IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 4, 2018, accepted May 13, 2018, date of publication May 17, 2018, date of current version June 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2837692

Big Data Analytics, Machine Learning, and Artificial Intelligence in Next-Generation Wireless Networks

MIRZA GOLAM KIBRIA, (Member, IEEE), KIEN NGUYEN, (Senior Member, IEEE), GABRIEL PORTO VILLARDI, (Senior Member, IEEE), OU ZHAO, KENTARO ISHIZU, AND FUMIHIDE KOJIMA, (Member, IEEE)

Wireless Systems Laboratory, Wireless Networks Research Center, National Institute of Information and Communications Technology, Yokosuka 239-0847, Japan

Corresponding author: Mirza Golam Kibria (mirza.kibria@nict.go.jp)

ABSTRACT The next-generation wireless networks are evolving into very complex systems because of the very diversified service requirements, heterogeneity in applications, devices, and networks. The network operators need to make the best use of the available resources, for example, power, spectrum, as well as infrastructures. Traditional networking approaches, i.e., reactive, centrally-managed, one-size-fits-all approaches, and conventional data analysis tools that have limited capability (space and time) are not competent anymore and cannot satisfy and serve that future complex networks regarding operation and optimization cost effectively. A novel paradigm of proactive, self-aware, self-adaptive, and predictive networking is much needed. The network operators have access to large amounts of data, especially from the network and the subscribers. Systematic exploitation of the big data dramatically helps in making the system smart, intelligent, and facilitates efficient as well as cost-effective operation and optimization. We envision data-driven next-generation wireless networks, where the network operators employ advanced data analytics, machine learning (ML), and artificial intelligence. We discuss the data sources and strong drivers for the adoption of the data analytics, and the role of ML, artificial intelligence in making the system intelligent regarding being self-aware, self-adaptive, proactive and prescriptive. A set of network design and optimization schemes are presented concerning data analytics. This paper concludes with a discussion of challenges and the benefits of adopting big data analytics, ML, and artificial intelligence in the next-generation communication systems.

INDEX TERMS Big data analytics, machine learning, artificial intelligence, next-generation wireless.

I. INTRODUCTION

In a service-driven next-generation network, a single infrastructure needs to efficiently and flexibly provide diversified services such as enhanced mobile broadband, ultra-reliable and low-latency communications and massive machine type communications. It should also support coexistent accesses of multiple standards such the fifth generation (5G), long-term evolution (LTE) and Wi-Fi. Also, it should coordinate a heterogeneous network with different types of base stations (BSs), for example, macro, micro, femto, pico BSs and diverse user devices as well as applications [1]. The challenge to efficiently operate a network capable of facilitating such flexibility while satisfying the demands from diversified services is enormous for a network operator. On top of this, the network operators face considerable challenges in extending their coverages and keeping up with the ever-increasing capacity demands with a limited pool of capital and scarcity of resources such as spectrum. Manual configuration for network planning, control, and optimization will make things even more complicated. Moreover, the human-machine interaction can, sometimes, be time-consuming, susceptible to error and expensive. Consequently, automation of various entities and functions of the cellular networks has been one of the principal concerns of the network operators in consideration of reducing the operational expenses.

Operators have been optimizing their networks all along, but even today, the prevailing approach is to independently optimize single key performance indicators (KPIs), or an

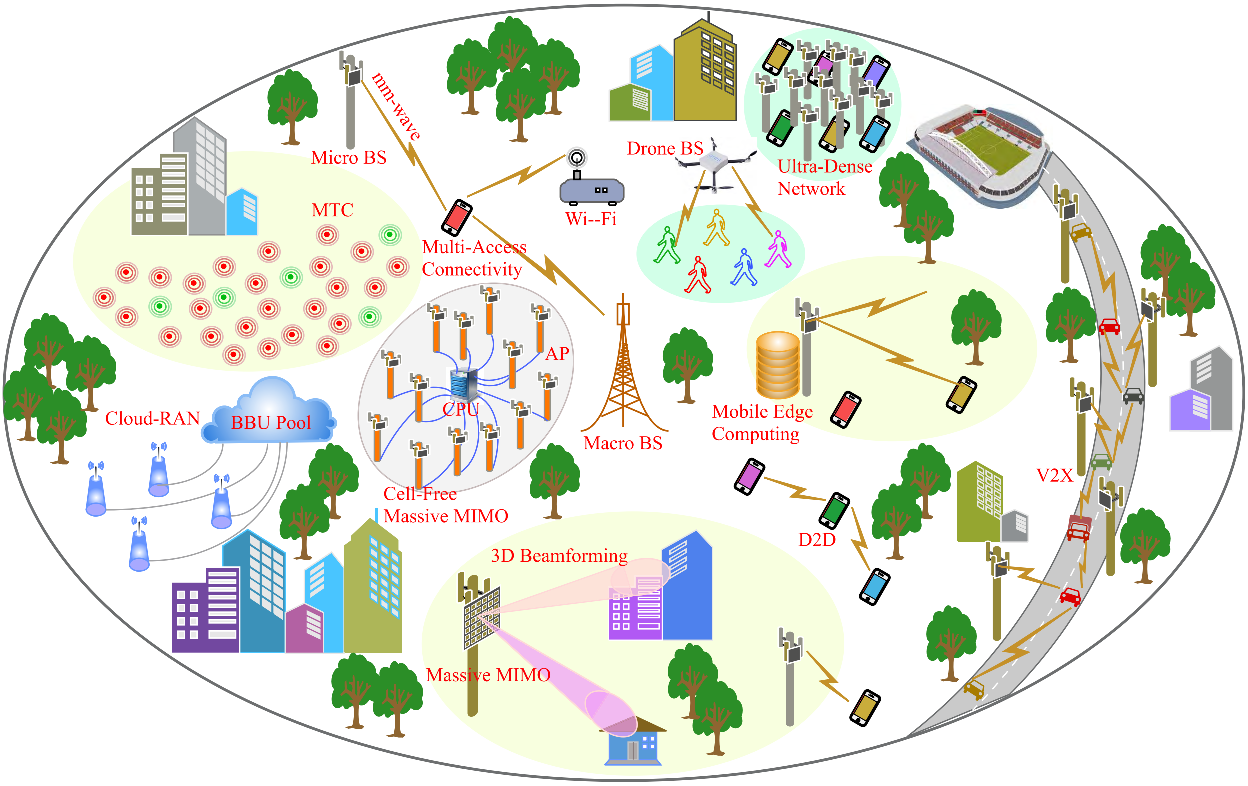


FIGURE 1. Graphical illustration of the next-generation communication system with some technological elements. Slicing, virtualization, (mobile) edge-computing, massive multiple-input-multiple-output (MIMO), 3-dimensional (3-D) beamforming, (ultra-dense) small cell networks, device-to-device(D2D) cell-free massive MIMO, multi-connectivity, cloud-radio access network (RAN), millimetre-wave, cloud-architecture/computing, etc., are the fundamental technologies to achieve the targets of the next-generation network.

element within the network independently [2], thus using a small number of data sources. The network operators mostly depend on KPIs accumulated at different locations/parts of the network to make decisions employing various data analysis tools. Network monitoring and optimization are still predominantly performed on old/recorded data, but this dramatically restricts their capacity. The network operators, in general, have/can have access to a vast amount of data from their networks and subscribers. With the appropriate analytics, big data can convey broader intuitiveness and understanding since it draws from multiple sources to reveal previously unknown patterns and correlations [3]. It benefits to acquire a thorough knowledge of various unknown values and delivers new measures in enhancing the performance from different levels of wireless networks.

