REVOLUTION OF SELF-ORGANIZING NETWORK FOR 5G MMWAVE SMALL CELL MANAGEMENT: FROM REACTIVE TO PROACTIVE

Jun Xu, Junmei Yao, Lu Wang, Kaishun Wu, Lei Chen, and Wei Lou

ABSTRACT

The mmWave frequency band has been widely accepted as an ideal carrier for the next-generation ultra-dense small cell communications to achieve the ambitious objective of 1000-fold increase in network capacity, but such a novel architecture also invokes tremendous anxiety for operators regarding the planning, configuration, maintenance, optimization, and troubleshooting. SON, introduced about 10 years ago, aiming to affiliate network management through addressing complexity and decreasing cost in an automatic way, nowadays has been extensively considered the only viable way to enable small cell deployments; however, as a technology traditionally designed and standardized for LTE-based systems, the conventional paradigm will no longer be suitable for mmWave-based cellular networks, which contain some peculiar characteristics, such as directivity of signal transmission and sensitivity to blockage and mobility, differing from their low-frequency counterparts and thereby bringing numerous unprecedented challenges. To smoothly migrate toward the 5G era, the legacy SON campaign eagerly demands a thorough revolution to think over the new features and cope with the new problems by defining new functionalities and proposing new algorithms. It is important, but not easy. In this article, we share some preliminary insights about the mmWave-oriented proactive SON paradigm, hopefully spurring further research in this area and accelerating its practical application in 5G.

INTRODUCTION

The fifth generation (5G) is coming! On December 21, 2017, the Third Generation Partnership Project (3GPP) successfully ratified the first 5G New Radio (NR) standard, which was an essential milestone to 5G deployment and laid a solid foundation for a global unified 5G system on a global market scale. 5G NR will be a revolutionary change compared to the prior generations of mobile technology. In addition to human-to-human connection with higher performance, more importantly, it is devoted to connecting everything for supporting more scenarios such as enhanced mobile broadband (eMBB), ultra-reliable and low latency communications (ULLRC), and massive machine type communications (mMTC). This is an ambitious vision, and while 3GPP took the first significant step, it merely completed the 5G non-standalone (NSA) mode, so there are still lots of innovations that need to be explored further.

Among the multiple objectives that the future 5G NR expects to realize, the most attractive one is no doubt the need to increase the network capacity by 1000-fold over that of the current Long Term Evolution (4G/LTE) standard, to satisfy the ever growing mobile traffic demand over the next 10 years. The general consensus is that this target can be hit through combined gains from three aspects: super broad bandwidth, extreme network densification, and massive multiple-input multiple-output (MIMO). On this basis, millimeter-wave (mmWave) technology, as a well-known enabler to facilitate high-speed transmission, breaks into the next-generation commercial communication systems from the early static point-to-point links. With a large amount of unused continuous spectrum in the mmWave band from 30 to 300 GHz, not only is the multi-gigabit data rate achievable, but the form factor of the antenna element also shrinks due to the millimeter-scale wavelength, which enables integration of a large antenna array into one device in favor of massive MIMO implementation. Besides, the mmWave coverage is naturally limited since the higher frequency suffers from severe free space propagation loss, whose positive side effect is to be able to reduce the inter-cell interference when the base stations (BSs) are deployed in an ultra-dense fashion. All these characteristics make mmWave both the ideal candidate to cater for the 5G performance requirements, and the best carrier to apply in the small cell deployments. Henceforth, it can be safely predicted that mmWave small cell networks will constitute a key ingredient in overcoming the 1000-fold improvement challenge associated with 5G.

However, new things always tend to generate new problems, not excepting the mmWave small cells. The most significant issue comes down to the management of the novel cellular systems. In the past, mobile systems were primarily managed by manpower for daily operations, including new site planning, fine-tuning of network parameters, and handling of outages. But later, as the node density has grown rapidly to address the exponential evolution of user traffic, especially with the advent of the small cell concept, the leg-

Digital Object Identifier:
10.1109/MWC.2018.1700420

Jun Xu, Junmei Yao, Lu Wang, and Kaishun Wu (corresponding author) are with Shenzhen University; Jun Xu and Lei Chen are with Hong Kong University of Science and Technology; Wei Lou is with Hong Kong Polytechnic University.

