SOFTMOBILE: CONTROL EVOLUTION FOR FUTURE HETEROGENEOUS MOBILE NETWORKS

TAO CHEN, HONGGANG ZHANG, XIANFU CHEN, AND OLAV TIRKKONEN

ABSTRACT

Heterogeneous mobile networks, HMNs, with flexible spectrum use, densified cell deployment, and multi-layer multiple types of radio access technologies, are expected to be key to meeting the 1000 times increase of mobile data traffic in 2020 and beyond. The increasing complexity in HMNs renders the control and coordination of networks a challenging task. The control frameworks of current cellular networks, which were previously designed for sparse network deployment, hit the wall for HMNs. HMNs need good separation of control and data planes, and call for novel control methods to handle the highly complex dynamics therein. In this article, we first briefly review the control planes of 2G to 4G cellular networks, and then identify their constraints to support HMNs. We analyze the complexity in HMNs and examine enabling control technologies for HMNs. We believe new thinking is needed for efficient control of HMNs. SDN is a promising technology to solve complex control problems in the Internet. Principle-based control methods applied in SDN are promising to solve control problems in HMNs. Several SDN approaches have been proposed for mobile networks. However, most of them are targeted at mobile core networks. We propose an SDN-based control framework named SoftMobile to coordinate complex radio access in HMNs. The main features of SoftMobile are low-layer abstraction, separation of control and data planes, and network-wide high-layer programmable control. Important research problems in SDN for mobile networks are highlighted. We believe SDN for mobile networks will be the controlling evolution of future HMNs.

Tao Chen andXianfu Chen are with VTT Technical Research Centre of Finland.

Honggang Zhang is with Zhejiang University.

Olav Tirkkonen is with Aalto University.

INTRODUCTION

Mobile networks tend to be extremely complex. The number of wireless devices is predicted to reach 7 trillion for 7 billion people by 2020. With more devices and emerging data-intensive services, mobile data traffic is estimated to increase by 1000-fold in 2010–2020 [1]. To keep pace with such enormous demand, the current available spectrum, spectrum efficiency, and cell density of mobile networks have to be extended by at least a factor of 10 [1]. Naturally, cells will tend to be smaller for more capacity and denser for sufficient coverage. Informa predicted that by 2016 the small cell deployment will reach 91 million, compared to 6 million in 2013. Indeed, next generation mobile networks will be heterogeneous, densified, and highly flexible on spectrum use, as shown in Fig. 1. The traditional border among cells and different radio access technologies (RATs) will disappear for seamless mobility, high capacity, spectrum and energy efficiency, and rich service support. In such a network, intense complexity of operation spans from the physical layer up to the application layer.

The key to taming such a level of complexity lies in the control and coordination innovation of mobile networks. We emphasize in this article cellular mobile networks because they will dominate mobile Internet access. The problem is that current control frameworks in cellular networks, originally designed for sparse cell deployment, are far from optimal for future extremely dense and highly heterogeneous mobile networks. Future heterogeneous mobile networks (HMNs) will feature flexible spectrum use, ad hoc deployment of densified small cells, multiple layers of RATs, and diverse fronthaul and backhaul solutions. Current distributed solutions such as Long Term Evolution (LTE) are not efficient for large-scale intercell coordination, while centralized solutions such as the Global System for Mobile Communications (GSM) and Universal Mobile Telecommunications System (UMTS) are not scalable and yield a severe latency problem. To orchestrate efficient radio access in HMNs, we need radically new thinking on the control framework design of HMNs.

Down-to-earth approaches by introducing additional functions are generally used in mobile networks to manage the increasing complexity. Less effort is spent to derive control principles from complexity. Only when simplicity is extracted from complexity and transformed to principles will complexity be tamed and innovation continuously fostered. Advances in computer science show extraordinarily good examples, for instance, in the evolution of programming languages and database design. Therefore, the radical change in the control design of HMNs, by following principle-based approaches, is highly required.

