

Special Issue on Wireless Big Data

Wireless big data: transforming heterogeneous networks to smart networks

Yudi Huang, Junjie Tan, Ying-Chang Liang*

University of Electronic Science and Technology of China, Chengdu 611731, China

*Corresponding author, email: ycliang@uestc.edu.cn

Abstract: In HetNets (Heterogeneous Networks), each network is allocated with fixed spectrum resource and provides service to its assigned users using specific RAT (Radio Access Technology). Due to the high dynamics of load distribution among different networks, simply optimizing the performance of individual network can hardly meet the demands from the dramatically increasing access devices, the consequent upsurge of data traffic, and dynamic user QoE (Quality-of-Experience). The deployment of smart networks, which are supported by SRA (Smart Resource Allocation) among different networks and CUA (Cognitive User Access) among different users, is deemed a promising solution to these challenges. In this paper, we propose a framework to transform HetNets to smart networks by leveraging WBD (Wireless Big Data), CR (Cognitive Radio) and NFV (Network Function Virtualization) techniques. CR and NFV support resource slicing in spectrum, physical layers, and network layers, while WBD is used to design intelligent mechanisms for resource mapping and traffic prediction through powerful AI (Artificial Intelligence) methods. We analyze the characteristics of WBD and review possible AI methods to be utilized in smart networks. In particular, the potential of WBD is revealed through high level view on SRA, which intelligently maps radio and network resources to each network for meeting the dynamic traffic demand, as well as CUA, which allows mobile users to access the best available network with manageable cost, yet achieving target QoS (Quality-of-Service) or QoE.

Keywords: wireless big data, cognitive radio network, network function virtualization, software defined network, machine learning, smart radio allocation, cognitive user access

Citation: Y. D. Huang, J. J. Tan, Y.-C. Liang. Wireless big data: transforming heterogeneous networks to smart networks [J]. Journal of communications and information networks, 2017, 2(1): 19-32.

1 Introduction

Future wireless communications are expected to support various emerging applications, including high speed mobile Internet, massive IoT (Internet-of-Things), industrial automation networks, unmanned vehicle networks, etc. Each application has different QoS (Quality-of-Service)/QoE (Quality-of-Experience) requirements in terms of transmission rate, coverage, capacity, reliability, latency, etc. To support such diverse and stringent requirements, it

is necessary to deploy different types of wireless networks in the same coverage area, such as Wi-Fi, UMTS (Universal Mobile Telecommunications System), 4G LTE (Long Term Evolution), and future 5G (5th Generation) networks, which form HetNets (Heterogeneous Networks). In such HetNets, each network is assisted with one or more specific RATs (Radio Access Technologies) operating in fixed radio spectrum, and there is usually no cooperation among different networks. While advanced technologies can be applied to improve the capacity of in-

dividual networks, a typical dilemma of such non-cooperative HetNets is that the capacities for each network are typically pre-determined but the traffic loads are distributed unevenly and dynamically, which makes it difficult for HetNets to meet the challenging demand from the dramatically increasing access devices, the consequent upsurge of data traffic, and user QoS/QoE.

In this paper, a framework to transform HetNets to smart networks is proposed for meeting the challenging demands of future wireless communications, by leveraging WBD (Wireless Big Data), CR (Cognitive Radio) and NFV (Network Function Virtualization) techniques.

In the past decades, CR^[1] technique had been developed for different wireless networks to share the radio spectrum in a dynamic manner. On the other hand, the development of SDR (Software Defined Radio) technology makes it possible for radio devices to have programmable and reconfigurable RATs^[2]. Besides the reconfigurability in spectrum domain and physical layer, programmable network layer using NFV has also been proposed^[3]. Through NFV, the network resources can be sliced into virtual resources to support different applications dynamically, e.g., massive IoT, vehicle networks and industrial wireless networks. By transferring hardware inflexibility into software reconfigurability, CR and NFV techniques can make radio resources and network functions effectively programmable.

While CR and NFV support resource slicing in spectrum, physical layer, and network layer, intelligent mechanisms for resource mapping and traffic demand prediction are highly desirable in order to maximize the potential of the smart networks. In this paper, WBD is used to design the intelligent mechanisms through using powerful AI (Artificial Intelligence) methods. Here, WBD is used to describe the “big data” resource specified in the field of wireless communications and networks, including traffic flows generated in devices to infrastructure communications, device to device communications, observed data from sensor networks, network traffic data collected in BS (Base Station), location in-

formation about mobile users, information related to device types. The computer science communities have found that patterns can be mined from the data if it is large enough and further the optimal decisions can be made^[4]. In wireless communications and networks, sensing techniques and large-scale computing infrastructures nowadays make it possible to obtain and process a variety of big data. The developed intelligent mechanisms will provide us the guidelines for effectively utilizing radio spectrum, air interfaces, as well as network resource. The study unlocks the power of knowledge from massive and heterogeneous data collected in HetNets and to apply the information to transform big data from burden to opportunities.

Despite the promise of WBD for HetNets, two related fundamental problems need to be addressed. The first one is how to design HetNets-specific AI methods for transforming raw data into actionable insight. AI methods such as machine learning are sensitive to the distribution and characteristics of data, thus general learning algorithms without taking into account domain knowledge may not be effective when applied to WBD. The second problem is how to evolve existing system to data-driven system. The implementation of AI methods in HetNets will inevitably introduce extra communication overhead for collecting distributed data. Also, some machine learning algorithms are time-consuming for latency constraint wireless applications, thus direct deployment of data-driven design into existing HetNets may introduce additional problems.

In this respect, in the proposed framework to transform HetNets to smart networks, WBD engine is the brain to control the operations of CR and NFV, as shown in Fig. 1, and an on-line/off-line hierarchical scheme is adopted for the tradeoff between latency constraints and computational complexity. For better showing the potential of WBD, SRA (Smart Resource Allocation) and CUA (Cognitive User Access) are investigated in details.

The rest of the paper is organized as follows. In section 2, a brief overview of enabling techniques for smart networks including CR, NFV and AI methods

is given. In section 3, we define WBD and give possible solutions to handle it based on its characteristics. In section 4, the framework of the smart networks is given in detail. The open problems of designing the smart networks will be discussed in section 5. Finally, we conclude the paper in section 6.