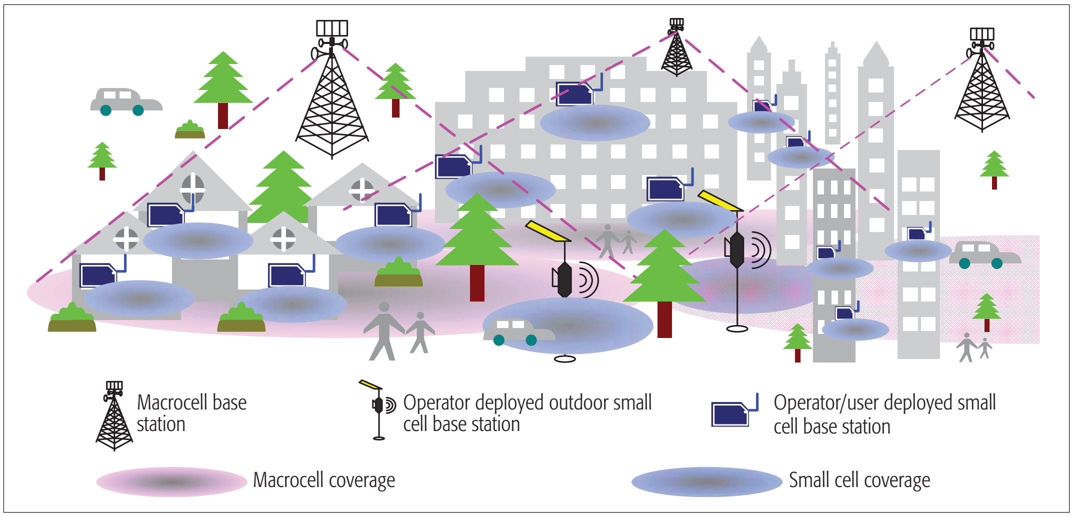

FIGURE 1. Future small cell deployment. For CAPEX saving, ordinary users are encouraged to deploy small cell BS in indoor facilities such as homes, offices, malls and so on. Combining with the operator deployed outdoor small cell BS and macro cell BS, they form a heterogeneous and dynamic network architecture.

SON itself is not a groundbreaking innovation. Its original emergence was in wireless sensor networks, ad hoc networks and automatic computing networks, where it had performed very maturely for a long time. Even in the context of wireless cellular networks, in the past 10 years, many outstanding research efforts were made at exploiting its full potentia.

acy management approaches resorting to manual efforts have gradually become inefficient for two reasons. The first one is due to the huge number of (BSs), which is an inevitability of ultra-dense small cell deployments and creates many workloads requiring manipulation. Correspondingly, the pressure from the capital and operational expenditure (CAPEX and OPEX) for the operators will substantially rise since they have to pay much more money for expensive skilled labors. The second one is attributed to the sharply increasing network complexity. The future system will exploit heterogeneous network (HetNet) architecture, mixing both macrocells and small cells with multiple radio access technologies (RATs) for extended coverage and capacity; in particular, some of the small cells will be deployed by users in an impromptu fashion with plug-and-play features, as illustrated in Fig. 1. The kinds of heterogeneity and dynamicity bring up substantial network parameters and various operational activities, which even go beyond the human capability to handle them well. Against this backdrop, understanding the desires of the operators for cutting cost and improving efficiency, some organizations such as the Next Generation Mobile Networks (NGMN) and 3GPP have successively proposed a new management mechanism, the self-organizing network (SON), to make the planning, configuration, maintenance, optimization, and healing of the future communication systems simpler and faster. Nowadays, it has been extensively accepted as the only viable way to achieve optimal performance in an economic manner to enable small cell deployments.

SON itself is not a groundbreaking innovation. Its original emergence was in wireless sensor networks, ad hoc networks, and automatic computing networks, where it has performed very maturely for a long time. Even in the context of wireless cellular networks, in the past 10 years, many outstanding research efforts were made to exploit its full potential, and even some rewarding industrial practices have been successfully used in the LTE system [1]. *Can we directly migrate these existing precious frameworks and algorithms to the mmWave small cells?* Regretfully, the answer may be negative. The radical reason is that the legacy SON mainly exploited a reactive paradigm, not responding until some events occur, which can work well in low-frequency systems, but not for the mmWave scenarios. On one hand, the original intention of exploiting mmWave is to achieve ultra-high-speed data transmission and create a perception of zero latency for the typical 5G applications including vehicle-to-vehicle, online gaming, virtual reality, uncompressed video, and so on; obviously, the response time caused by the reactive mechanism cannot be avoided in physics, and will degrade the quality of experience (QoE) and quality of service (QoS), not to mention the intolerable expenses for responding channel changes induced by the intrinsic mmWave transmission characteristics. On the other hand, recent progress in deep sensing, machine learning, and big data analytics provide us the capabilities of both inferring environmental information and predicting future trends, which opens the chances of taking proactive measures in advance to compensate for the time overhead troubling the mmWave landscape. Therefore, to support 5G mmWave small cell management, it is not only necessary for but also capable of revolutionizing the legacy SON paradigm from *reactive* to *proactive*.

The revolution is disruptive, embracing both gigantic opportunities and huge challenges, but rare pertinent contributions can be found in the literature. Although some works involve SON for 5G, to the best of our knowledge, they are not really aimed at dealing with mmWave-based systems. On the eve of the 5G rollout, it is important to develop a comprehensive understanding of mmWave small cells, and explore the key techniques toward the proactive SON paradigm. In this article, after describing the background and discussing the motivation, we share some preliminary insights about this topic, followed by a survey about the contemporary state-of-the-art designs. We point out several open issues and potential research directions before the conclusion is drawn.

Background and Motivation

The mmWave small cell network is not a simple upgrade of the microwave predecessor, by increasing additional spectrum or substituting it with advanced radio technologies. It exhibits some peculiar transmission traits, violating the prerequisite long existing in traditional SON designs. In this section, we first present the new features of mmWave communications, and then explain the legacy SON definitions, accompanied by remarks about why a revolution is needed.