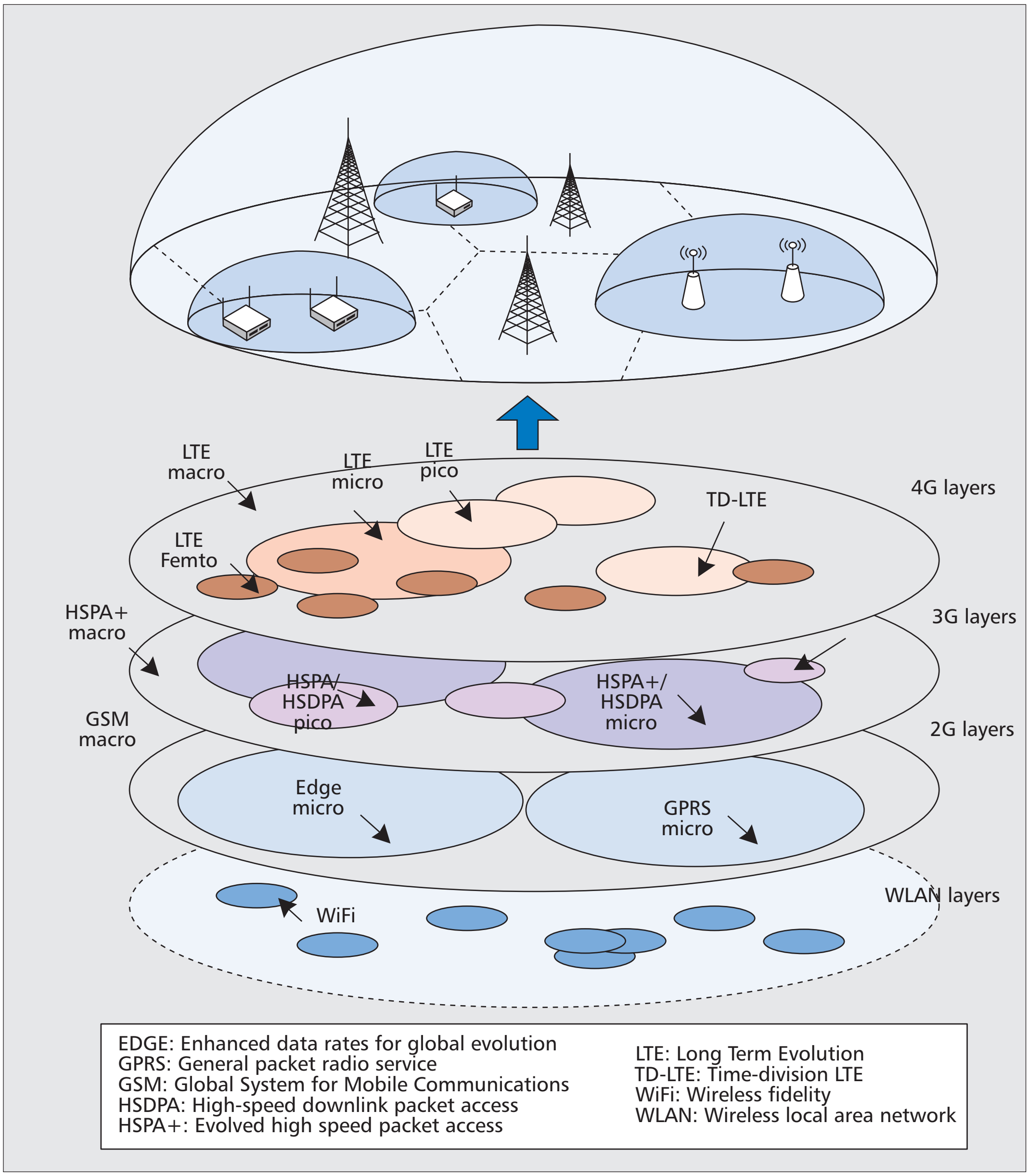

Figure 1. Future heterogeneous mobile networks.

As we know, GSM is based on circuit switching and originally designed for voice traffic. To support data services, General Packet Radio Service was introduced, which adds two control units, Serving GPRS Support Node and Gateway GPRS Support Node, at CN for packet switching and connection to external IP networks.

It is the aim of this article to survey current control solutions for mobile networks, and to provide new thinking on the control design for HMNs. We start with a brief summary of the control plane evolvement in cellular networks, followed by a complexity analysis in HMNs, and then providing an overview of current enabling techniques for the control of HMNs. We argue that while these enabling techniques are necessary components for the efficient control of future HMNs, fundamentally new thinking on the control design is needed. New control approaches should put intercell coordination in the inherent design, and use the proper network abstraction to allow principle-based control. Similar approaches, known as software defined networking (SDN), have been proposed for the Internet and successfully implemented data center networks. Considering the unique features and functionalities of HMNs, we propose SoftMobile, an SDN-based control framework, to manage the complex radio access of HMNs. We believe principle-based control approaches will be the key for the success of future HMNs.

Control Plane Evolution of Mobile Networks

In general, a telecommunication system comprises three planes: the control plane, which carries control signaling; the data plane, which carries user traffic; and the management plane, which carries network management traffic. The design of the data plane can simply follow the layered structure of the open systems interconnection (OSI) model, which has already provided good abstraction for wired and wireless systems. The design of a control plane varies since it has to tightly fit the nature of physical systems. It is particularly challenging to design the control plane for mobile systems as in a base station

Expecting OFDMA to remain mainstream in future mobile networks, we foresee great challenges in joint scheduling among small cells, due to the need for near-real-time scheduling information exchange among a large number of cells.

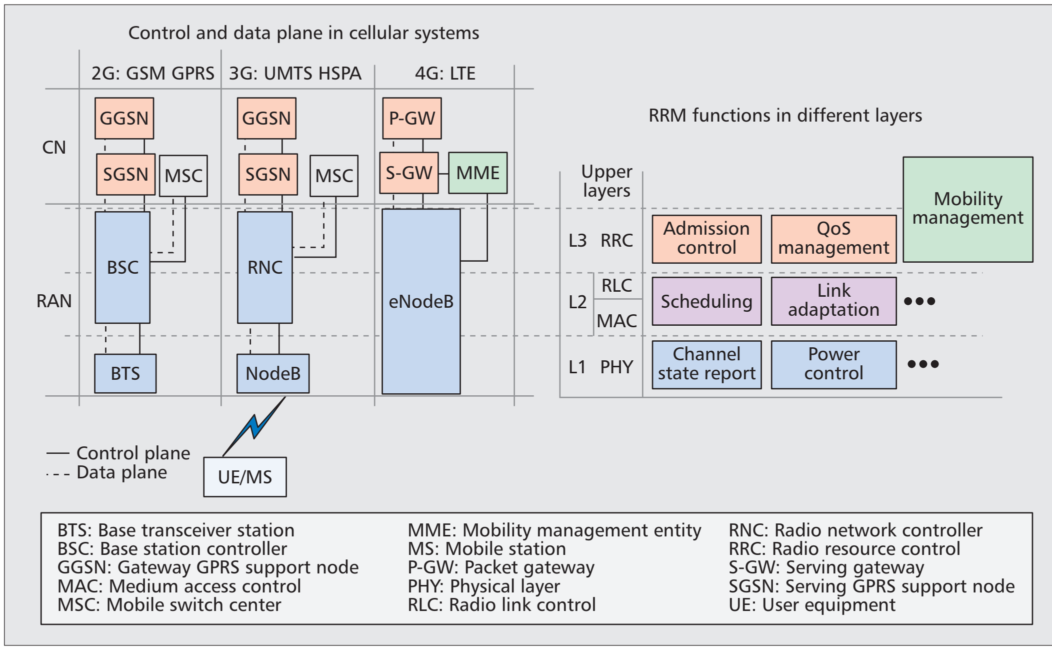

Figure 2. System architecture of GSM, UMTS, and LTE, and location of RRM functions in protocol layers.

alone there are hundreds to thousands of parameters to be controlled.

