BIG DATA SYNERGY IN 5G WIRELESS NETWORK

[1]Hassan Adam, [2]Mr. Sunil Ghildiyal

[1]Student, [2]Assistant Professor (Ph.D.)
[1]Department of Computer Science and Engineering,
[1]Uttaranchal University, Dehradun, India

Abstract: This article explores the complementary and synergistic features of 5G wireless networks and big data. An outlook of its interaction is provided at first, which includes both big data driven networking and also big data assisted networking. Big data driven networking makes use of varied resources such as caching, computing and communication in 5th generation wireless networks to aid big data provided services and applications by catering for big data's attributes such as velocity, volume, variety and value. While big data assisted networking leverages big data methodologies to collect wireless big data and extract in-depth information regarding the users and networks to enhance operation and network planning. To further demonstrate the mutual benefits of both, two case studies on big data assisted edge content caching and network aided data acquisition are provided. At last, some noteworthy open research issues are also examined.

I. INTRODUCTION

With the spread of mobile devices and evolving IoTs, the connected devices grow in numbers and will easily reach around 50 billion in the near future by the year 2020. At the same time, a large number of multimedia services are also blooming rapidly. As a result of these factors, data generation is also growing at an a very fast pace. Every 60 seconds, YouTube users post four hundred hours of new video, while Instagram users create two and a half million posts. According to a recent report by IBM,

a. In each day, two and a half quintillion bytes of data is generated. It is predicted that the data amount generated in the year 2020 will be forty trillion gigabytes, which is forty-four times greater than that in the year 2009.

b. With such a data generation, we are entering the time of big data which includes the following big "V" attributes: velocity, value, variety, and volume.

Volume depicts the amount of data produced which are enormous, while variety suggests great differentiation in data structure and the type of data, eg., audio, video, geospatial data, smart meter reading, and logging. Velocity indicates the collection of data and its processing that should be carried out in a timely fashion to explore its potential worth and value. These data are usually associated with a large array of applications, such as advertisement, healthcare, trading of stocks, and smart-grids. Moreover, such value of data also opens up new chances to explore its great potential worth. To manage the complexity which is high and massive scale, it is kept at cloud computing data centers due to its high computing power and storage facilities. Many a times, the cloud or the data center usually happen to be in remote locations, which can be distant from the original data sources. Hence, there is a need to perform the following procedures:

i) Data-acquisition: gathering of various data;

ii) Data pre-processing: perform various operations on the unprocessed data, like data-aggregation, reduction in size and ciphering,

iii) Data transportation: move data towards data centers,

iv) Data analyzing: Mining of data or applying analytics to get the hidden value of data which can assist decision making process.

1.1 Expansion of the Next Generation Wireless Network

Increased by rising diverse applications and mobile data, the system of wireless networks is developing towards the next generation of wireless networks (5G), it is achieved by incorporating colossal computing resources, caching, and communication. To house large amount of data traffic, infrastructures of communication like small cell base stations are expected to be deployed more densely, for coverage improvements and increasing the capacity of the network. To support low-latency services and mitigate backhaul congestions, various techniques of caching are employed to accommodate the contents which are prevalent in either core network or network edge. By establishing contents at the proximity of users, latency can be reduced and also redundant-transmissions can be evaded. In inclusion to this, edge computing or mobilized cloud computing is also incorporated into wireless networks, by expanding wireless domains by cloud resources, facilitating computation-intensive applications, e.g., augmented reality (AR) and (IG). In conclusion, the fifth-generation wireless network would be a convergence network, withholding computing capabilities, communication and caching.

Fifth-generation wireless networks plays a crucial role in the big data processing chain due to its ever-present coverage as well as computing proficiencies and also in-network storage. For example, mobile devices or Internet of Things can help to collect data, while the edge computing/caching can perform local processing/storage, including data aggregation and compression. Fifth-generation wireless networks can transport data towards data centers as it consists of core network segments and radio access network (RAN). Hence, it can act as the communicating bridge between data centers and data sources. To better support big data, the attributes of big data must be taken into account for network operation and design. To cater for variety and velocity, provisioning of differentiated networking should be performed, by integrating relevant resources of network, to fulfill varied requirements of services in terms of security, latency and reliability. To house big data volume amount, capacity of the network should be considerably improved.