Fundamentals of mmWave Communications

The most prominent feature that mmWave propagation faces is severe path attenuation thanks to the property of ultra-high-frequency radio. This disadvantage is remedied by integrating large-scale steerable antenna arrays as metal patterns on printed circuit boards to achieve high-gain beamforming, which implicates that mmWave communications are essentially directional. Another significant characteristic of mmWave transmission is extreme poor diffraction because millimeter-scale wavelength holds weak capability to diffract around obstructions, which signi-

Release	Work Item	Features
Rel.-8	SA5-SON concepts and requirements	SON concepts and requirements
Rel.-8	SA5-Self establishment of eNBs	Self-configuration
Rel.-8	SA5-SON Automatic Neighbor Relation (ANR) list management	ANR, physical cell identity (PCI) assignment
Rel.-9	SA5: Study of SON related operation and management (OAM) interfaces for home eNBs (HeNBs)	OAM interfaces for HeNBs
Rel.-9	SA5: Study of self-healing	Self-healing
Rel.-9	SA5: OAM aspects of automatic radio network configuration data preparation	Automatic radio network configuration data preparation
Rel.-9	SA5: OAM aspects of self-organization management	Self-optimization, comprised of mobility robustness optimization (MRO), mobility load balancing (MLB), inter-cell interference cancellation (ICIC)
Rel.-9	RAN3: Self-organizing networks	Coverage and capacity optimization (CCO), MRO, MLB, random access channel optimization (RACH opt.)
Rel.-10	SA5: Self-optimization continuation	Self-coordination, self-optimization (MRO, MLB, ICIC, RACH opt.)
Rel.-10	SA5: Self-healing management	Cell outage detection (COD), cell outage compensation (COC)
Rel.-10	SA5: Energy saving in radio networks	Energy saving (ES)
Rel.-10	RAN2-3: LTE SON enhancements	CCO, ES, MLB, MRO enhancements
Rel.-11	SA5: SON management	SON management
Rel.-11	SA5: LTE SON coordination management	SON coordination
Rel.-11	SA5: Inter-RAT ES management	AM aspects of ES management
Rel.-11	RAN3: Further SON enhancements	MRO, minimization drive test (MDT) enhancements
Rel.-12	SA5: Enhanced network management (NM) centralized CCO	Enhanced NM centralized CCO
Rel.-12	SA5: Multi-vendor plug and play eNB connection to the network	Multi-vendor plug-and-play eNB connection to the network
Rel.-12	SA5: Enhancements of distributed MLB	OAM aspects of distributed MLB
Rel.-12	SA5: Energy efficiency related performance measurements	Energy efficiency related performance measurements
Rel.-12	SA5: Het-Net management/OAM aspects of network sharing	HetNet/network sharing
Rel.-12	RAN2-3: Next generation SON for evolved UMTS terrestrial radio access network (EUTRAN)	SON per user equipment (UE) type, active antennas, small cells
Rel.-12	RAN2-3: ES enhancements for EUTRAN	ES
Rel.-13	RAN2-3: Enhanced NM centralized CCO	CCO
Rel.-13	SA5: Study on distributed MLB SON function	MLB
Rel.-14	RAN: SON for active antenna system (AAS)-based deployments	ES

TABLE 1. Evolution of SON in 3GPP.

fies that mmWave communications are inherently sensitive to blockage.

Thus, to realize successful communication with mmWave between transmitter and receiver, their beams have to align at first. This procedure is referred to as beam training, and is generally triggered either at initial access or after blockage and mobility occur due to directivity. It is worthwhile to mention that searching for the best beam pairs is a time-consuming behavior. Although camping on each beam direction only takes a few seconds, the overhead increases almost quadratically with antenna array dimension (e.g., 48.4 ms for a 16 × 16 phased-array antenna and 785.7 ms for a 32 × 32 one [2]); besides, beam training is often invoked only at scheduled time slots, and as such, has to wait long after the triggering events arise. Thus, when mmWave communications are applied in mobile environments, random blockage and human mobility are apt to break off the link for a comparatively long time before beam training is finished, thereby causing very bad QoE and QoS to those latency-sensitive applications.

But the directivity also brings an extra merit. Through exploiting angle diversity, mmWave can effectively avoid the co-channel interference, and hence drive mmWave-based communication systems to enter the noise-limited regime rather than the interference-limited domain.

In a nutshell, directivity is the most distinguishing feature for mmWave signal transmission, while sensitivity to blockage and mobility is the most severe challenge for its application in mobile cellular systems, and interference-free is the biggest advantage except for ultra-high speed when it is used to construct a communication network.

What Is Legacy SON?

In brief, SON is a network management automation technology to address complexity and decrease expenditure. Compared to the past human-based methodology, SON primarily aims to reduce human intervention in network design and eliminate human error in network operation.

SON is an umbrella term, including a series of functionalities that could be applied in various aspects for different goals. According to the phases of the life cycle of a cellular system, they can be categorized into the following three parts.

- Self-configuration: the ability of the network to configure itself automatically when nodes are added, deleted, or modified
- Self-optimization: a recurring and automated process for the dynamic tuning of network parameters for optimal performance in changing conditions
- Self-healing: a function used in automatic compensation of network node failures to restore service where it has been degraded

SON was first identified by NGMN as one of the key design principles for next-generation networks in 2008. Later, 3GPP also introduced it in LTE Release 8, and continuously expanded it until the latest Release, 14. Table 1 lists its evolution in terms of the 3GPP Work Items (WIs), including the related features [1]. It can be seen that 3GPP has taken heterogeneity, dynamicity, and small cells into consideration since Release 12, but still under the confinement of LTE systems.