The evolution of control planes from the 2G to 4G cellular networks was driven by service needs and advances in RATs. Note that a cellular network is normally divided into two parts: the radio access network (RAN) for radio access, and the core network (CN) for the interconnection of RANs, accounting, billing, packet switching, and management. Hereinafter, we focus on the control plane design for RAN. Figure 2 illustrates simplified system architecture, protocol layers, and location of key radio resource management (RRM) functions in GSM (2G), UMTS (3G), and LTE (4G). Since Worldwide Interoperability for Microwave Access (WiMAX) is similar to LTE, it is not listed in Fig. 2.

As we know, GSM is based on circuit switching and originally designed for voice traffic. To support data services, general packet radio service (GPRS) was introduced, which adds two control units, the serving GPRS support node (SGSN) and gateway GPRS support node (GGSN), at the CN for packet switching and connection to external IP networks. This design for packet data services is inherited in UMTS and LTE, with changes only in the naming of control units and some function implementations. The common feature in GPRS, UMTS, and LTE is the heavy use of GPRS Tunneling Protocol (GTP) tunnels, which puts a circuit-switching-oriented approach on top of packet switching networks. This, along with hard state signaling, is identified as the main barrier to releasing the full potential of future mobile networks [2].

The RAN of GSM uses a centralized control approach to implement RRM functions. The base station controller (BSC) is the real control unit in the GSM RAN, which implements the control functions at the link and network layers, and provides the radio control for up to hundreds of base transceiver stations (BTS). A BTS simply implements the physical layer functions and can be regarded as a dumb base station. The advantage of this design is better radio resource allocation and handover control. UMTS inherits the RAN design of GSM, but the names of BSC and BTS are changed to radio network controller (RNC) and NodeB respectively.

Significant architecture changes occur in LTE, which adopts a pure IP-based CN and a flat RAN design. In a RAN the previous separated RRM functions are aggregated into a single node, the eNodeB. The immediate advantage is the abrupt reduction of the RAN access latency to less than 10 ms and the end-to-end delay to less than 300 ms. The new design also provides agility and scalability for RRM, which is suitable for OFDMA. However, the problem associated with this distributed control approach is coordination efficiency among eNodeBs. Although the X2 interface enables direct over-the-air signaling between eNodeBs, intercell coordination does not scale well when the network becomes small and dense.

To understand the difference of various mobile networks on RRM, in Table 1 we provide a simple comparison of spectrum access methods, places for resource allocation, interference management techniques, places to support mobility, and cell cooperation. WiFi is included here because it is becoming increasingly important for mobile Internet.

Complexity Challenges in Heterogeneous Mobile Networks

Due to flexible spectrum access schemes, the use of advanced physical layer techniques, the demand for high-volume data traffic, and the

	Spectrum access	Place for resource allocation	Interference management	Place to support mobility	Cell cooperation
GSM	Neighboring cells use different spectrum, TDMA in cell	Centralized at BSC, no joint scheduling among cells	No strong interference between cells	Centralized at BSC + MSC	No
UMTS	Neighboring cells use same spectrum, CDMA in cell	Centralized at RNC, limited joint scheduling among cells	Load control/Power control	Centralized at RNC + MSC	Limited
LTE	Neighboring cells use same spectrum, with flexible sharing, OFDMA/SC-FDMA in cell	Local at eNodeB, flexible joint scheduling among cells	ICIC/eICIC	Distributed at eNodeB + MME	X2 interface to support flexible cell cooperation, e.g., CoMP
WiMAX	Neighboring cells use same spectrum, OFDMA in cell	Local at BS, limited joint scheduling among cells	Similar to LTE	Distributed at BS and supported by mobility IP	Limited
WiFi	APs share channels, CSMA/CA for channel access	Local at AP, no joint scheduling among APs	CSMA/CA to avoid interference	No	No

List of abbreviations: AP: access point; BS: base station; CDMA: code-division multiple access; CoMP: Coordinated multipoint transmission/reception; CSMA/CA: carrier sense multiple access with collision avoidance; ICIC: intercell interference coordination; eICIC: enhanced ICIC; OFMDA: orthogonal frequency-division multiple access; SC-FDMA: single-carrier frequency-division multiple access; TDMA: time-division multiple access.

Table 1. Comparison of mobile networks on resource allocation and control implementation.

need for mobility, the increasing complexity in HMNs spans accordingly from the physical to the application layer.

The complexity in the physical layer comes from the strong need to handle intercell interference [3]. The frequency reuse of one in LTE systems has created serious interference problem at the cell edge. ICIC techniques like fractional frequency reuse (FFR) and soft FFR (SFR) have been intensively studied. More advanced techniques, such as coordinateed multipoint (CoMP) transmission/reception, massive multiple-input multiple-output (MIMO), and interference alignment, rely on joint signal processing among multiple nodes. The ad hoc and densified deployment of small cells in HMNs provides users rich accessibility but more complex interference scenarios. Since the cell density is expected to increase by 10 times [1], dynamic CoMP would be a standard feature to deal with interference and mobility in HMNs. In this case, the physical layer will heavily rely on the networked feature to function properly, yielding an extremely complex coordination problem.

