste or degradation in the provided service within the monitored area. Finally, a control decision can be taken by the service provider against that phenomenon. The *Cloud* system represents the top core and it has two different functionalities which are data processing and the overall system operations control for the whole city. The *Cloud* core is a high capacity system with the ability to process large amounts of data aggregated from the lower cores. It is connected directly to *MEC* cloud systems through a high speed backbone network to reduce data transfer latency. But on the other hand, processing large scale data in the *Cloud* core will put extra cost in terms of time and data communication.

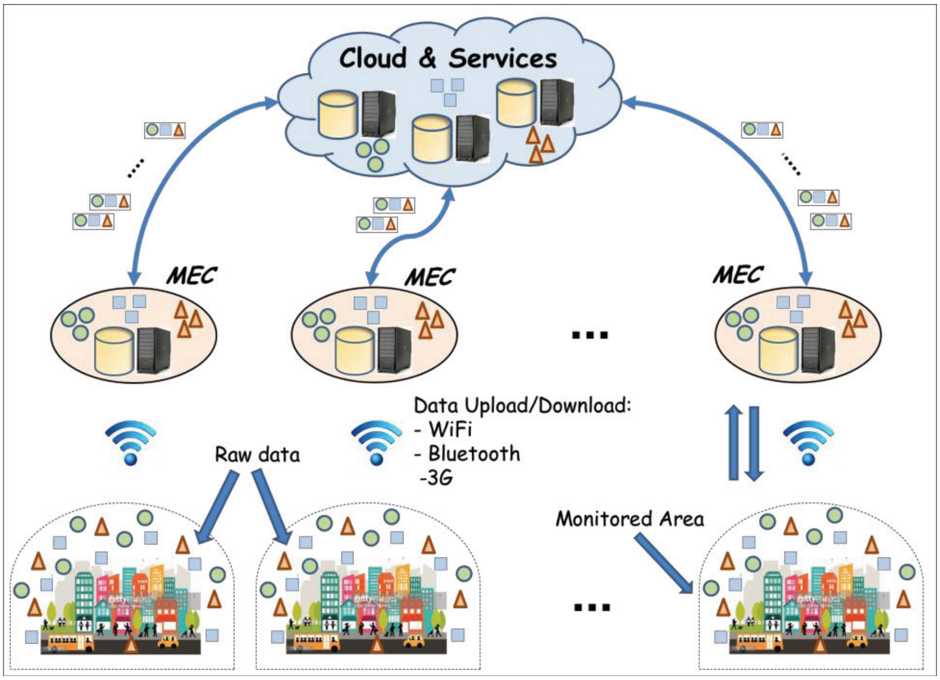

Figure 1: Smart city data management infrastructure using mobile edge computing

4. Multi-Core Data Processing

The goal of this section is to determine at which core the collected data should be processed, and then the downlink action be taken, in order to provide more reliable system for upcoming smart city. An optimization problem should be considered to determine the best processing core for a large scale smart city system. In addition, the optimization problem should consider the computed strategy to may include processing the collected data in a decentralized fashion which we called computing at the edge or processing at the edge. In this strategy, the collected data is processed in the lower core closer to the source of data. But the problem with that, an arrangement between different monitored areas need to be considered in order to have any alternative solution or action. On the other hand, processing the whole *MEC*s data in centralized *Cloud* center will cost less processing delay, but extra communication cost. Therefore, processing the collected data in *MEC* should provide an optimum solution for reliable upcoming smart city. In this case the transferred data size and processing delay are less compared with processing the whole data in a center *Cloud* core system. The main contribution of the proposed data management infrastructure for upcoming smart cities is to minimize the computation delay, from the source of the collected data in *MA* sensors to the *Cloud* service provider.

5. Conclusion and Future Work

In this paper, we proposed a novel cloud support data management infrastructure for upcoming smart cities that overcome the grand challenges associated with such systems and to improve civil services include water and waste, health care, government services, energy, traffic management and transportation. The system is built on the top of the resources rich and easy to deploy cloud computing system.