Why Is Legacy SON Unfit for mmWave Small Cells?

Ultra-low latency is the most essential appeal for using mmWave small cells. But the practical use confronts two-fold challenges. For one thing, the legacy SON methodology employs a reactive mechanism to network operations, that is, the optimization and healing functionalities kick in only after some problems have been detected through drive tests, customer complaints, or operation and management (OAM) reports [3]. This process will induce inescapable delay, which violates the latency requirement of the mmWave communications. For another thing, the vulnerability to both blockage and mobility implied in the mmWave transmission characteristics also hinders its application in small cells. Consequently, both factors demand proactive approaches to satisfy user QoE through addressing the latency difficulties by taking actions beforehand.

In addition, mainstream legacy SON optimizations such as MLB, MRO, ICIC, CCO and so on, mainly utilize the methods like adjusting transmit power and antenna tilt, as well as channel borrow, and so on. Nevertheless, due to the capability of directional transmission with pencil-beam in mmWave communications, the concept of cell boundary becomes blurred, and the interference problems are fewer, thereby causing these classical optimization methods to not be effective anymore [4]. As a result, it is also necessary to find new solutions for QoS enhancement based on the peculiar traits of mmWave.

Revolution of SON for MmWave Small Cells

The above analysis has clarified that a revolution of SON is imperative for adapting to the characteristics and requirements of mmWave small cells. The next question is: *how do we implement a proactive SON?* Instead of the time-consuming procedure of observing, diagnosing, and triggering, the new paradigm must be able to anticipate the potential situation beforehand and process it online. This can be done by either tracking the user behavior through the smart sensing hardware embedded in the user equipment (UE) or inferring the network-level intelligence from the immense volume of traffic data to predict the risk in its infancy, and then take preemptive actions to resolve the issue before it occurs. Correspondingly, smart sensing, big data analytics, and machine learning are the key enablers to transform SON from being reactive to proactive. Although there are some works that also propose to apply proactive SON in 5G ultra-dense networks [5], the mmWave characteristics are not fully exploited in those studies. In this section, we focus on the techniques that are exclusive to mmWave small cells. First, the proactive mobility and blockage management mechanism is discussed. Second, a big data empowered proactive SON framework is introduced, and we put the emphasis on the mmWave-oriented self-optimization methods. Last, a case study regarding machine learning based cell outage detection is elucidated.

Proactive Mobility and Blockage Management

The issue brought by occasional mobility and blockage in mmWave communications is closely related to the time-consuming beam training procedure. Numerous beam training strategies have

In a nutshell, directivity is the most distinguishing feature for mmWave signal transmission, while sensitivity to blockage and mobility is the most severe challenge for its application in mobile cellular systems, and being interference-free is the biggest advantage except for ultra-high speed when it is used to construct a communication network.

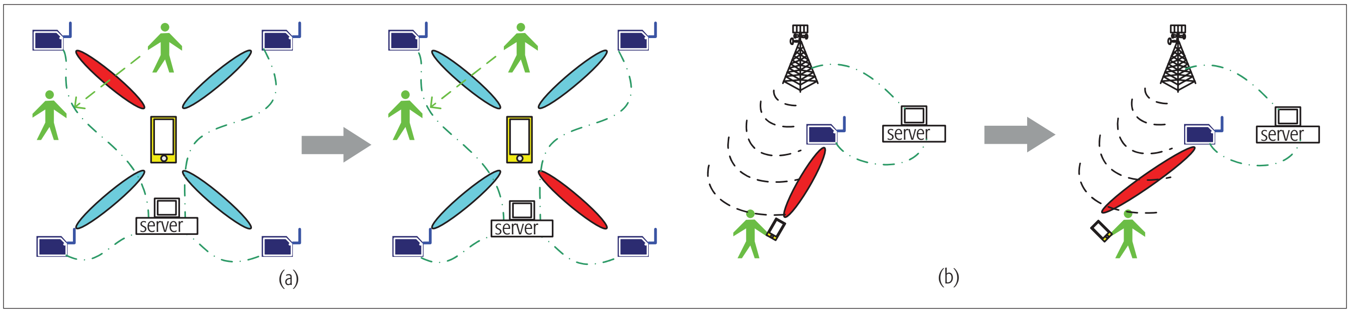

FIGURE 2. Proactive blockage and mobility management: a) multiple beams are preserved during initial access phases, but only one is selected for data transmission. If a blockage occurs at the primary link, another candidate beam will be activated immediately; b) the sensors embedded in the UE will report its location and orientation to the macro BS through omnidirectional signals. Then the server will proactively derive the best beam pairs in the mmWave small cell according to the assistance information.

been proposed to mitigate the problem, but none of them can implement the perception of zero latency, since they are always reactive to channel changes, not affording any preventive mechanism to reduce the likelihood of link outage. An ideal solution should be able to find a perfect deployment method that can make an mmWave link survive under blockage and mobility, but it is either unrealistic or too tricky for ordinary users if taking the complicated environmental factors into account. Therefore, a set of proactive schemes would be the second best solution for this biggest challenge holding mmWave back in cellular systems. But it is hard to overcome it using only one BS. There are two potential solutions to address this problem, as shown in Fig. 2, and both of them have to be executed under the coordination of SON.