The complexity in the medium access control (MAC) layer lies in dynamic scheduling of wireless resources. Thanks to the flexible spectrum allocation in LTE, joint frequency and time scheduling becomes a necessary approach to provision radio resource and mitigate intercell interference. Expecting OFDMA to remain mainstream in future mobile networks, we foresee great challenges in joint scheduling among small cells, due to the need for near-real-time scheduling information exchange among a large number of cells. Current distributed solutions adopted by LTE macrocells may not scale for small cells. A centralized solution seems as a better option. However, total centralized MAC solutions, like those in GSM and UMTS, have proven their limitations in performance. For mobile data traffic, latency is one of the key performance indicators. For the access delay of 1 ms required by 5G and beyond, it is necessary to put the scheduling close to radio links. It is highly likely that a hybrid approach which mixes distributed and centralized control will be applied to HMNs. However, this would not be a simple combination but based on new control principles.

A remarkable feature of HMNs would be the heavy use of node cooperation to exploit the spatial diversity gain. The advanced features at lower layers (e.g., CoMP and joint scheduling) are dependent on the network layer for information exchange. There is a strong need for a virtual network, formed by over-the-air links or backhaul links across CNs, to provide a virtualized backbone for networked features at the physical and MAC layers. By virtualization we mean the network is set up by logic links on top of physical networks. This virtual network should be scalable, adaptable to the backhaul capacity and latency requirements of control functions. It imposes a significant challenge on the control plane design of HMNs.

The fast penetration of cloud computing, social networks, and machine-to-machine (M2M) communications in mobile networks will make large-scale use of every kind of network application possible in HMN. These applications will inevitably have direct impact on the performance of cellular networks. This implies that the control plane for RANs should be aware of charac-

By abstraction, the complex distributed routing problem turns into a simplified network graph problem. Moreover, by abstraction, it enables the programming of control functions at high level and thus provides unlimited potential for control innovation.

teristics of these applications and be elastic for different applications.

Current Enabling Technologies for Control Problems in HMNs

To satisfy complex control requirements in HMNs, three technologies will play important roles: cross-layer design, self-organizing networks (SONs), and cognitive radio networks (CRNs).

Cross-layer design can be defined as any design that violates the layered reference communication model [4]. Because the layered reference model was not originally designed for wireless networks, cross-layer design is often used in wireless systems to improve performance. It provides direct feedback across protocol layers and thus enables quick adaptation. This is important for HMNs as lower-layer controls like interference coordination and joint scheduling are all time-critical tasks relying on fast feedback from other layers. In HMNs, control crossing the physical link, network layer, and even application layer will be necessary for the efficient use of wireless resources. Cross-layer design should be standardized in HMNs for various key control functions (e.g., interference coordination, traffic steering, mobility, and node cooperation). Therefore, there is a need to revisit relevant cross-layer designs currently available in mobile networks, and to make them optimal for HMNs.

The SON has been an ever growing important technique to handle the complexity in LTE networks. At the core of the SON is the autonomous adaptation of network functions [5]. Those functions can be roughly divided into self-configuration, self-optimization, and self-healing. Of all SON features, self-optimization receives more attention due to the ability to self-tune the network performance. Self-optimization deals with load balancing, capacity, coverage, and interference control in LTE networks. Considering the exponentially increasing dynamics among different layers of networks, we expect the SON to be the key technology in HMNs. However, the SON for small-cell-enhanced HMNs needs further investigation. Due to the complexity in HMNs, we may need to understand fundamental control problems in HMNs and develop new SON principles.

Cognitive radio (CR) has been widely studied for over a decade. So far, the main focus of CR studies has been on spectrum sharing in dynamic spectrum access (DSA) scenarios [6]. However, the basic principles, which are represented by the cognitive cycle, have been well extended from radio to network-wide. A close look at the cognitive cycle in CR [7] reveals that it is essentially an adaptive cycle with learning capabilities. The key advantages of cognitive radio are awareness, intelligence, learning, adaptability, reliability, and efficiency [8]. Learning plays an important role in CRNs as it enables awareness and intelligence to combat complexity, and to tune networks to desired states [7].

We believe that both CR and SONs will be critical technologies in the future HMN. It is natural to apply CR in the future HMN, as it will have to employ flexible spectrum use as a capacity booster. The intelligence offered by CRN and the adaptability of SON wills help HMNs to handle complex network-wide resource allocation.

Control Plane Evolution

The SDN is widely regarded as the next generation control solution for large-scale IP networks, because it tackles the control problem in the Internet by extracting fundamental control principles from the complexity of network dynamics. The essential idea of the SDN lies in the proper abstraction of the control problem. By abstraction, the complex distributed routing problem turns into a simplified network graph problem. Moreover, by abstraction, it enables the programming of control functions at a high level and thus provides unlimited potential for control innovation.

Mobile networks face similar control problems as the Internet. At the edge of mobile networks, radio resources are controlled by individual base stations almost independently. Since radio resources tend to be widely shared in HMNs, it is beneficial to coordinate wireless resources as a whole by regions among overlapped cells, as in SDNs. Recently, research has been conducted to introduce SDNs to mobile networks. At the present time, available solutions can be roughly divided into three types: sdirectly applying SDNs for wired networks to mobile CNs [9]; adopting SDN principles at mobile CNs [10, 11]; and adopting SDN principles at mobile CNs and RANs [12]. We briefly introduce them in the following.