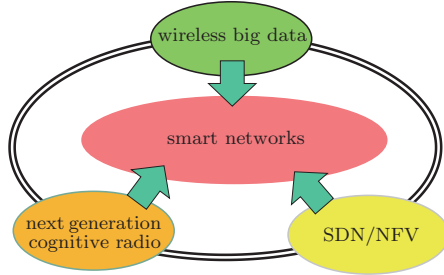


Figure 1 Enabling technologies of the smart networks

2 Brief overview of enabling techniques for smart networks

In this section, the enabling technologies for smart networks are briefly reviewed, including CR, NFV, and data-driven AI methods.

2.1 CR

CR was first proposed in Ref. [5], and became popular after FCC (Federal Communications Commission) recommended it as a promising solution for supporting dynamic spectrum access^[6]. Since then, a lot of researches have been carried out^[7-9]. Here, we focus on two CR paradigms related to HetNet: OSA (Opportunistic Spectrum Access) and CSA (Concurrent Spectrum Access). In the OSA model, a spectrum band can be reused by secondary systems only if it is idle. In the CSA model, the spectrum band is shared by primary system and secondary system while secondary system has to obey certain constraints for protecting the primary system.

The cognitive cycle of CR consists of spectrum awareness, analysis, decision, and spectrum exploitation. The loop will be repeated until the system fully reaches the optimal state. Spectrum awareness is powered by spectrum sensing^[10], which aims to

tell whether the specific spectrum is idle. Analysis and decision focus on the solution of the most effective resource allocation^[11]. Although CR has been investigated for over a decade, practical imperfection is unavoidable^[12]. The imperfection mainly lies in the channel/noise uncertainty, signal uncertainty, noise/channel correlation and transceiver design. To cope with the problem brought by uncertainty, some robust schemes, such as eigenvalue based spectrum sensing, have been proposed, and a good survey is given by Sharma et al.^[12].

Reconfiguration of operational parameters, such as waveform, transmitting power, modulation, carrier frequency, bandwidth, coding, etc. is the enabler for CR. Self-adaptive reconfiguration of operational parameters are implemented so as to reuse the available spectral opportunities^[2]. In Ref. [13], SDAI (Software Defined Air Interface) is proposed as a framework of 5G air interface. In SDAI, modules and related parameters are adapted in real time to flexibly fit all sorts of applications for 5G, and to realize agility and efficiency.

2.2 NFV

Wireless networks are supported by RAN (Radio Access Network) and CN (Core Network), where the RAN is in charge of user access and related service for each base station while the CN controls the interconnection among base stations and Internet. In traditional wireless networks, physical equipments such as routers, base stations and switches are coupled heavily with the software running on them. Such coupling makes it difficult to upgrade services through updating evolving software or network protocols. To solve this problem, SDN (Software Defined Network) which utilizes virtualization technology to abstract away the softwares originally running on physical devices is proposed to offer a new paradigm to manage networking services^[3]. The main purpose of SDN is to decouple the physical network forwarding devices and the CPFs (Control Plane Functions)^[14].

NFV has been promoted and advocated by

many industrial organizations^[15], and standardization organizations^[16]. NFV for wireless networks^[17] is different from wired networks. Ref. [18] proposes a substrate for virtualizing wireless resource in cellular networks. Based on that, Ref. [19] proposes a new framework called wireless virtualization to slice the resource for wireless communication. Specifically, IEEE 802.11 based virtualization, 3GPP LTE based virtualization and IEEE 802.16 based virtualization have their own challenges and solutions^[19,20]. Ref. [3] presents a good survey on this topic.

2.3 AI methods

Machine learning is the core topic of artificial intelligence. The machine learning model is expected to abstract hidden knowledge from the observed data, and then the model can be used to predict the unknown data with the learned knowledge. The data utilized to train the model is defined as training data. The most remarkable difference between machine learning and traditional algorithmic approaches lies in the generation of models. Machine learning is an data-driven method, for its models are entirely determined by data^[21]. From the angle of training modes, machine learning techniques can be classified into supervised learning, unsupervised learning and reinforcement learning. In supervised learning, each training data has a label as the supervisor to indicate the ground truth. On the contrary, data in unsupervised learning does not have the label. Though the data in reinforcement learning also has no label, its model can interact with environment, which generates feedback data to train the model.

Much research has been done on applying machine learning to wireless networks for intelligence. Ref. [22] proposes a construct called knowledge plane (KP). In that paper, the author stated that the traditional network carries the data without knowing what is transmitted, making the network can only make low level decisions on packets. Thus, network should have an AI techniques based KP. The main purpose of embedding KP into network is to maintain a high level view of users' purpose in order to re-

alize fault diagnosis, automatic reconfiguration, support for overlay networks and intrusion detection. In Ref. [23], a wide range of machine learning algorithms for KP are surveyed. According to the formulation of algorithms, Ref. [23] categorizes the machine learning algorithms into classification & regression, acting & planning and interpretation & understanding. Based on Ref. [23], Ref. [24] proposes the definition of cognitive networks, and explains that the cognitive networks should utilize hybrid observations from different elements of network to make decisions and to output actions that can be implemented to globally optimize the networks. Ref. [25] surveys the existing applications of machine learning algorithms on CRN (Cognitive Radio Networks).

However, there are some problems with the literatures. First, the existing application of AI techniques is within a single network, which is not suitable for complex heterogeneous networks. Second, existing work focuses on applying machine learning algorithms in CRN but there is little research on CR-specified AI methods. NLP (Natural Language Processing), CV (Computer Vision) and speech recognition are most successful applications of machine learning algorithms, however, all these three fields design new domain-specified algorithms which inversely influence the machine learning community, such as CNN (Convolutional Neural Network) in CV and LSTM (Long Short-Term Memory) in NLP. Up to now, no CR specified AI tool is designed.

3 Wireless big data

WBD can be seen as the big data specified in the field of wireless networks, which comes from a wide range of sources. Specifically, WBD contains, e.g., traffic flow generated in device to infrastructure communication, device to device communications, observed or sensed data from sensor networks, network traffic data collected in base station and stakeholders, GPS information about human and device, power consumption of communications and sensing, and some other hardware related information, as shown in Fig. 2.

