2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress

A Cloud-based Network Architecture for Big Data Services

Ming Zhao*, Arun Kumar†, G. G. Md. Nawaz Ali*, and Peter Han Joo Chong‡
*School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore
†Department of Electrical and Computer Engineering, National University of Singapore, Singapore
‡Department of Electrical and Electronic Engineering, Auckland University of Technology, New Zealand
Email: zhaoming@ntu.edu.sg; elearun@nus.edu.sg; nawaz.ali@ntu.edu.sg; peter.chong@aut.ac.nz

***Abstract*—In recent years, due to the development and pervasiveness of information and communication technology (ICT), data is generated at a tremendously increasing rate from various types of data sources. This massive amount of data is widely referred as Big Data, which contains large amount of hidden rich information. By uncovering the valuable information, Big Data services are envisioned to bring huge potential for commerce, business and research. In fact, series of sophisticated processes are involved in Big Data services. However, there exists a structural gap, which is holding back the development of Big Data services. In this paper, a four-layer cloud-based network architecture is proposed to support Big Data services. The proposed network architecture provides a systematic and efficient approach for Big Data access, storage and retrieval. Specifically, the processes for supporting Big Data services are categorized into data transfer, data collection, data processing and data retrieval. Furthermore, a cloud-based architecture is presented, followed by a closer look at the network architecture for Big Data collection at device level.**

***Index Terms*—Big Data services, cloud-based networks, data transfer, data collection, data processing, data retrieval.**

I. Introduction

New information and communication technologies (ICTs) continue to penetrate our everyday lives, as higher numbers of people and things are getting connected. With the advancement of ICTs, information sources are becoming highly diverse, ranging from embedded devices to daily-updated news on the social media. As much as ICT is everywhere, so is the data due to the emergence of Internet of Things (IoT) [1], [2]. Along with this trend, the concept of Big Data [3], [4] arises and it typically refers to huge amount of data sets that are difficult to collect, store, analyse and process with traditional data analytics and management tools [5]. Due to the broad information and prospective opportunities, Big Data has drawn considerable attention from both industries and academias in recent years. Importantly, Big Data is of extreme importance to boost the productivity in businesses and scientific disciplines. In fact, the Wireless World Research Forum (WWRF) has recently predicted 7 trillion wireless devices for 7 billion people by 2017 [6], resulting in around a thousand devices for every human. The amount of data that these 'things' have is a huge source of knowledge and innovation, which has high economical value. Global enterprises like Google and Amazon have already shown that it is possible to transform data to economic value by producing novel and popular services based on intelligent use of massive data sets. However, it is extremely difficult to seamlessly and effectively collect, store, search, share, process, analysis, and visualize big data due to its indistinctive nature and sophistic processes associated with it.

Majority of reported works [7], [8], [17] on Big Data in academia and enterprises can be categorized into two approaches. First, how to efficiently collect and store it, and second, how to make sense of it. While both problems are undoubtedly important, the problem of data gathering is severely overlooked. One of the reasons for this is that the data we are talking about is still largely produced and stored in purpose, e.g., due to activities in social media or various processes and events in enterprises and industry. The world of data is radically changing along with IoT. In future, multitude of sources for any type of data will exist everywhere. Most of this data will not be associated with an explicit process for collection and storing. In addition, those 'things' having the data or measuring the data are also mostly wireless things (sensors, devices or RFID tags) having largely no permanent connection to gateways (GWs) for easy data delivery to data centres. In addition, those 'things' usually have limited energy resources. In the era of Big Data, the existence of a centralized data collector and efficient data gathering technique is one of the major concern to utilize the great potential of Big Data.

Considering that the Big Data services involve sophisticated processes [10], [11], [12], [13], it is difficult to design efficient approaches to support Big Data services without understanding its structure. With a proper structure describing the processes of Big Data services, protocols and algorithms can be developed for individual layers, significantly reducing the complexities to support Big Data services. To fortify these challenges, this paper proposes a four-layer network architecture for Big Data services. With the proposed architecture, Big Data services can be attained with low resources and quick data processing. In addition, fast and real-time data storage and access can be achieved. Specifically, this paper first presents the general architecture of Big Data analysis. Then, the nature of Big Data and its requirements on the communication support are introduced. Subsequently, a four-layer network architecture, including data transfer, data collection, data processing and data retrieval are introduced, followed by the introduction of a cloud-based architecture for Big Data services and the network architecture for Big Data collection at device-level.