1.2 Big Data's Role

This can also help in ameliorating the operational functioning of wireless networks. Through analytics in big data, insight or knowledge in the data can be extracted, which can aid proficiently handling and managing of network resources, improve users' experience and boost the revenue. For instance, from the perspective of the users, the pattern of usage, habit, pattern of mobility, the preferences of users can be obtained, helping to improve the experience of the user as a whole and providing context-aware services. From the perspective of the network, by unearthing the spatio-temporal traffic distributions, the balancing of network traffic can be achieved.

1.3 Overview and Conclusion

Here, our intent is to provide an outlook of the interaction between fifth-generation wireless networks and big data, and to present their complementary and synergistic characteristics. We first show big data driven networking and examine how fifth-generation wireless networks can house big data's various attributes (i.e. variety, velocity and volume) by exploiting varied resources (i.e., computing, communication along with caching). Following, we detail big data assisted networking to demonstrate that big data analytics can be used to enhance the user experience and network efficiency. Use cases of big data in fifth-generation wireless networks are briefed big data attributes' effects on network operations shall be discoursed. Also, two cases are provided to discuss the mutual advantages. Then, some research scopes are conversed.

II. BIG DATA DRIVEN NETWORKING

Hefty volume of data is initially accumulated from numerous sources, after that the data is pre-processed followed by carrying through networks to the storage facilities, in data centers, data analytic procedures such as machine-learning, stochastic modelling, and data-mining are performed to unearth the knowledge and insights. Here, we will first show the development of fifth-generation networks, and describe how wireless network is tangled in the chain of processing of big data; then discourse how big data is supported by wireless networks, by making use of big data's attributes.

2.1 5G: Convergence of Computing, Caching and Communication

Fifth-generation wireless network is envisioned to include various resources to back numerous services and heavy traffic. 5G will be described by the convergence of computing abilities, caching and communication.

2.1.1 Computing

The core focus in this regard is to offload users' computation resource demanding jobs to the cloud. To achieve this, various cloud platforms including cloudlet, mobile edge computing, femto-cloud computing, mobile cloud computing, are either currently under deployment or planned to be, at various places in wireless networks. With on-demand computing power facilitation, low-cost, and scalability, cloud-computing permits wireless networks to competently be able to handle computation-intensive applications like AR.

2.1.2 Caching

In-network caching is one of the possible solutions to extenuate backhaul congestions. Small cell-based stations are closely deployed to house heavy traffic, hefty load will be executed on the backhaul, which can decrease redundant transmissions to remote servers and stock prevalent contents nearer to end-user. Caches can be installed in core networks or RAN or both, to achieve dissimilar trade-offs in performance to cost ratio.

2.1.3 Communication

Fifth-generation wireless network will reveal numerous varieties in resources and infrastructure of communication. Varied infrastructure including various small cell base stations will be compactly installed; Diverse communication assets will be combined, including various bands of spectrum and diverse communication methods will be deployed to increase effectiveness.

2.2 5G Wireless Networks: The Bridge

Making use these varied assets accompanied by the ever-present coverage, fifth-generation wireless networks may also work as a connecting path in-between data centers and sources of data, by enabling the procurement of data, pre-dispensation, and carriage. (explained in Figure 1)

2.2.1 Data Acquisition

IoTs can produce large amount of data or gather massive data produced by numerous in-built sensors (e.g., sound sensors, light and humidity sensors, and temperature sensors). Internet of Things enable corporal objects to trade data and allows to be able to be managed effectively, resulting in enormous connected devices, like automobile, smart houses and remote-controlled drones. Along with this, prevalent telephony devices such as smartphones also have ample potential to collect varied types of data, which is allowed by various in-built sensors. This sprouting data procurement stage is also known by the term mobile crowdsensing. This allows mobile crowdsensing to be an interesting choice for cost-effective large-scale data collection from mobile devices.