OpenRoad

OpenRoad [9] was proposed by the same team at Stanford University who invented the SDN concept. It is simply the mobile version of SDN, which includes similar basic components as SDN, for instance, OpenFlow for control, FlowVisor for slicing the network into isolated virtual networks, and NOX as the network operation system to support programmable control. The difference is that Simple Network Management Protocol (SNMP) is used to configure parameters in air interfaces. OpenRoad is designed to support WiFi and WiMAX, with the main purpose of allowing different control algorithms to run concurrently in one network. It is a concept remaining at the experiment level. It does not handle the control complexity in mobile CNs nor that in RANs.

SoftCell

SoftCell [10] applied the design principles of SDN to redesign the control plane of mobile CNs. It proposed a single controller to govern the control plane of the whole core network. The main purpose of SoftCell is to improve scalability and flexibility of mobile CNs. By orchestrating the forwarding tables of switches in CNs, traffic flows traversing on a given path are controlled according to service policies, and thus is capable of distributing flows among packet processing middleboxes for load balancing and scal-

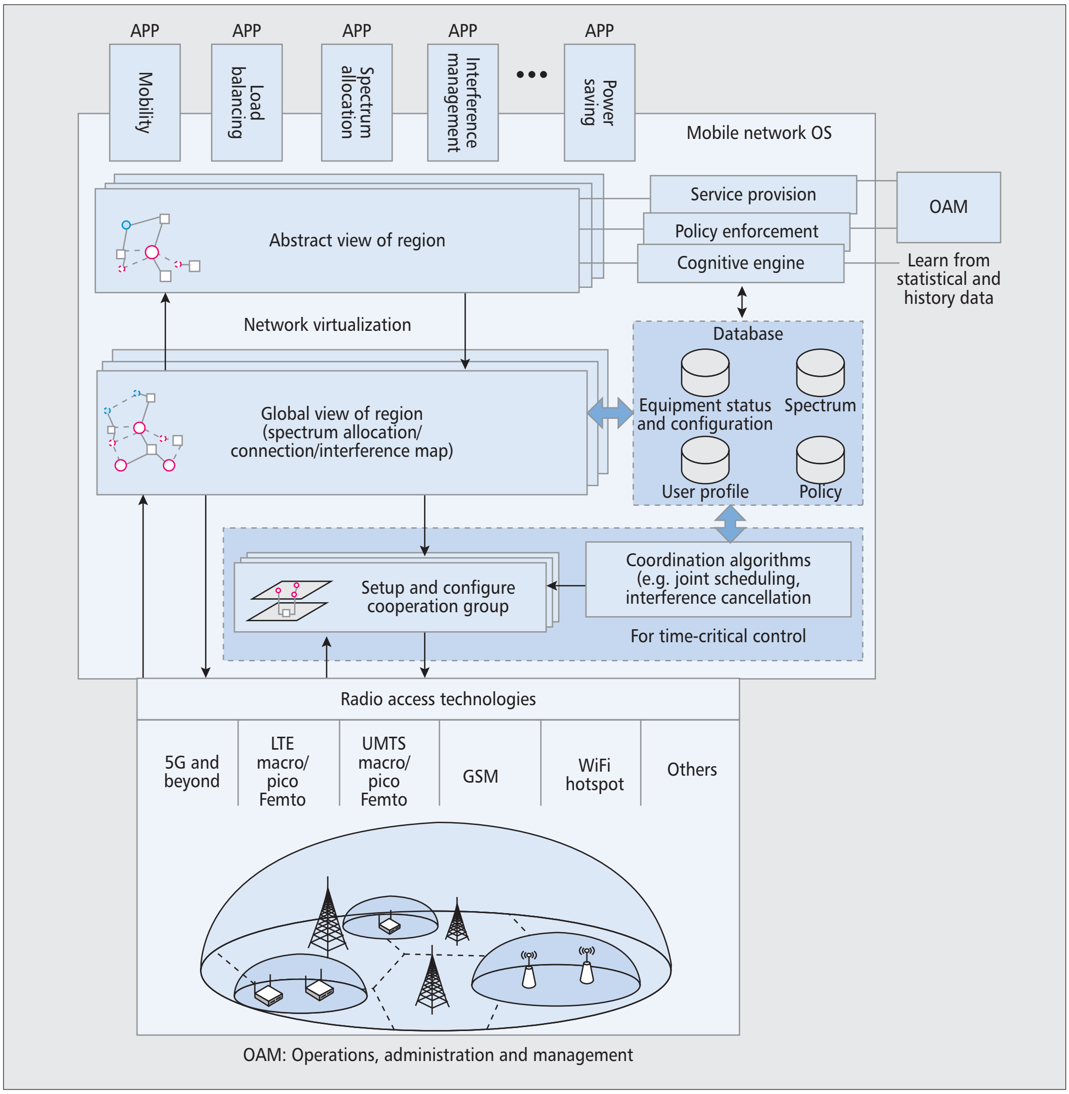

Figure 3. SoftMobile: SDN-based control framework for radio access networks.

Inspired by SDN for Internet, we identify the following features in the control plane of a SDN based HMN: proper abstraction, separation of control and data plane, open Application Programming Interfaces (API), programmable model, SON features and network intelligence.

ability. In order to improve scalability, it moves the fine-grained packet classification from previously a single place at the P-GW to edge switches located in base stations.

SoftRAN

SoftRAN [12] is a concept proposed by Gudipati *et al.* at Stanford University. It abstracts a RAN into a big virtual base station, which performs resource allocation, mobility, load balancing, and other control functions at a single place for all cells in the RAN. It turns the distributed control plane of LTE into a centralized software defined control plane, and takes advantage of full knowledge of the network for global optimization of that network. Obviously, a totally centralized solution for a does not scale well. To solve this problem, it proposes to keep time-critical control functions remaining at local base stations. To date, SoftRAN is the only approach dealing with both mobile core and edge networks. However, it has several hard problems to solve. For instance, due to the high level of centralization, it may need high capacity on backhauls connecting base stations and the central control point. This limits its application. Moreover, latency would be another problem in the control plane.

