Big Data Driven Optimization for Mobile Networks towards 5G

Research · November 2015

DOI: 10.13140/RG.2.1.2389.1923

6 authors, including:

Kan Zheng
Beijing University of Posts and Telecommunica…

Kuan Zhang
University of Waterloo

Periklis Chatzimisios
Alexander Technological Educational Institute …

Kan Yang
University of Waterloo

Big Data Driven Optimization for Mobile Networks towards 5G

Kan Zheng*, *Senior Member, IEEE,* Zhe Yang*, Kuan Zhang♯, Periklis Chatzimisios‡, *Senior Member, IEEE*, Kan Yang♯, and Wei Xiang†, *Senior Member, IEEE,*

* Intelligent Computing and Communication (IC^2) lab
Key laboratory of Universal Wireless Communication, Ministry of Education
Beijing University of Posts & Telecommunications
Beijing, China, 100088
♯ Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
‡Department of Informatics
Alexander Technological Educational Institute of Thessaloniki (ATEITHE)
P. O. Box. 141, 57400 Sindos
Thessaloniki, Greece
† College of Science
Technology & Engineering Division of Tropical Environments and Societies
James Cook University Cairns
QLD 4870 Australia
E-mail: kzheng@ieee.org

Abstract

Big Data offers a plethora of opportunities to mobile networks operators for improving the quality of service. This paper explores various means of integrating Big Data analytics with network optimization towards the objective of improving the user quality of experience. We first propose a framework of Big Data Driven (BDD) mobile network optimization. We then present the characteristics of Big Data that are collected not only from user equipments but also from mobile networks. Moreover, several techniques in data collection and analytics are discussed from the viewpoint of network optimization. Certain user cases on the application of the proposed framework for improving the network performance are also given in order to demonstrate the feasibility of the framework. With the integration of the Fifth Generation (5G) emerging mobile networks with Big Data analytics, the quality of our daily mobile life is expected to be tremendously enhanced.

Index Terms–Big Data, Network optimization, Heterogeneous networks.

I. Introduction

A massive number of mobile phones are in wide use and produce massive amounts of data every day. This brings about a profound impact on society and social interaction as well as creating tremendous challenges for Mobile Network Operators (MNOs). The volume, velocity and variety of the data from both mobile users and communication networks have been exploding exponentially [1]. Therefore, Big Data are already in our mobile life and will be further entrenched by the upcoming Fifth Generation (5G) cellular communications in the near future [2].

Big Data in mobile networks need to be extensively analyzed in order to retrieve interesting and informative information. It provides unprecedented opportunities for MNOs to understand the behavior and requirements of mobile users, which in turn allows for intelligence real-time decision making in a wide range of applications. By analyzing this data, mobile networks can actually provide and support different smart services. However, the nature of Big Data presents vast challenges in relation to data mining, mobile sensing and knowledge discovery [3]. New technologies are required to handle Big Data in a highly scalable, cost-effective, and fault-tolerant fashion [4] [5]. Researchers are currently investigating new Big Data analytical techniques in order to discover previously unknown patterns and knowledge from the collected data.

Apart from mobile users, MNOs can also benefit from Big Data. They can easily obtain massive volumes of data, which are generated by mobile devices owned by their customers as well as by various network elements in their networks. All the data can be jointly explored in order to improve the network operation efficiency. The challenges lie not only on the huge volumes of data but also on the non-homogenous structure that is often related with incomplete and ambiguous information. Therefore, it is imperative to have a good knowledge of the unique characteristics of Big Data in mobile networks, which is crucial for the optimization of 5G mobile networks.

With recent advances in data analytics, Big Data based mobile network optimization has attracted intensive efforts from researchers worldwide [6] [7]. Big Data analytical techniques can provide MNOs with deep insights into the networks before making informed decisions. For example, these analytical techniques can help MNOs to monitor and analyze various types of data as well as event messages in the networks. Intelligence and important insights can be extracted from both instantaneous and historic data. Useful information, such as the correlation between user behaviors and network traffic, can help MNOs to not only make decisions based on long-term strategies, but also to optimize resource allocation so as to minimize deployment and operational costs. Furthermore, MNOs are expected to play a key role in the standardization of 5G networks. However, a critical challenge is to understand the requirements of utilizing Big Data analytics to provide user services with personalized Quality of Experience (QoE), and to enable highly efficient resource utilization in 5G networks.

In this article, we propose a generic framework in an effort to support a variety of Big Data Driven (BDD) mobile network optimization methods. The proposed framework enables practitioner engineers to utilize the data from both the network and users when optimizing networks, in lieu of only user data. The BDD schemes can enhance QoE performance and leverage the investment of the entire network. We also discuss the features of the data collected from both users and operators, followed by a study of certain data analytical schemes. We finally present a case study of BDD mobile network optimization to further validate the proposed framework.

II. Framework of Big Data driven mobile network optimization

The Heterogeneous Network (HetNet) is well suited for MNOs to make more efficient use of spectra in densely populated areas as they move towards 5G deployment, which is the main focus of this paper. HetNets usually consist of two layers, i.e., the macro-cell and small-cell layers, where the former provides mobility while the latter boosts coverage and capacity [8]. Such a layered network architecture can enable both large coverage and high capacity, and thus provides users with enhanced QoE. However, although the use of small cells can improve the capacity of the entire network, it is not able to support the adaption of various network resources according to their time-changing traffic characteristics. In order to enhance operational efficiency in network infrastructure under varying environments, MNOs are encouraged to adjust network traffic requirements and improve resource allocation efficiency through the use of intelligence and analytics based on Big Data.

As illustrated in Fig. 1, the proposed BDD network optimization framework includes: 1) Big Data collection; 2) storage management; 3) data analytics; and 4) network optimization.