2.2.2 Data Pre-processing

By in-network computing capabilities and caching either in the core network or Radio Access Network, fifth-generation wireless networks can help in facilitating pre-processing and data storage. When the raw data is accumulated from the preceding segment, before its ready to be transmitted the data needs to be stored and pre-processed first, in order to reduce the redundancy, the data is compressed to reduce its size and also aggregated. Also, the data which is collected locally in a geographical zone can also be examined to promote context-awareness or location-awareness, to aid location-based services (eg., AR) and to also enhance the user experience. The place where to stock and process the data rests on the applications and beset scale/area. If we take an example of a base station which can process the data collected in its coverage to get the variation in patterns of the traffic in sphere of time for better processes. Whereas computing and caching facilities in Radio Access Network (having many base stations), can process the data accumulated in a comparatively larger physical area to stabilize traffic between base stations. It is of vital necessity to determine what, how, and where to stock and process the pertinent data set, based on the resource restrictions service necessities.

2.2.3 Data Transportation

Data will be transmitted from Radio Access Network to the data centers passing through the core network. In this phase, the variety in techniques, resources, and the infrastructure of the network as well as the variety in data should be wisely examined. Fifth-generation wireless networks surrounding RAN and segments of core network is able to send data towards data centers for examination. Various varied sets of data (for e.g., geospatial data and logging data) linked with different target applications can have different necessities in terms of reliability, end-to-end delay, and security. Varied assets extending from access to core networks should be given to safeguard processing in real-time, confidentiality of data, consistent communication and integrity safety during carriage.