978-1-5090-4065-0/16 $31.00 © 2016 IEEE
DOI 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.118

The reminder of this paper is organized as follows. Section II introduces the essential components for Big Data services. Section III presents the proposed network architecture for Big Data services. Section IV concludes this paper.

II. Essences of Big Data

This section introduces a general architecture for Big Data Analysis. Then, the nature of Big Data is defined as 5Vs. Specifically, *versatility* is defined as the fifth's V in the perspective of communication support.

A. Architecture for Big Data Analysis

Nowadays, Big Data applications essentially require effective analyses of large amount of datasets [14]. Fig. 1 shows the general architecture of Big Data analysis which comprises three main domains; *Big Data Collecting*, *Big Data Storing* and *Big Data Processing*[15]. 1) *Big Data Collecting*: massive increase in smart devices and simultaneous proliferation in various types of IoT applications, e.g., smart grid and healthcare, strains the expanding source of big data. This essentially drives research on networking towards an efficient data collection approach [16]. In addition, the massive amounts of Big Data requires efficient data retrieval and storage solutions. 2) *Big Data Storing*: distributed storage systems are on the way to be widely deployed to provide necessary capacity and high performance. Specifically, cloud-based data storages are regarded as promising technologies to store the data in a distributed and efficient manner. 3) *Big Data Processing*: representation and transformation of Big Data in a meaningful way is a vital approach to analyse the information used for detection, monitoring and prediction [17]. Furthermore, detailed analysis can be done by increasing the rate of data stream or scalability of transformed data stream.

B. Big Data Services with Versatile Communication Support

Initially, big data is characterized by three Vs, namely, volume, velocity and variety. Subsequently, the concept of big data is extended to four Vs, i.e., volume, velocity, variety and value [4]. The definition of this 4Vs essentially highlights the necessity and the meaning of Big Data. Big Data refers to diverse, complex, large scale datasets which requires new technologies and techniques to collect, process and integrate in order to discover the great potential value of it.

Volume refers to large amount of different datasets generated by different types of sources. This property of Big Data brings great potentials and challenges for Big Data services. Uncovering hidden information from this large amount of data relies on efficient data analysis.

Velocity refers to constant accumulation of new data. The data collection speed is diverse due to variety of applications. Furthermore, the dynamics of transmission medium and hash environment make it difficult to predict the speed of generated data.

Variety refers to different kinds of data, such as text, image, video and audio etc. These datasets are generated by diverse sources, ranging from sensors, mobile devices, surveillance systems, e-health applications to different types of social networks.

Value is of paramount importance in Big Data. It refers to the process of uncovering the valuable hidden information in the large amount of datasets.

Big Data is generated from various types of networks, such as sensor networks, wireless local area networks (WLANs), cellular networks. Providing Big Data services requires effective and efficient solutions to deal with heterogeneous networks. Importantly, versatile communications support over heterogeneous networks is essential for Big Data services. In order to support Big Data services across large numbers of embedded devices, the amount of autonomously transferred data is tremendously large. Consequently, Machine-to-Machine (M2M) communications [18] is one of the key enabler for Big Data services. M2M communications refers that large numbers of embedded devices efficiently share information and collaboratively make decisions with little or no human intervention. In addition, a universal communication platform is also highly demanded. Effective Big Data services require seamless connection and efficient communication among het-