The value that analytics brings to optimization comes from expanding the range of data sources and taking a customer-centric, quality of experience (QoE)-based approach to optimizing end-to-end network performance. In widening the variety of data sources, analytics requires more effort than traditional optimization, but it also provides a unified and converged platform for multiple targets of optimization. Now, within the 3rd generation partnership project (3GPP), network data analytics (NWDA) has been introduced to deliver slice and traffic steering and splitting (between 3GPP and non-3GPP access) related analytics automatically [4]. The European telecommunications standards institute (ETSI) has created the industry specification group called experimental network intelligence (ENI) that defines a cognitive network management architecture based on artificial intelligence (AI) techniques and context-aware policies. The ENI model helps the network operators in automating the network configuration and monitoring process.

From the operational expenses point of view, the system needs to be smart, self-aware, self-adaptive and must be able to run the network services economically and manage and operate the networks autonomously [5]. Conventional reactive maintenance is no more efficient. With big data analytics, the predictive and proactive maintenance of the network elements can be performed. With the volume of the data, the speed of data flowing in and the range and type of data sources, the network even go beyond prediction, i.e., it can assist and or prescribe the operations and maintenance unit with decision options and impacts of the actions, etc. Machine learning (ML) and AI can help in uncovering the unknown properties of wireless networks, identify correlations and anomalies that we cannot see by inspection, and suggest novel ways to optimize network deployments and operations.

II. DRIVERS AND EVOLUTION OF ANALYTICS IN NEXT-GENERATION WIRELESS SYSTEMS AND COMPUTATIONAL INTELLIGENCE

A. DRIVERS TO ADOPTION OF ANALYTICS

The ever-increasing complexity of the networks and complicated traffic patterns make the big data analytics appealing and very important for the network operators. The network operators were earlier very cautious about the adoption of big data analytics. However, multiple drivers are turning the network operators conservative stance towards the comprehension that in-depth optimization of the networks and the services are essential for near future. As a consequence, there exists a consistent and rational commitment to capturing a thorough knowledge and understanding of the network dynamics and make the best use of them through optimization. Three predominant drivers strengthening the adoption of big data analytics [2] can be identified as *Cost and Service* drivers, *Usage* drivers and *Technology* drivers. In the following, we discuss them in more details.

1) COST AND SERVICE DRIVERS

The subscribers, in general, are more demanding but less eager to raise the wireless payout. In such environment, there is an urgent need for optimization of the usage of network resources. Furthermore, the network-centric service model is transforming into a user-centric service model based on the QoE. As a result, the network operators need to better understand the QoE and its relationship with the network's KPIs. Also, the network operators need to retain its customers. As a result, the network operators need to (i) manage its traffic based on service and application, (ii) improve efficiency to retain profit margins, (iii) improve network performance and QoE without increasing cost and (iv) keep churn as low as possible, etc.

2) USAGE DRIVERS

The traffic patterns, subscriber equipment, and subscribers' profiles are all heterogeneous. In a user-oriented service model, analytics supports the network operators maintain and regulate traffic types, wireless devices, and subscribers diversely based on the network operators' strategies and each's requirements. Furthermore, the wireless traffic load is growing faster than the capacity, and the network operators are facing severe challenges to increase network capacity cost-effectively. Therefore, intensifying the resource utilization is required. Analytics take the network load into account and helps the network operators to manage network traffics more efficiently in real time.

3) TECHNOLOGY DRIVERS

The next-generation wireless networks have many technology components such as network resource virtualization, edge-computing, mobile edge-computing, network-slicing, etc. It integrates multiple air-interfaces, network layers and accommodates a range of use-cases. The network operators need robust analytics framework to orchestrate the virtualized network resources efficiently. The analytics also help the network operators to balance the centralized and distributed functionality. The data analytics facilitates the network operators to figure out the most competent way to slice the network and traffic, i.e., the number of slices, splitting traffic across slices, etc., which depend on the type of traffic and how varies over time and space.

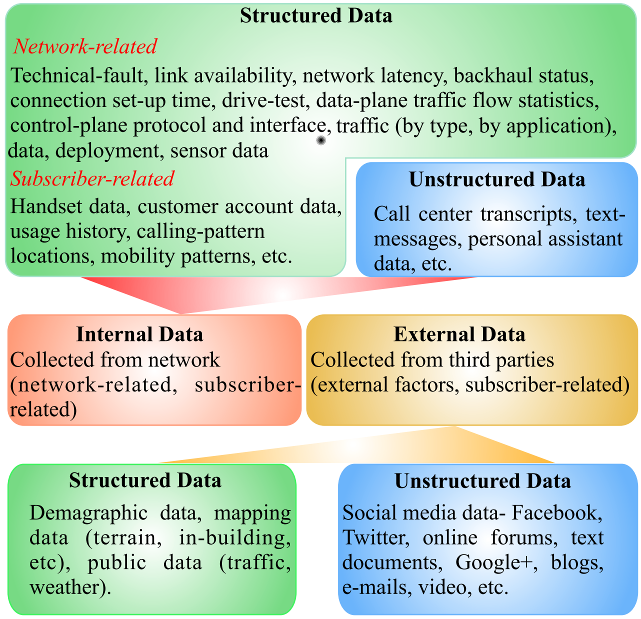


FIGURE 2. Data sets and the sources of data available to the network operators for big data analytics, machine learning and artificial intelligence.

B. TYPES OF ANALYTICS

There exists a succession of evolution in big data analytics, starting from descriptive analytics to diagnostic analytics to predictive analytics, and excelling towards prescriptive analytics as shown in Fig. 3, out of which three (descriptive, predictive and prescriptive analytics) are dominant. The network operators currently are in descriptive phase and use mainly the visualization tools to get insights on what has happened, the network performance, traffic profile, etc. The network operators can make use of the diagnostic analytics to figure out the root-causes of the network anomalies and find out the faulty KPIs and network functions/elements. To get the diagnostic analytics, the analytics tool employs techniques like drill-down, deep learning, data discovery, correlations, etc.

