INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS
Int. J. Commun. Syst. 2016; **29**:1645–1657

Service-oriented 5G network architecture: an end-to-end software defining approach

Mao Yang[1,*,†], Yong Li[2], Bo Li[1], Depeng Jin[2] and Sheng Chen, Fellow, IEEE[3]

[1]*School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China*
[2]*Department of Electronic Engineering, Tsinghua University, Beijing, China*
[3]*Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK*

SUMMARY

With the ever-increasing mobile demands and proliferation of mobile services, mobile Internet has penetrated into every aspect of human life. Although the 4G mobile communication system is now being deployed worldwide, simply evolving or incrementally improving the current mobile networks can no longer keep the pace with the proliferation of mobile services. Against this background, aiming to achieve service-oriented 5G mobile networks, this article proposes an end-to-end software defining architecture, which introduces a logically centralized control plane and dramatically simplifies the data-plane. The control plane decomposes the diversified mobile service requirements and, correspondingly, controls the functions and behaviors of data-plane devices. Consequently, the network directly orients towards services, and the devices are dynamically operated according to the service requirements. Therefore, the proposed architecture efficiently guarantees the end-to-end QoS and quality of experience. The challenges and key technologies of our architecture are also discussed in this article. Real traces-based simulations validate the performance advantages of proposed architecture, including energy efficiency and the whole performance. Copyright © 2015 John Wiley & Sons, Ltd.

Received 23 July 2014; Revised 21 December 2014; Accepted 21 January 2015

KEY WORDS: 5G mobile networks; end-to-end architecture; software-defined network; service-oriented; network functions virtualization

1. INTRODUCTION

With the rapid growth of mobile demands and ever-increasing mobile applications, mobile Internet has become an irreversible trend that profoundly changes human life. Over recent years, the 4G mobile communication system is being deployed and operated worldwide, providing us with advantages in mobile access capacity, data transmission rates, and so on, in comparison to 3G and other old versions of mobile networks. At the same time, 4G in turn dramatically stimulates the mobile demands and services. Consequently, simply evolving and incrementally enhancing the existing mobile networks will no longer meet the ever-increasing mobile demands and services because the current mobile and wireless networks are witnessing a series of following profound expectations and changes.

(1) *Voice-oriented to media-oriented.* As a heritage of the old telecoms network, the traditional mobile communication system naturally concentrates on the voice services. Whereas, the data transmission rates are continuously accelerating, and the use of smartphones is rapidly increasing, which compel the mobile Internet to provide media-oriented services such as video streaming, online

---

*Correspondence to: Mao Yang, School of Electronics and Information, Northwestern Polytechnical University, Xi'an, 710072, China.

[†] E-mail: yangmao@nwpu.edu.cn

game, cloud services, social media, e-economy, and so on. Correspondingly, the services of mobile networks witness the change from voice-oriented to media-oriented [1].

(2) *Increasingly proliferating and multifarious services*. Recently, mobile services and applications proliferate significantly and become increasingly multifarious. For example, *apps* R& D has become one of the most booming markets, and service providers regard service innovations as their core competitiveness. Part of the reason lies in the yearnings of better life, which continuously creates more applications with increasingly demands for better services. Meanwhile, the increasingly powerful smartphones and the new Internet technologies such as cloud computing speed up this trend.

(3) *From pipelines to service-oriented*. Both the current mobile communication system and Internet offer the pipelines that are occupied by the bit streaming for the end-to-end communications. These pipelines neglect the differences among various services as they support these wide-ranging and extremely different services with the same granularity: bit. This pipelined design method, deeply rooted in the inherent telecoms network design, nevertheless has brought great successes to mobile communication and Internet, because the traditional services are comparatively limited. However, it gradually becomes out-of-date and impedes the progress for future mobile networks. Over recent years, network operators' cost keeps growing caused by the ever-increasing demands while their revenues remaining stagnate [2]. As a result of this stagnating revenues, service providers lack motivation for diversified and innovative services, and this has a direct impact on the end users who may hardly experience high service qualities they are expecting. In order to address the poor user experiences, 'service-oriented' is strongly necessary for future mobile networks.

Unfortunately, because the current mobile networks, including 4G, are based on this somewhat out-of-data pipelined design that has the difficulty to satisfy wide-ranging and extremely different service requirements, simply evolving or incrementally improving the current design can hardly meet the previously mentioned expectations and changes. This calls for the research and deployment of the next generation mobile networks, specifically 5G, which are capable of providing an efficient and service-oriented mobile communication network. In other words, not only does 5G need to offer higher network capacity and data transmission rate but also it should provide higher QoS and quality of experiences (QoE) from the design. QoS and QoE are two relevant, both oriented towards the services, but not identical concepts. QoS focuses on the network resources and reflects the matching degree between the service requirements and the resources provision, whose metrics are typically several network parameters such as bandwidth, latency, and jitter [3]; while QoE focuses on reflecting the satisfaction of end users towards the services offered. Because QoS and QoE focus on different objectives, they cannot guarantee each other [4]. Therefore, it is crucial for 5G to introduce a revolutionary architecture in order to guarantee both the QoS and QoE. However, achieving this objective is challenging as many factors deeply influence QoS and QoE. The network and its constituency of diverse and different devices should have the capacity of identifying services and dynamically adjusting their behaviors to cooperatively support services. Moreover, the network should provide flexible and programmable interfaces for the service providers to launch and modify their service requirements.