The first solution is through multiple-BS association. In traditional cellular networks, one UE is associated with only one BS, in general depending on which BS can bring the strongest received signal strength to it. When it detects a stronger BS, or the channel quality falls below some threshold, it will disconnect with the current BS and reconnect with another one. This procedure is called handover, and is inherited from off-the-shelf mmWave standards such as IEEE 802.11ad and IEEE 802.15.3c. However, working at a low frequency band, handover can bring less overhead because the BS broadcasts its existence in an omnidirectional way, which can guarantee that the UE receives the system information in a timely manner, which is not the case for mmWave communications due to beam training. Hence, this solution lets the UE register at multiple BSs in advance at the initial access phase, but only select one to transmit data, so even if there are some obstacles occasionally occurring on the link, it can quickly activate the next link through the previously prepared beam pairs with another BS. Figure 2a depicts this scheme.

The second solution is through out-of-band assistance. Beam training in mmWave systems is fulfilled by exchanging control messages, which usually occupy less load compared to data transmission. According to the characteristics of directional transmission, before the aligned beam pairs are found, the acquisition of the control messages is also blind. To fight against this drawback, we can utilize the various sensors embedded in the UE to infer its location and orientation, and then use the HetNet concept and dual connectivity, that is, employing the microwave working frequency band of the macrocells, to upload this information to the SON engine to speed up beam training by reducing blindness. Because other radio band rather than mmWave alone is used, it is called out-of-band assistance. Figure 2b illustrates such an example. In this way, UE mobility can be tracked quickly, at the expense of multiple transceivers being integrated into one device, increasing hardware complexity and energy consumption.

The embryo of these ideas can be found in some state-of-the-art designs. Haider *et al.* [6] devised Beam Sounding in 2016, a short control-frame exchange prior to data transmissions, to proactively search for alternate fail-over sectors in advance, which can potentially be used upon link breakage. Sur *et al.* [7] proposed BeamSpy in the same year to predict the quality of alternative beams by only inspecting the channel response of the current beam used by the receiver. It is a model-driven framework, mainly leveraging two fundamental properties in mmWave communications: channel sparsity and spatial correlation. And in 2017, Sur *et al.* [8] again presented WiFi-assisted 60 GHz wireless networks. To resist the timing overhead of the reactive interface switching approach used in off-the-shelf multi-band devices, the authors designed a novel platform as an IEEE 802.11-compliant system to provide high-speed and robust connectivity over 60 GHz and WiFi multi-band devices in dynamic indoor environments. Identically, in the same year, Wei *et al.* [9] proposed Pia, a pose information assisted platform, to implement seamless coverage and mobility support, which uses the ideas of both multi-BS association and out-of-band assistance; the same authors also realized a framework, E-Mi [10], to facilitate robust mmWave network deployment by sensing ambient reflectors. Adari *et al.* [11] invented low-cost electronic mirrors to implement cordless virtual reality. They used those mirrors to replace the multiple high-end mmWave BSs, but the idea is similar.

Big Data Empowered Proactive SON Framework

Another more generic way to offer proactive capabilities in SON for 5G is with the help of big data analytics. Figure 3 describes such a framework. In the following, we illustrate the components of the framework step by step.

Data Collection: This step is used to gather management data from the mmWave small cell networks. Nowadays, operators or third-party over-the-top (OTT) companies can exploit the

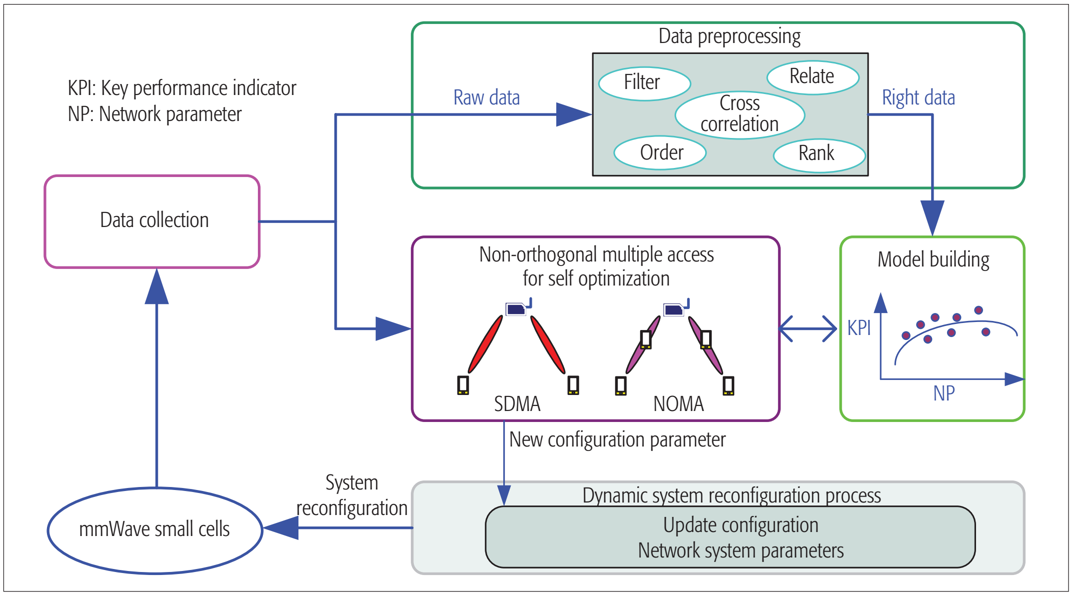

FIGURE 3. A big data empowered proactive SON framework.