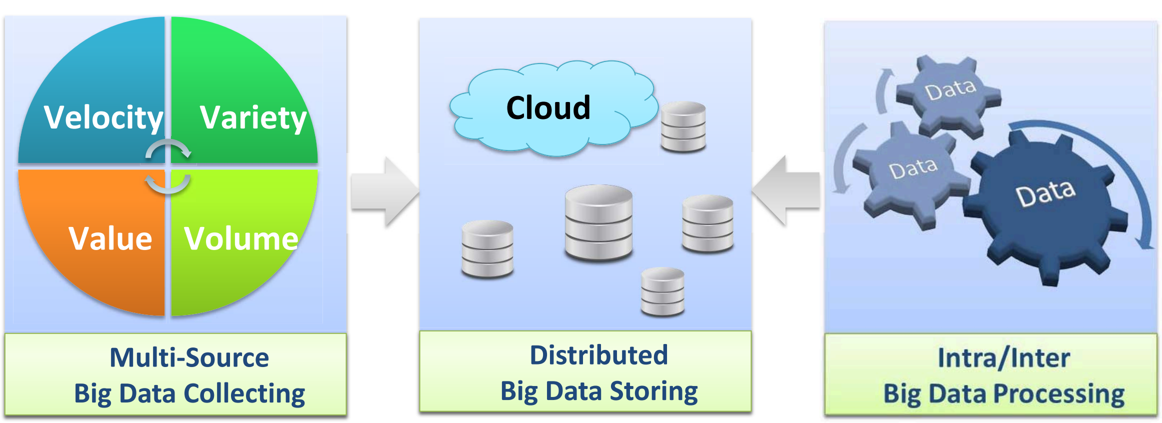

Fig. 1 General architecture of big data analysis.

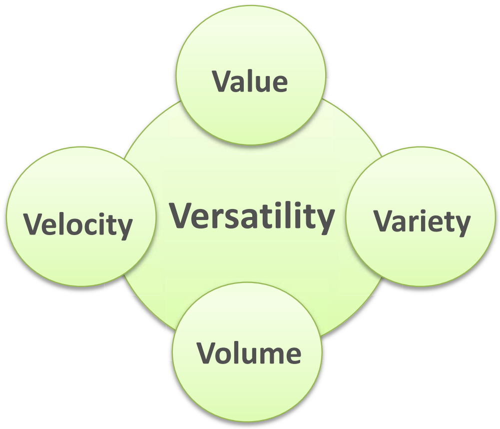

Fig. 2 Four Vs of Big Data with versatile communications support.

erogeneous networks. In fact, versatility of communication network is crucial to support Big Data services, as shown in Fig. 2. The main challenges of converging heterogeneous networks for efficient data transfer and collection are related to different access schemes and protocols.

III. Proposed Network Architecture for Big Data Services

In this section, the proposed four-layer network architecture is presented. Then, the proposed network architecture is further introduced with focus on cloud-based structure. In addition, this section provides a closer look into the network architecture for data collection.

A. Four-Layer Network Architecture for Big Data Services

The proposed four-layer cloud-based network architecture for Big Data services is shown in Fig. 3. The four layers are *data transfer*, *data collection*, *data processing* and *data retrieval*. Specifically, in the data transfer layer, communication relies on multiple wireless interfaces among different devices. Based on the data exchange and relay in the first layer, efficient routing protocols can be designed so as to collect massive datasets by the gateways in the data collection layer. Subsequently, the collected data is processed at the gateway and the cloud, which is the third layer. The fourth layer is the data retrieval that essentially bridges the stored and real-time data for the Big Data applications. With the support of these four-layers, Big Data services can be finally enabled. The detailed information of the proposed four layers are as follows:

Data Transfer: Ubiquitous connectivity among large numbers of devices is a fundamental requirement for Big Data services, which relies on efficient data transfer. Device-to-Device (D2D) communications [19] has been defined as an efficient approach to transfer data among wireless devices. Specifically, the local peer-to-peer data transmission plays an important role in D2D communications. Several essential components are involved in the data transfer layer of Big Data services. The first component is devices. Devices are usually different in terms of size, computational capabilities and resources. The second component is wireless technologies. Different wireless technologies, such as Wifi, ZigBee, Bluetooth, 3G/4G can be utilized to transfer data among devices. Various applications of IoT can be achieved with the help of the wireless technologies. The third component is wireless spectrum, which is a limited resource shared by multiple wireless technologies. In order to support Big Data services, dynamic spectrum access is vital for devices to choose different wireless access technologies and operating spectrum. Cognitive radio techniques that allow devices to reuse channels in cellular or other bands, can be applied to support the dynamic spectrum access. Importantly, as heterogeneous networks are involved in Big Data services, unified framework is crucial to address the data transfer with multiple wireless interfaces in a coherent way to achieve an optimal performance given resource constraints.