Predictive analytics is an excellent tool for making predictions. Note that it can never report or be precise about what will happen, however, predictive analytics can only produce forecasting about what might occur, for example, future locations of the subscribers, future traffic pattern and network congestion, etc. Predictive analytics deliver predictions events based on the real-time and archived data by

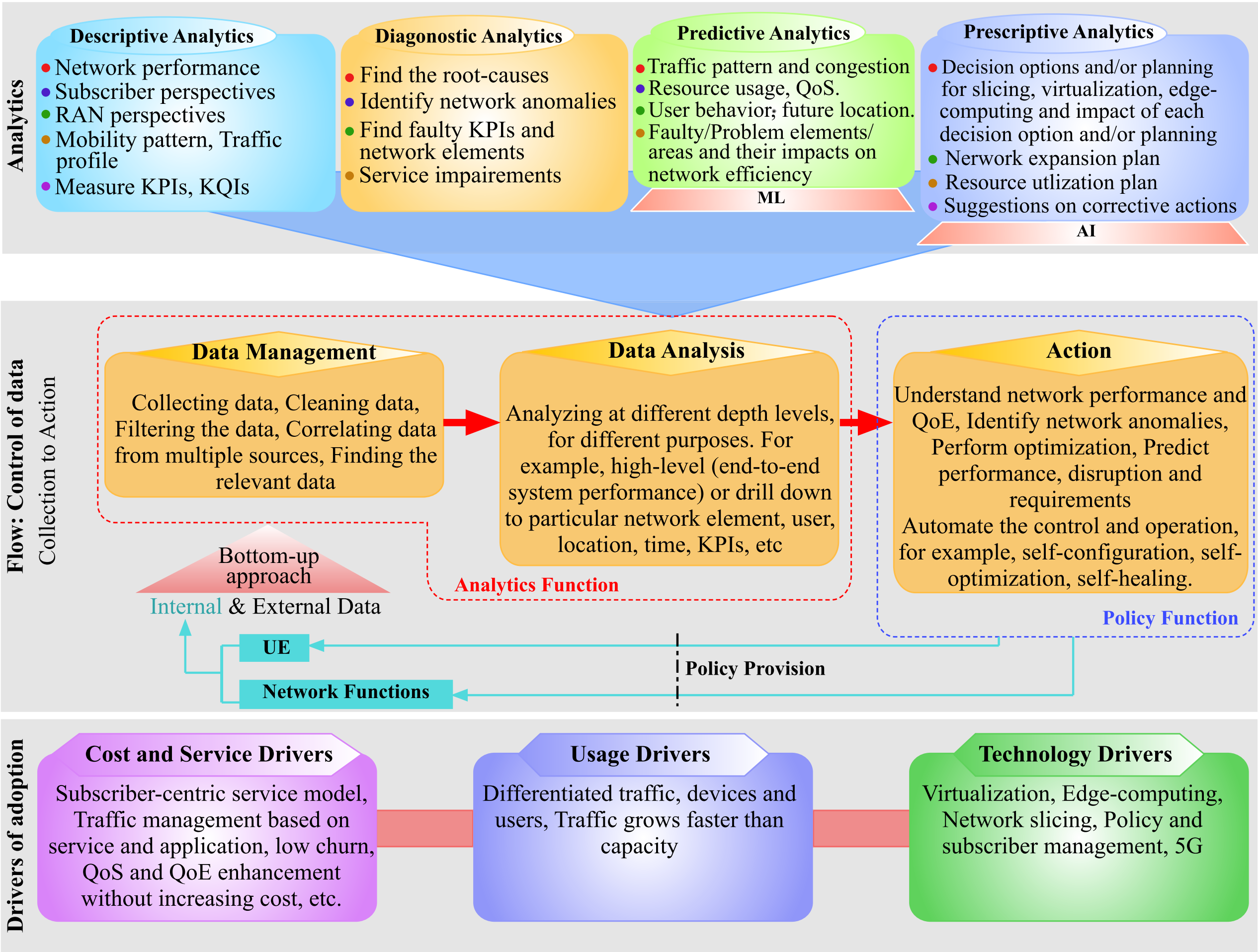


FIGURE 3. Drivers from different directions that strengthen the case for the adoption of analytics in next-generation communication systems. The flow of processes/actions of the network operator that employs analytics. Here, the Analytics Function collects the data and analyze them. The Policy Function obtains the analytics reports produced by the Analytics Function and may dynamically and intelligently deliver analytics-based policy rules for UE and the network functions.

making use of various statistical techniques such as machine learning, data mining, modeling as some analytical process and game-theoretic analysis. Prescriptive analytics goes steps ahead of just predicting the future events by suggesting decision options for slicing (i.e., how to slice, how many slices), virtualization, edge-computing, etc., along with the implications of each decision option. Therefore, the prescriptive analytics need a useful predictive model, actionable data and a feedback system for tracking down the results generated by the action taken. The decision options (e.g., for network expansion, resource usage) are produced considering the network operators preferences, system constraints (backhaul, fronthaul, spectrum, transmission power), etc. Prescriptive analytics can also suggest the finest course of actions for any pre-defined target, for example, of a particular KPI.

The network operators have access to large amounts of data which can be categorized into two classes such as internal data and external data as shown in Fig. 2. The internal data corresponded to data belonging to the network operators and produced in the network, which is network related and subscriber related. The external data is collected from the third parties. Both the internal and external data can be further classified into two categories, which are structured data and unstructured data. The structured is stored in a relational database, i.e., each field in the database has a name, and the relationship between the areas are well-defined. On the other hand, the unstructured data (for example, call center transcripts, messages, etc.) is not usually saved in a relational database. Comprehensive coverage of the features and sources of mobile big data can be found in [6] and [7].

C. COMPUTATIONAL INTELLIGENCE

network operators have access to a collection of data sets (i.e., these data can be highly dimensional, heterogeneous, complex, unstructured and unpredictable) that are so large and complex that the traditional data processing and analysis approaches cannot be employed due to their limited processing space and processing time. Computational intelligence, a set of nature-influenced computational techniques and methods, play a very crucial role in the big data analysis [8].

It enables the analytics agent to process and analyze the historical and real-time data computationally, and eventually finds out and explain the underlying patterns, correlations, as well as to intensely understand the specific tasks. The computational analysis tools and methodologies convert the network operators' massive amount of raw data (unprocessed, structured/unstructured) into meaningful data/information.