The ideas of software defined network (SDN) efficiently address these challenges. SDN [5, 6], an innovative paradigm, is one of the latest and hottest topics in networking. It advocates separating the control plane and data plane and abstracting the control functions of the network into a logically centralized control plane. This dramatically simplifies the network devices that can be flexibly programmed via open interfaces. Consequently, SDN not only enables the centralized control plane to make the global decision and to schedule the behaviors of data-plane devices according to the service requirements but also easily provides abundant open interfaces for the service providers to flexibly program. There are several existing studies focused on the QoS of future network or software defined network [7–10]. Differently, this article addresses aforementioned challenges from the global network's perspective. Therefore, aiming to establish a new paradigm for service-oriented 5G mobile network, in this article, we propose an end-to-end software defining architecture. Our architecture possesses three innovative features.

- *Service-oriented*. The architecture directly orients towards the mobile services and focuses significantly on guaranteeing the QoS and QoE. The centralized control plane decomposes the services requirements and controls the behaviors of data-plane devices to effectively support these diversified services. The control plane also offers abundant open application programming interfaces (APIs) to the service providers for meeting their diverse service requirements and innovations.
- *End-to-end*. Our proposed architecture achieves the end-to-end solution by covering both the core network (CN) and radio access network (RAN). Specifically, the CN introduces the *network function virtualization* (NFV) and middleboxes, which make it easy to deploy various functionalities to enhance the QoS and QoE, while the RAN regards the *cross-layer software defining* as one of its key technologies, which enriches and extends the service enhancement methods to the physical layer.
- *Software defining*. The architecture possesses a logically centralized control plane. All the data-plane devices in both CN and RAN are dramatically simplified and controlled by this control plane. The control plane makes the global decision by decomposing the service requirements, employs the specific functionalities by NFV, and controls the behaviors of each data-plane device. Therefore, our service-oriented 5G architecture is software defined.

The remainder of this article is organized as follows. We first outline the challenges faced by the current mobile networks and corresponding opportunities provided by software-defined networks. Next, we present our 5G architecture overview and detail the network function virtualization-based CN and cross-layer software defining RAN. Then, we interpret the key technologies and challenges of the architecture and present several simulation results. Finally, we conclude the article by summarizing our contribution.

2. SYSTEM CHALLENGES AND DESIGN INSIGHTS

As mobile Internet has become an irreversible trend, the future mobile network, specifically 5G, needs to orient towards network services to meet the proliferating mobile Internet demands. As simply evolving the traditional mobile communication architecture can hardly achieve this objective, we study the challenges facing the future mobile network and investigate the opportunities that SDN will provide, in order to naturally envisage how the future mobile network should look like.

Challenge 1: Service-oriented. The traditional pipelined method increasingly hinders the future mobile networks for all the network entities: service providers, network operators, and end users. Service-oriented network and corresponding operation model will efficiently break this stalemate. However, it is almost impossible to achieve this objective by just simply evolving the current mobile networks, including 4G, because the pipelined method is deeply rooted in their inherent design from their birth. Moreover, diverse services require a more flexible and configurable network. Unfortunately, the current mobile network and Internet are extremely closed and ossified.

Opportunity: SDN introduces a logically centralized control plane while dramatically simplifying the data-plane devices – it makes the whole network and all devices controllable and programmable. SDN can supply the service providers with plenty of open APIs by abstracting the underlying functions. To guarantee the QoS and QoE of these different services, SDN can allocate sufficient network resources and can configure optimal paths with appropriate function models according to the service characteristics and network state.

Challenge 2: Mobile environment. Mobile wireless environment is much more complicated, vulnerable, and fast-varying than the wired scenario. Consequently, the QoS and QoE enhancement methods for the wired Internet are mostly unsuitable for the mobile networks. For example, although the bandwidth, latency, and other required network characteristics can be ensured in the wired part of the network, serious inter-cell interferences conspicuously deteriorate the service quality. As a result, the RAN is considered as the bottleneck for achieving desired QoS and QoE, and hostile mobile environment imposes a design challenge.