Data Collection: Based on efficient data transfer in the first layer, robust, reliable, scalable and energy-efficient communications protocols are required to collect/gather large amount of data at gateways to support Big Data services. The collection of large volume of data seeks efficient wireless solutions as they are cost-effective, easy-to-install, scalable and flexible. However, direct communication is rarely possible in large scale networks due to limited range of wireless radio signals. Thus, data has to be relayed through multi-hops before it reaches the destination. Consequently, routing plays a crucial role in big data collection. In recent years, a paradigm shift of routing from small-scale networks to large convergecast networks that consist of multitude of devices has taken place [20]. Importantly, in the proposed architecture, data collection takes place in either reactive or proactive manner. Specifically, on-demand data collection can be triggered by an event or users' request. In this case, real-time data is collected by the gateway. Normally, latency is an important performance metric in the on-demand data collection.

Data Processing: In the proposed architecture, the cloud can handle data storage and data processing. Meanwhile, part of data collection, processing and services are kept out of the cloud. Specifically, the cloud plays an important role in Big Data services by collecting, processing and storing necessary and frequently needed data, as well as managing the customers' requests when needed. Rather than being handled by the cloud, part of requests will be handled by distributed processing units closer to users. These processing units may ask the cloud to provide necessary data and analysis results when needed [21]. More importantly, it will be responsible for the activation of a certain portion of the data collected. The collected data, together with the data from the cloud if any, will be either forwarded to the users directly or be processed by the unit and then with the processed results will be sent to the users, depending on the nature of the service the user has requested. In case, where the processing unit cannot process the user's request, the request may be forwarded to the cloud for better understanding and handling.

Data Retrieval: The top layer of Big Data services is the data retrieval. Two types of data, i.e. stored data and real-time data are retrieved to support the applications of Big Data services. In the proposed architecture, the cloud provider can have access to all the stored data, imposing serious threat privacy. In fact, the cross-domain data retrieval with security techniques has attracted considerable attentions with the development of Big Data services.

B. Cloud-based Network Architecture for Big Data Services

This section further introduces the proposed four-layer network architecture with a focus on cloud-based support for Big Data services. An overview of the proposed cloud-based network architecture for on-demand data services is shown in Fig. 4. At the edge of the network, the GWs collect data from respective wireless sensor networks (WSNs). This data collection is bounded by certain data collection policies. The collected data is transmitted from respective access points (APs) of the WSNs to the GWs through a wireless link (e.g., Wi-Fi and cellular networks (3G/4G)) or a wired link (e.g., Ethernet). All GWs are connected to a cloud infrastructure via a nearby switch node (SN, typically consisting of an optical cross-connect (OXC) and an IP router) of the network. In a cloud infrastructure, a number of data centers (DCs) provide large storage and computing capacities, and are connected to different SNs in a distributed manner. These GWs can also retrieve the data from the cloud. Each GW contains a number of software components and servers. These servers are divided into three classes: 1) application/web servers such as wireless application protocol (WAP), Web, FTP, etc, which can retrieve or deliver data from/to users; 2) administration and access control servers, which are used to perform authentication and access control to protect privacy of users; 3) memory cache servers, which are used to store popular data temporarily, to expedite the data access by users and avoid network congestion. This research work consists of three integrated parts each of which is briefly described below.

1) Network architecture design: A three-level network model is designed in the proposed architecture: cloud-based infrastructure, GWs and WSNs. The cloud-based infrastructure offers computing, networking and storage capability, along with distributed data analytic and processing. The GWs with advanced data processing capabilities connect to the cloud infrastructure and to nearby WSNs. In order to further enhance the functionalities of the network architecture, several research issues can be addressed, including the designing of various functional modules and coordination between these modules in the network model, defining of the data flow and control flow between the upper and lower levels, and dynamic adjustment of functional modules according to available resources and different application requirements.

2) Data processing and delivery techniques: As shown in Fig. 4, each GW is equipped with advanced data processing capabilities and can offer a number of application-related services. To avoid network congestion and to facilitate data access, data are classified into different types (e.g., video, image, text, real time or non-real time) for different applications. A novel data processing framework will be developed for GWs, based on data classification techniques for identifying data type and application scenarios. Each