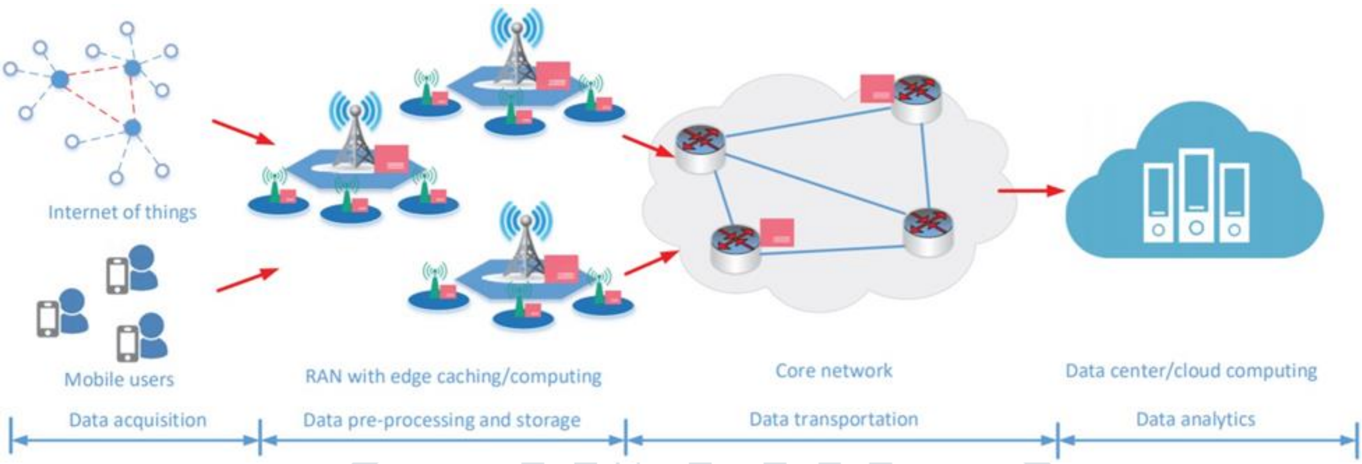

Figure 1. Illustration of big data driven networking

2.3 Networking for Big Data

To be able to efficiently allow handling of big data and extract its useful value to a good extent, big data's attributes such as variety, volume, and velocity should be effectively accommodated by fifth-generation wireless networks. Firstly, one of the five Vs - velocity requires the data to be gathered, pre-processed, and delivered at a quick pace. Secondly, big data volume requires network capacity to be significantly increased. The methods to boost the capacity of wireless network mainly involves spectrum extension, network densification, and spectrum efficiency improvement. The elementary idea is to increase the number of spectrum resources, improve spatial spectrum reuse, improve spectrum utilization. Lastly, the data variety, which is also linked with distinct applications, should be aided in data transportation and processing by satisfying the matching requirements. Practically, volume of big data is mostly undertaken by the installment of network, whereas variety and velocity are aided through effective network process operation, that is, it determines how to efficiently make use of the resources and infrastructure to meet varied requirements. In the chain of processing in big data, storage mediums, computing, and communication are required at suitable positions in between the data center and the data sources, to facilitate data pre-processing, data procurement and storing. It is of utmost importance to cater for the variety and velocity, to provide tailored end-to-end networking which is also service-oriented, and combines relevant varied functions and resources, conferring to distinct demands from different use cases and applications which uses big data. In core, it provides appropriate networkwide assets through the path from data centers to data sources, in order to facilitate the big data chain of processing. When multiple applications which involve big data (for eg., analysis of finance, power-grid, healthcare) should be made to work over the same infrastructure at the same time, the network-wide assets have to be allocated effectively.

A developing resolution for networking which is oriented with service is slicing of network, which over one physical infrastructure, generates different slices, spreading from core network domain to the access domain. The importance for network slicing is to divide network-wide varied resources for distinct slices to support different use cases efficiently. Varied slices would function autonomously and share the same assets and resources. A network slice consists of a set of network assets for a given use case or application. Which can also be tailored to cater the matching end-to-end obligations of service such as reliability, latency, etc. Henceforth, it's of significance yet interesting to efficiently make use of network assets while fulfilling the requirements of service of distinct use cases and applications.

Along the growth of network function virtualization (NFV), it can greatly simplify network slicing mechanism. With NFV in the core network, numerous network functions which are virtualized (for eg., deep-packet examination, splitting of network traffic, data collection, and firewalls) that the streams of data require to go through, and can be composed, which is referred to as service function chain (SFC). NFV allows network functions to be formed in environments which are virtualized, instead of needing a dedicated hardware platform, thus significantly improving network flexibility and also scalability. Based on the use cases or service demands of application, an SFC can be created when required, by constituting various virtual network functions and determining their steps of execution. Various SFCs can be formed for distinct use cases and applications rooted over the same physical infrastructure. Also, virtualization allows the capability of virtual functions to adapt to the dynamics in networks and requirements of the service or dynamically scaled out. In radio access network, if network function virtualization is made available, radio access network slicing may be carried out by creating virtual radio access network observations, including virtual base stations for distinct slices. Or else, resource scheduling-based network slicing needs to be done. For example, at a definite radio access network constituent such as a base station, by effective scheduling, its assets can be shared by or distributed to various slices. Also, by wisely distributing assets, the service demands of distinct slices can also be fulfilled.

Network slicing depends on the network-wide asset management. In short, it primarily goals on how to integrate computing resources, communication and caching as a whole to provide service-oriented networking for big data use cases. An allowing platform is the growing software defined networking (SDN). When having a unified control and network programmability provided by software defined network, network-wide resource management can be drastically simplified to facilitate network slicing. In

RAN, resource distribution for different slices can be achieved in a unified way by software defined networking controllers, and in the core networks, software defined networking controllers with network hypervisor is able to schedule varied data flows to create virtual networks or distinct slices. With software defined networking platform, radio access networking slicing and server function chain may be arranged to have full end-to-end network slicing.

III. BIG DATA ASSISTED NETWORKING

Big data can be smoothed out by fifth-generation wireless network also by the identical token, wireless big data has an ample competence in enhancing user experience and network performance. With the information obtained by examining wireless big data, knowledge or the intuitive features can be obtained, like mobility patterns, temporal and spatial traffic distributions and user distribution; network efficiency can be significantly enhanced. Here, we will firstly show the advantages by using features of wireless big data. Following, we will converse how big data would provide added benefits network operation to achieve those advantages.

3.1. Wireless Big Data: Opportunities

The opportunities could be used to enhance the network execution with regards to network management, operation, service quality improvements and deployment. The following table gives some use cases and examples of wireless big data.