DOI: 10.1002/dac

Opportunity: Recently, software-defined RAN has become an attractive extension of SDN because RAN provides ubiquitous wireless connectivity to mobile end users, for example, OpenRF [11], OpenRadio [12], SoftRAN [13], OpenRAN [14], and MobileFlow [15]. In this article, we propose a software defining cross-layer architecture for RAN. Consequently, to guarantee the QoS and QoE of the RAN, the SDN controller is able to control the network layer scheduling such as bandwidth and offloading down to the physical technologies such as beamforming and interference canceling.

Challenge 3: Flexible and realtime QoS and QoE enhancements. Different services require completely different QoS and QoE strategies. For example, Voice over Internet Protocol (VoIP) calls for low latency and jitter as well as echo canceling, while online video requires enough bandwidth and, probably, bitrate adaptation and effective buffering. Even trickier, the service requirements and network state are continuously varying. Therefore, providing flexible and realtime QoS and QoE enhancements becomes more complex and challenging.

Opportunity: NFV [16], which offers a useful complement for SDN, is able to create plenty of middleboxes with various functions such as bitrate adaptation, firewall, and security model. Therefore, in this article, the CN controller will determine the appropriate paths along which the service flows need to transmit and then flexibly and dynamically will place several specific middleboxes in the proper positions across the paths. Middlebox controlled by SDN in the CN is also studied in [17, 18] and [19].

Challenge 4: End-to-end guarantee. QoS and QoE guarantee provided by the future mobile network must be end-to-end. Otherwise, just part of it, for example, just the CN guaranteeing its services, is far from sufficient. Therefore, the QoS and QoE guarantee needs to cover all the entities: service providers, network operators, and end users, as well as both the CN and RAN. This demands for efficient corporations among them, which is an extremely a difficult task.

Opportunity: SDN introduces unified and scalable protocols for the network elements to cooperate with each other. Specifically, in this article, a service coordinator is proposed in our architecture to cooperate the rules and actions of the CN and RAN controllers based on the service providers' requirements, network state, and the end users' attributes. Consequently, it facilitates guaranteeing the end-to-end QoS and QoE.

3. END-TO-END SOFTWARE DEFINING 5G ARCHITECTURE

3.1. Overview

In order to achieve the service-oriented objective and to address the challenges discussed in the previous sections, we propose an end-to-end software defining architecture, as depicted in Figure 1. There are three entities in the architecture: service providers offering mobile services, network operator taking charge of the whole network, and devices, end users requiring and enjoying the services. To meet the end-to-end service requirements, this architecture covers both RAN and CN whose network devices are dramatically simplified by this software defining approach.

Consistent with the key principles of SDN, the architecture separates the control plane from the data plane and introduces a logically centralized control plane. The data-plane devices controlled by the centralized control plane focus on the data processing. The control plane not only makes the rules and controls the behaviors of network devices but also abstracts the functions of underlying data-plane network and further provides open APIs to various service providers. Therefore, the service providers may launch customized service requirements by using these abstract open APIs, while the control plane rapidly responds to these requirements. The control plane consists of three sub-control modules. The CN and RAN controllers are specifically responsible for the CN and RAN, while the service coordinator takes charge of the events related to both the CN and RAN simultaneously.

Core network NFV is introduced into the CN. Then, the devices including the forwarding devices and other functionalities are implemented in the unified middleboxes. These middleboxes are controlled and configured by the CN controller. The CN controller determines the optimal paths to transmit the service packets and selects the most appropriate middleboxes along the paths to guarantee the service enhancement functions.