There are several SDN for mobile solutions proposed by industry. Ravi *et al.* from NEC proposed the network virtualization substrate (NVS) combined with CellSlice to virtualize wireless resources in cellular networks and enable multiple virtual mobile network operators on a single physical network [13]. The main idea is to use the flow concept as in the SDN to isolate wireless network resources among virtual networks. Kempf *et al.* from Ericsson introduced an OpenFlow controller in the LTE CN to separate control and data planes in the core, and move core control functions into cloud architecture [14]. Pentikousis *et al.* from Huawei proposed a flow-based forward model for mobile CNs [15]. The idea is to use SDN-based virtualization to help roll out new network features and reduce time to market for new services.

SoftMobile: SDN for Mobile RAN

Inspired by SDN for Internet, we identify the following features in the control plane of an SDN-based HMN: proper abstraction, separation of control and data planes, open application

With network virtualization functions, the regional view of cells is further abstracted to the high level abstracted view, which is used by control applications. In addition, we propose a cognitive engine to learn from the history configuration and statistics of user data for the better configuration of the network.

programming interfaces (APIs), programmable model, SON features, and network intelligence. Based on these features, we propose a novel control plane structure, SoftMobile, as illustrated in Fig. 3, for future HMNs. We first describe the abstraction of the control problem in HMNs, followed by the control plane design.

Abstraction of Control Plane — Modularity by abstraction of a problem is the efficient way to fight against complexity. The control problem in HMNs can be decomposed to three sub-problems: how to distribute state information among cells for efficient coordination, how to configure cells in a coherent way, and how to deal with real-time intercell cooperation. Associated with these three problems are three abstractions: distribution abstraction for the distribution of cell state information, specification abstraction for coherently configuring a large number of correlated cells, and cooperation abstraction that allows time-critical control for intercell cooperation implemented on a common, flexible, and open structure.

Because of spectrum sharing, cells need information exchange from neighbors to make proper control decisions. Thus, a natural solution to the first problem is to provide a regional view of cells in the control plane, in which we abstract the state information distribution by an annotated network graph. To form the regional view, cells periodically report their necessary state information to a centralized database. The regional view of cells is generated by extracting information from the database. A mobile network operating system (OS) can be developed to generate the network graph and perform control functions through APIs. It is similar to that in the SDN for Internet, but in HMNs we do not need a single global view of all cells, as the coordination may only occur in densified cell areas. It is wise to just generate the network graph of cells on demand by targeted regions.

In the second problem, we have to think how to configure cells when from the regional view the optimal configuration for cells are derived by high-level control applications. Since a base station may have hundreds to thousands of parameters to control, it is infeasible for a centralized point to configure every detail of a cell. To solve this kind of problem, at the high level of the control plane we use behavior-based control. That is, we only define at the high level of the control plane what behaviors the low layer may have, and map behaviors to detailed configurations at the low layer through the device-dependent middle layer. We split the control plane into three layers. The bottom layer resides in individual cells, taking charge of the detailed control in the cell. The middle layer oversees all relevant cells and provides the regional view of cells to the high layer. The top layer virtualizes the regional view of the network and provides an API to control applications in order to implement control algorithms. The top and middle layers use predefined behaviors to guide the control of lower layers. The real implementation of control is hidden at the bottom layer, after mapping selected behaviors from the middle layer to detailed control commands.

To enable certain cooperative functions among cells, it is necessary to have efficient information exchange for real-time intercell coordination. The third problem is how real-time coordination is abstracted. There are diverse needs and different implementations for the cooperation at this level. It affects the method of information exchange and the placement of decision making. To abstract this level of control, we propose to use the open interface and network virtualization. The open interface defines a standardized way of information exchange, and provides a common language for cells to exchange information. Network virtualization maintains a logic control layer to set up signaling paths for distributed and centralized control approaches. It turns multihop physical connections virtually into a simple star connection and automatically maintains connection paths during the cooperation process.

Control Plane Building Blocks — We have defined three abstractions for the control problem in HMNs. The distribution and specification abstraction cover semi-static control of the network, while the cooperation abstraction supports fast control of intercell cooperation. The reference SoftMobile control framework based on these abstractions is depicted in Fig. 3.

With the necessary function extension, the network equipment is able to report status information of cells to a centralized database. The centralized database is also the repository for policies and spectrum information of the network. The regional view of cells, according to the correlation of the cells, is generated by extracting the information from the database. This view includes spectrum allocation, connection of user devices, and interference map among cells.

With network virtualization functions, the regional view of cells is further abstracted to the high level abstracted view, which is used by control applications. In addition, we propose a cognitive engine to learn from the history configuration and statistics of user data for better configuration of the network.

On top of the control framework there are programmable control applications for different control purposes. When a control is made by control applications, it is mapped to control commands at the device level from the network virtualization layer.

Time-critical control is also supported by modules, as shown in Fig. 3. These modules take charge of cooperation among cells, and interpret the requirements and set up the signaling path for the cooperating nodes. It receives control instructions from the high level of the control plane and responds to set up control signaling paths. Depending on the requirements on delay, signaling volume, and control method, the path may be over the air or from backhauls of cells. After setup of the control paths, the involved nodes use signaling to configure properly for cooperation. The principle is to provide an open platform for real-time intercell cooperation, not limited to specific cooperation algorithms.