Table 1.
For improving network performance

Big data examples	**Network performance**
Statistics of channel	Channel modelling, power-control
Spectrum usage	Spectrum sharing, MAC and utilization of unlicensed band
Topology dynamics	Black-hole detection, loop and routing
Traffic statistics	Network utilization and load balancing
Network monitoring data	Faults detection, diagnostics, trouble shooting
User distribution and mobility pattern	Infrastructure deployment and seamless handoff
User usage pattern	Anomaly detection and context-aware services
Network traffic and system logs	Intrusion detection and fraud detection systems

3.1.1 Network Management

Various equipment used in networking can produce alarms, tracking data etc. Through data analytics and data mining, diagnostics in real-time can be carried out to automatically detect faults in the network, irregular behaviors, and recognize the matching reasons. The data collected from various sensors and probes used in network can provide information in real-time about the network. Following this, relevant procedures can be conducted to recuperate from the faults. Moreover, the heavy network data may also be used to coach prediction models to forecast future events to be occurred in the network, also active actions may be carried out in advance to prevent service failures and faults in network. Therefore, reliability of network can be considerably enhanced without plenty manual efforts for upkeep.

3.1.2 Network Optimization

User generated requests along with network traffic demonstrate high dynamics in distinct geographical areas over the time. The temporal and spatial distribution obtained by appropriate data sets can enhance network operation and deployment.

3.1.2.1 Network Operation

In the operation phase, by analyzing network data in real time, operations may be efficiently adjusted to improve the effectiveness. For example, with data mining, the pattern of traffic requirements over time may be acquired, also small cell base stations can be dynamically switched on or off to save power. Also, if cache size is limited, only prevalent contents are kept to provide users in area. It is to be noted that popularity of various contents may vary over time at different places. By examining the past recorded user request information, the varying content popularity can be cultured to update the contents which are cached efficiently, as so to increase content hit rate.

3.1.2.2 Network Deployment

When deploying the base stations, the space related traffic load data gained from data analysis may assist to govern the number and the fitting locations of base stations, to be able to lessen installment costs while provisioning certain standards and quality of service (QoS). Moreover, when deploying edge caches, if it is possible to get the data of content requests, the cache size furnished at base stations may also be enhanced to attain cost-effectiveness and meet the obligatory content hit ratio.

3.1.3 Improved Quality of Experience

Along with the data from the network, distinct users' data usage profile also shows individual features, like daily using habits, pattern of mobility, and content request preference. Examining those data has the capability to provide customized and also context-aware service to enhance the experience of the user. For example, by the course data, users' pattern of mobility may be studied so as to facilitate seamless handovers, eg., by pre-storing the needed contents of the predicted path. Moreover, by examining the user's pattern of usage, the situation may be recognized, like the satisfaction of the user, communication scenarios, perceived quality of service, and the running applications. Following, content delivery or context-aware asset allocation may be conducted, eg., switch user to a different wireless network (wireless fidelity or cellular) and regulate transmission parameters associated with modulation, coding and broadcast power.

3.2 Big Data Assisted Operation

Practically, the state of the network is continually varying, due to variations in traffic produced from users and different events in network like congestion and link failure. Manually reconfiguring can be burdensome, prone to errors and inconvenient. A promising networking architecture, software defined networking can obtain an agile network management, in which logical software defined networking controllers reconfigure and dynamically control the underlying infrastructure by open interfaces. Usually, software defined networking functions in a three-phase loop which are given as follows:

i) Abstraction of network;
ii) Decision making by controller; and
iii) Enforcement of policy.

Events of network including changes in topology, traffic statistics and diverted packets are reported automatically to the controllers. Abstraction in network gathers the information on the state of network to software defined networking controllers through control channels. After that, software defined networking controllers allows to make well informed administration decisions linked to asset allocation, network configuration, etc. Considering the amount and scale of information, various big data methodologies may be employed to aid informed decision making. By analytics in big data, deep knowledge of the states of network may be obtained and various events may be correctly predicted to direct the decisions of software defined networking controllers. And lastly, those SDN controllers made controlled decisions will be obligated to substrate networks through application programming interfaces (APIs). By incorporating software defined networking with big data, networks would operate in an automated fashion and overall optimization can be enabled.