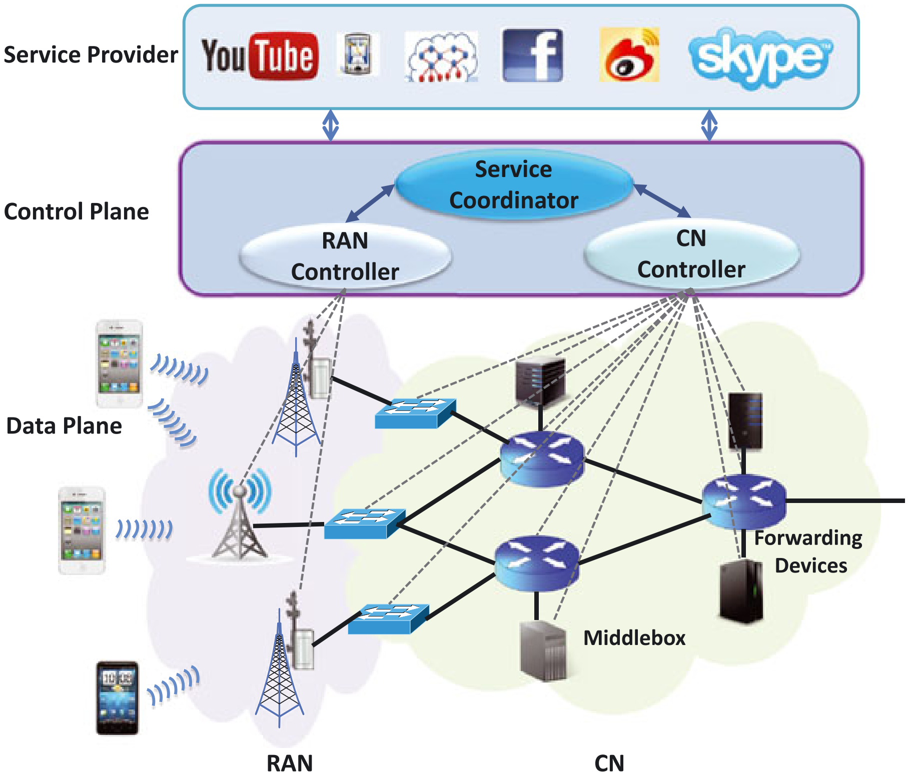


Figure 1. Overview of our proposed software defining end-to-end 5G architecture.

Radio access network We propose the software defining cross-layer RAN architecture. Here, ‘cross-layer’ refers to the fact that the RAN controller, acting from the angle of global perspective, controls not only the network layer service enhancement technologies such as dynamic bandwidth adjusting and offloading but also the physical layer technologies such as beamforming and adaptive coding.

Service coordinator Because the 5G network needs to guarantee the end-to-end service requirements, both the CN and RAN must cooperate with each other. For example, when the network needs to deploy a new service or to modify the priority of an existed service, which relates to the SDN operations in both CN and RAN, the service coordinator makes the coordination decision, and then the CN and RAN controllers add or modify the corresponding rules and take the actions in their own devices. The cooperation decision launched by the service coordinator may be passive, reactive, or proactive. A passive cooperation decision indicates the requirements from service providers, and a reactive one indicates that the service coordinator is scheduling the policies according to the real-time network state, while a proactive one indicates that the coordinator is adjusting the strategies according to its prediction of the network state. The cooperation events are typically fewer for the service coordinator than those for the CN and RAN controllers.

Service providers The centralized control plane offers plenty of open APIs for the service providers. This enables the service providers to ‘programme’ the network to efficiently guarantee their QoS and QoE requirements.

End users The end users initiate various service requests, and the network will select the best access network or even multiple networks simultaneously to support the last hop.

The mobile services are offered by the service providers, initiated by the end users, and carried out by the network operators. Therefore, the control plane needs to establish an open and flexible protocol deployed in all the data-plane devices. The ‘rule-action’ strategy utilized by the wired SDN is equally applicable to the architecture. Each data-plane device matches the pre-installed rules to the match fields of the incoming service flow. If so, it executes the actions corresponding to the matched rules. Otherwise, the flow will be dropped or sent to the controller. However, the data-plane devices are quite different to those found in the wired SDNs, and they include forwarding devices, access devices, and more diversified middleboxes, resulting in significant complexities of both rules and actions in these devices.

3.2. Core network architecture

The CN not only carries the end-to-end traffic among the end users or between the end users and the Internet but also implements a number of data-plane functionalities. The traditional CN such as the long-term evolution centralizes nearly all these functionalities in the edge gateways, referred to as P-GWs. This makes the CN inefficient and complicated and may result in network delay and congestion.

In our proposed architecture, the CN and its data-plane devices are not only software defined but also implemented by the NFV, as illustrated in Figure 2. The NFV virtualizes the diverse classes of network functions into various building blocks that may be connected or chained together to create the specific functions. Specifically, it creates various middleboxes with the specific functions such as routing, content cache, bitrate adaptation, and firewall. In other words, nearly all the network devices in the CN are implemented by NFV. Consequently, each middlebox function is defined by the controller and is implemented by NFV. Then, the controller can route the flows through the appropriate middleboxes, via efficient network paths, to achieve the service requirements. These middleboxes enable deeply programmable and customizable network devices.

More specifically, the CN controller centralizes the control functions of the whole network and its devices, while the NFV manager is integrated with the controller and is under its control. Given a specific incoming service flow, the controller first analyzes its QoS and QoE requirements and parameters such as service type, security level, bandwidth, duration time, and user attributes. Then, the controller calculates the most appropriate paths for this service and determines the necessary middleboxes along the paths. After that, the controller sets the rules and actions for the corresponding devices. Next, the data-plane devices have the ability to process the flows by 'rule-action' control protocols, which will be explained in Section 4. In this way, the flows can be correctly supported by the CN with the guaranteed QoS and QoE.