The collection of Big Data can be achieved from User Equipments (UEs), the Radio Access Network (RAN), the Core Network (CN), and the Internet Service Providers (ISPs). The events that occur at UEs are collected either through user applications or via control signaling. At the RAN evolved NodeB (eNB), the cell-level data (including the exchanged signaling over the air) and instantaneous measurement reports are collected. Meanwhile, MNOs possess huge amounts of data relating to user bearers/services in the CN. When the cell size becomes smaller in HetNets, the number of eNBs increases. As this trend continues, network data may explode and impose a great burden on data collection. Furthermore, Big Data storage infrastructure needs to have scalable capacity as well as scalable performance. Thus, storage management needs to be simple and efficient so that storing and sorting of Big Data can be easily achieved.

After data are collected and stored, another big challenge for MNOs is how to process such huge volumes of data. The collected data are multi-source, heterogeneous, real-time and voluminous. For this reason, data analytics and knowledge extraction techniques are required to process the data and convert it into actionable knowledge. Consequently, this knowledge can be used to design adaptive schemes for network optimization.

Data analytics enables MNOs to manage networks and provide services to customers in a systematic manner. Not only the network measurements but also the application/service status for each region can be monitored and analyzed over time. The BDD network optimization functions are capable of analyzing Big Data to identify problems, and to decide what/how to optimize the appropriate level, e.g., the user, cell or service. The improvement measures based on the optimization results are then implemented by the control functions in the RAN. Moreover, user-level optimization can be performed. In particular, for users closely located in the same cell, optimization can be customized for each user depending on its service class. Furthermore, the BDD network optimization functions are able to predict traffic variations either in a local area or over the network coverage and eventually help to improve the network and user performance.

III. Features of Big Data and Data Analytics

A. Big Data in mobile networks

Data can be roughly divided into two types, i.e., users' and network operators' data. A comprehensive analysis of both types of data is able to provide valuable insights, which can be used by MNOs for network optimization. MNOs can analyze data in order to perform network planning, spectrum allocation, resource management and so on.

1) User data: The data collected from UEs are highly related to the user's profile and behavior, which offers a great deal of insights for users, such as their location, mobility and personal communication behavior/pattern. With the rapid expansion of mobile networks and the enormous increase of the deployed smart mobile devices, excessive amounts of data are generated from the applications installed in the users' mobile devices. Application level data has become one of the primary sources of mobile Big Data.

2) Operator data: Data collected by operators is mainly sourced from their CN and RAN. The CN has abundant bearer/service data regarding, e.g., network performance information, successful calls and usage index per application. On the other hand, there is a large amount of data in RANs including cell information (e.g., eNB configuration information, resource status information, interference information, handover reports, mobility information, fault status, link utilization, call drop ratio), signaling messages exchanged between the eNB and UE (e.g., RRC messages for connection establishment and handover), and radio signal measurements (e.g., reference signal received power, reference signal received quality, and so on). The key features of these two kinds of data are summarized in Table I.

Once a MNO has data collected from different sources, the next challenge is how to efficiently utilize them. The data generated from all the sources need to be processed and converted into actionable knowledge, which can be used to design adaptive algorithms and optimization strategies for improving network performance. Advanced techniques in data collection and analytics are essential for network optimization.

B. Analytical schemes for Big Data in mobile networks

In mobile networks, collecting raw data is the first step of a Big Data analysis. For example, MNOs can collect data from the mobile users who share/download information associated with their mobility. However, the user location information may not be obtainable if the mobile users disable localization on their mobile devices. Alternatively, the location information obtained directly from the eNBs may not be precise due to the inaccuracy of localization techniques. Localization errors and environmental interferences are often big barriers to the usability of Big Data. Moreover, the battery of the user device may be depleted, and thus the desirable information cannot be collected in certain times. Towards this end, data mining, filtering and extraction techniques are developed with the objective of removing interference or useless data, which belong to the so-called error-prone classification schemes. However, it is still a major challenge to obtain useful information from incomplete, redundant and uncertain Big Data. One promising solution for data mining over such data is multi-source dynamic data mining, since data are usually collected from various sources [9].

With an unprecedented increase on collected data, both users and network operators require effective data analysis and prediction tools to enable fast response and real-time classification. Currently, various Big Data applications offer both predictive and prescriptive analytics with powerful machine learning techniques, such as the Support Vector Machine (SVM) and deep learning. Machine learning reflects the emerging advance of multivariate statistics, pattern recognition, data mining and some other advanced data analytics or prediction. It plays the utmost role if deep and predictive insights are required to uncover hidden knowledge from data sets that are large, diverse and fast changing. In general, accuracy, scale and speed are the main metrics in evaluating machine learning methods.

Another important machine learning technique is sequence classification, which is applicable to traffic analysis and user behavior classification. Although there are various feature selection techniques, it is still challenging to effectively classify the feature sequences in a Big Data set due to the data volume and dimensionality of potential features within the sequences. The SVM technique is proven to be effective in classifying featured sequences since it attempts to assign a given sequence into one category or other over a feature space and identify the maximum-margin hyper-plane to differentiate two classes. Given two sequences a and b, a similarity function $s(a, b)$ can be used as the kernel function for classification.

However, it is important to determine the feature space and the kernel function, such as the k-spectrum kernel, string kernel, polynomial-like kernel, or kernels derived from probabilistic models.