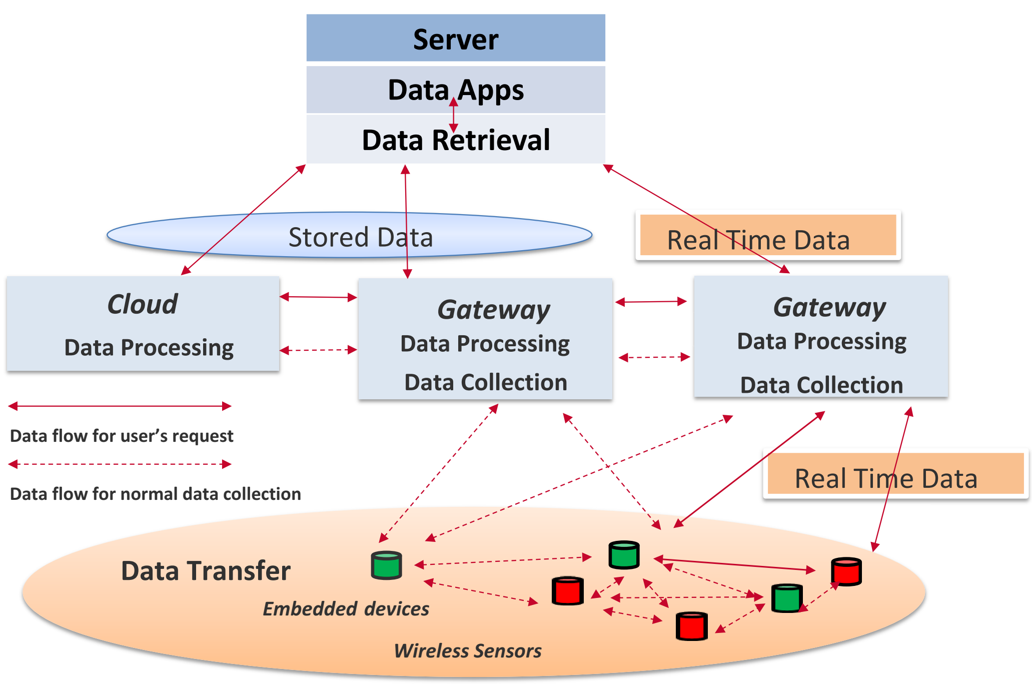

Fig. 3 The proposed four-layer network architecture for Big Data services.

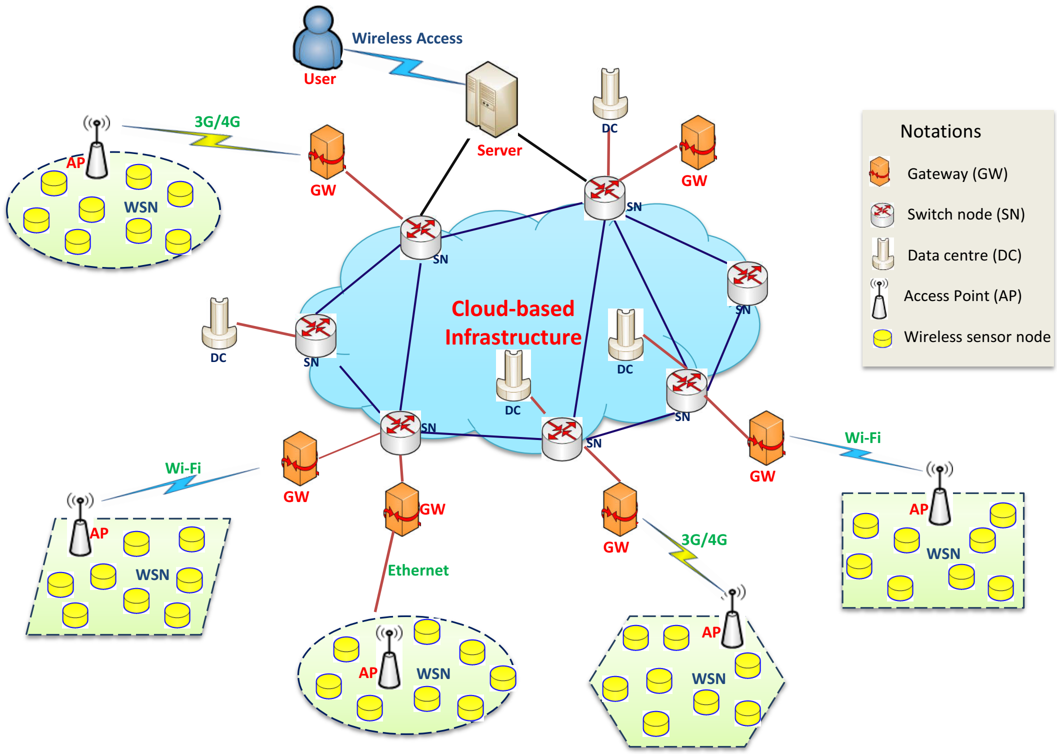

Fig. 4 The proposed cloud-based network architecture for Big Data services.