For example, as depicted in Figure 2, the voice services need to guarantee the security, and, therefore, these service flows will be transmitted through the middleboxes that implement the encryption function. Moreover, to guarantee the voice experience, the echo cancelling middlebox is deployed across the paths. Similarly, an online video service flow is transmitted though the firewall and bitrate

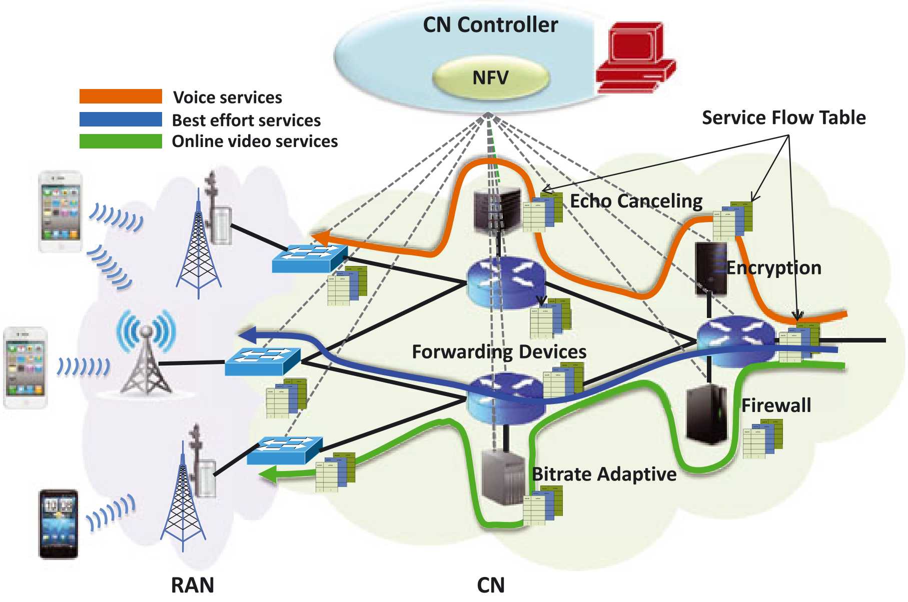


Figure 2. Illustration of the proposed core network architecture.

adaptive middleboxes, while a best-effort service will just go though the forwarding devices without any other functionalities.

With the aforementioned mechanism, the end-to-end traffics and network functionalities are processed and implemented distributively in the entire CN by means of programming, not just in the P-GWs as is the case in a traditional CN. Therefore, it significantly enhances the programmability, flexibility, and evolvability of the CN.

3.3. Cross-layer software-defined radio access network architecture

The RAN offers ubiquitous wireless connectivity to mobile end users, which significantly enriches the Internet services. RAN is also often the bottleneck in a mobile network, because of the vulnerable and fast-varying mobile wireless environment. Consequently, software defining approach in a mobile wireless scenario is much more complex than in a wired networking, and it is essential to extend the software defining approach from the traditional higher layer into the lower layer to include the physical layer technologies. We introduce a software defining cross-layer RAN approach in this article. In our proposed architecture, according to the service requirements, the centralized control plane takes charge of the entire network, from the network layer down to the physical layer.

Network layer Different services call for quite different network layer guarantees. In addition, the wireless access resources are limited, and the network layer needs to provide dynamical resource and strategy scheduling for guaranteeing the required QoS and QoE. In our architecture, the RAN controller dynamically allocates the network resources according to the service requirements. Taking the bandwidth as an example, the realtime services such as VoIP and an online game, have higher priorities to obtain the bandwidth resources and to assure the latency, while the streaming media services such as online videos occupy much more bandwidth, but we may select multipaths to support them and may dynamically adjust the bandwidth with the aid of video buffering strategy. The needs of the delay-insensitive services such as bulk transmissions are more elastic, and we may adjust the bandwidth of these services to guarantee the maximum delay tolerance. Moreover, the RAN controller can easily achieve the data offloading by forwarding optimizing. For example, as depicted in Figure 3, when one access device is incapable of providing sufficient bandwidth because of the high requirements such as video streaming, the software-defining controller may simply ask the neighboring access device to offload these services with very little programming effort.

Physical layer Centralized controlling of the physical layer technologies in mobile networks is innovative but difficult to implement. As illustrated in Figure 3, the RAN controller that has the global information guides the access devices to run the most appropriate physical layer technologies and determines the best parameters to guarantee the QoS and QoE. For example, when an end user at the coverage boundary of the two access devices is experiencing very serious interference, the RAN controller can require the in-service device to operate beamforming to enhance its signal detection performance, and it may also ask the interfering device to initiate interference management to reduce the effect of the interference. To effectively utilize the adaptive coding modules integrated in the access devices, the controller can dynamically schedule the coding type, can vary the code rate, and can adjust the code length for different services to pertinently enhance the QoS and QoE. Furthermore, because the controller possesses the global view, it can easily make the appropriate rules to facilitate cooperative communications, which further increases the network capacity and reduces the interferences.