Unlike traditional learning methods, under the consideration of shallow-structured learning architectures, deep learning emerges by using supervised and/or unsupervised methods to automatically learn hierarchical (or multiple levels of) representations in deep architectures for classification. Due to the recent unprecedented growth of data in mobile networks, tremendous efforts have been focused on effective and scalable parallel algorithms for training deep models. Although we often tend to collect abundant unlabeled data, it is not a straight-forward process if traditional models are utilized. It is necessary to utilize a deep belief network with a deep architecture to capture the feature representations from not only the labeled but also unlabeled data. The deep belief network exploits pre-training for unsupervised learning, and adjusts strategies for supervised learning, which will eventually lead to the establishment of a learning model. In particular, this incorporates unsupervised learning to obtain the data distribution without the aid of labeled data, and performs supervised fine tuning to improve both the newly-added classification layers and the pre-trained layers. A typical architecture of the deep belief network consists of a stack of Restricted Boltzmann Machines (RBMs), which is a probabilistic generative model to learn a joint probability distribution of training data, associated with several additional layers for discrimination tasks. Within a RBM, there are usually two layers, where any node is fully connected to all nodes in the other layer but does not connect to any node in the same layer. As a result, every node is independent of any other nodes in the same layer, leading to a possibility of training the generative weights of each RBM with Gibbs sampling [10].

IV. CASE STUDIES FOR BDD MOBILE NETWORK OPTIMIZATION

In Fig.2, we firstly present the traffic loads of several typical applications, i.e., WeChat, news and on-line music services, under three typical urban scenarios, i.e., business, restaurant and residence zones during various hours of every day for one week. All the data are collected in a big city in Northeastern China. It is clear that the traffic varies either within one day or between different days, i.e., the time-varying characteristics. Also, the traffic loads in different zones are not the same, i.e., geographical variations. Moreover, the traffic of each application has its own special features, i.e., service-related characteristics.

Therefore, it is better to take the traffic characteristics into account when deploying and developing a mobile network. In order to provide a deep and detailed investigation of BDD optimization, several case studies are discussed in detail in this section in order to answer why and how to apply Big Data to network optimization. These only give some preliminary hints for MNOs to improve network performance, which still needs a significant amount of solid work before practical usage.

A. Resource management in HetNet

MNOs should be aware of their long-term deployment objectives in terms of network capacity, coverage, the number and locations of the base stations, etc. They also need new resource allocation strategies in order to fulfill different traffic demands/requirements across the entire coverage area. To achieve these goals, MNOs have been monitoring the network Quality of Service (QoS) through driving tests with smart phones. Measurement results are gathered from selected smart phones or specific driving testing phones in their networks, which are analyzed by specialized software. However, this is not cost effective attributed to excessive time and human resources, and is also inaccurate due to the limited test samples.

Thus, the use of Big Data analytics can provide a new way to tackle these problems. The network analytics involves monitoring, analyzing real-time and history data across users, mobile networks and service providers. There are several stages that BDD approaches can help MNOs to deploy and operate their networks more efficiently, which are detailed in the following.

1) Network planning: In most traditional deployment cases, the sites of the eNBs are not optimized due to a lack of sufficient statistical data. By tracking mobile devices, their detailed activities can be recorded to provide real-time information about where, when, and what information of the mobile users in the network. A feasible solution is to make use of both the network and the anonymous user data including dynamic position information and other various service features. Consequently, massive volume, velocity, and variety of data need to be processed by advanced analytics techniques, which can transform the data into actionable knowledge. In order to well understand the traffic trends, it is imperative to analyze the data in relation to the corresponding content and events.

Given the actionable knowledge inferred from the big data sets, the MNOs are able to make wise decisions on where and how to deploy the eNBs in the networks. This also allows them to predict the traffic trends and to prepare plans for future investment.

2) Resource allocation: By utilizing data analytics, the resource requirements changing from one location to another in a specific period becomes predictable. In addition to the network data, the behavioral and sentiment analyses from social networks and other sources are to be taken into account in an effort to predict where and how users may use the mobile network, For example, when a social event such as a a marathon takes place in a city, some places like the streets in the race route may attract large crowds of people, resulting in potential congested traffic in these locations during the event. Hence, with this predicted information from data analytics, the operators can allocate more radio resources to the "hotshot" in such a way that the peak traffic can be absorbed smoothly without sacrificing the user QoE.

Mobile users often travel from one place to another around the city, e.g., work in the central business district during the day time and probably live in an outskirt suburb at night. This causes the traffic of each cell to fluctuate significantly during the different times of the day, which is dubbed the "tide effect". If the resources are allocated to each cell with a fixed configuration, resource utilization must be underestimated, and the users in the hotspot are difficult to obtain a good QoE during peak hours. On the other hand, a great deal of resources may be wasted in idle times in low traffic locations. The current and history data can be utilized by data analytics to predict traffic for high-density areas in the networks. Then, with the Cloud RAN architecture [11], predictive resource allocation in centralized baseband units may help to accurately serve the right place at the right time, i.e., knowing when and where peak traffic arises, causing minimum disruptions to services.

3) Interference Coordination: Within a HetNet that has small cells, interference coordination among macro and small cells has to be carried out in the time domain in lieu of the frequency domain, e.g., the enhanced Inter-Cell Interference Coordination (eICIC) scheme in LTE-Advanced [12]. Such a scheme enables efficient resource allocation among interfering cells, and improves inter-cell load balancing in the HetNet. The essential principle behind eICIC is that it allows a macro cell eNB (MeNB) and its neighboring small cell eNBs (SeNBs) to transmit data in separated subframes, i.e., be kept orthogonal in the time domain, especially avoiding the interference from MeNB to SeNBs. Thus, when communicating with cell-edge UEs, the SeNBs use the subframes that are orthogonal with their neighboring macro-cell, thus avoiding potential interference from the MeNB. Meanwhile, the SeNBs can transmit to the UEs in their cell center in any subframe regardless whether the MeNB is transporting data at the time.

