Towards Real-Time Analytics for Mobile Big Data using the Edge Computing

Ibrar Yaqoob, Umer Majeed, and Choong Seon Hong
Department of Computer Engineering, Kyung Hee University, Yongin, 446-701 Korea
{ibraryaqoob, umermajeed, cshong}@khu.ac.kr

Abstract

The extraordinary proliferation of digital devices has led to the production of enormous amounts of heterogeneous data called mobile big data (MBD). However, obtaining insights into the flood of MBD in a real-time manner has become a crucial issue. Coping with this issue, we propose a theoretical framework based on the edge computing, which envisions bringing computation closer to the user's proximity. The proposed framework comprises three layers, where each layer is dependent upon the layers above it. The qualitative analysis results reveal the proposed framework can enable real-time analytics for the business environments.

1. Introduction:

Recent years have witnessed an astonishing rate of increase in spatiotemporal network-level data generated by broadband mobile networks, known as mobile big data (MBD). MBD analytics is a process of finding meaningful patterns and extracting knowledge from the flood of mobile users' data collected at the network-level or the application-level [1]. It is reported in [2], currently, over five billion people are using smartphones. In addition, it is predicted that 50 billion devices will be connected to the Internet by the end of 2020 (44 times greater than that in 2009). On the other hand, mobile application vendors are collecting massive amounts of data implicitly or explicitly using crowdsourcing schemes. Figure 1 shows the growth rate of mobile data traffic and the number of mobile phone Internet users, which are putting an intolerable pressure on the network systems [2]. The massive amount of data is currently mostly being processed and analyzed in the central cloud, which will be inadequate in the future because of the latency issues [3][4]. Thus, a necessity is arising to process and analyze the data in a real-time manner. This study aims to address this problem by proposing a framework based on the edge computing concept which envisions bringing computation closer to the user's proximity.

The contributions of this study are two folds: (1) We investigate the recent research works directed at the MBD era in terms of analytics. (2) We propose a framework for enabling real-time analytics for MBD.

These contributions are provided in the separate sections from 2-3. We summarize our study in section 4.

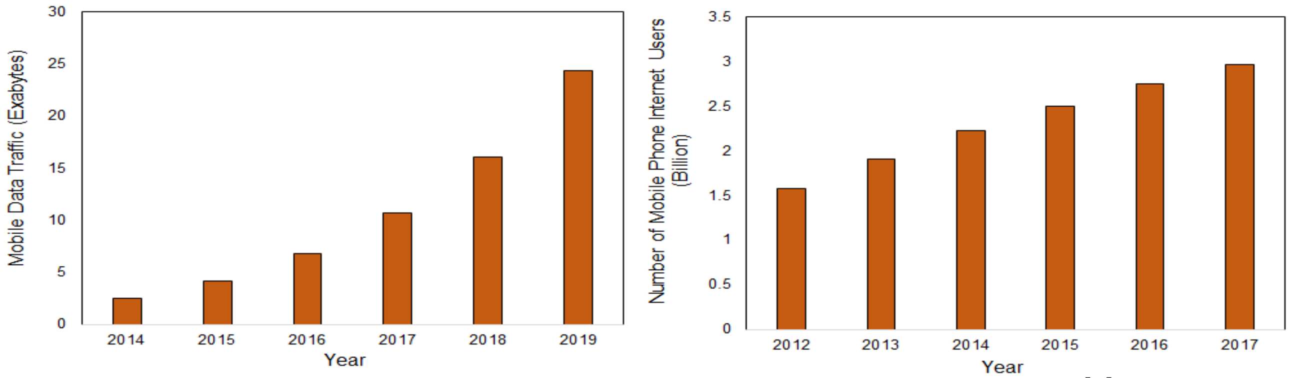

Figure 1: The growth rate of mobile data traffic and number of mobile phone Internet users [2]

2. Recent Advances:

This section critically investigates the recent studies conducted on the MBD era in terms of analytics.

The study [5] has proposed a scalable Spark-based framework for enabling deep learning in MBD analytics. The framework helps to add value by learning essential characteristics of the collected raw MBD. The framework was validated through a large-scale system. Although the study focuses on MBD analytics, how to perform analytics in a real-time manner is one of the remaining concerns.

In [6], an anomaly detection solution based on machine learning technique was presented for MBD. The solution analyzes the users' activities at the different time and location through k-means and hierarchical clustering techniques. The user activities causing unusual high traffic were considered as anomalies. These anomalies were verified with the ground truth information, i.e., there was a soccer match at that location and thus the traffic rate was very high than on the usual days. Detecting the anomalies in an advanced manner can lead towards an efficient resource allocation. The study further revealed that utilization of big data analytics solutions can help to analyze the users' contextual information, such as mobility pattern, traffic pattern, to name a few. The extraction of such information helps in developing intelligent resource allocation algorithms which result in the efficient resource utilization. Despite many advantages of the presented work, the solution overlooks the high latency issue while extracting invaluable information from MDB.

The work [7] has demonstrated that machine learning algorithms can help to identify interference issues in 4.9 and fifth generation mobile networks. The networks are collecting billions of measurement samples used for resource allocation and link adaptation. These measurements provide details about the service quality experienced by the customers at their locations, thus, applying big data analytics methods can help to extract the useful information. Later on, optimization methods can be applied based on the extracted information that can help to lower the interference levels and improve the throughputs and data rates in the mobile networks. Although this study helps to fix several indispensable issues in the mobile networks, the study does not focus on how the massive amounts of network generated data can be analyzed in a real-time manner.