For feature selection, data-size and feature space compliance, active-incremental-manifold-imbalance learning on big data, uncertainty modeling, sample selection, classification/clustering, etc., many tools and methodologies can be applied for big data analysis. For example, fuzzy logic, neural algorithms, rough sets, swarm intelligence, evolutionary computing, stochastic algorithms, physical algorithms, immune algorithms, learning theory, probabilistic methods are the tool and methodologies that the network operators' big data analytics agents can employ for computationally processing and analyzing the available data.

In general, for big data analytics, the network operator can follow two distinct approaches, namely top-down approach and bottom-up approach [9]. In the top-down approach, the network operators define their targets to be achieved or problems to be resolved, and then decide what data sets are required. Whereas, in the bottom-up approach, the network operators already have access to massive amounts of data and then exploit the big data on hand to get the insights. The top-down approach delivers incremental benefits and it is very challenging to execute. It also, in most of the cases, does not bring on surprising and adventitious results. On the other hand, the bottom-up approach facilitates a more outright and transparent view of the network performance, subscribers' behaviors, resource utilization, etc., and may bring on completely new opportunities for the network operators. The bottom-up approach is also likely to capture the subscribers' perspectives the RAN perspectives and may beget new business opportunities for the network operators.

III. MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE FOR MANAGING COMPLEXITY

The ML and AI are two compelling tools that are emerging as solutions for managing large amounts of data, especially for making predictions and providing suggestions based on the data sets. They are, however, very often appear to be used interchangeably in spite of some parallels. ML is sometimes brought up as a subspace of AI based on the concept that we can let the machines learn for themselves by providing them access to large amounts of data. On the other hand, AI is the widespread and broader perception of devices becoming capable of carrying out tasks in an intelligent way. Compared to the generalized AI (a generalized AI system, in theory, can handle any task), applied AI is more suitable for next-generation communication systems as the applied AI system can be devised to adeptly controlling and optimizing the wireless networks. Unlike ML models, AI models reach out the world, accustomed to the changes and rebuild themselves [2]. While ML is excellent for predictive analytics [10], AI goes beyond predictions and prescribe plans/suggestions with implications to realize a benefit.

Managing wireless networks that grow in size and complexity becomes very difficult since there is need to integrate new elements and technologies to benefit from the technological advances. The amount of data such large and complex networks produces is too large and too complicated. Machine learning and artificial intelligence are useful for analytics as they can extract valuable information from the raw data and generate insightful advice and predictions. ML and AI are expected to assume the primary role in the development and evolution of analytics, but analytics will not reduce to them. ML is developed mainly from AI, hence the two overlap. ML has tools to extract relevant information, suggestions, and predictions from the data sets that are too large and too complex. AI has a broader scope: to replicate human intelligence or some aspects of it and other cognitive functions. The differentiation among big data analytics, ML, and AI, and their mutual dependence relationships are discussed in details in [2].

Furthermore, for non-recurring events, there is no historical data to rely on. Hence the real behavior of the network will diverge from the predictions [2]. The ML and AI are becoming potential to help network operators to address areas which are new, and there is no historical data, or too complicated to understand with traditional approaches. The ML and AI tools can correlate multiple sources of data and find what is relevant. They may also reveal interrelations and dependencies that were not previously identified because their automated mechanisms have the capability of anatomizing and inspecting data more intensely and more methodically. Although human expertise is useful in confining the focus to produce solutions and to manage complex problems, it has limited capability in finding new answers and insights. The future of wireless networks will undoubtedly rely on AI. In [11], the authors have provided a panned overview of the range of wireless communication problems and issues that can be efficiently addressed using AI while delivering detailed examples for the use-case scenarios.

IV. DATA-DRIVEN COVERAGE AND CAPACITY OPTIMIZATION OF NEXT-GENERATION CELLULAR WIRELESS NETWORKS

The conventional network-centric architecture cannot capture all of the nuances that can affect service quality. Mobile operators need solutions that provide them with an analysis capability that captures all the information relating to the network and subscribers into a single enterprise geolocation platform that can help remove the assumptions involved in fault isolation and reduce mean time to repair. The network operators are suitably positioned to exploit big data analytics because of their access to huge amounts of data. The big data analytics engine/agent can produce/predict the following analytics based on its data, primarily from two sources, such as the network data and the subscriber data, which are then exploited to design and optimize the network.

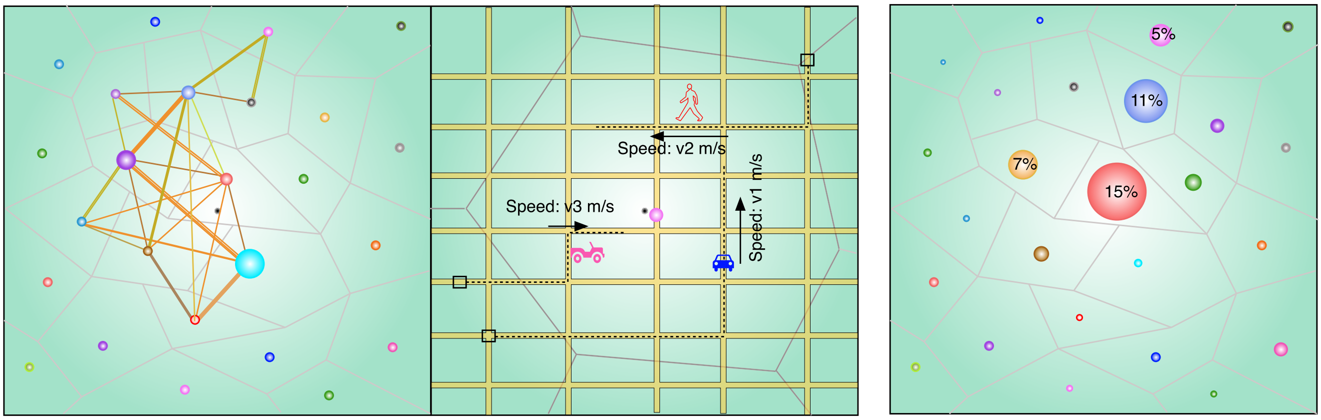


FIGURE 4. Left: Example of the trajectory of a mobile user who visited the vicinity of during an observational period. The circles correspond to the serving BSs, i.e., give the approximate locations. The gray lines depict the Voronoi lattice. The width of the line between two BSs is proportional to how often the user moves between the BSs, and the sizes of the circles are proportional to the resources/traffic consumed. Middle: Example of the small-scale (e.g., within few seconds/minutes) trajectories of the user for more accurate/exact location of the user. Right: The circles represent the BSs and sizes of the rings are proportional to the traffic loads on the BSs, i.e., the traffic pattern and congestion status in the network can be measured. Note that results here do not come from real field experience.