For the sake of eICIC implementation, a special type of subframes is defined, namely the Almost Blank Subframe (ABS), which carries no data but only minimum control information, e.g., the reference signal, the mandatory system information and so on. Thus, no interference to the data signals occurs, while the interference caused by the control signals can also be mitigated. In an LTE system, one radio frame consists of ten subframes, each of which can be used as either a normal subframe or an ABS by the eNB except subframes 0 and 5. The decision on how to configure ABS subframes is made by the network operator.

However, the determination of an appropriate ABS ratio of the macro cell to the small cells depends on many factors, e.g., the service types, the traffic load in the given area, etc. As it is well known, the service behaviors in small cells vary with time. Moreover, the traffic patterns of individual services also change. Thus, the inter-cell interference does not remain constant. Therefore, the optimal ABS ratio essentially changes dynamically with time.

In a BDD system, network analytics can be used to optimize the allocation of radio resources. Resource allocation can be made to adapt to both environmental and traffic changes based on information gained from data analytics. In order to enable a quick response, some BDD optimization functions can be deployed at the MeNB so that they can collect and analyze eNB-originated raw big data in time, e.g., the characteristics of service and traffic features. Consequently, the performance of each cell and the users can be optimized. This can be done by periodically processing raw data to obtain statistics and automatically detect traffic variances, targeting to predict ICIC optimized parameters such as the ABS ratio.

Moreover, a global optimization process can jointly optimize the location and the traffic demands of the users of multiple eNBs. For instance, a certain SeNB can be deactivated in order to avoid the interference to its nearby SeNB, which might have the larger throughput due to the higher Signal-to-Interference-Noise Ratio (SINR). Additionally, a reduction in energy consumption may be another optimization objective to be taken into account.

B. Cache server deployment in the mobile CDN

Cellular networks are currently experiencing an explosive growth of data traffic. The Content Delivery Network (CDN) has been considered by MNOs as an efficient delivery method for popular contents such as blockbuster movies. The main purpose of having their own CDNs is to reduce operational costs while providing good support to their core businesses. It is important to locate distributed cache servers in the CDN as close to end user as possible in order to shorten response time and also reduce delivery costs, e.g., a distributed cache server working together with a central cache server in a hierarchical CDN [13]. However, the cache access rate on the distributed cache server might be lower than the one of the central cache server. Sometimes the distributed cache server even needs mobile users data traffic to traverse the associated central cache server through the backhaul link in the event of improper placement. Therefore, it is vital to choose the optimum location for the cache servers in the hierarchical CDN. In this section , only the RAN is in our primary interest since it has unique features compared with fixed CDNs.

As shown in Fig. 3, it may be beneficial to co-locate the distributed cache server with the MNO's radio access network, which enables content distribution more efficiently on the network edge. Thanks to the hierarchical structure of the heterogeneous network having small cells, MeNB cell site is another good location for the distributed cache server, since it usually locates in the center of the local network. Due to the backhaul capability among the eNBs or from the eNBs to the CN that is expected to be significantly enhanced in 5G networks, there is little concern with the traffic load and latency of backhaul transmission. Thus, not all the MeNBs need to be deployed with an individual distributed cache server. Moreover, a distributed cache server can be deployed alongside a SeNB if needed. Besides the cost of storage and streaming equipment, the features and load of traffic in a given area are among the important factors that determine the optimal placement of a cache server. After collecting the data relating to all the relevant factors in the coverage area over a long period of time, cluster analysis can be used as a feasible method in data analytics to help the MNOs deploy cache servers in the RAN, i.e.,

- *Pre-processing:* Not all raw data are suited for analysis so that some data have to be eliminated. For example, incomplete and redundant data should be filtered out. Consequently, the main attributes can be selected from the remaining data, e.g., the traffic load, the service type, backhaul usage, and the latency of retrieval.

- *Clustering:* Each eNB location is regarded as a point in the groups. Each attribute determines the value of one dimension of the point. Thus, there are lots of points, i.e., eNB locations, which are represented by a multi-dimensional vector. According to the specific clustering principle, all the points can be partitioned into two non-overlapping groups, i.e., one is for cache server deployment while the other is not.

The analytics capabilities are built into the hierarchical CDN by utilizing a collective intelligence data architecture. Each cache server has a monitor agent to collect log information. This monitor agent sends log status information to the function block of data analytics, which then determines when/what content to outsource and where to place the replicas.

Content of high popularity is more likely to be placed on the cache servers in order to improve the cache access ratio. As shown in Fig. 4, popularity usually depends on not only on the content itself but also on the users. Moreover, user mobility may cause the content in the cache to change frequently, resulting in inefficiency in content caching. Therefore, the data analytics function needs to analyze the data related to both content and users in order to accurately determine or predict content popularity.

C. QoE modeling for network optimization

Typically, various services and applications are managed by using a set of QoS parameters (e.g., packet loss, delay, and jitter). However, management can be more efficient when the quality as perceived by end users, i.e., QoE is taken as the optimization objective instead of QoS. Towards this end, automatic and accurate estimation of the QoE in real-time is the first step. Data analytics can help with QoE modelling and monitoring in a diverse heterogeneous environment, which is essential for global network optimization.

As shown in Fig. 5, the data needed for estimating QoE comes from both the network and users. Besides the technical factors, various non-technical factors exist that may influence QoE results, e.g., device type, user emotion, habit, expectation, and so on. Thus, in QoE evaluation, it is useful to create an individual profile for each user, which is a user model representing users preferences, habits, and interests. A user does not usually like to spend much time to answer questions for creating a profile model. As an alternative, a user profile can be built and monitored using data analytics with implicit information gathered by a profile collection engine, which is installed at mobile devices. The activities of users are tracked and compared to identify similarities and differences. For example, the output of the motion detector in the profile collection engine may include (but not limited) the number of clicks and the scrolling on the screen. In the emotion detector, a user emotion may be extracted from a detected user behavior with affective computing techniques [14]. Meanwhile, network data including QoS parameters are collected through the measurement and signaling in the network. All the data are stored in a database for further processing.