GW can adopt different data collection policies, depending on the application requirements, on-demand by adjusting parameters such as collection objections, collection scope, collection duration, and collection sampling frequency, etc. Novel data analysis and filtering techniques will be developed for extracting valuable information from large amount of raw data without compromising the privacy and security. Different approaches are adopted, depending on data types, to store/distribute the massive amount of data in an efficient manner. Specifically, data can be cached in GWs (to expedite the result data access), or stored in the cloud. Also, different data types should adopt different data distribution policies (e.g., centralized distribution, content delivery network, or proxy-server cache) and different data delivery approaches (e.g., multicast, peer-to-peer, or machine-to-machine). Different data types will be treated differently so that they can be efficiently distributed and delivered to users. And, for this purpose, novel data distribution and delivery techniques will be developed.

3) Network and resource management framework: Unified network and resource management framework is required, which includes three modules: computing task scheduler, network and storage resource manager and distributed/parallel data processing (such as data mining, data visualization, and data transfer). The management framework should be implemented in a distributed manner to make it scalable and robust. The design goal of the management framework is to dynamically balance the competing needs of various resources as various applications compete for a variety of resources including sensors, networks, computing resources, storage, energy, and wireless spectrum. So that the functionality, robustness, utility, and quality of service are guaranteed. Meanwhile, an energy-efficient resource discovery scheme, such as cognitive radio technology is also necessary so that various resources can be reconfigured dynamically and on-demand data services can be fulfilled efficiently.

C. A Closer Look at the Network Architecture for Big Data Collection (at device domain)

As mentioned earlier, Big Data services rely on efficient data collection approach. The huge benefits of Big Data services calls for robust, scalable, reliable, and energy-efficient routing protocol to support the data collection. The routing over low-power and lossy links (ROLL) working group from Internet Engineering Task Force (IETF) standardized the IPv6 Routing Protocol for Low Power and Lossy Networks (RPL) [22], [23] as the infrastructure protocol for IoT. RPL forms the network topology in a way that data collection at the gateway is optimized.

Targeting the networks consist of large number of energy-constraints devices operating in dynamic and lossy wireless mediums, RPL constructs a robust network topology over unreliable wireless links. Specifically, RPL forms the network topology as a destination oriented directed acyclic graph (DODAG), as shown in Fig. 5. The root node acts as the gate-

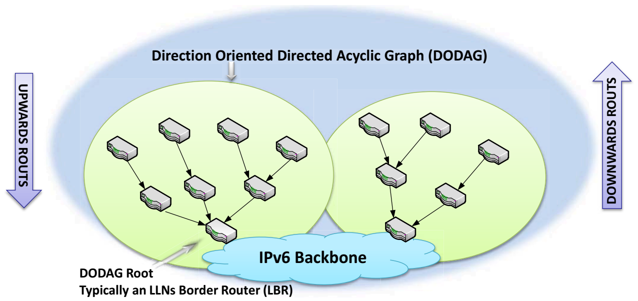

Fig. 5 Network architecture constructed by RPL.

way bridging internal networks with the external IPv6-based networks. RPL provides different traffic support including multipoint-to-point (MP2P), point-to-point (P2P) and point-to-multipoint (P2MP). Each node in the network has a preferred parent node that forwards data towards the root/gateway. RPL is implemented with trickle algorithm [24], so that RPL can set up the network in a fast, low-overhead, distributed manner.

IV. Conclusion

This paper proposed a four-layer network architecture for Big Data services, which described the process structure involved in Big Data services. In addition, the protocols and the algorithms are introduced which can be developed for each layers. This paper has presented a general architecture of Big Data analysis and services, and illustrated the requirements of versatile communication support. With detailed introduction of the proposed four-layer network architecture, including data transfer, data collection, data processing and data retrieval, a cloud-based architecture of Big Data services and a network architecture for Big Data collection are presented. The proposed network architecture provides a systematic and efficient approach for Big Data services.

With rapid proliferation of big data in diverse fields, numerous opportunities and challenges are brought before the researchers and engineers. In future, we plan to implement the infrastructure routing protocol with the proposed architecture for Big Data service. Experimental studies will be conducted in Network Simulator 3 (NS-3) to evaluate the performance of the proposed architecture.