- **Subscriber Profile (SP)**: In this context, the subscriber profile consists of the device profile, service-level agreement (SLA), subscriber's affordability (price per unit of data-rate), quality of service (QoS)/policy, behavioral profile, etc. It plays a vital role in the abovementioned controlling and optimization process. The priority of the subscriber in the network is defined in the subscriber profile when resource allocation, congestion control and traffic offloading is performed. Behavioral data provides information how the user behaves in using various applications/services. For example, how frequently and when the user makes video/audio call and the average length of the call duration? Through analytics, we can speculate on a lot of these user attributes.
- **Subscriber Perspective (Sub-P)**: Subscriber perspective is an attribute/measure that associates network operator's network activity with the user's SLA, pricing, QoS, QoE, etc., and it delivers a subscriber-centric outlook of the network for analytics [12]. Subscriber perspective is, in general, defined by the *Cost Over Quality Ratio*, which sometimes gets polished through a variety of attributes linked to the requested service class and perceived friendliness to the service, i.e., QoS violation, delay violation, etc. It enables the network operator to measure or make a perception about the RAN quality from the subscriber point of view and put them in a better position to provide a high QoE.
- **RAN Perspective (RAN-P)**: RAN perspective is a measure that provides the network operator the subscriber-centric RAN quality, i.e., the RAN performance from the subscriber's point of view [13]. The user equipment's view of the signaling information such as signal strength, error codes, available networks, etc., are beneficial to the network operator for analytics. From the user's predicted trajectory, spatial deployment of the BSs and signaling metrics, the network operator can generate heat map for coverage and determine the RAN quality. Advanced cell mining that statistically analyzes the performance data enables the network operator to identify radio cell irregularities and other negative syndromes via anomaly (i.e., SLA violation) diagnosis and trend study of the time series data, and control traffic and RAN congestion problems. With RAN Perspective, the full end-to-end subscriber experience can be measured regarding *Service Availability* and correspondingly mapped to the exact location in the network. The network operator can also use *Subscriber Satisfaction Coefficient* to define the RAN perspective. Note that the signaling metric cannot be easily retrieved from mobile gateways or retrieved by network probes. An efficient retrieving method is discussed in [13] that uses SIM-based applet stored in users devices to collect the signal strength and quality metrics. Thus the subscriber devices act as network probes in measuring the RAN perspective.
- **Subscriber Mobility Pattern (SMP)**: To guarantee the QoS requirements and to efficiently maintain resources utilization, traffic offloading and routing, knowing the mobility information of a user in advance is very crucial. Human travel pattern analyses reveal that people travel along specific paths with reasonably high predictability [14]. The trajectory of a mobile user can be predicted based on user's present location, the movement direction and the aggregate history of SMP. It is possible to predict the spatiotemporal trajectory (trajectory with both spatial and temporal information), i.e., not only the mobile user's future location but also the time of arrival and the duration of stay can be predicted. Mobility pattern is based on user positioning, which can be estimated using the signals from the cellular system.

FIGURE 5. Some data analytics and their applications in control and optimization of next-generation wireless communication systems.

- **Radio Environment Map (REM)**: The network operators can better plan, build, control and optimize their networks conforming to the spatiotemporal radio atmosphere, through prediction of radio signal attenuation. Many schemes have been developed that give the network operators the means to predict the distribution of radio signal attenuation at different operating frequencies and in many different radio environments. The radio map along with the mobile user's predicted trajectory facilitates the prediction of average channel gains. There are several different methods to construct the radio map, for example, radio map based on drive test measurements, radio map based on measurements through user terminal equipped with global positioning system [15].
- **Traffic Profile (TP)**: To attain as well as predict the network's congestion status, tempo-spatial traffic load variation needs to be known, i.e., the knowledge of temporal traffic trace, BS spatial deployment and BSs' operating characteristics (transmission power, height, etc.) are critical. The authors in [16] report that the network's traffic load dynamics demonstrates periodical characteristics over days and hours, thus implying high predictability of the traffic load. The traffic profile along with the SMP can be used to estimate and predict the traffic arrival rate, congestion status of the network with required time resolution/granularity.

It is very crucial to have a sturdy, well-balanced load-distributed cellular system in a dynamic network and radio environment with mobile users using bursty applications and services. The next-generation network can employ the systems analytics, user, and service analytics, radio analytics for control and optimization of the network [17] in the following scenarios.

- **Resource Allocation Strategy** Advanced resource allocation is very crucial for enhancing the spectrum and power utilization efficiency of the communication systems. Leveraging the big data analytics based prediction ability in optimizing the resource allocation has been reported to be very advantageous. With the help of SMP and TP, the network operator can approximate the typical resource usage per-cell per-user in the network. Because (i) the average channel gains are predictable from the trajectory of the user and (ii) based on content popularity, user's behavioral profile and currently running application, the preferred contents can be predicted even before the individual users put forward their service requests. As a result, with the help of big data predictive analytics, the operators can predict changes in the users' service demands and thus can manage and optimize the resource allocation in real time. Integrated backhaul and access in mmWave.
- **Subscriber-Centric Traffic Routing** Providing the best QoE as the end users' subjective perception is one of the most important requirements. Service delay, jitter affect the mobile users' QoEs very severely. Data-driven solutions can deliver traffic to different users depending on their subscription profile, types of applications, and preferences. A QoE-aware network continuously adopts the changing environment to provide acceptable QoE. The SMP, the network utilization profile, and TP can help the operator to devise efficient routing protocol while considering the backhaul load, the SLA, and the corresponding cost. Depending on users preferences and interests, and currently running application, the system can proactively cache the favorite content, and use the backhaul route that is closer to the local caching server.
- **Subscriber-Centric Wireless Offload** Due to an exponential surge in mobile data traffic carrier over macrocell layer, the network operators are more and more finding out approaches to optimize the traffic in the network while ensuring seamless connectivity and minimum guaranteed QoS to its subscribers. Traffic offloading from macrocell layer to small cell layer (specifically towards WiFi networks) is a great way to relieve congestion in the macro layer and enhance the overall network throughput. Blindly offloading the mobile users may result in dissatisfaction of the subscribers of the higher tier and breaching SLAs. Therefore, it is necessary to devise practical solutions that aid the network operators to decide and offload mobile users to WiFi, based on user profile and network congestion conditions. Data-driven contextual intelligence originated from correlating the customer profile (types of application, spending pattern, SLA) with SMP, TP, and REM, can decide which customer should be offloaded, and even to which small cell/WiFi the customer needs to be unloaded.
- **Optimized Cell Placement** Small cell placement plays a vital role in defining the capacity of a heterogeneous network. Strategic small cell placement is crucial in

areas where subscribers concentrate while taking care of coverage goals, radio frequency interference issues and its potential in relieving congestion from the macro layer by offloading traffic. Rapidly placing the small cells at the very best locations is a complex problem as the number of small cells is much larger. Traditional macro cell layer management tool and even the self-organized networking tool may not compensate for improper cell placement. However, the data-driven solution can efficiently administer the small cell placement issue exploiting the knowledge of long-term user density, traffic intensity. Data-driven solutions incorporating the long-term TP, REM can devise optimized dynamic small cell placement strategy that identifies key locations where small cells need to be deployed and re-arranged to enhance the network capacity, minimize interference and improve the traffic offloading capability. The 3D geolocator tool that uses predictive ''fingerprinting'' algorithms to locate traffic hotspots can simplify the cell placement task.