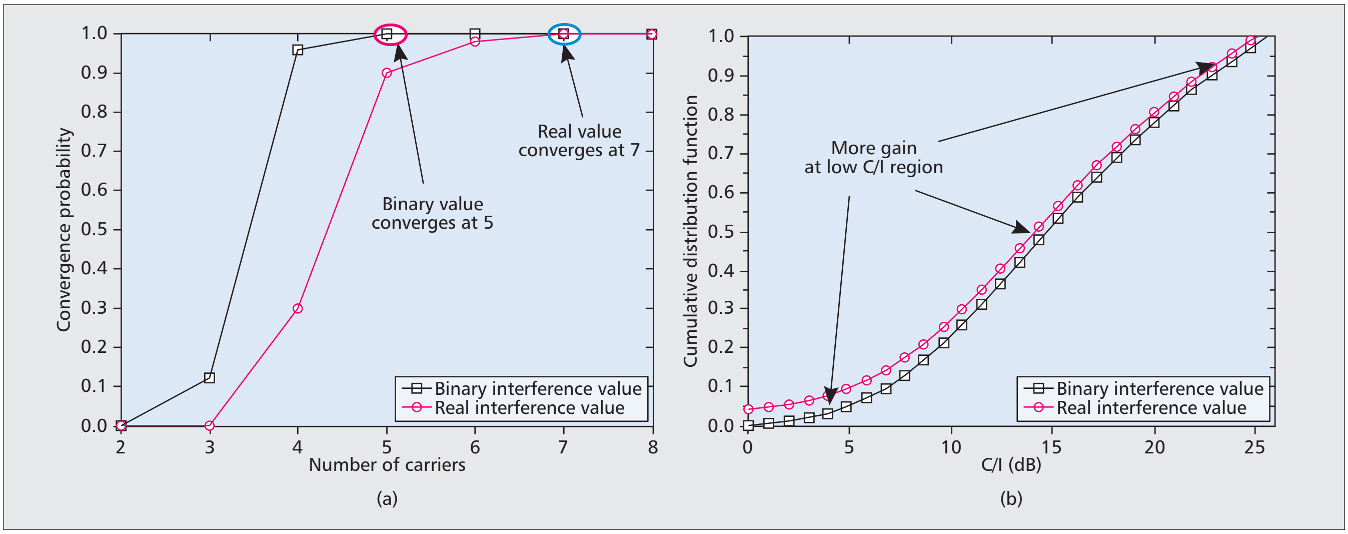

Figure 4. Performance comparison of primary component carrier selection algorithms based on an abstracted binary interference link and a real interference link in a picocellular network with 96 eNodeBs [16]. It is a typical frequency assignment problem in cellular networks. It shows that using the abstracted binary link can help to avoid local optimal and thus provide even better performance: a) the number of primary component carriers required by the network; b) carrier-to-interference (C/I) ratio after carrier selection.

Research Directions

SDN for HMNs is a promising control evolution for mobile networks. We admit that SDN for wireless networks could take similar design principles as SDN for wired networks (i.e., network abstraction, separation of control and data plane, virtualization, and programmable control). Meanwhile, we also expect fundamental differences in low-layer modeling, network abstraction, high-level programmable modeling, and resource virtualization. These topics are critical for the success of a control plane design for HMNs, and thus need deep investigation.

Low-Layer Modeling

It is critical to model the physical and link layers in a proper way so that the network abstraction can be built on with sufficient openness as well as accurate control on different physical layer implementations. We believe network graphs will be widely used in the wireless SDN approach to represent low-layer network reality. One research problem is how to model link states in a network graph with sufficient granularity. As illustrated in Fig. 4, the study showed that for the frequency assignment problem in LTE networks, abstracted binary interference links can even provide optimal results comparable to those with real value interference links [16]. To abstract the low layers, we need to first obtain fundamental understanding of behaviors of specific low-layer implementation and requirements to coordinate those behaviors. For instance, for dynamic CoMP in HMNs, the high-level control plane needs to know which base stations are selected according to their channel gain with the targeted mobile device.

Network Abstraction

The ultimate goal for control in HMNs is to extract control principles from complex radio access and apply them to the new control plane design. Network abstraction is a key to apply control principles in HMNs. The way to abstract HMNs is different from that in SDN for the Internet. In HMNs the spectrum sharing among different network entities has to be taken into account in the abstraction. The abstraction should consider control operations in different timescales. This raises the question of implementing certain control functions at local or central points, and thus needs careful studies. An ideal SDN-based control plane for HMNs should act like a network OS to shield high-level control applications from low-layer implementations through well developed APIs.

Programmable Control

It is desirable to enable programmable control in HMN, in which control functions are implemented based on well defined APIs from low layers. The benefit is obvious as it provides the most flexible way to define or change control behaviors without the need to modify control implementation in the hardware. Programmable control, combined with different service policies from operators, can enable very flexible control functions for different groups of network entities or end users. It allows fast deployment of new control algorithms. For control applications in HMNs, we believe spectrum allocation, mobility, traffic steering, and energy efficiency are key topics to be explored.

Virtualization

In addition to improve the utilization of wireless resources, there is also a strong need to use the new control approach to support virtual mobile network operators that share same physical networks with incumbent mobile network operators. For mobile CNs, the network slicing approach originally supported by SDN for wired networks, provides a solution to slice CNs to several isolated virtual CNs. But to enable an

The principles applied in SDN for Internet will be a valuable reference to design the control plane for HMN. At the same time, we need to capture the complexity in HMN and derives its own control principles and mechanisms. Surely, more fundamental researches are highly expected in this emerging area.

end-to-end solution, wireless resources in HMNs should be virtualized properly so that they can be sliced for virtual mobile networks. Slicing does not simply mean spectrum isolation, but logical isolation of wireless resources. It should allow spectrum sharing among virtual mobile networks, but provide good isolation of capacity and service provisioning. Moreover, we should consider how to integrate SDN for HMNs with SDN for the Internet for a smooth end-to-end control solution.