References

1. Z. Khan, A. Anjum, K. Soomro, and M. A. Tahir, "Towards cloud based big data analytics for smart future cities," *J. Cloud Comput.*, vol. 4, no. 1, pp. 1–11, 2015.
2. J. P. Park, C. H. Yun, H. S. Jung, and Y. W. LEE, "Mobile Cloud and Grid Web Service in a Smart City," in *Proceedings of The Fifth International Conference on Cloud Computing, GRIDs, and Virtualization*, 2014, pp. 20–25.
3. M. Strohbach, H. Ziekow, V. Gazis, and N. Akiva, "Towards a big data analytics framework for IoT and smart city applications," in *Modeling and Processing for Next-Generation Big-Data Technologies*, Springer, 2015, pp. 257–282.
4. D. Kossmann, T. Kraska, and S. Loesing, "An evaluation of alternative architectures for transaction processing in the cloud," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, 2010, pp. 579–590.
5. A. A. Soror, U. F. Minhas, A. Aboulnaga, K. Salem, P. Kokosielis, and S. Kamath, "Automatic virtual machine configuration for database workloads," *ACM Trans. Database Syst. TODS*, vol. 35, no. 1, p. 7, 2010.
6. C. Jardak, J. Riihijärvi, F. Oldewurtel, and P. Mähönen, "Parallel Processing of Data from Very Large-scale Wireless Sensor Networks," in *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*, New York, NY, USA, 2010, pp. 787–794.
7. K.-H. Lee, Y.-J. Lee, H. Choi, Y. D. Chung, and B. Moon, "Parallel data processing with MapReduce: a survey," *AcM SIGMoD Rec.*, vol. 40, no. 4, pp. 11–20, 2012.
8. Y. Chen, S. Alspaugh, and R. Katz, "Interactive Analytical Processing in Big Data Systems: A Cross-industry Study of MapReduce Workloads," *Proc VLDB Endow*, vol. 5, no. 12, pp. 1802–1813, Aug. 2012.
9. N. Grozev and R. Buyya, "Inter-Cloud architectures and application brokering: taxonomy and survey," *Softw. Pract. Exp.*, vol. 44, no. 3, pp. 369–390, 2014.
10. R. Buyya, R. Ranjan, and R. N. Calheiros, "Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services," in *Algorithms and architectures for parallel processing*, Springer, 2010, pp. 13–31.
11. D. Bernstein, E. Ludvigson, K. Sankar, S. Diamond, and M. Morrow, "Blueprint for the Intercloud - Protocols and Formats for Cloud Computing Interoperability," in *Fourth International Conference on Internet and Web Applications and Services, 2009. ICIW '09*, 2009, pp. 328–336.
12. B. Rochwerger, D. Breitgand, E. Levy, A. Galis, K. Nagin, I. M. Llorente, R. Montero, Y. Wolfsthal, E. Elmroth, J. Caceres, M. Ben-Yehuda, W. Emmerich, and F. Galan, "The Reservoir model and architecture for open federated cloud computing," *IBM J. Res. Dev.*, vol. 53, no. 4, pp. 4:1–4:11, Jul. 2009.
13. A. Celesti, F. Tusa, M. Villari, and A. Puliafito, "How to Enhance Cloud Architectures to Enable Cross-Federation," in *2010 IEEE 3rd International Conference on Cloud Computing (CLOUD)*, 2010, pp. 337–345.
14. A. J. Ferrer, F. HernáNdez, J. Tordsson, E. Elmroth, A. Ali-Eldin, C. Zsigri, R. Sirvent, J. Guitart, R. M. Badia, K. Djemame, and others, "OPTIMIS: A holistic approach to cloud service provisioning," *Future Gener. Comput. Syst.*, vol. 28, no. 1, pp. 66–77, 2012.
15. M. Quwaider and Y. Jararweh, "Cloudlet-based Efficient Data Collection in Wireless Body Area Networks," Simulation Modelling

Practice and Theory, Volume 50, January 2015, Pages 57-71
16. M. Quwaider and Y. Jararweh, “An efficient big data collection in Body Area Networks,” in *2014 5th International Conference on Information and Communication Systems (ICICS)*, 2014, pp. 1–6.
17. S. Biswas and M. Quwaider, “Remote Monitoring of Soldier Safety through Body Posture Identification using Wearable Sensor Networks,” *SPIE Def. Secur. Symp. Multisens. Multisource Inf. Fusion Archit. Algorithms Appl.*, pp. 1–14, 2008.
18. M. Quwaider and Y. Jararweh, “A cloud supported model for efficient community health awareness,” *Pervasive Mob. Comput.*, 2015.