In order to obtain efficient and timely network management in software defined networking, various data regarding the network is gathered. Acquisition of data, its pre-processing, and analysis are essentially in-built in the software defined networking operational loop. The three Vs – variety, velocity and volume of the gathered data is of significant status for decision making of software define networking controllers. Firstly, the volume influences control decisions by determining its value, such as network utility, fairness and efficiency. The more volume of abstraction of network, the improved state awareness the controller may have, and therefore improved control decisions may be expected. Secondly, control channel possessing the velocity should be guaranteed. It assists to quicker responsiveness to events in the network. In bad cases which are very few, unacceptable latency or the delay might make the control decisions invalid. Lastly, variety governs the network control's granularity. If a network information which is varied in nature can be gathered, more precise management and control may be obtained.

IV. CASE STUDIES

The following segment demonstrates the synergies of wireless networks and big data with two case studies. We will first show how big data acquisition could be enhanced with the help from of network edge assets. After that, we will examine the usefulness of big data analytics on edge caching shown as follows.

4.1 Edge Assisted Spatial Data Acquisition

With the expanding storage resources and computing powers to the proximity of mobile users, edge computing may considerably benefit location-based data procurement, and space related crowdsourcing, where various employees are drafted to gather space related data. Firstly, depending on the edge platform which is also distributed, spatial crowdsourcing may be achieved in a context-aware and precise manner. Secondly, by pre-processing the gathered data on the edge of network, it aids to decrease the traffic problem in core networks. Lastly, by providing processing resources edge computing may mobile users with a more reliable and also more secure services. For instance, edge platform may be utilized for the finding of malicious users during data gathering. In space related crowdsourcing with a restricted cost, it happens to become a necessity to choose workforces to collectively attain tasks related to crowdsourcing. With help of stage of edge, the workers' past recital information could be stocked and utilized to enhance worker selection. In order to validate it, we conduct an experiment based on the online dataset from Yelp. Yelp publishes various reviews of customers on local businesses, the reviews are crowdsourced and this dataset comprises of two million seven hundred thousand reviews from six hundred and eighty-seven thousand users on eighty-six thousand business in 10 cities. We use this information to match spatial crowdsourcing use cases, in which Yelp users are considered as workers and the reviewing local business are counted as spatial tasks. The more descriptions the worker puts in the review comments, the higher utility will be achieved. We distribute the dataset into daily occurrences and only a subset of workers can be selected to review local business each day. Edge platform can store and examine users' past information to forecast worker's future performance on the task of reviewing. So as, worker selection can be enhanced in space related crowdsourcing, letting to better platform utility.

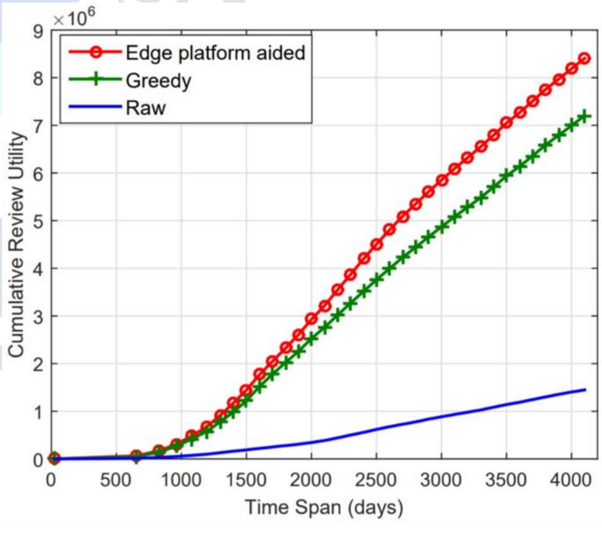

Figure 2. Contrast of growing utility of collaborative platform.

As provided in Figure 2, platform of edge assisted worker selection can significantly enhance the functioning of the space related crowdsourcing. This is due to the fact that, past historical data plays a crucial role, the platform is allowed to make fairly good prediction on worker's future role and then make decisions which are well-versed.

4.2 Big Data Assisted Edge Caching

In this process, the data which is cached should be updated dynamically as the popularity of the original contents change over time. To examine how popular the contents are, we selected the dataset of a random set of videos from the broadcasting site YouTube between the year 2013 and 2017. In Figure 3 daily views of a video since it was uploaded is shown. The normalized status score of this video is also provided in graph based on the comparison with different randomly crawled videos. In the graph, both views which occurred daily and score of popularity varies continuously across the total span of time. Hence, it is needed the the contents of the cache to be updated periodically.