Conclusion

In this article, we provide an overview of control problems in HMNs. To deal with the increasing complexity in HMNs, we believe SDN-based control is a promising approach to solve control problems in HMNs. The principles applied in SDN for the Internet will be a valuable reference in the design of the control plane for HMNs. At the same time, we need to capture the complexity in an HMN, and derive its own control principles and mechanisms. Surely, more fundamental research is highly expected in this emerging area.

Acknowledgement

The work from the authors of VTT and Aalto University was supported by Tekes EWINE-D project (DN40141/13). The work by Honggang Zhang was supported by the National Basic Research Program of China (973Green, No. 2012CB316000), the Key (Key grant) Project of Chinese Ministry of Education (No. 313053), the Zhejiang Innovative Research Team Project (No. 2013TD20), the Key Technologies R&D Program of China (No. 2012BAH75F01), and the grant of "Investing for the Future" program of France ANR to the CominLabs excellence laboratory (ANR-10-LABX-07-01).

References

[1] Nokia Siemens Networks, "2020: Beyond 4G Radio Evolution for the Gigabit Experience," White Paper, Feb. 2011.

[2] B. J. Kim and P. S. Henry, "Directions for Future Cellular Mobile Network Architecture," *First Monday*, vol. 17, no. 12, 2012.

[3] X. Ge *et al.*, "Capacity Analysis of a Multi-cell Multi-antenna Cooperative Cellular Network with Co-Channel Interference," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, 2011, pp. 3298–3309.

[4] V. Srivastava and M. Motani, "Cross-layer Design: a Survey and the Road Ahead," *IEEE Commun. Mag.*, vol. 43, no. 12, 2005, pp. 112–19.

[5] O. G. Aliu *et al.*, "A Survey of Self Organisation in Future Cellular Networks," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 1, 2013, pp. 336–61.

[6] B. Wang and K. R. Liu, "Advances in Cognitive Radio Networks: a Survey," *IEEE J. Selected Topics in Signal Processing*, vol. 5, no. 1, 2011, pp. 5–23.

[7] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A Survey on Machine-Learning Techniques in Cognitive Radios," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 3, 2013, pp. 1136–59.

[8] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE JSAC*, vol. 23, no. 2, 2005, pp. 201–20.

[9] K.-K. Yap *et al.*, "OpenRoads: Empowering Research in Mobile Networks," *ACM SIGCOMM Comp. Commun. Rev.*, vol. 40, no. 1, 2010, pp. 125–26.

[10] X. Jin *et al.*, "SoftCell: Scalable and Flexible Cellular Core Network Architecture," *9th ACM Conf. Emerging Networking Experiments and Technologies*, 2013, pp. 163–74.

[11] L. E. Li, Z. M. Mao, and J. Rexford, "Toward Software-defined Cellular Networks," *Software Defined Networking '12*, 2012, pp. 7–12.

[12] A. Gudipati *et al.*, "SoftRAN: Software Defined Radio Access Network," *2nd ACM SIGCOMM Wksp. Hot Topics in Software Defined Networking*, 2013, pp. 25–30.

[13] R. Kokku *et al.*, "CellSlice: Cellular Wireless Resource Slicing for Active RAN Sharing," *Commun. Sys. and Networks '13*, 2013, pp. 1–10.

[14] J. Kempf *et al.*, "Moving the Mobile Evolved Packet Core to the Cloud," *Wireless and Mobile Computing, Networking and Communications '12*, 2012, pp. 784–91.

[15] K. Pentikousis, Y. Wang, and W. Hu, "Mobileflow: Toward Software-Defined Mobile Networks," *IEEE Commun. Mag.*, vol. 51, no. 7, 2013.

[16] F. Ahmed *et al.*, "Distributed Graph Coloring for Self-Organization in LTE Networks," *J. Electrical and Computer Engineering*, vol. 2010, 2010, pp. 5:1–10.

Biographies

Tao Chen [S'05, M'10, SM'13] (tao.chen@vtt.fi) received his B.E. degree from Beijing University of Posts and Telecommunications, China, in 1996, and his Ph.D. degree from the University of Trento, Italy, in 2007, both in telecommunications engineering. He is currently a senior researcher at VTT Technical Research Center of Finland. His current research interests include dynamic spectrum access, energy efficiency and resource management in heterogeneous wireless networks, software defined networking for 5G mobile networks, and social-aware mobile networks.

Honggang Zhang is a professor at Zhejiang University, China, as well as the International Chair Professor of Excellence for Université Européenne de Bretagne (UEB) and Supélec, France. He is also an Honorary Visiting Professor at the University of York, United Kingdom. He served as the Chair of the Technical Committee on Cognitive Networks of the IEEE Communications Society from 2011 through 2012. He is currently involved in research on green , and was lead Guest Editor of *IEEE Communications Magazine* Feature Topics on Green Communications.

Xianfu Chen received his Ph.D. degree in signal and information processing from Zhejiang University, China, in 2012. He is presently a senior scientist with VTT Technical Research Centre of Finland. His research interests cover various aspects of wireless communications and networking, with emphasis on green communications, applications of artificial intelligence to wireless communications, and software-defined networking.

Olav Tirkkonen (olav.tirkkonen@aalto.fi) received his M.Sc. and Ph.D. in theoretical physics from Helsinki University of Technology, Finland. Currently he is associate professor in communication theory in the Department of Communications and Networking at Aalto University, Finland. His current research interests are in coding theory, multi-antenna techniques, as well as cognitive and heterogeneous cellular systems.