A machine learning engine is then used to establish the relationship between the influencing factors and the QoE through artificial intelligence. Machine learning techniques enable ever more accurate decision making over time, even when the data sets are incomplete or new situations arise. For some typical approaches such as the neutral network model used in the Pseudo-Subjective Quality Assessment (PSQA) assessment method [15], the QoE model has to be trained before being used for performance evaluation. The analysis of large data sets leads to insights into the users' real experience, which may need to incorporate social data.

Data analytics is able to discover what MNOs need to know, which impacts QoE across devices, services and network resources. Then, network optimization functions can promptly find the cause of problems and choose the best action accordingly. In general, the network optimization objective is to maximize QoE for users with the proper resource allocation, while minimizing the costs of infrastructure through data analytics.

V. Conclusion

This article elaborated on potential benefits of exploiting Big Data in mobile network optimization. A Big Data driven framework for mobile network optimization was proposed, in which efficient data analytics is deemed as the key enabling technique for reducing deployment costs and increasing network efficiency. We also studied the features of Big Data from both the users' and operators' perspectives, which can be used together for mobile network optimization. Moreover, three case studies of the BDD schemes were presented to demonstrate possible new solutions to improving the performances of mobile networks towards 5G.

However, Big Data driven schemes also pose significant challenges. First of all, how to collect the completed Big Data not only from users and MNOs is limited by the techniques as well as the market and policy. Moreover, the communication overhead and latency caused by using the data analytics have to be investigated in order to find the feasible solutions for applying BDD optimization. It is very important to find ways to evaluate the costs of introducing the data analytics in emerging 5G networks while comparing with the expected performance enhancement, thus to achieve the good tradeoff. Therefore, many challenges need to be properly addressed in order to maximize the entire network performance and thereby to ensure high returns on investment by the MNOs.

ACKNOWLEDGMENT

This work is funded in part by the National High-Tech R&D Program (863 Program 2015AA01A705), the China Natural Science Funding under grant 61271183, National Key Technology R&D Program of China (No. 2015ZX03002009-004) and Fundamental Research Funds for the Central Universities (No. 2014ZD03-02).

REFERENCES

[1] M. Musolesi, "Big mobile data mining: good or evil?," *IEEE Internet Computing*, pp. 78 -81, vol. 18, no.1, 2014.

[2] F. Boccardi, R. W. Heath, A. Lozano, T.L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74 -80, Feb., 2014.

[3] R.K. Lomotey, and R. Deters, "Towards knowledge discovery in Big Data," *IEEE International Symposium on Service Oriented System Engineering (SOSE), 2014*, pp. 181 -191, 2014.

[4] L. Gu, D. Zeng, P. Li, and S. Guo, "Cost minimization for Big Data processing in Geo-distributed data venters," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 3, pp. 314-323, Sept.,2014.

[5] L. Lei, Z. Zhong, K. Zheng, J. Chen, and H. Meng, "Challenges on wireless heterogeneous networks for mobile cloud computing," *IEEE Wireless Communications*, vol. 20, no. 3, pp. 34 -44, June, 2013.

[6] S. Samulevicius, T.B. Pedersen, and T.B. Sorensen, "MOST: mobile broadband network optimization using planned spatio-temporal events," *in proc. IEEE VTC-Spring*, pp. 1- 5, May, 2015.

[7] A. Ramaprasath, A. Srinivasan, and C. Lung, "Performance optimization of big data in mobile networks," *in proc. IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp.1364 - 1368, May, 2015.

[8] K. Zheng, Q. Zheng, P. Charzimisios, W. Xiang, and Y. Zhou, "Heterogeneous vehicular ntworking: a survey on architecture, challenges and solutions," *IEEE Communication Surveys & Tutorials*, pre-printed, June, 2015, DOI.10.1109/COMST.2015.2440103.

[9] X. Wu, X. Zhu, G. Wu, and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp.97- 107, 2014.

[10] X. Chen, and X. Lin, "Big Data deep learning: challenges and perspectives," *IEEE Access*, vol.2, pp.514- 525, 2014.

[11] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M.S.Berger, and L. Dittmann, "Cloud RAN for mobile networksłA technology overview," *IEEE Communications Surveys & Tutorials*, vol. 17, no.1, pp. 405 -426, 2015.

[12] K. Pedersen, Y. Wang, S. Strzyz, and F. Frederiksen, "Enhanced inter-cell interference coordination in co-channel multi-layer LTE-Advanced networks," *IEEE Wireless Communications*, vol. 20, no. 3, pp. 120- 127, 2013.

[13] Z. Su, Q. Xu, H. Zhu, and Y. Wang, "A novel design for content delivery over software defined mobile social networks," *IEEE Network*, vol. 29, no. 4, pp. 62 -67, 2015.

[14] B. Zhu, and H. Li, "Designing finger movement on mobile phone touch screen for rich emotional expression," *in proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)'2014*, pp. 1 -6, 2014.

[15] K. Zheng, X. Zhang, Q. Zheng, W. Xiang, and L. Hanzo, "Quality-of-experience assessment and its application to video services in LTE networks," *IEEE Wireless Communications*, vol. 22, no. 1, pp. 70 -78, 2015.

Biography

Kan Zheng (SM'09) is currently a full professor in Beijing University of Posts & Telecommunications (BUPT), China. He received the B.S., M.S. and Ph.D degree from BUPT, China, in 1996, 2000 and 2005, respectively. He has rich experiences on the research and standardization of the new emerging technologies. He is the author of more than 200 journal articles and conference papers in the field of wireless networks, Internet-of-Things (IoT) and so on. He holds editorial board positions for several journals. He has also served in the Organizing/TPC Committees for more than ten conferences such as IEEE PIMRC, IEEE SmartGrid and so on. Now He is the chair of IEEE Computer Society STC Internet-of-Everything (IoE).