Globally, the RAN controller analyzes and decomposes the service requirements into multidimensional sub-requirements. It then translates these service sub-requirements into network layer and physical layer requirements. Finally, it sets the corresponding rules and actions for the forwarding devices and access devices. Furthermore, the controller can dynamically adjust the rules and actions to response to the realtime network state. For example, the RAN controller notices that a VoIP service not only is asking for the bandwidth and latency guarantee but also is suffering from unacceptably high bit error rate (BER). The controller may offload the bulk transfer and may decrease the transmit rate of an online video to free some bandwidth and other resources for the

DOI: 10.1002/dac

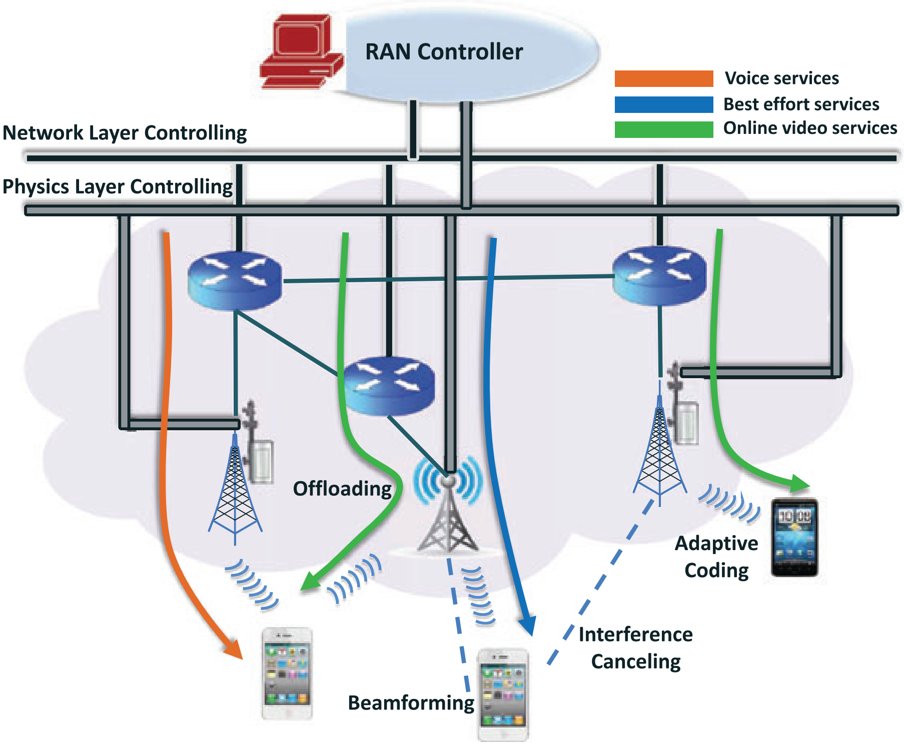


Figure 3. Illustration of the proposed cross-layer radio access network architecture.

battled VoIP service, while initiating beamforming and interference cancelling to reduce the BER of this VoIP service.

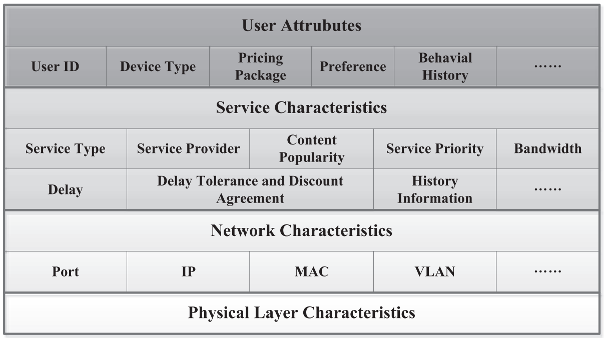
4. CHALLENGES AND KEY TECHNOLOGIES

4.1. Fine-grained control protocol

The control protocol is one of the most critical components in SDN. OpenFlow [20, 21] is a most common realization of the SDN control protocol in wired networking, which has attracted great attention of both academia and industry. However, the control protocol for end-to-end service-oriented mobile networking is quite challenging, and the IP-based OpenFlow protocol is inadequate. The control protocol should reflect the mobile service requirements. Specifically, the 'rules' and match fields of the protocol should cover the mobile service characteristics. Service quality and user experience in the mobile Internet are related to many factors that call for a more fine-grained and elastic control protocol. In addition, unlike the wired networking where forwarding devices are nearly the unique device type, the data-plane devices in mobile networks have diversified functionalities and are much more complex than in the wired scenario. Moreover, controlling the physical layer technologies in RAN is extremely difficult because of the difficulties in abstracting the physical layer and the fast-varying mobile environments. Therefore, fine-grained control protocol is one of the key technologies for the proposed architecture.