Cell outage detection is regarded as one of the fundamental functionalities in self-healing. Its purpose is to autonomously detect cells in an outage state, where cells are inoperable and cannot provide any service due to hardware or software failures as well as wrong configurations.

powerful data acquisition platforms to collect almost all wireless data from ordinary users, the radio access network (RAN), the core network (CN), and service providers [5]. In this example, the data may include the amount of users per cell, user locations, and so on.

Data Preprocessing: This is also called data cleaning. Unprocessed massive data are called raw data, which can contain useless or redundant information. Through ranking, filtering, ordering, relating, and cross-correlating, these raw data can be transformed into useful correct data. On the basis of these correct data, the key performance indicator (KPI) can be extracted and correlated with the corresponding network parameter (NP). For the convenience of post-processing, multiple KPIs can be merged through dimension reduction techniques.

Model Building: The correct data are utilized by this component via statistics analysis to derive the functional relationship, known as the model, between the KPI and the NP. The statistics analysis approach can be regression, classification, or clustering. To fit the dynamic change in 5G small cells, the model built here is also needed to dynamically adjust coefficients in an automatic manner.

Self-Optimizing Algorithm: This is part of the SON engine and receives the model outcomes as inputs. As mentioned in the previous section, due to the blurring of the cell boundary in mmWave small cells, some legacy optimization methods are obsolete. Instead, some novel approaches can be used to improve the network overall performance. For example, multiple access (MA) technologies are very important for each generation of cellular networks to enhance capacity, but the traditional orthogonal styles cannot exert full potential due to the directional attribute of the mmWave band. As a result, non-orthogonal MA technologies become more appropriate. The typical methods include space-division multiple access (SDMA) [13], which enables multiple UEs from different directions to transmit data simultaneously, and power domain non-orthogonal multiple access (NOMA) [14], which allows different users to share channels at the same time even when they transmit in the same directions. Whether SDMA or NOMA is selected, the key problem is the scheduling algorithm. As shown in Fig. 3, the required information for the scheduling algorithm comes from the model result.

Dynamic System Reconfiguration Process: The new configuration parameters can be injected into the cellular system to update configurations and improve the system performance. Note that it is an online method, so low latency can be guaranteed compared to the offline processing mechanism used in the legacy SON paradigm.

MACHINE LEARNING BASED CELL OUTAGE DETECTION

Cell outage detection is regarded as one of the fundamental functionalities in self-healing. Its purpose is to autonomously detect cells in outage state, where cells are inoperable and cannot provide any service due to hardware or software failures as well as wrong configurations. Cell outage leads to decreased capacity and coverage gap, but detecting outaged cells is nontrivial due to the lack of a triggering mechanism. Moreover, it is hard for the network management functions to detect an outage cell directly if the outage is caused by wrong configurations. It often takes hours or days before successful detection, and in some cases it remains unknown until customer complaints are filed.

This problem becomes more stringent in mmWave applications because ordinary blockage can degrade the cell performance in addition to the other factors. Some researchers have proposed using learning machines to fulfill the job. But machine learning tools need large amounts of mobile traffic data, especially control data or signaling data. Regretfully, in mmWave systems, the off-the-shelf devices are usually black boxes about the physical layer information, which prevents the employment of a machine learning approach.

Scalabrin *et al.* [15] recently proposed an interesting solution to the issue. They found that the

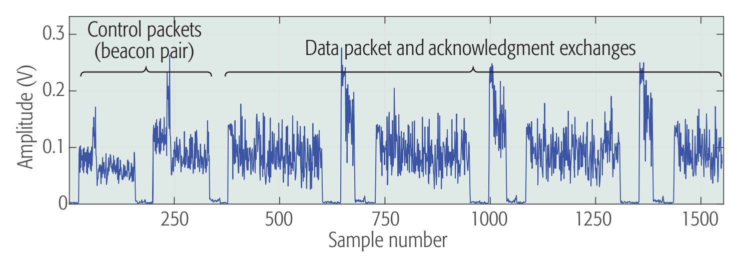

FIGURE 4. Energy trace example of a data burst starting with a pair of beacons and ending with the acknowledgment exchanges.

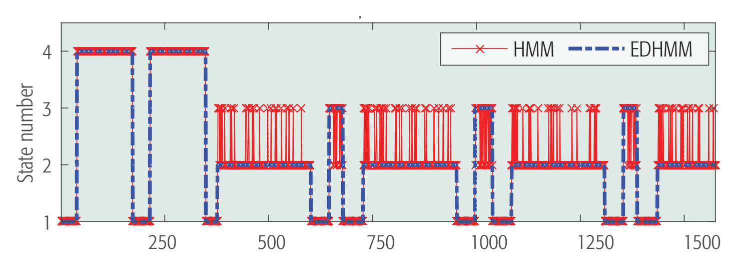

FIGURE 5. Validating result comparison between HMM and EDHMM.

energy trace can reflect the network behavior, as shown in Fig. 4, where the data transmission always starts with a pair of beacons and ends with acknowledgment exchanges. This information can infer the network dynamics. For example, a missing Acknowledgment after a data packet indicates a deafness issue [4], while overlapping packet frames hint at a collision. Therefore, through installing some sniffers in the network environments, this information can be collected for further analysis. The authors use a combination of template training and an explicit duration hidden Markov model (EDHMM) to train the model, and the learning result is validated in Fig. 5. It shows that the proposed machine learning framework can correctly identify all frames in the trace such as data packets, Acknowledgments, beacons, and inter-frame spacing. Moreover, it also proves the necessity of the EDHMM approach. The normal HMM method wrongly classifies many of the samples. In contrast, the EDHMM classifies all samples correctly, even in the case of varying data packet lengths.