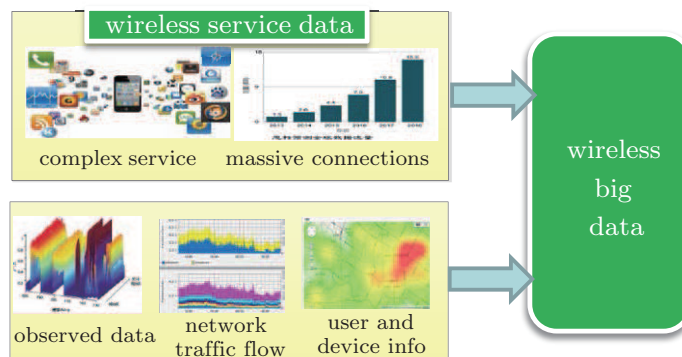


Figure 2 Definition of wireless big data

WBD implies informative hidden knowledge about the wireless environment and can help to meet the challenges in today's wireless networks if utilized correctly. Spectrum data collected by wireless sensor network, for example, can provide deeper understanding about coverage map of base station^[26]. Also, WBD can help to predict the future^[27]. When a user is predicted to handover to another base station using wireless localization, resource provisioning can be made to occur in the new base station. Further, WBD can help to design next generation wireless networks. Through comprehensive analysis of WBD from all aspects of wireless networks, some unseen problems can be identified for performance improvement.

However, big data analysis in wireless networks is a typical multi-type multi-objective task, which is totally different from that in CV and NLP. For example, the data from different base stations or different stakeholders can be significantly diverse. To better understand and utilize WBD, we will investigate it from the perspective of decentralization, heterogeneity, imperfectness as well as highly temporal-spatial variation.

3.1 Decentralized generation and storage

Distributed storage and parallel analysis in data center have been well studied in data science since the scale of data is far beyond the capability of single server^[28]. In the fields like CV and NLP where big data has played a significant role, data is collected

and dumped into hundreds of wired servers and further analyzed in parallel.

Different from CV and NLP scenarios, WBD is generated and stored by lots of separate entities. Unfortunately, to collect a full copy of the data is quite costly as the related entities are even not all connected with a wired link. The data exchange could be expensive^[29]. Therefore, current big data based machine learning techniques are not fully suitable for analyzing WBD and thus a distributed learning technique is highly desirable. It has been studied that the accuracy of distributed learning and the corresponding required communication overhead are interrelated^[30]. The granularity and structure are the tunable components to find an appropriate trade-off between communication overhead and learning accuracy. For example, the traffic flow patterns can be learned in a cloud data center by collecting all flow data from each base station. Or, the network flows can also be learned preliminarily in base station and only learned results are sent to data center. The former solution can absolutely learn better but at a higher cost.

Although densely deployed sensors can bring higher accuracy, it is too expensive so that sensors in wireless monitoring network have to be sparsely implemented. Reconstructing accurate wireless environment from sparse sensors in a wide area is of significant importance for WBD^[26]. Matrix completion^[31] is a powerful tool to recover the matrix with only several known elements by convex optimization. Therefore the wireless environment can

be reconstructed without using densely deployed sensors.

3.2 Heterogeneity

Data generated by different entities in the wireless networks differs a lot. For example, historical billings of a specific user and estimation of channel information are respectively generated in user's equipment and base station. However, billings can tell if it is a heavy or important user. Therefore, by combining the channel information and historical billings, the network has the ability to decide the most suitable network for the user. Besides, data from one entity also has apparent difference. Let us take the user device for example, the streaming video flow looks independent from its historical locations. Conversely, with the track, network can predict the moves of devices and provide proactive handoff and video pre-caching for better experience.

Some methods have been proposed to tackle the problem of heterogeneous data fusion^[32]. Tensor based learning^[33] and multi-view learning^[34] are two most promising ways among them. Tensor is the matrix with higher order. Compared with the matrix based methods, tensor based approaches better sketch the data when it comes from more than two views. As for multi-view learning, it is originally designed to use unlabeled data to enhance the performance of supervised learning^[21], and further extended recently to handle the multi-source learning problem^[35]. The consensus principle and the complementary principle are two basic idea in multi-view learning. The former implies that results from different views should agree with each other to improve the learning performance. The latter indicates that some complementary information can be extracted from multiple views for using as much information as possible.

3.3 Incompleteness and inconsistency

Deficiency in completeness and consistency of data evidently causes inaccuracy to decision making^[36], which definitely lowers the QoE of users. Imperfect

data mostly results from the various data providers with diverse reliability^[37] as well as the denial of accessing data due to, e.g., privacy protection.

For the first reason, robustness can be ensured by building trust relationships^[22]. Credence values are determined according to the past accuracy or other standards so that the uncertainty in low-trust data should be fixed. The conflicting or missing data and errors can thus be fixed by the reliable sources. For instance, in the smart networks scenario where the system has to obtain the real time system load of CDMA (Code Division Multiple Access) system for resource allocation, base station reports its load with an interval while the traffic of gateway is real-time, thus credit of actual system load for gateways should be higher than that for base stations. As for the second reason, it is an emerging problem incorporating multiple networks held by several stakeholders only concerning their own interest. Service providers want to collect data as much as possible while users are in the opposite for the privacy sake and even creates their own cloud platforms through such as OwnCloud. And for the companies which already have user data, it is also difficult for data sharing because legally this needs the authorization of users and the cost of data gathering also lowers the willingness of sharing. This problem remains to be solved and changed from a zero-sum game to a win-win game, demanding the participation of every company and individual involved.

However, there has been literature trying to solve this problem by adapting to the environment of limited information. Ref. [38] aims to integrate HetNets by reinforcement learning with limited feedback and noisy information. Game theory based learning and sequential decision making are also investigated in Refs. [39] and [40] respectively for CRN where players can hardly obtain complete and instantaneous knowledge of the network.