Zhe Yang received his B.S. degree from Shandong University, China, in 2014. He is currently a Ph.D. candidate in the Intelligent Computing and Communication (IC^2) lab, Key Lab of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications (BUPT). His research interests include performance analysis and data mining in heterogeneous wireless networks.

Kuan Zhang received the B.S. degree on Communication Engineering and the M.S. degree on Computer Science from the Department of Information Science Engineering, Northeastern University, Shenyang, China, in 2009 and 2011, respectively. He is currently working toward the Ph.D. degree in Broadband Communications Research (BBCR) group, Department of Electrical and Computer Engineering, University of Waterloo, Canada.

Periklis Chatzimisios (SM'13) serves as an Associate Professor at the Computing Systems, Security and Networks (CSSN) Research Lab of the Department of Informatics at the Alexander TEI of Thessaloniki (ATEITHE), Greece. Recently he has been a Visiting Academic/Researcher in University of Toronto (Canada) and Massachusetts Institute of Technology (USA). Dr. Chatzimisios is involved in several standardization activities serving as a Member of the Standards Development Board for the IEEE Communication Society (ComSoc) (2010-today) and lately as an active member of the IEEE Research Groups on IoT Communications & Networking Infrastructure and on Software Defined & Virtualized Wireless Access. He is the author/editor of 8 books and more than 100 peer-reviewed papers and book chapters on the topics of performance evaluation and standardization activities of mobile/wireless communications, Quality of Service/Quality of Experience and vehicular networking. His published research work has received more than 1500 citations by other researchers. Dr. Chatzimisios received his Ph.D. from Bournemouth University (UK) (2005) and his B.Sc. from Alexander TEI of Thessaloniki (Greece) (2000).

Kan Yang received his B. Eng. degree from University of Science and Technology of China in 2008 and his PhD degree from City University of Hong Kong in August 2013. He is currently a postdoctoral fellow of Broadband Communications Research (BBCR) group in Department of Electrical and Computer Engineering at University of Waterloo, Canada. He was a visiting scholar in State University of New York at Buffalo in 2012. His research interests include Cloud Security and Privacy, Big Data Security, Cloud Data Mining, Cryptography, Social Networks, VANET, Smart Grid, Wireless Communication and Networks, Distributed Systems, etc.

Wei Xiang (SM'10) received the B.Eng. and M.Eng. degrees, both in electronic engineering, from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunications engineering from the University of South Australia, Adelaide, Australia, in 2004. He is currently a Full Professor in the College of Science, Technology and Engineering at James Cook University, Cairns, Australia. During 2004 and 2015, he was an Associate Professor with the School of Mechanical and Electrical Engineering, University of Southern Queensland, Toowoomba, Australia. He was a co-recipient of the Best Paper Awards at 2015 WCSP and 2011 IEEE WCNC. He is an IET Fellow. He has been awarded several prestigious fellowship titles. He was named a Queensland International Fellow (2010-2011) by the Queensland Government of Australia, an Endeavour Research Fellow (2012-2013) by the Commonwealth Government of Australia, a Smart Futures Fellow (2012-2015) by the Queensland Government of Australia, and a JSPS Invitational Fellow jointly by the Australian Academy of Science and Japanese Society for Promotion of Science (2014-2015). In 2008, he was a visiting scholar at Nanyang Technological University, Singapore. During Oct. 2010 and Mar. 2011, he was a visiting scholar at the University of Mississippi, Oxford, MS, USA. During Aug. 2012 and Mar. 2013, He was an Endeavour visiting associate professor at the University of Hong Kong. His research interests are in the broad area of communications and information theory, particularly coding and signal processing for multimedia communications systems.

TABLE I
KEY FEATURES OF THE USER AND OPERATOR DATA.

Feature	User data	Operator data
Objective/Subjective	• Highly influenced by the subjective feelings or personal preferences	• Measured by the network objectively without involving human factors
(Non)-structured	• Various data formats including the semi-structured and non-structured data (e.g., locations, logs, and sensor data)	• Mainly structured data generated according to specific given protocols
Privacy	• High privacy is required since users are not willing to disclose their personal information	• Usually internal use for network operators without sharing with others
Energy limitation	• Data accuracy constrained by device energy consumption • Accuracy adaptively controlled to save energy	• No energy limitation for main-powered network devices
Redundancy	• High correlation and redundancy in the event of a large number of users located in popular locations during a specific period of time	• Usually high correlation and redundancy because data are coherently processed across the different layers of the network
Distribution	• Usually fragmentary and discontinued in time and space	• Usually periodical and uniform distribution in time
Reliability	• Low reliability due to changing user numbers and locations. •Pre-processing is needed to filter noise and maintain data integrity	• Usually high reliability because data are mostly from signaling and control information in networks • Instable due to varying dynamics, heterogeneity and the large scale of the networks
Controllability	• Difficult to control in terms of data rates, sizes, collecting moments and so on	• Easily collectable by the operators through specific network interfaces and measurement devices