OPEN ISSUES AND RESEARCH DIRECTIONS

The mmWave serves as a new radio; meanwhile, SON is also a developing technology in cellular networks. Their combination generates many open issues. We choose three of them as research directions.

MACHINE LEARNING BASED SON DESIGN

With the prosperity of machine learning in recent years, many SON researchers have sensed that it is a powerful tool to be used in self-optimization and self-healing. However, it faces three challenges. First, there are too many machine learning models, including *K*-nearest neighbor (KNN), support vector machine (SVM), artificial neural network (ANN), principal component analysis (PCA), model-based and model-free reinforcement learning, deep learning networks, and so on; how to make proper selections in accordance with different objectives is a big problem. Second, all machine learning based design methodologies depend on enormous amounts of data to accomplish training and verification; but where can we find these real data? Operators may be reluctant to open them, and even if they are willing, not all physical data is available at the chip level, especially for mmWave chipsets. Therefore, sometimes we have to install some sniffers to capture air signals, which also increases the difficulty. Third, all exploited machine learning based algorithms must be able to converge in reasonable time; otherwise, the QoS or QoE will be impacted. However, that is not easy.

SELF-COORDINATION IN 5G SON

When concurrently running in the same network, different SON functionalities tend to cause parametric and objective-based conflicts, which will deteriorate the overall gains of SON. Hence, self-coordination among different SON functions are required to ensure stable network operation, but so far it remains an under-addressed problem even for 4G [5]. This phenomenon will be aggravated in the 5G era given the complexity of the envisioned network architecture, so the analysis of potential conflicts brought by the huge numbers of SON functionalities and the design of an applicable self-coordination framework are really challenging.

The efficiency of the self-coordination framework is related to the SON deployment. The 3GPP has proposed different architectures for different SON implementations, ranging from centralized SON (C-SON) to distributed SON (D-SON). If C-SON is selected, SON functions are deployed in the operation and maintenance center (OMC) or in the network management system (NMS)as part of the operation and support system (OSS). Its advantage is embracing global information about KPIs and powerful computing resources to run complicated optimization algorithms, and its disadvantage is longer delays due to the slow response. If, instead, D-SON is used, where the SON functions are distributed in the control plane across the edges of the network, typically in the BSs, the interaction between the SON function and the local SON coordinator will be via internal vendor-specific interfaces, with much lower latency characteristics. From the perspective of 5G mmWave small cells, the D-SON architecture is more appealing, but it can only obtain sub-optimal SON gain since it can only see limited information and computing resources. To get a better trade-off between gain and latency, the hybrid deployment is worth further research.

SON EMULATOR

Currently, although operators have great desire for SON, they also have huge concerns that such a disruptive technology could generate enormous revenue losses if not properly matured. For example, in France, a recent 24-hour network outage cost $13.5 million and $27 million in repairs and compensation to customers, respectively, so it is a legitimate subject for operators. How to enhance their confidence so that SON can really be put into practice is an important issue Maybe an emulator would be the best answer. We suggest that this is an urgent research direction.

CONCLUSION

In this article, we discuss the vision of mmWave small cells in the future 5G era, and present the peculiar features in mmWave communications. Then we claim that the legacy SON design philosophy, which is oriented to low-frequency cellular networks, will not be suitable for the mmWave small cell system any longer due to the different channel characteristics; therefore, a revolution is demanded to adapt to the new radio and new network. We share our preliminary insights about how to achieve a proactive SON that are exclusive to mmWave small cells, and investigate some state-of-the-art designs. Finally, we discuss some important open issues and possible research directions. We hope that our work can pave the way toward making SON technology reach a practical solution in 5G mmWave small cell management.

ACKNOWLEDGMENTS

This research was supported in part by China NSFC Grants 61472259, 61601308, 61702343, and 61502313, Guangdong Natural Science Foundation 2017A030312008, Guangdong Provincial Science and Technology Development Special Fund Project no. 2017A010101033, Shenzhen Science and Technology Fund nos. JCYJ20170302140946299 and JCYJ20170412110753954, Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China (Grant no.161064), Guangdong Talent Project 2014TQ01X238, 2015TX01X111, and GDUPS (2015).

REFERENCES

[1] J. Moysen *et al.*, "From 4G to 5G: Self-Organized Network Management Meets Machine Learning," 2017; arXiv:1707.09300v1.

[2] S. Sur *et al.*, "BeamSpy: Enabling Robust 60 GHz Links Under Blockage," *Usenix Conf. NSDI*, 2016, pp. 193–206.

[3] B. Keshavamurthy *et al.*, "Conceptual Design of Proactive SONs Based on the Big Data Framework for 5G Cellular Networks: A Novel Machine Learning Perspective Facilitating a Shift in the SON Paradigm," *IEEE Conf. System Modeling and Advancement in Research Trends*, 2017, pp. 298–304.