3.4 Temporal-spatial variation

Intelligent decision making in the smart network demands the real-time training and responds from

WBD analysis system. Real-time big data analysis in fact has been investigated long before. Ref. [28] demonstrates a structure of real-time big data system and advocates the principle of design, Ref. [41] suggests to apply it to build a smart city. Nonetheless, the data in WBD, such as spectrum information, often has drastic variation of probability distribution as the change of time and space. Chaos will be caused if the training data is simply scaled up as the traditional way does. Hence, applying the same mechanism in Ref. [28] to build a WBD based decision making system induces severe degradation of performance^[42].

Of course, the problem can be possibly solved in tabular manner, in which data is categorized and carefully labeled for different scenarios. However, complex wireless environment is hard to enumerate, and it is worth noting that labeling the data is time-consuming and extremely costly. Two bright ways to work out this problem are transfer learning^[42] and incremental learning^[43]. Transfer learning believes that though data distribution varies, some high level knowledge remains the same. Thus these schemes try to turn the historical data into a transferable or incremental knowledge by off-line training. With new arriving data, the system is able to obtain new knowledge to adapt to the new environment in a short period of time. It will become the core technique for the WBD-driven decision making.

3.5 Streaming feature

The behavior of wireless networks is often modeled as the Markov process^[44], implying that data originated in wireless networks is not independent but correlated in stream manner. Further, phenomena in wireless networks always simultaneously affect multiple data streams, which implies the designed algorithm for single data stream is not effective for extracting complete information about an event.

Time series mining^[45] and forecasting^[46] are two powerful tools for finding frequent patterns of data stream. Ref. [47] proposed a PCA (principle component analysis) based algorithm to find correlations

in multiple data stream. Tensor based algorithms^[48] are also suitable for multiple streaming mining.

4 The framework of smart networks

Smart networks are expected to provide the following four key features: efficient resource utilization, dynamic network optimization, intelligent service provisioning and enhanced user experience, and these features are achieved by WBD engine in conjunction with CR and NFV techniques. The WBD based smart decision making guides CR techniques and SDN enabled NFV to reconfigure radio resource, infrastructure resource, as well as network functions.

WBD engine is actually an adaptive big data engine specialized for wireless big data due to its special characteristics. It consists of four important functions, namely wireless data acquisition, wireless data management, wireless data analysis and smart service provisioning. Wireless data acquisition concerns the deployment of data collecting agents for reporting real time wireless environment as well as the storage of the data. Wireless data management is responsible for transforming unorganized and unstructured data from various sources, like spectrum data and traffic flows, into structured and unified data in convenience of analysis. Wireless data analysis provides the hidden information of WBD and helps to make intelligent decisions for smart service provisioning.

Logical structure, action stream of smart networks and the deployment of WBD engine are shown in this section for detailed description of smart networks' framework. In the end, the cases of SRA and CUA are exhibited to demonstrate the power of smart networks.

4.1 Logical structure of smart networks

Smart networks are logically decomposed into application component, network component and resource component, as shown in Fig. 3. The application component is connected to network component by access controller. Similarly, the resource component

provides service to network component through resource controller. Application component consists of all the users to access smart networks, ranging from UE (User Equipment) to IoV (Internet of Vehicles). Since different types of users have different QoE demands, the users in application component are clustered based on their QoE demand and corresponding resource requirement. For example, applications in industrial automatic control system often require the communication latency less than 10 ms, while the amount of data to be transmitted is relatively small. In contrast, multimedia streaming applications are latency tolerant but bandwidth consuming. Also, users can switch from one cluster to another.

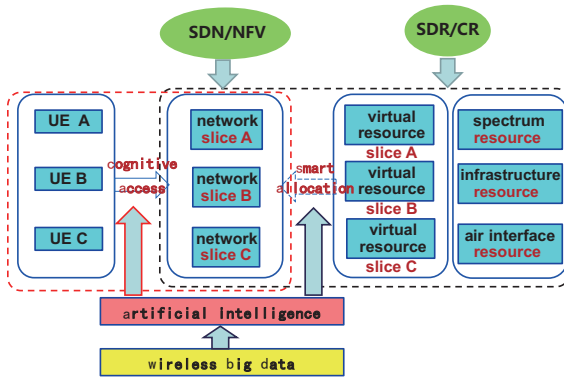


Figure 3 Composition of smart networks

Before the start of the detailed discussion of resource component, we first clarify the meaning of accessing to wireless network. If a device intends to get the wireless service, an AP (Access Point) and the corresponding radio frequency as well as the time slots are then all chosen by the device. Next, a routing path^[49] from the chosen AP to target service is constructed. Thus the access request demands not only the AP and radio resource, but also network functions, routers, switches and corresponding bandwidth in backhaul and fronthaul. The resource component contains radio resource and infrastructure resource. The radio resource is sliced in both time domain, frequency domain, and spatial domain for further multiplexing. While infrastructure resource consists of AP, antennas, processor hardware and

routers. When the resource controller between resource component and network component receives the request from networks, corresponding resource will be appropriately provided to each network in the form of virtual resource slice.

Network component bridges the gap between physical resource and users. Due to NFV, CPFs (Control Plane Functions) in network component are abstracted away from originally coupled forwarding devices into a logically centralized network function controller. The CPFs of wireless network can be categorized into RAN functions and CN functions. The CPFs in RAN consist of admission control, inter and intra cell radio resource management, radio resource scheduling and handoff management, etc., while mobility management, charging rules function, packet gateway etc. are the counterparts in CN. Because of different latency constraints of different applications, network function controller is hierarchical to control the latency constraints, ranging from BS controller to NET controller, where NET controller is in charge of QoE provisioning and cooperation within HetNets. Distributed BS controller is on the bottom of the hierarchy for radio resource management, synchronization and access control in BS. The BS controller is commanded by RAT controller, which is responsible for the network to allocate resource to distributed BS. Because the corporation of HetNets can be both centralized and distributed, the NET controller can be a central substrate to give instructions or a set of mechanism for different networks to cooperate. For instance, networks held by the same mobile service provider may have a centralized NET controller, however NET controller for networks from different operators can only be distributed.