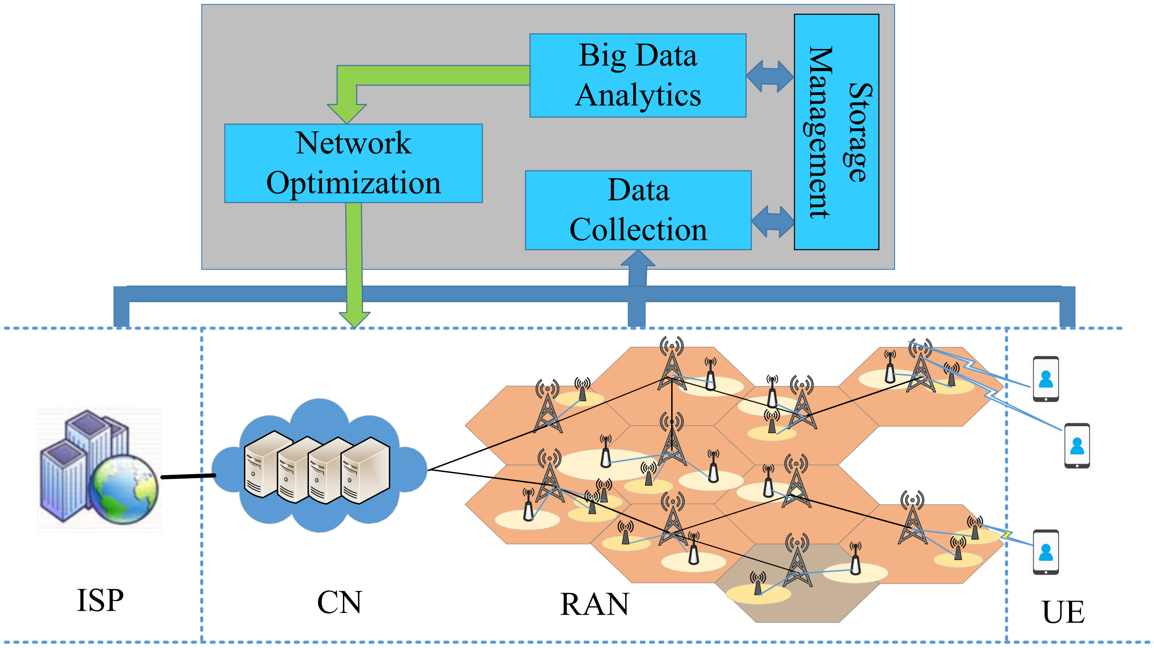
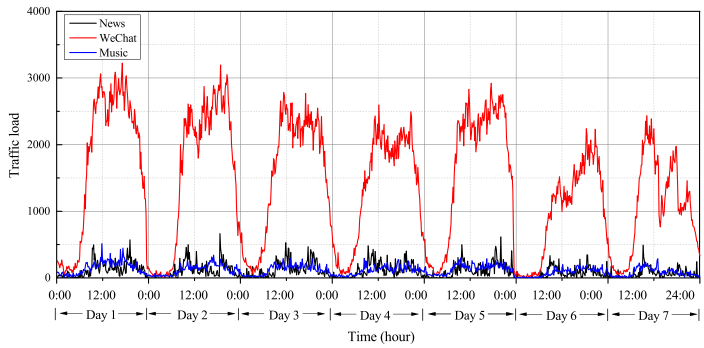
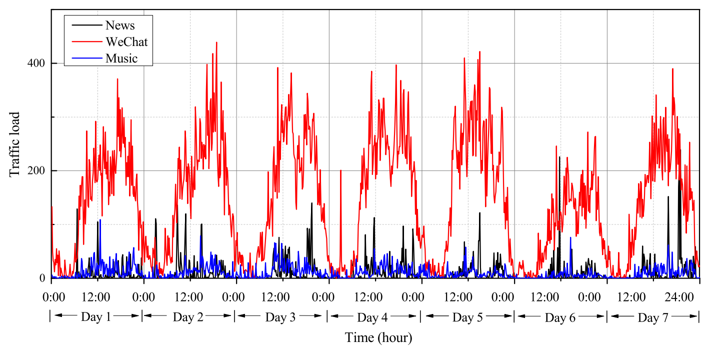


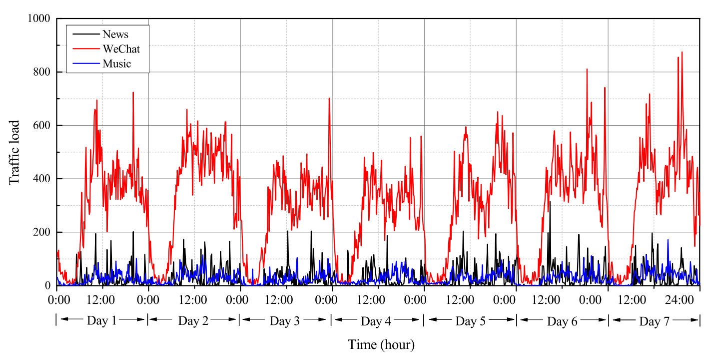
Fig. 1. Illustration of the proposed BDD network optimization framework.



(a) Business zone.



(b) Restaurant zone.



(c) Residence zone.

Fig. 2. Measured traffic loads of several applications in typical urban scenarios.

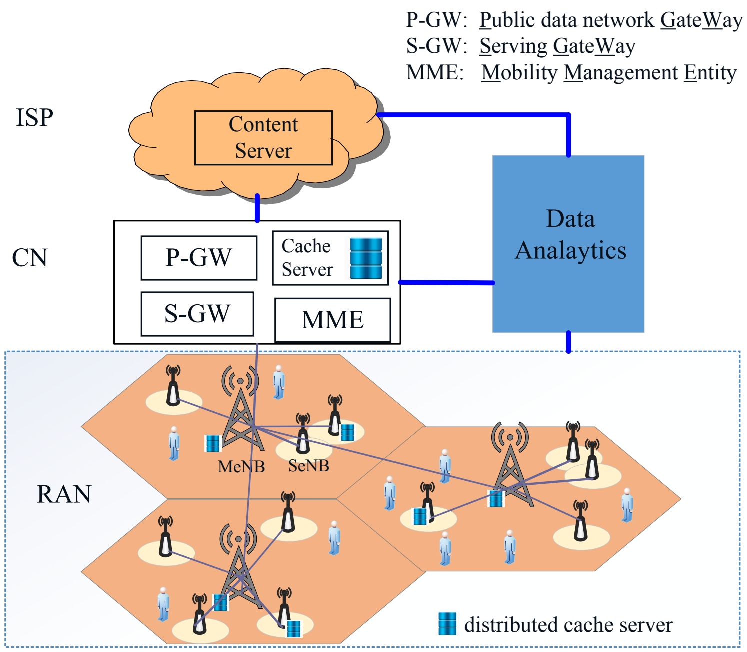


Fig. 3. Illustration of the BDD cache server deployment.