It is to be noted that the ambiguity of future content popularity makes caching which is based on popularity an interesting problem. From data analytics on video view dataset, prediction of future content popularity may also be made, to improve the content hit rate. In the provided case, to predict content hit rate, we use a simple linear model. Specifically, let $\boldsymbol{h}_{c,t}$ be the predicted hit rate of content $\boldsymbol{c}$ at time slot $\boldsymbol{t}$, and $x_{c,t} \in \mathbf{R}^d$ be the d-dimensional historical statistical vector. Then, we have

$$\mathbf{E}[h_{c,t}|x_{c,t}] = x^t_{c,t}\,\theta^*_c + \eta_t$$

where $\theta^*_c \in \mathbf{R}^d$ acts as the featured parameter vector of content $\boldsymbol{c}$, also η_t is the random noise. Therefore, by applying standard ordinary least square linear regression, an unbiased approximation on the feature vector θ^*_c of each content is made, and then make accurate future content popularity prediction. The accuracy of the prediction vastly depends on the amount of historical data. Figure 4 depicts the performance of big data aided edge caching. It is visible that utilization of big data can considerably enhance the cumulative content hit rate on the edge. Also, the mathematical model of content popularity remains to be explored to further enhance the prediction accuracy.

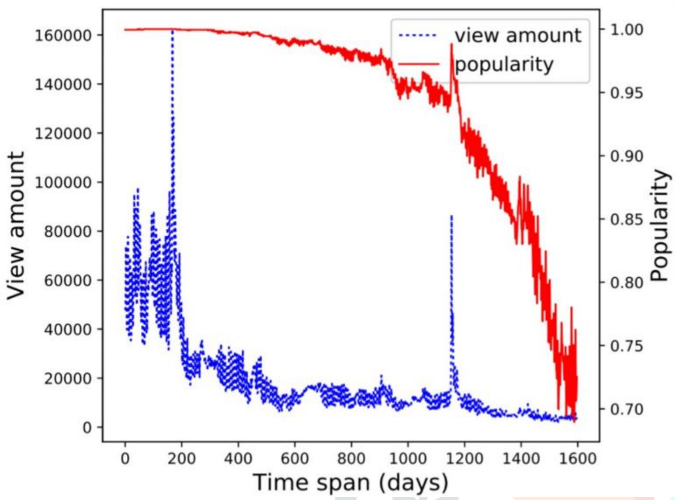

Figure 3. *Data statistics of the daily views and popularity graph of a YouTube video*

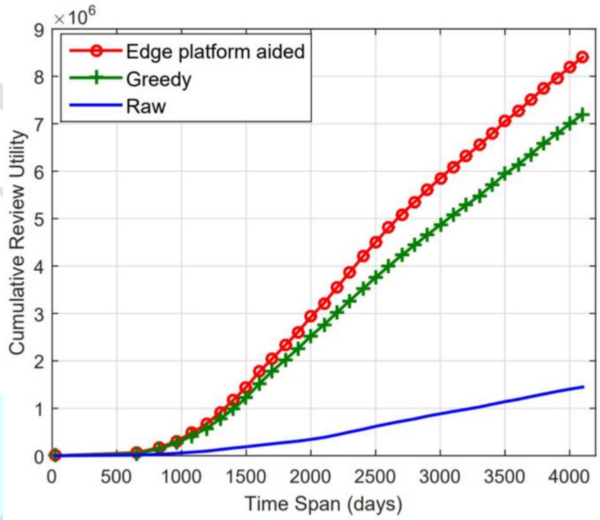

Figure 4. *Data statistics of the contrast of the big data assisted caching with other schemes with regard to long term hit rate.*

V. OPEN ISSUES

This segment discusses some open research issues.

Figure 4. *Data statistics of the contrast of the big data assisted caching with other schemes with regard to long term hit rate.*

5.1 Big Data Aided Network Framework

To get the profits of big data and produce valuable prospects, a big data supported framework is needed to examine and make effective use of the wireless data. It is seen that big data can bring great prospects to networks, the current wireless infrastructure is mainly designed for the delivery of information. The framework expects to combine big data chain proficiently into the network, by efficiently analyzing, collecting, storing and processing data, to enhance network operation. It expects to reject the non-useful data and place storage and processing assets at appropriate places.