WBD engine is a pervasive structure in the smart networks. Roughly speaking, all decision makings in the smart networks are based on the WBD engine, especially in access control and allocation control. It's worth noting that the WBD engine is not a substantiality as a software or a protocol layer, but an embedded data enabled framework within each element of the smart networks. In practical scenarios, WBD engine embedded in access controller

learns the matching policy between user demand and network selection. Also, the matching policy between network demand and resource allocation is learned by the WBD engine in allocation controller. Thus when new access request arrives, these two controllers can quickly react and match the appropriate network and resource. The WBD engine is not a substantial structure but indicates that all elements in the smart networks can leverage data-driven methods to continuously update the knowledge from data, and that is the reason why we call it smart.

4.2 Action stream of smart networks

The actions of smart networks can be concluded as perception, learning, reasoning, memory and adaptation, as shown in Fig. 4. Especially in the perception step, distributed sensors continuously monitor the wireless environment. Besides, GPS traces and network flows are also stored for further learning. Then the WBD engine combines existing knowledge and the data to be processed in the learning step. After learning from historical data, WBD engine is capable of giving instructions for networks to automatically adapt. These actions again generate new data as part of WBD for network to perceive and learn.

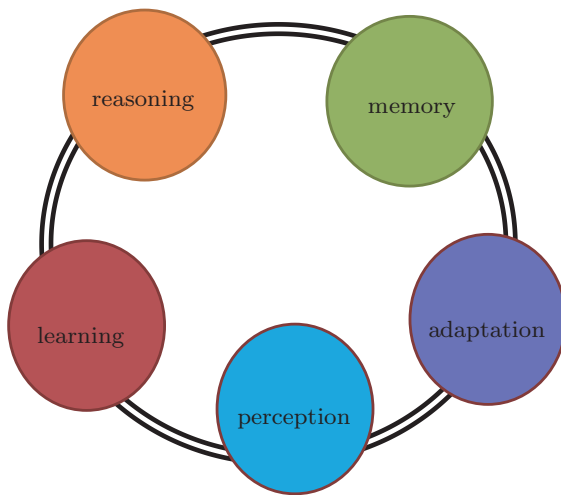


Figure 4 Action stream of smart networks

In a scenario where the UE and BS both sense

the wireless environment and take actions to interact with each other, data is generated during interactions and is learned by WBD engine, and then the UE and BS controllers remember the knowledge about matching relations between QoE demand, network selection, and corresponding resource allocation. When new users come in, the knowledge in memory will give instructions to both UE and BS for selection process, and again the interactions from new users generate new data for WBD engine to update the knowledge. Through this data-driven methods and adaptive resource allocation, more users can be accessed without decrease of experience, as shown in Fig. 5.

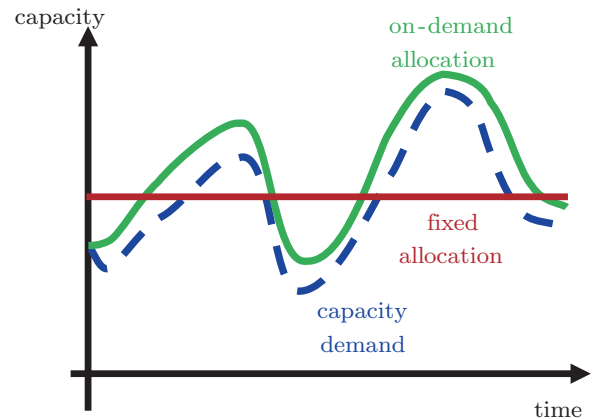


Figure 5 Capacity improvement by on-demand physical resource allocation

4.3 WBD engine

WBD is expected to reveal the intrinsic feature of wireless environment. For example, interference and jamming analysis in hot spot can help better achieve interference cancellation and network selection. Also, health care application can benefit from the analysis of wireless reflected signals. Complex data mining is time-consuming and computation costly, which thus should be placed in cloud centers. However, wireless networks are expected to react and adapt quickly when environment changes. Thus, a on-line off-line hierarchical framework is proposed in the smart networks, as shown in Fig. 6.

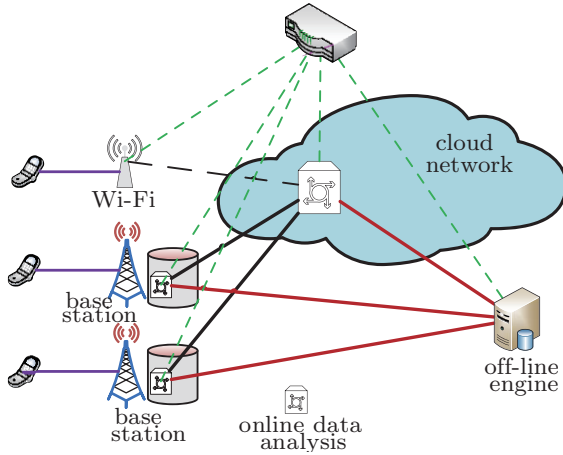


Figure 6 Wireless big data engine framework

In the smart networks, off-line data engine, deployed in cloud center, is responsible for complex analysis to extract knowledge from raw data. While the on-line engine is deployed in the edge of networks for quick reaction to new collected data. The motivation behind the on-line off-line framework is that learning knowledge from raw data is time consuming and stays closely related to the scale of data. Thereafter, it's somewhat reasonable to assume that time complexity of nonlinear data mining algorithm (e.g., the support vector machine^[21]) without prior knowledge is $O(M^2N)$, where M is the number of training data and N is the dimensionality of each data. In wireless big data scenarios, M will always more than 10^7 , and N is larger than 1 000. It can be seen that the time consuming is unacceptable for real time analysis. Although so, with prior information and Teacher-Student interactions, learning speed and required training data can be significantly cut down^[42,50]. Thus, we deploy complex mining algorithm in off-line engine to learn high level knowledge and further migrate it to the on-line engine, where the learning process in on-line engine is speed up to meet the latency constraints with prior knowledge and smaller M . It is worth noting that the off-line engine will not always be running due to the energy consumption and cost constraints, instead, it will operate periodically. The computation in on-line engine heavily depends on the knowledge sent by

off-line engine. To accelerate the reaction time of the smart networks, on-line engine will unify the learned knowledge in off-line engine, as explained in section 3.3, with new arriving data to adapt to the environment. In conclusion, the off-line engine is responsible for complex and time-consuming computation to extract knowledge for on-line engine to update, so that complete computation is not needed in on-line engine.