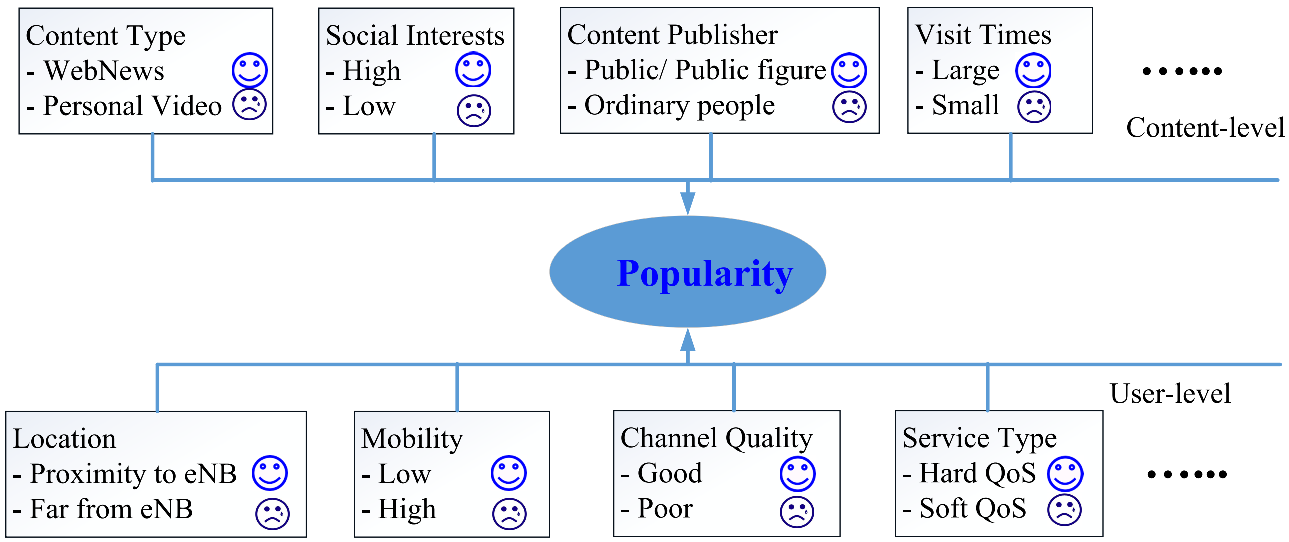


Fig. 4. Different factors influencing content popularity in the mobile CDN.

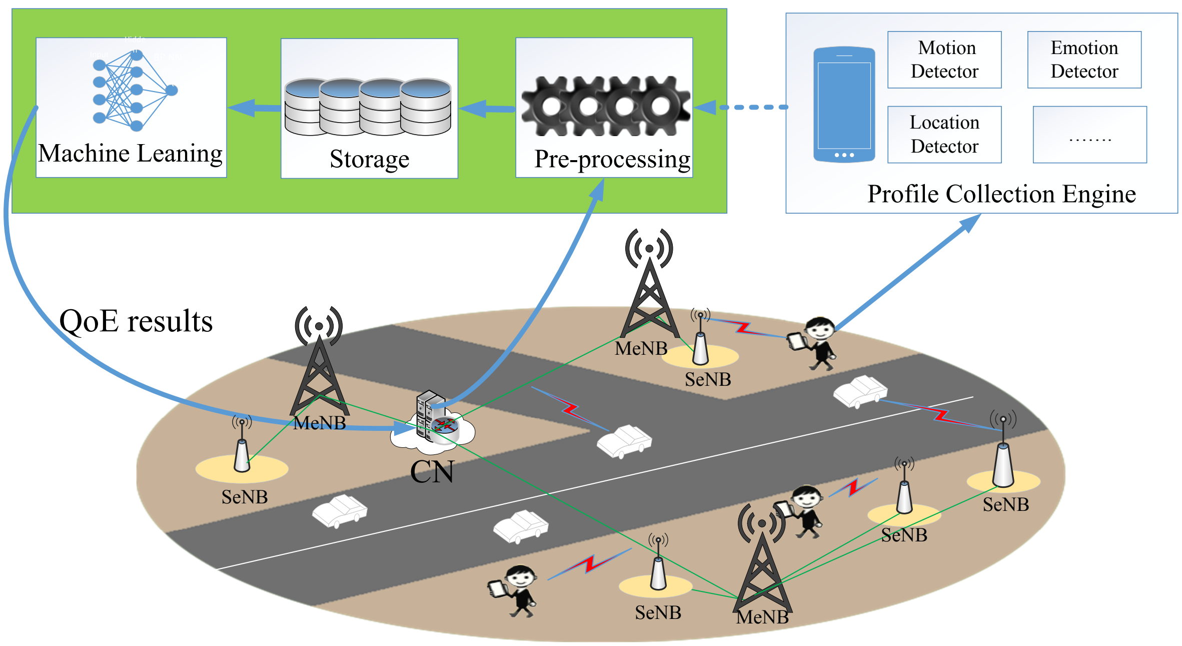


Fig. 5. Illustration of QoE modeling driven by Big Data.