- **Radio Access Network Congestion Control** The combination of limited network resources and ever-growing demands result in unavoidable RAN congestion, which degrades users' quality of experience. Expansion of existing RAN provides a solution to this problem, but it is expensive. A flexible, as well as a cost-effective solution, is to deploy a proactive policy control mechanism preventing deficiency of RAN resources. Smart congestion control solution considering location information, the load level of network elements and users' service level agreements can deliver perceptibility at a particular sub-cell level and caters priority to some set of subscribers based on their tiers. The congestion events are short-lived (typically congestion occurs at busy times of the day) and users future locations are predictable. With the help of data-driven predictive analytics incorporating the correlation between SMP, radio map and traffic profile, advanced proactive RAN congestion control mechanism can be deployed where the occurrence of RAN congestion is predicted. RAN congestion controlling can be done in many ways, for example, by reducing the QoS for subscribers belonging to the lowest tier of users, rejecting new session establishment, terminating specific sessions.
- **Advanced Load Balancing** Note that the profile of mobile users and traffic in each cell is distinctive, and the patterns change from time to time. When some users disassociate with one cell and move to the neighboring cell, the network's traffic load distribution, i.e., the traffic profile may change severely, and as a consequence, some cells in the network may get overloaded causing service downgrade. Currently, the load balancing methods employed by the network operators are almost manual, thus not efficient, and at the same time, they are not accurate enough. Predictive analytics by data mining and correlating the network and subscriber data such TP and SMP can not only help in understanding the cells' current load situation, but also in identifying the heavily loaded parts of the network and predicting the traffic variation in advance. Consequently, the network operators can perform advanced load balancing and cell planning by adding capacity, expanding the coverage of unloaded or lightly loaded cells to unburden the neighboring overburdened/overloaded cell. The data-driven advanced load balancing will enable the network operators to optimize the utilization of available network resources.
- **Advanced Beamforming** Beamforming is an integral technology component in next-generation communication systems for enhancing the coverage and data rates. A BS with multiple antennas can generate many beams simultaneously [18]. Under static beamforming (fixed beam pattern without beam-steering), for a mobile user, the quality of the serving beam may deteriorate, and hence a different beam from the same BS (from same sector or different sector) or an adjacent BS that serves the user well needs to be selected. The ML can help the serving node to choose the best beam for the user dynamically. The ML also enables dynamic switching ON/OFF the beams based on TP and SMP for energy and interference minimization. Holographic beamforming[1] (with electronic speed beam-switching/beam-steering) [19] along with data analytics and ML can help in dynamically rerouting the traffic, dynamic adjacent cell access, steering coverage where it is needed to accommodate usage patterns, for example, rush hour traffic, events, etc., as shown in Fig. 6.
  Fig. 6(a) contemplates adaptive/dynamic rerouting of traffic when there is an obstacle or physical interference between two communicating nodes. The rerouting path can be dynamically selected based on TP and other information such as resource availability. Dynamic adjacent cell assistant in Fig. 6(b) with holographic beamforming facilitates serving a distant user outside the general coverage area of the assisting BS when the original serving BS has bandwidth shortage/overloaded or becomes non-functional. Depending on the received information and making use of TP and REM, the holographic beamforming antenna can dynamically configure a high directivity beam towards the distant user [20]. Similarly, dynamically steering coverages where it is required as shown in Fig. 6(c) combined with 3D geolocator tool and configuring long-range, high-capacity links along with electronic speed beam-switching to provide access in motion as shown in Fig. 6(d) can be enhanced by analytics from internal data, external data and ML.

Apart from the control and optimization scenarios mentioned above, accurately and efficiently accomplishing the

[1] In holographic beamforming, the complex propagating wave across surface scattering antenna or the transmitting aperture becomes a holographic profile, i.e., the collective profile across the antenna-array elements represents the desired hologram for transmission.

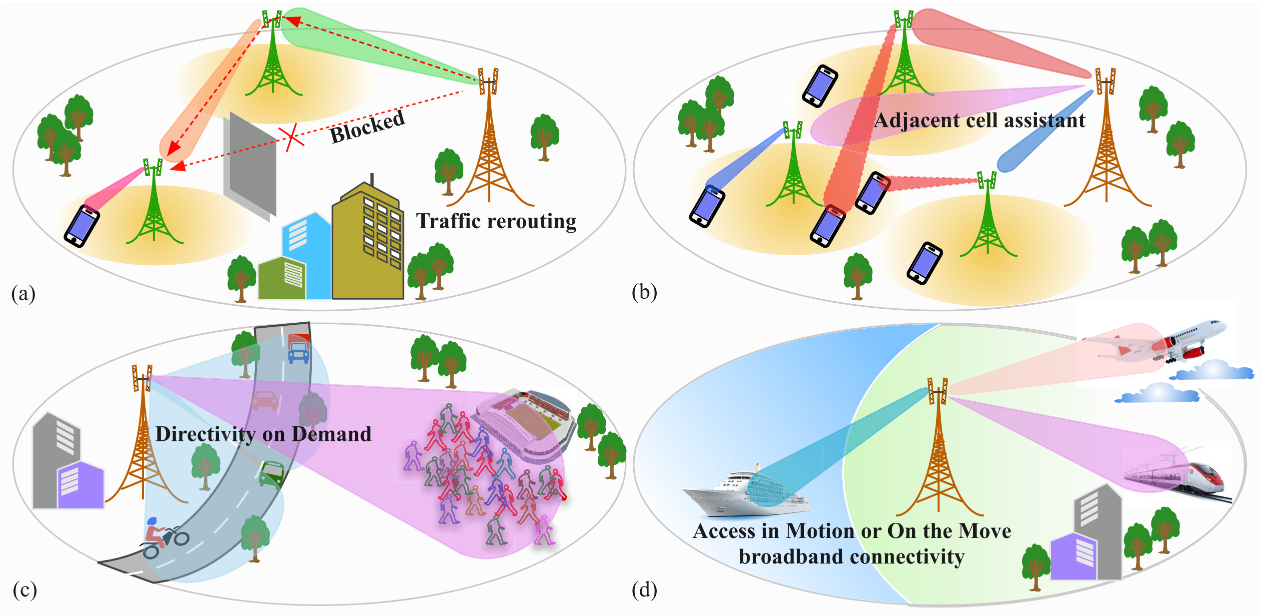


FIGURE 6. Data analytics and ML, and AI techniques can be used in analog, digital, and hybrid beamforming in terms of generating the optimal beam patterns, dynamically selecting the most suitable beam, and performing beam-steering operation. In this figure, holographic beamforming has been taken as the use-case scenario.