[4] H. S. Ghadikolaei *et al.*, "Millimeter Wave Cellular Networks: A MAC Layer Perspective," *IEEE Trans. Commun.*, vol. 63, no. 10, 2015, pp. 3437–58.

[5] I. Ali *et al.*, "Challenges in 5G: How to Empower SON with Big Data for Enabling 5G," *IEEE Network*, vol. 28, no. 6, 2014, pp. 27–33.

[6] M. K. Haider *et al.*, "Mobility Resilience and Overhead Constrained Adaptation in Directional 60 GHz WLANs: Protocol Design and System Implementation," *ACM Intl. Symp. MobiHoc*, 2016, pp. 61–70.

[7] S. Sur *et al.*, "BeamSpy: Enabling Robust 60 GHz Links Under Blockage," *Usenix Conf. NSDI*, 2016, pp. 193–206.

[8] S. Sur *et al.*, "WiFi-Assisted 60 GHz Wireless Networks," *ACM Intl. Symp. MobiCom*, 2017, pp. 28–41.

[9] T. Wei *et al.*, "Pose Information Assisted 60 GHz Networks: Towards Seamless Coverage and Mobility Support," *ACM Intl. Symp. MobiCom*, 2017, pp. 42–55.

[10] T. Wei *et al.*, "Facilitating Robust 60 GHz Network Deployment By Sensing Ambient Reflectors," *Proc. Usenix Conf. NDSI.*

[11] O. Adari *et al.*, "Enabling High-Quality Untethered Virtual Reality," *Proc. Usenix Conf. NSDI*, 2017, pp. 531–44.

[12] S. Han *et al.*, "Big Data Enabled Mobile Network Design for 5G and Beyond," *IEEE Commun. Mag.*, vol. 55, no. 9, Sept. 2017, pp. 150–57.

[13] C. Fan *et al.*, "Learning-Based Spectrum Sharing and Spatial Reuse in mm-Wave Ultra Dense Networks," *IEEE Trans. Vehic. Tech.*, 2017. DOI 10.1109/TVT.2017.2750801.

[14] R. Ruby *et al.*, "Enhanced Uplink Resource Allocation in Non-Orthogonal Multiple Access Systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, 2018, pp. 1432–44.

[15] M. Scalabrin *et al.*, "Millimetric Diagnosis: Machine Learning Based Network Analysis for mm-Wave Communication," *Proc. IEEE Symp. WoWMoM Networks*, 2017, pp. 1-9.

BIOGRAPHIES

JUN XU received his Ph.D. and M.E. degrees in computer architecture from the Institute of Computing Technology, Chinese Academy of Sciences in 2011 and 2005, respectively. He is currently a research associate in the College of Computer Science and Software Engineering, Shenzhen University (SZU), China, and a postdoctoral fellow in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology (HKUST), China. His research interests include artificial intelligence, big data analytics, wireless communication, network management, computer architecture, and mobile computing.

JUNMEI YAO received her B.E. degree in communication engineering from Harbin Institute of Technology, China, in 2003, her M.E. degree in communication and information systems from Harbin Institute of Technology, China, in 2005, and her Ph.D. degree in computer science from Hong Kong Polytechnic University in 2016. She is currently a research associate in the College of Computer Science and Software Engineering, SZU. Her research interests include wireless networks, wireless communications, and RFID systems.

LU WANG received her B.S. degree in communication engineering from Nankai University, Tianjin, China, in 2009, and her Ph.D. degree in computer science and engineering from HKUST in 2013. She is currently an assistant professor with the College of Computer Science and Software Engineering, SZU. Her research interests include wireless communications and mobile computing.

KAISHUN WU received his Ph.D. degree in computer science and engineering from HKUST in 2011. After that, he worked as a research assistant professor at HKUST. In 2013, he joined SZU as a distinguished professor. He has co-authored two books and published over 90 high-quality research papers in international leading journals and premier conferences, like IEEE TMC, IEEE TPDS, ACM MobiCom, and IEEE INFOCOM. He is the inventor of six U.S. and over 60 Chinese pending patents. He received the 2012 Hong Kong Young Scientist Award and the 2014 Hong Kong ICT Award: Best Innovation, and the 2014 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award.

LEI CHEN [M] received his B.S. degree in computer science and engineering from Tianjin University, China, in 1994, his M.A. degree from the Asian Institute of Technology, Thailand, in 1997, and his Ph.D. degree in computer science from the University of Waterloo, Canada, in 2005. He is now an associate professor in the Department of Computer Science and Engineering, HKUST. His research interests include crowdsourcing on social networks, uncertain and probabilistic databases, web data management, multimedia and time series databases, and privacy.

WEI LOU is currently an associate professor with the Department of Computing, Hong Kong Polytechnic University. He has been involved intensively in designing, analyzing, and evaluating practical algorithms on a theoretical basis, and building prototype systems. His research interests include wireless networking, mobile cloud computing, mobile ad hoc and sensor networks, and smart city technologies. His research work is supported by several Hong Kong GRF grants and NSFC grants.

From the perspective of 5G mmWave small cells, the D-SON architecture is more appealing, but it can only obtain sub-optimal SON gain since it can only see limited information and computing resources. To get a better trade-off between gain and latency, the hybrid deployment is worth further research.