Nonetheless, a possible solution to this challenge can be implemented in our architecture, as the 'rule-action' strategy is still an appropriate strategy for realizing the required fine-grained control protocol. The match fields of this fine-grained control protocol contain multi-dimension information including user attributes, service characteristics, network features, and physical layer properties, as illustrated in Figure 4. Thus, the centralized control plane can set the rules and actions and the data processing behaviors of each device according to the multi-dimensional requirements. Moreover, to efficiently control the physical layer technologies, the access devices may be implemented by a building-blocks method, similar to the case of software-defined radio.

DOI: 10.1002/dac



User Attrubutes					
User ID	Device Type	Pricing Package	Preference	Behavial History	……
Service Characteristics					
Service Type	Service Provider	Content Popularity	Service Priority	Bandwidth	
Delay	Delay Tolerance and Discount Agreement		History Information	……	
Network Characteristics					
Port	IP	MAC	VLAN	……	
Physical Layer Characteristics					

Figure 4. Schematic of the proposed fine-grained control protocol.

4.2. Scalability

As the natural consequence of the centralized and fine-grained controlling, scalability becomes one pressing challenge for the proposed architecture. Scalability contains two levels. The first level is network scalability. The mobile services should be provided for anybody in anytime and anywhere. Consequently, there are multiple CNs and especially many RANs distributed over a large area. The distributed controllers and coordinators must cooperate effectively in order to provide the seamless services.

The second level is controlling scalability. The end-to-end service-oriented architecture efficiently guarantees the QoS and QoE, but it also leads to a large amount of controlling transmission overhead and a high load of the centralized controller. A possible solution is to design a physically distributed controlling architecture, that is, multiple controllers, which does not conflict with the logically centralized method. Actually, our proposed control plane architecture, including CN controller, RAN controller, and a service coordinator, has taken a step towards this solution.

4.3. Open application programming interfaces

The control plane needs to provide plenty of open APIs for the service providers, which is far from a trivial task. Providing open APIs calls for abstracting diverse network functions, which is obviously challenging. Specifically, the open APIs should be fine-grained and adaptable in order to satisfy the ever-changing multifarious and innovative service requirements. At the same time, it is necessary for the open APIs to be user-friendly and easy to program. Simultaneously meeting these two somewhat conflicting requirements is often difficult.

4.4. Mobility

Mobility is an inherent property of mobile networks, and any future mobile network architecture must support mobilities. Therefore, efficiently supporting mobility is a pressing issue for our architecture, as it directly affects the QoS and QoE. Mobility leads to many challenges to the control plane design, including usual seamless handovers and interference managements. Fast mobility will test the realtime response property of the centralized controlling to the limit.

4.5. Evolvability

Proliferating services require increasingly innovative technologies, including both the network-layer and physical-layer technologies. How to facilitate a smooth evolving and fast deployment is a key challenge. The middleboxes must have the capability of fast and easily deploying various innovative technologies, while the access devices must be able to effectively implement new and innovative physical layer technologies. Naturally, this is challenging. Fortunately, our proposed architecture is software-defined and is controlled by programming, which lays a solid foundation for network evolution and innovation.

DOI: 10.1002/dac

5. PERFORMANCE EVALUATION

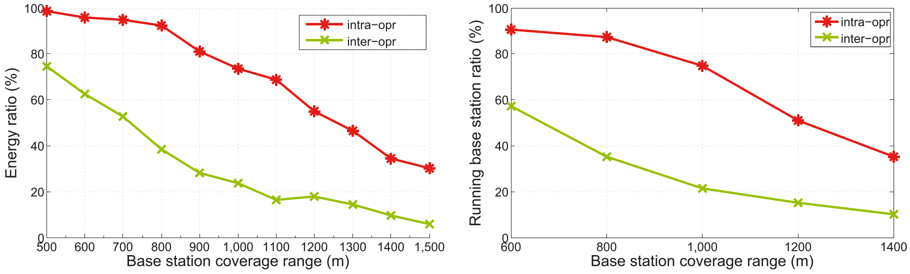
In this section, we evaluate our proposed architecture by simulations with real traces.

5.1. Energy efficiency

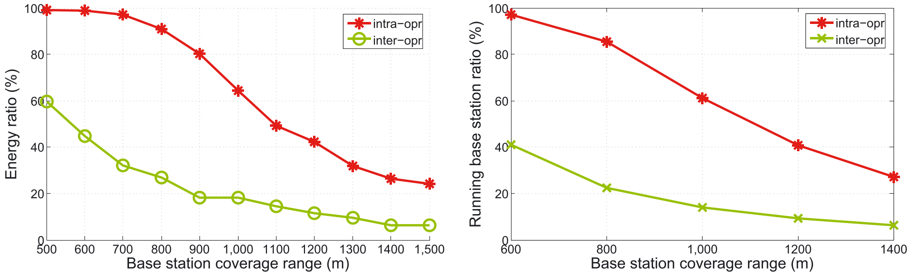
We utilize one online real dataset called ‘Sitefinder’ in order to truly evaluate the performance. ‘Sitefinder’ is an open dataset that is supported and continuously released by Ofcom that is a government organization of the UK [22]. The data collected by ‘Sitefinder’ is provided by multiple big mobile operators such as T-Mobile, Orange, and Vodafone, and the dataset is updated every several months. The information in ‘Sitefinder’ includes operator information, communication standard, base station information, location information, and power information. Finally, we select $100km^2$ geographic area including 414 base stations from Manchester. The communication standards of these base stations are either Universal Mobile Telecommunications System (UMTS) or Global System for Mobile communication (GSM).