maintenance of the network elements, backhaul monitoring (potential bottlenecks in backhaul networks). Fronthaul management and orchestration, intelligent network slicing, energy optimization, monitoring of critical network health variables are some of the key pain and troublesome issues the network operators are very often challenged with. With big data analytics, the predictive maintenance of the network elements can be performed. The predictive maintenance inspects the operational status of the network elements through sensors in real-time. With the help of big data analytics, the potential risks can be identified. Thus the possible faults are found earlier. This helps the operation and maintenance team of the network operator to become proactive to work out predictive maintenance planning.

Furthermore, there are plenty of other applications of data analytics, ML, and AI in next-generation communication systems. For example, the network operators can employ analytics for obtaining useful insights about the physical layer [21] and the medium access control (MAC) layer. Also, the optimal constellations in interference channels where the optimal schemes are unknown, the best beamformer, pre-caching/buffering, the most suitable forward error correction code, the optimum MAC protocols, predictive scheduling, etc., can be performed. Intelligent wireless network architecture, RAN optimization regarding transmission control protocol (TCP) window optimization, mobility management optimization can also be achieved through the use of big data analytics [22]. Data-aided transmission, network optimization, for example, channel modeling, multiple user access and novel applications such as unmanned aerial vehicle/drone communications, smart grid, etc., have been discussed in [23].

V. CHALLENGES AND BENEFITS

Although employing big data analytics for control and optimization of wireless networks is very attracting to the network operators, it comes with some challenges. The process of managing and leveraging of a massive amount of data, designing algorithms for dynamic and efficient processing of sizable data sets and then exploiting the insights from the data analytics in networks can pose unique challenges. The prime concerns for the network operators emerge from the extent of effort, skills, and workforce needed to manage and operate a big data platform. However, the most critical and challenging task is more likely to stem from the loss of direct control that the network operators still have over the wireless network. The loss of direct control is incurred from the combination of automation and real-time operations within the big data analytics framework. However, the considerable complexity of the next-generation networks makes the automation inevitable, and handover or relinquish that level of direct control is imperative. On top of these, a substantial investment is necessary.

Despite all the challenges, the network operators are more considerate towards data analytics platform since the challenges are outweighed by the benefits. The big data analytics infuses efficiency into the provisioning of services and end-to-end network. Analytics facilitates the network operators to gain from the better planning, increased utilization of network resources, efficient maintenance of the network elements and lower operating costs. It gives the network operators the flexibility to define and execute their network utilization strategy. It helps the operators to make new service and offer plans that are suited to subscribers' needs. Although the network operators are already performing these kinds of service provisioning, analytics delivers richer insights. Analytics helps to improve subscriber management and policy implementation. With the aid of natural language processing and interfacing with the smart digital assistants in the user devices, autonomous customer care can be facilitated. Another benefit the network operators get by employing analytics is differentiation, which is compelling to strengthening an network operator's market positioning. Analytics can support the network operators to employ new techniques to traffic handling such as network-slicing (i.e., the way to slice the network) and edge-computing (i.e., the way to balance centralized and distributed functionality).

VI. CONCLUSION

We consider a data-driven next-generation wireless network model, where the network operators employs advanced data analytics, ML and AI for efficient operation, control, and optimization. We present the main drivers of big data analytics adoption and discuss how ML, AI and computational intelligence play their important roles in data analytics for next-generation wireless networks. We present a set of network design and optimization schemes with respect to data analytics. Finally, we discuss the benefits and challenges that the network operators encounter in adopting big data analytics, ML, and AI in next-generation wireless networks.

ACKNOWLEDGEMENT

This research was conducted under a contract of R&D for Expansion of Radio Wave Resources, organized by the Ministry of Internal Affairs and Communications, Japan.

The first author would like to thank Monica Paolini from Senza Fili Consulting for the insightful comments.

REFERENCES

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

[2] M. Paolini. *Mastering Analytics: How to Benefit From Big Data and Network Complexity*. Accessed: Nov. 2, 2017. [Online]. Available: http://content.rcrwireless.com/20170620_Mastering_Analytics_Report

[3] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 190–199, Oct. 2015.

[4] *Study on Access Traffic Steering, Switching and Splitting Support in the 5G System Architecture*, document TR 23.793 V0.1.0, 3GPP, Aug. 2017.

[5] S. Han, C.-L. I, G. Li, S. Wang, and Q. Sun, "Big data enabled mobile network design for 5G and beyond," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 150–157, Sep. 2017.

[6] X. Cheng, L. Fang, L. Yang, and S. Cui, "Mobile big data: The fuel for data-driven wireless," *IEEE Intenet Things J.*, vol. 4, no. 5, pp. 1489–1516, Oct. 2017.

[7] X. Cheng, L. Fang, X. Hong, and L. Yang, "Exploiting mobile big data: Sources, features, and applications," *IEEE Netw.*, vol. 31, no. 1, pp. 72–79, Jan./Feb. 2017.

[8] A. Engelbrecht, *Computational Intelligence: An Introduction*, 2nd ed. New York, NY, USA: Wiley, 2007.

[9] O. Acker, A. Blockus, and F. Pötscher. *Benefiting From Big Data: A New Approach for the Telecom Industry*. Accessed: Dec. 5, 2017. [Online]. Available: https://www.strategyand.pwc.com/reports/benefiting-big-data and https://www.strategyand.pwc.com/media/file/Strategyand_Benefiting-from-Big-Data_A-New-Approach-for-the-Telecom-Industry.pdf

[10] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[11] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah. (2018). "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks." [Online]. Available: https://arxiv.org/abs/1710.02913

[12] S. A. Kyriazakos and G. T. Karetsos, *Practical Radio Resource Management in Wireless Systems*. Boston, MA, USA: Artech House, 2004.

[13] Procera Networks. *RAN Perspectives: RAN Analytics & Enforcement*. Accessed: Oct. 13, 2017. [Online]. Available: https://www.proceranetworks.com/hubfs/Resource%20Downloads/Datasheets/Procera_DS_RAN_Perspectives.pdf?t=1481193315415

[14] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the limit of predictability in human mobility," *Sci. Rep.*, vol. 3, p. 324, Nov. 2013.