4.4 Smart resource allocation and cognitive user access

CR and NFV techniques integrate and slice the fixed radio and network resource of HetNets into a resource pool, enabling the dynamic resource allocation among HetNets. Along with the WBD engine, smart networks have the capability to predict the load distribution of HetNets and thus to make proactive resource allocation, which is called SRA (Smart Resource Allocation). For a better demonstration of SRA, we present a simple example of HetNets containing a CDMA system and a LTE system here. The CR techniques investigated by Ref. [51] introduces a scheme for the spectrum sharing between CDMA system and LTE system, which realizes the dynamic radio resource allocation between two networks. In the meanwhile, NFV takes care of the consequent network resource allocation, such as building more instances of control modules like handoff management, mobility management, policy, etc. for the network with extra users and release the redundant instances for the network with reduced users^[20]. However, this scheme requires that some parameters, like the loads of CDMA system^[51], should be obtained in real time. Rather than passively and continuously listening to the delayed reports from base stations, prediction of needed parameters provided by WBD engine are preferred, which helps the networks to make proactive adaptations.

Apart from SRA, CUA is another important feature of smart networks. CUA concerns the problem about which network is the best choice for a smart device who is able to adopt multi-RAT. The opti-

mal network access selection should consider lots of factors, therefore, it is not always wise to select the network with the latest technologies though SRA ensures the scalability of networks. For instance, to simultaneously achieve the required rates, lowest latency and least energy consumption, the smart decision is based on the real-time system parameters like loads, packet loss, signal strength as well as the real-time user devices parameters like battery level, application and location^[52]. However, real-time parameters are quite difficult to obtain and thus the existing schemes mostly focus on trial based learning^[52-54], which is passive and inefficient. With the intelligence provided by WBD through reinforcement learning, smart networks realize CUA by prediction, proactive handoff and proactive resource allocation along with SRA. Compared to the past schemes, CUA shows its advantages in proactivity and efficiency.

5 Challenges and open problems

The proposed smart network is not a complete off-the-shelf solution, a lot of research is needed to make it a reality. In this section, some fundamental and open problems that are important for the success of smart network are discussed.

5.1 Definition of the knowledge

In section 3.3, the knowledge is deemed the key for deploying WBD in practice, however, it is not easy to fully understand the definition of the knowledge. In wireless context, what can be learned and what should be learned are not defined. Also, quantitative relationship between knowledge and accuracy in wireless networks stays unclear. To control the tradeoff between analysis accuracy and communication overhead, how much knowledge that current task needs is essential.

5.2 Tradeoff between analysis accuracy and communication overhead

More data often implies higher learning accuracy, however, they also indicates stringent storage and

transmission requirement. Careful tradeoff between them is fundamental to the success of the smart networks. It is worth noting that different hierarchies and different elements of wireless networks have different accuracy demand and computing power constraints. Formulation and optimization of this problem is a challenging issue.

5.3 Inference rather than correlation

Large portion of research in machine learning community is about how to find correlation or unexplainable results, however, the operators holding wireless networks need not only unexplainable correlation, but also causality that human can understand. Recent criticism on deep learning^[55] says that the DNN (Deep Neural Network) is a black box that human cannot understand. How to design wireless networks specified inference tools is important for the operators and the whole community to adopt data-driven network design.

5.4 Integrating networks from different operators

Incompleteness and inconsistency problem of WBD has been discussed in previous sections and some existing research has pointed out the directions^[56]. Yet, since wireless networks in a coverage area are usually held by several stakeholders with competing business, efficient cooperation between them is difficult from commercial perspective. How to design solutions to persuade competitive operators to cooperate is hard but important for all research on virtualization of wireless networks.

5.5 Legacy support and scalable for future

Existing radio resource allocation scheme is fixed spectrum allocation, thus isolation and slicing of physical resource in the smart networks must support legacy system as much as possible to reduce the expenditure when evolving to the smart networks. Also, the framework of the smart networks should

be scalable to new techniques and even next generation cellular networks.

5.6 Security and privacy

Privacy and security of data collection and processing is important for practical implement. Traditional research on data privacy mainly concerns the protection of data generators during the process in database, and differential privacy technique is the main solution. Because the open environment of wireless communication, privacy and security problem of wireless data processing is far more difficult^[57], and further research is required.

6 Conclusions

To meet the unprecedented requirements of wireless services in terms of data traffic, massive access, and user QoE, in this paper, we proposed an architecture for smart networks with intelligent mechanisms supported by wireless big data and AI methods, in conjunction with cognitive radio and network function virtualization techniques. The network capacity can be boosted using smart resource allocation, while the QoE or QoS of users can also be enhanced using cognitive user access. Possible artificial intelligence methods are reviewed, and some fundamental problems of great importance for future research are pointed out.