Acknowledgment

This research is funded by the Republic of Singapore's National Research Foundation (NRF) through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

References

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[2] N. Bessis and C. Dobre, "Big data and internet of things: a roadmap for smart environments," *Springer International Publishing,* vol. 546, ed. 1, Switzerland, 2014.

[3] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: a technology tutorial," *IEEE Access*, vol. 2, pp. 652–687, 2014.

[4] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of big data on cloud computing: Review and open research issues," *Information Systems*, vol. 47, pp. 98–115, 2015.

[5] C. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[6] S. Babar, P. Mahalle, A. Stango, N. Prasad, and R. Prasad, "Proposed security model and threat taxonomy for the internet of things (iot)," *Recent Trends in Network Security and Applications*. Springer, pp. 420–429, 2010.

[7] M. Chen, S. Mao, Y. Zhang, and V. C. Leung, "Big data storage," *Big Data Springer,*, pp. 33–49, 2014.

[8] S. Madden, "From databases to big data," *IEEE Internet Computing*, no. 3, pp. 4–6, 2012.

[9] C. Ji, Y. Li, W. Qiu, U. Awada, and K. Li, "Big data processing in cloud computing environments," *In the 12th IEEE International Symposium on Pervasive Systems, Algorithms and Networks (ISPAN),* pp. 17–23, 2012.

[10] H. Demirkan and D. Delen, "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud," *Decision Support Systems*, vol. 55, no. 1, pp. 412–421, 2013.

[11] Z. Zheng, J. Zhu, and M. R. Lyu, "Service-generated big data and big data-as-a-service: an overview," *In the Proceedings of IEEE International Congress on Big Data (BigData Congress)*, pp. 403–410, 2013

[12] B. Geerdink, "A reference architecture for big data solutions-introducing a model to perform predictive analytics using big data technology," *International Journal of Big Data Intelligence*, vol. 2, no. 4, pp. 236–249, 2015.

[13] A. K. Mandal, S. Changder, A. Sarkar, and N. C. Debnath, "Architecting Software as a Service for Data Centric Cloud Applications", *International Journal of Big Data Intelligence*, pp. 77-92, Vol. 6(1), 2014

[14] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *Journal of Parallel and Distributed Computing*, vol. 74, no. 7, pp. 2561–2573, 2014.

[15] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *IEEE Network*, vol. 28, no. 4, pp. 46–50, 2014.

[16] X. Yi, F. Liu, J. Liu, and H. Jin, "Building a network highway for big data: architecture and challenges," *IEEE Network*, vol. 28, no. 4, pp. 5–13, 2014.

[17] C. Ji, Y. Li, W. Qiu, U. Awada, and K. Li, "Big data processing in cloud computing environments," *In the Proceedings of the12th IEEE International Symposium on Pervasive Systems, Algorithms and Networks (ISPAN)* , pp. 17–23, 2012.

[18] T. Bourgeau, H. Chaouchi, and P. Kirci, "Machine-to-machine communications," *Next-Generation Wireless Technologies*. Springer, pp. 221–241, 2013.

[19] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, 2012.

[20] T. Watteyne, A. Molinaro, M. G. Richichi, and M. Dohler, "From manet to ietf roll standardization: A paradigm shift in wsn routing protocols," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 4, pp. 688–707, 2011.

[21] S. J. Chen, J. Y. Huang, C. T. Huang and W. J. Wang "SAMEVED: A System Architecture for Managing and Establishing Virtual Elastic Data Centers", *International Journal of Grid and High Performance Computing*, pp. 27-42, Vol. 5(2), 2013.

[22] O. Gaddour and A. Koubâa, "Rpl in a nutshell: A survey," *Computer Networks*, vol. 56, no. 14, pp. 3163–3178, 2012.

[23] M. Zhao, H. Y. Shwe, and P. H. J. Chong, "Cluster-parent based rpl for low-power and lossy networks in building environment," *In the Proceedings of the 12th IEEE Annual Consumer Communications and Networking Conference (CCNC)*, pp. 779–784, 2015.

[24] P. A. Levis, N. Patel, D. Culler, and S. Shenker, "Trickle: A self regulating algorithm for code propagation and maintenance in wireless sensor networks," Computer Science Division, University of California, 2003.