[15] R. Atawia, H. S. Hassanein, and A. Noureldin, "Fair robust predictive resource allocation for video streaming under rate uncertainties," in *Proc. IEEE Globecom*, Dec. 2016, pp. 1–6.

[16] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 56–61, Jun. 2011.

[17] A. Banerjee, "Advanced predictive network analytics: Optimize your network investments & transform customer experience," Heavy Reading, New York, NY, USA, White Paper, Feb. 2014.

[18] W. Hong *et al.*, "Multibeam Antenna Technologies for 5G Wireless Communications," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6231–6249, Dec. 2017.

[19] E. J. Black *et al.*, "Broadband surface scattering antennas," U.S. Patent 0 187 123 A1, Jun. 2017.

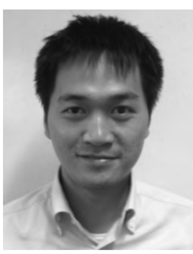
[20] E. J. Black *et al.*, "Methods and systems for communication with beam-forming antennas," U.S. Patent 0 373 181 A1, Dec. 2016.

[21] T. O'Shea, and J. Hoydis. (2017). "An introduction to deep learning for the physical layer." [Online]. Available: https://arxiv.org/abs/1702.00832

[22] C.-l. I, Q. Sun, Z. Liu, S. Zhang, and S. Han, "The big-data-driven intelligent wireless network: Architecture, use cases, solutions, and future trends," *IEEE Veh. Technol. Mag.*, vol. 12, no. 4, pp. 20–29, Dec. 2017.

[23] Q. Lijun, Z. Jinkang, and S. Zhang, "Survey of wireless big data," *J. Commun. Inf. Netw.*, vol. 2, no. 1, pp. 1–18, 2017.

MIRZA GOLAM KIBRIA (S'11–M'14) received the B.E. degree from Visveswaraiah Technological University, India, in 2005, the M.Sc. degree from the Lund Institute of Technology, Lund University, Sweden, in 2010, and the Ph.D. degree from Kyoto University, Japan, in 2014, all in electrical engineering. In 2014, he joined the National Institute of Information and Communications Technology, where he is currently a Researcher with the Wireless Systems Laboratory, Wireless Network Research Center, Yokosuka Research Park, Japan. His research interests include resource allocation optimization, wireless signal processing, shared spectrum access communications, small cell networks, and stochastic geometry. He was a recipient of the Japanese Government (Monbukagakusho, formerly known as Monbusho) Scholarship for his Ph.D. study. He was a recipient of the IEICE WBS Student Paper Award in 2013, the IEEE WPMC Best Paper Award in 2015, and the Young Researcher's Encouragement Award from the Japan chapter of the IEEE Vehicular Technology Society in 2012.



KIEN NGUYEN (SM'16) received the B.E. degree in electronics and telecommunication from the Hanoi University of Science and Technology, Vietnam, in 2004, and the Ph.D. degree in informatics from the Graduate University for Advanced Studies, Japan, in 2012. In 2014, he joined the National Institute of Information Communications Technology, Japan, as a Researcher. His research interests include novel, software-based, and evolvable networking technologies for the next generation of mobile networks and Internet of Things. He is a member of the IEICE.

GABRIEL PORTO VILLARDI (S'07–M'09–SM'12) received the B.E. degree in electrical engineering with emphasis in telecommunications from the Federal Center of Technological Education of Rio de Janeiro, Rio de Janeiro, Brazil, in 2002, and the M.E. and Ph.D. degrees (Hons.) as a Japanese Government (Monbukagakusho) Scholar in physics, electrical, and computer engineering from Yokohama National University, Yokohama, Japan, in 2006 and 2009, respectively. In 2009, he joined the National Institute of Information and Communications Technology, where he is currently a Senior Researcher with the Wireless Systems Laboratory, Yokosuka Research Park, Yokosuka, Japan. His current research interests span several areas in wireless communications, such as communications theory, statistical modeling, multiple-input multiple-output, physical-layer design for white space cognitive radios, space–time codes, fault tolerance, and energy reduction issues in sensor networks. He has been actively contributing to the IEEE 802.19.1, IEEE 802.22b, IEEE 802.11af, and IEEE I900.6a task groups since 2010, toward the standardization of TV white space and cognitive radio technologies. He is currently a Voting Member of the IEEE 802.22, IEEE 802.11, IEEE 802.15, and IEEE 802.19 standardization working groups. He has been serving as the Secretary for the IEEE 802.22 Working Group and the IEEE 802.22b Task Group since 2014. From 1999 to 2000, he received the Coordination for the Improvement of Higher Education Personnel/Institute of International Education Scholarship to pursue his studies at Clemson University, Clemson, SC, USA.

OU ZHAO received the B.E. degree in electronics and communication engineering from the Nanjing University of Posts and Telecommunications, China, in 2005, and the M.E. and Ph.D. degrees in electronic engineering from the Graduate School of Informatics, Kyoto University, Kyoto, Japan, in 2014 and 2016, respectively. He is currently a Researcher with the Wireless Networks Research Center, National Institute of Information and Communications Technology. His major research interests include resource allocation, wireless signal processing and its hardware implementation, machine learning, and big data analytics. He received the Young Researcher's Encouragement Award from the Japan chapter of the IEEE Vehicular Technology Society in 2013.

KENTARO ISHIZU received the Ph.D. degree in computer science from Kyushu University, Japan, in 2005. He has been working with National Institute of Information and Communications Technology (NICT), Japan, for 13 years. He has been dedicated to R&D on heterogeneous wireless networks, cognitive radio systems, and TV white space systems. He has been leading NICT's TV white space trials at various locations in the world. He is currently managing the TV white space projects of NICT. One of the developed systems were sent to the disaster area of the great earthquake in the eastern Japan on March 11, 2011, and contributed to recover from the damage of network access environment. He also has been involved in wide area of international standardizations, including the IEEE 1900.4, IEEE 802.11, and IEEE 802.21.

FUMIHIDE KOJIMA (S'96–M'99) received the B.E., M.E., and D.E. degrees from Osaka University, Osaka, Japan, in 1996, 1997, and 1999, respectively, all in electrical communications engineering. He is currently the Director with the Wireless Systems Laboratory, Wireless Network Research Institute, National Institute of Information and Communications Technology, Yokosuka, Japan. In 1999, he joined the Communications Research Laboratory, Ministry of Posts and Telecommunications, where he has been involved in research on various topics, such as intelligent transportation systems, radio-over-fiber multimedia transmissions, mobile ad hoc emergency networks, wireless grid systems (smart utility networks), and medium access control protocol for communications systems.