References

- [1] J. Mitola, G. Q. Maguire. Cognitive radio: making software radios more personal [J]. *IEEE personal communications*, 1999, 6(4): 13-18.
- [2] L. B. Thiagarajan, S. Attallah, Y.-C. Liang. Reconfigurable transceivers for wireless broadband access schemes [J]. *IEEE wireless communications*, 2007, 14(3): 48-53.
- [3] R. Mijumbi, J. Serrat, J. L. Gorricho, et al. Network function virtualization: state-of-the-art and research challenges [J]. *IEEE communications surveys and tutorials*, 2015, 18(1): 236-262.
- [4] C. Luo, J. Zeng, M. Yuan, et al. Telco user activity level prediction with massive mobile broadband data [J]. *ACM transactions on intelligent systems and technology*, 2016, 7(4): 63.
- [5] J. Mitola III, G. Q. Maguire Jr. Cognitive radio: making software radios more personal [J]. *IEEE personal communications*, 1999, 6(4): 13-18.
- [6] FCC Spectrum Policy Task Force. Report of the spectrum Efficiency Working Group [R]. ET Docket, 2002.
- [7] S. Haykin. Cognitive radio: brain-empowered wireless communications [J]. *IEEE journal on selected areas in communications*, 2005, 23(2): 201-220.
- [8] Y. C. Liang, K. C. Chen, G. Y. Li, et al. Cognitive radio networking and communications: an overview [J]. *IEEE transactions on vehicular technology*, 2011, 60(7): 3386-3407.
- [9] J. Lunden, V. Koivunen, H. V. Poor. Spectrum exploration and exploitation for cognitive radio: Recent advances [J]. *IEEE signal processing magazine*, 2015, 32(3): 123-140.
- [10] A. T. Hoang, Y. Zeng, Y. C. Liang, R. Zhang. A review on spectrum sensing for cognitive radio: challenges and solutions [J]. *EURASIP journal on advances in signal processing*, 2010, 2010(1): 1-15.
- [11] R. Zhang, Y. C. Liang, S. Cui. Dynamic resource allocation in cognitive radio networks [J]. *IEEE signal processing magazine*, 2010, 27(3): 102-114.
- [12] S. K. Sharma, T. E. Bogale, S. Chatzinotas, et al. Cognitive radio techniques under practical imperfections: a survey [J]. *IEEE communications surveys and tutorials*, 2015, 17(4): 1858-1884.
- [13] Q. Sun, I. Chin-Lin, S. Han, et al. Software defined air interface: a framework of 5g air interface [C]//*Proceedings of IEEE Wireless Communications Network Conference*, 2015, 6-11.
- [14] B. Han, V. Gopalakrishnan, L. Ji, et al. Network function virtualization: challenges and opportunities for innovations [J]. *IEEE communications magazine*, 2015, 53(2): 90-97.
- [15] M. Chiosi, D. Clarke, P. Willis. et al. Network functions virtualisation: an introduction, benefits, enablers, challenges and call for action [R]. Darmstadt-Germany: SDN and OpenFlow World Congress, 2012.
- [16] G. ETSI. Network functions virtualisation (nfv): architectural framework [J], *ETSI GS NFV*, 2013, 2(2): V1.
- [17] C. Liang, F. R. Yu, X. Zhang. Information-centric network function virtualization over 5g mobile wireless networks [J]. *IEEE network*, 2015, 29(3):68-74, 2015.
- [18] R. Kokku, R. Mahindra, H. Zhang, et al. Nvs: a substrate for virtualizing wireless resources in cellular networks [J]. *IEEE/ACM transactions on networking*, 2012, 20(5): 1333-1346.
- [19] C. Liang, F. R. Yu. Wireless virtualization for next generation mobile cellular networks [J]. *IEEE wireless communications*, 2015, 22(1): 61-69.
- [20] V. Yazici, U. C. Kozat, M. O. Sunay. A new control plane for 5g network architecture with a case study on unified handoff, mobility, and routing management [J]. *IEEE communications magazine*, 2014, 52(11): 76-85.

- [21] K. P. Murphy. Machine learning: a probabilistic perspective [M]. USA: MIT press, 2012.
- [22] D. D. Clark, C. Partridge, J. C. Ramming, et al. A knowledge plane for the Internet [C]//Proceedings of Special Interest Group on Data Communication, 2003, 3-10.
- [23] T. G. Dietterich, P. Langley. Cognitive networks: towards self-aware networks [M]. USA: Wiley, 2007.
- [24] R. W. Thomas, D. H. Friend, L. A. Dasilva, et al. Cognitive networks: adaptation and learning to achieve end-to-end performance objectives [J]. IEEE communications magazine, 2006, 44(12): 51-57.
- [25] N. Abbas, Y. Nasser, K. El Ahmad. Recent advances on artificial intelligence and learning techniques in cognitive radio networks [J]. EURASIP journal on wireless communications and network, 2015, 2015(1): 1.
- [26] S. Chouvardas, S. Valentin, M. Draief, et al. A method to reconstruct coverage loss maps based on matrix completion and adaptive sampling [C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing , 2016, 6390-6394.
- [27] Z. Lu, G. De Veciana. Optimizing stored video delivery for mobile networks: the value of knowing the future [C]//Proceedings of IEEE Conference on Computer Communications, 2013, 2706-2714.
- [28] N. Marz, J. Warren. Big data: principles and best practices of scalable realtime data systems [M]. USA: O Reilly Media, 2015.
- [29] Y. Zheng, L. Capra, O. Wolfson, et al. Urban computing: concepts, methodologies, and applications [J]. ACM transactions on intelligent systems and technology, 2014, 5(3): 1-38.
- [30] M. F. Balcan, A. Blum, S. Fine, et al. Distributed learning, communication complexity and privacy [C]//Proceedings of Conference on Learning Theory, 2012, 23: 26.1-26.22.
- [31] E. J. Cands, B. Recht. Exact matrix completion via convex optimization [J]. Foundations of computational mathematics, 2009, 9(6): 717-772.
- [32] Y. Zheng. Methodologies for cross-domain data fusion: An overview [J]. IEEE transactions on big data, 2015, 1(1): 16-34.
- [33] E. E. Papalexakis, C. Faloutsos, N. D. Sidiropoulos. Tensors for data mining and data fusion: models, applications, and scalable algorithms [J]. ACM transactions on intelligent systems and technology, 2016, 8(2): 16:1-16:44.
- [34] C. Xu, D. Tao, C. Xu. A survey on multi-view learning [J]. Preprint arXiv:1304.5634, 2013.
- [35] K. Crammer, M. Kearns, J. Wortman. Learning from multiple sources [J]. Journal of machine learning research, 2008, 9(12): 1757-1774.
- [36] O. Kwon, N. Lee, B. Shin. Data quality management, data usage experience and acquisition intention of big data analytics [J]. International journal of information management, 2014, 34(3): 387-394.
- [37] H. Jagadish, J. Gehrke, A. Labrinidis, et al. Big data and its technical challenges [J]. Communications of the ACM, 2014, 57(7): 86-94.
- [38] J. Lundn, S. R. Kulkarni, V. Koivunen, et al. Multiagent reinforcement learning based spectrum sensing policies for cognitive radio networks [J]. IEEE journal of selected topics in signal processing, 2013, 7(5): 858-868.
- [39] B. Wang, Y. Wu, K. R. Liu, et al. Clancy. An anti-jamming stochastic game for cognitive radio networks [J]. IEEE journal on selected areas in communications, 2011, 29(4): 877-889.
- [40] K. Liu, Q. Zhao. Distributed learning in multi-armed bandit with multiple players [J]. IEEE transactions on signal processing, 2010, 58(11): 5667-5681.
- [41] R. Kitchin. The real-time city? big data and smart urbanism [J]. GeoJournal, 2014, 79(1): 1-14.
- [42] S. J. Pan, Q. Yang. A survey on transfer learning [J]. IEEE transactions on knowledge and data engineering, 2010, 22(10): 1345-1359.
- [43] D. A. Ross, J. Lim, R.-S. Lin, et al. Incremental learning for robust visual tracking [J]. International journal of computer vision, 2008, 77(1-3): 125-141.
- [44] K. W. Choi, E. Hossain. Estimation of primary user parameters in cognitive radio systems via hidden markov model [J]. IEEE transactions on signal processing, 2013, 61(3): 782-795.
- [45] T. C. Fu. A review on time series data mining [J]. Engineering applications of artificial intelligence, 2011, 24(1): 164-181.
- [46] N. I. Sapankevych, R. Sankar. Time series prediction using support vector machines: a survey [J]. IEEE computational intelligence magazine, 2009, 4(2): 24-38.
- [47] S. Papadimitriou, J. Sun, C. Faloutsos. Streaming pattern discovery in multiple time-series [C]//Proceedings of International Conference on Very Large Data Bases, 2005: 697-708.
- [48] J. Sun, D. Tao, S. Papadimitriou, et al. Incremental tensor analysis: theory and applications [J]. ACM transactions on knowledge discovery from data, 2008, 2(3): 1-37.
- [49] S. Yu, M. Liu, W. Dou, et al. Networking for big data: A survey [J]. IEEE communications surveys and tutorials, 2016, 19(1): 531 - 549.
- [50] V. Vapnik, R. Izmailov. Learning using privileged information: similarity control and knowledge transfer [J]. Journal of machine learning research, 2015, 16: 2023-2049.
- [51] S. Han, Y. C. Liang, B. H. Soong. Spectrum refarming: a new paradigm of spectrum sharing for cellular networks [J]. IEEE transactions on communications, 2015, 63(5): 1895-1906.
- [52] O. Mammela, P. Mannersalo. Cognitive wireless access selection at client side: performance study of a q-learning approach [C]//Proceedings of IEEE Network