The concept of employing edge computing for the Internet of things (IoT) was introduced in [8]. In this context, this study has proposed a new architecture called GigaSight, which is a hybrid cloud architecture that employs a decentralized cloud computing platform in the form of cloudlets. The study envisions utilizing data analytics techniques on the cloudlets, which are near to the user proximity (only the end results along with the metadata are required to send to the cloud). Although this work enables real-time analytics for the IoT paradigm, the work is in its infancy.

Unlike the above-mentioned works, we propose a theoretical framework, which envisions enabling real-time analytics for MBD using the edge computing.

3. Proposed Framework:

Figure 2 illustrates the proposed framework which can enable the real-time analytics for MBD by performing computation closer to the user's network.

The framework is comprised three layers. The first layer is responsible for collecting and transmitting the data to the above layer, whereas, the second layer is responsible for applying different data analytics tools and techniques at the edge of the network to extract the meaningful information and expose the hidden patterns which help the mobile network operators to improve the network efficiency. Additionally, the second layer is responsible for transmitting meta-data to the cloud. The third layer manages the results along with meta-data, i.e., data ID, Edge device IP address, owner/company details, location, and time stamp information. The proposed framework can provide certain benefits as follow:

(a) The proposed framework can help in reducing the time of data analytics. However, in the typical

environment, the data need to be processed in the central cloud, which prolongs the data analytics time.
(b) The second benefit of the framework is scalability of the analytics. (c) The framework reduces bandwidth costs. (d) Performing analytics at the edge for MBD can ensure high privacy and security as devices are in the user's proximity.

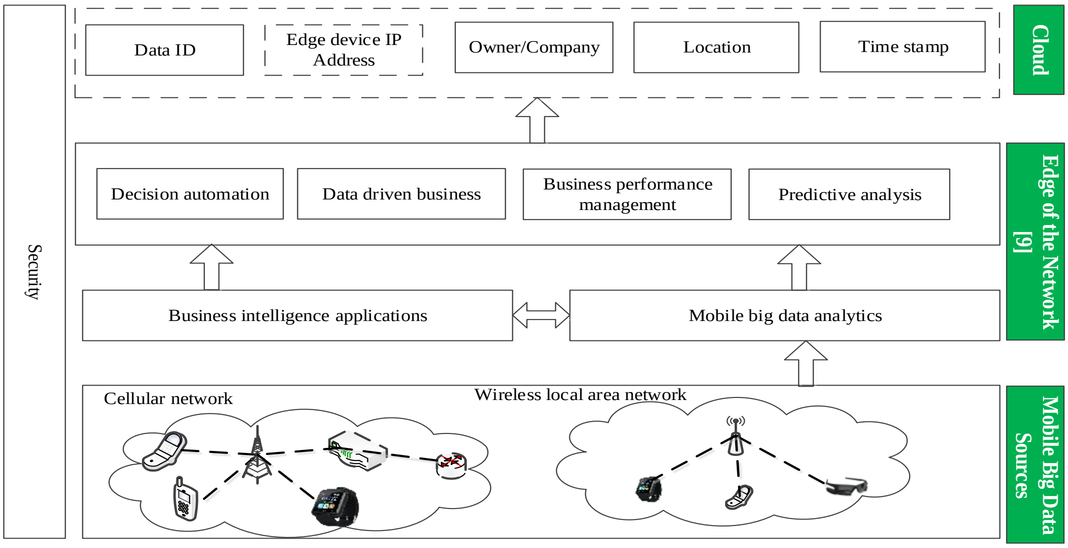

Figure 2: An illustration of the proposed framework

4. Summary:

Tremendous advancements in smartphone technology have enabled the people to share their data, resulted in the MBD era. Collected MBD is unprofitable unless some appropriate analytics techniques are utilized to extract meaningful information and expose the hidden patterns in a real-time manner. In this paper, we initially explored and critically analyzed the recent advances made in the MBD era in terms of analytics. Then, we proposed a theoretical framework which helps in enabling real-time analytics for MBD. In the future, we are planning to implement and validate our framework using a large-scale system.

Acknowledgment:

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2018-2015-0-00742) supervised by the IITP (Institute for Information & communications Technology Promotion).
*Dr. CS Hong is the corresponding author.

[References]

[1] Yazti, Demetrios Zeinalipour, and Shonali Krishnaswamy,"*IEEE 15th International Conference on Mobile Data Management*, Brisbane, QLD, 2014, Vol.1, pp. 1-2.
[2] Ahmed, Ejaz, et al. "Recent Advances and Challenges in Mobile Big Data." *IEEE Communications Magazine* 56.2 (2018): 102-108.
[3] Su, Zhou, Qichao Xu, and Qifan Qi. "Big data in mobile social networks: A QoE-oriented framework." *IEEE Network* 30.1 (2016): 52-57.
[4] Wang, Kun, et al. "Mobile big data fault-tolerant processing for ehealth networks." *IEEE Network* 30.1 (2016): 36-42.
[5] Alsheikh, Mohammad Abu, et al. "Mobile big data analytics using deep learning and apache spark." *IEEE Network* 30.3 (2016): 22-29.
[6] Parwez, Md Salik, Danda B. Rawat, and Moses Garuba. "Big Data Analytics for User-Activity Analysis and User-Anomaly Detection in Mobile Wireless Network." *IEEE Transactions on Industrial Informatics* 13.4 (2017): 2058-2065.
[7] Chiu, Peter, et al. "Big Data Analytics for 4.9G and 5G Mobile Network Optimization," *IEEE 85th Vehicular Technology Conference (VTC Spring)*, Sydney, NSW, 2017, pp. 1-4.
[8] Satyanarayanan, Mahadev, et al. "Edge analytics in the internet of things." *IEEE Pervasive Computing* 14.2 (2015): 24-31.
[9] Hashem, Ibrahim Abaker Targio, et al. "The role of big data in smart city." *International Journal of Information Management* 36.5 (2016): 748-758.