5.2 Trade-offs among Communication, Caching and Computing

Fifth-generation wireless networks facilitate varied computation, caching and communication resources. It is of vital importance to efficiently utilize those varied assets to aid varied big data use cases. Also, the intermediate or final results of computation process may need to be stored as valued asset. Stocking all those data can invite high storage cost, also deleting those data may require re-computation when needed. It should be noted that there exist trade-offs among computing, caching and communication resources. For instance, extra computation resources can be traded to reduce the communication load. To help different big data applications cost-effectively, thorough investigation is essential to reveal the trade-off relationship among these varied assets, to provide direction for provisioning of resources.

5.3 Cooperative Edge Caching/Computing

By installment of mobile edge computing and mobile edge caching, the data gathered from distinct sources can be efficiently stored and pre-processed. Various caches may cooperatively form a dispersed storage system, and also distributed edge computing may provide parallel computing capabilities for data processing. Due to the non-uniform data load distribution in both temporal and spatial domains, cooperative edge caching and cooperative edge computing is a promising solution to storing and processing large amount of data in a cost-efficient manner.

5.4 Customized Networking for Big Data

Slicing of network or server function chain has the capability to efficiently support varied big data use cases and services, by creating networking which is service oriented over the physical infrastructure of network. To fulfill the service requirement, the end-to-end networking solution can be personalized, with regards to of resource provisioning, functions which are embedded and resource topology. Also, multiple slices or server function chains should be well arranged to make efficient use of networking resources and fulfill their respective requirements. Moreover, network slicing or server function chain should adapt to dynamics in service requests and network conditions.

5.5 Security and Privacy

It is seen that big data can learn the knowledge from the large amount data through analytics, it also raises privacy and security concerns. To explore the great value, the data sources should be trusted and genuine. During processing and transportation, the data should be prohibited from being modified or altered. Moreover, to protect data from untrustworthy entities, only authorized entities should have access to the data. In addition, data mining on personal user data can violate user's privacy such as habit and location information. For the spread of big data methodologies, security and privacy issues needs to be well addressed.

VI. CONCLUSION

In this report, we have investigated the complementary and synergistic features of fifth-generation wireless network and big data. By exploring computing abilities, caching and communication. Fifth-generation networks can enhance processing chain of big data. With slicing of network, varied assets can be integrated to furnish the attributes of big data. Also, with big data methodologies, insights on users and networks can be provided to enhance network operation and network design. Moreover, big data processing chain can be unified into software defined network operation loop for overall network optimization. This article can also shed the light on the research regarding synergies between big data and 5G wireless networks. To accelerate the pace of big data driven and big data assisted networking, significant research activities are anticipated in this interesting scope.

REFERENCES

[1] S. Wang, X. Zhang, Y. Zhang, W. Wang, L. Wang, and J. Yang, "A survey on mobile edge networks: convergence of computing, caching and communications," IEEE Access, DOI: 10.1109/ACCESS.2017.2685434, 2017.

[2] S. Zhang, N. Zang, P. Yhang, D. Chen, B. Liang, W. Zhuang, and X. Shen, "Software defined networking enabled wireless network virtualization: challenges and solutions," IEEE Network, volume 31, number 5, pp. 42 – 49, 2017.

[3] A. M. Medhat, A. Elmangoush, T. Taleb, G. A. Carella, T. Magedanz, and S. Covaci, "Service function chaining in next generation networks: state of the art and research challenges," IEEE Communications Magazine, volume 55, number 2, pp. 216–223, 2017

[4] D. Kreutz, P. E. Verissimo, F. M. Ramos, C. E. Rothenberg, S. Uhlig, and S. Azodolmolky "Software-defined networking: A comprehensive survey," Proceedings of the IEEE, volume 103, number 1, pp. 14–76, 2015

[5] N. Ruan, M. Li, Q. Qian, H. Zhu, L. Yu, and X. Liang, and "SPFM: Scalable and privacy-preserving friend matching in mobile cloud," IEEE Internet of Things Journal, volume 4, number 2, pp. 583–591, 2017.