- Operations and Management Symposium, 2014: 1-4.
- [53] H. Tabrizi, G. Farhadi, J. Cioffi. A learning-based network selection method in heterogeneous wireless systems [C]//Proceedings of IEEE Global Communications Conference, 2011: 1-5.
- [54] H. Tabrizi, G. Farhadi, J. Cioffi. Dynamic handoff decision in heterogeneous wireless systems: Q-learning approach [C]//Proceedings of IEEE International Conference on Communications, 2012: 3217-3222.
- [55] M. Jordan, T. Mitchell. Machine learning: trends, perspectives, and prospects[J]. *Science*, 2015, 349(6245): 255-260.
- [56] L. Sorber, M. Van Barel, L. De Lathauwer. Structured data fusion [J]. *IEEE journal of selected topics in signal processing*, 2015, 9(4): 586-600.
- [57] S. Yu. Big privacy: challenges and opportunities of privacy study in the age of big data [J]. *IEEE access*, 2016, 4: 2751-2763.

About the authors



hyd1123camel@gmail.com)

Yudi Huang was born in 1994. He received the B.E. degree in University of Electronic Science and Technology of China (UESTC) in 2016. He is now a master candidate in communication and information system. His research interests include cognitive radio and statistical machine learning. (Email:



Junjie Tan was born in 1994. He received the B.E. degree in University of Electronic Science and Technology of China (UESTC) in 2016. He is now a master candidate in communication and information system. His research interests include cognitive radio and reinforcement learning. (Email: tan@kuspot.com)



Ying-Chang Liang (F'11) [corresponding author] is a professor in the University of Electronic Science and Technology of China (UESTC), China, and also a professor in the University of Sydney, Australia. He was a principal scientist and technical advisor in the Institute for Infocomm Research (I2R), Singapore.

His research interest lies in the general area of wireless networking and communications, with current focus on applying artificial intelligence, big data analytics and machine learning techniques to wireless network design and optimization. Dr

Liang was elected a fellow of the IEEE in December 2010, and was recognized by Thomson Reuters as a highly cited researcher in 2014, 2015 and 2016. He received IEEE ComSoc's TAOS Best Paper Award in 2016, IEEE Jack Neubauer Memorial Award in 2014, the First IEEE ComSoc's APB Outstanding Paper Award in 2012, and the EURASIP Journal of Wireless Communications and Networking Best Paper Award in 2010. He also received the Institute of Engineers Singapore (IES)'s Prestigious Engineering Achievement Award in 2007, and the IEEE Standards Association's Outstanding Contribution Appreciation Award in 2011, for his contributions to the development of IEEE 802.22 standard. Dr Liang is now serving as the chair of IEEE Communications Society Technical Committee on Cognitive Networks, an associate editor of IEEE Transactions on Signal and Information Processing over Network, and an associate editor-in-chief of the World Scientific Journal on Random Matrices: Theory and Applications. He served as founding editor-in-chief of IEEE Journal on Selected Areas in Communications C Cognitive Radio Series, and was the key founder of the new journal IEEE Transactions on Cognitive Communications and Networking. He has been an (associate) editor of IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, and IEEE Signal Processing Magazine. Dr Liang was a distinguished lecturer of the IEEE Communications Society and the IEEE Vehicular Technology Society, and has been a member of the board of Governors of the IEEE Asia-Pacific Wireless Communications Symposium since 2009. He served as Technical Program Committee (TPC) Chair of CROWN'08 and DySPAN'10, Symposium Chair of ICC'12 and Globecom'12, General Co-Chair of ICCS'10 and ICCS'14. He serves as TPC Chair and Executive Co-Chair of Globecom'17 to be held in Singapore.