During these years, energy efficiency problem has captured increasing attentions, and, consequently, there exist several valuable studies, for example, [23–25]. In this article, we address the energy efficiency problem through network architecture perspective. Tide effect is obvious for mobile communications, for example, at midnight, the network traffic is quite low compared with daytime. This results in lots of energy wastes because the static energy operation consumption and cooling system energy of base stations are considerable. We introduce software defining, and this centralized controlling makes it easy to optimize the energy efficiency solution and to control the network devices by programming. Therefore, we introduce two energy efficiency strategies, each of which needs to first guarantee the 100% area coverage. The first one is called intra-opr that means the controller intelligently shuts down some base stations of each mobile operator. The second one is called inter-opr that means the controller optimizes and shuts down the base stations across multiple operators, which requires operator cooperation.

Figure 5 illustrates the energy efficient performance of these two strategies with the base station coverage range varying from $500m$ to $1.5km$. Specifically, Figure 5(a) and (b) depicts the GSM scenario, while Figure 5(c) and (d) corresponds to the UMTS situation. We can obtain that whether UMTS or GSM, the energy efficient performance is getting increasingly better as the base station coverage range increases. As Figure 5(a) and (c) shows, when the coverage range is short, $500m$ as



(a) Energy efficiency for GSM (b) Running base station for GSM



(c) Energy efficiency for UMTS (d) Running base station for UMTS

Figure 5. Energy efficiency performance.

an example, very little energy will be saved for strategy intra-opr because it is difficult to achieve 100% area coverage. When the coverage range gets larger, $1.2km$ as an example, the network may save 44% of energy by adopting intra-opr and even 82% by adopting inter-opr for GSM as shown in Figure 5(a). While in the UMTS scenario, as Figure 5(c) shows, it saves 58% of energy for intra-opr and 87% for inter-opr. The inter-opr strategy may obtain much better energy-efficient performance because more base station can be shut down because of the operator cooperation controlled by the controller. Figure 5(b) and (d) depicts the base station shutting down results for GSM and UMTS, respectively, which are consistent with energy efficient results. We can apparently observe that more base stations will be shut down with the increase of coverage range. In Figure 5(b) and (d), when the coverage range is short, $0.6km$ as an example, almost all base stations, 92% for GSM and 97% for UMTS, are running in order to guarantee the area coverage. Comparatively, it just needs 52% and 41% of running stations for GSM and UMTS, respectively, when the coverage range becomes $1.2km$. Similarly, inter-opr performs better than intra-opr because of the cooperations among multiple operators.

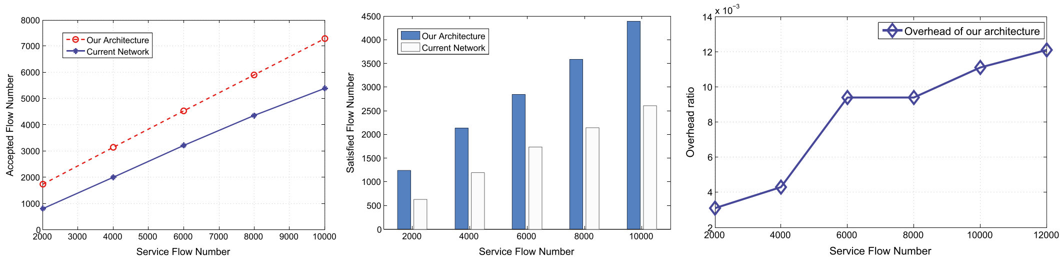
5.2. Whole performance

In this subsection, we focus on the dynamically bandwidth scheduling, and we design a topology according to [26, 27]. This topology possesses three layers: access layer, aggregation layer, and core layer. The access layer consists of clusters of N_A base stations interconnected in a ring [27]. The aggregation layer have N_B groups, each of which comprises N_B aggregation switches connected in full-mesh. Half of the aggregation switches in each group are connected to equivalent access layer clusters, while the other half of the switches are connected to the equivalent switches in the core layer. The core layer contains N_B^2 core switches connected in full-mesh; half of which are connected to the aggregation switches. In the simulation, as shown in Table I, we set the parameters $N_A = 10$ and $N_B = 8$. Therefore, there are 320 base stations, 64 aggregation switches, and 64 core switches.

On top of this topology, we need to generate service flows. We consider four service types: voice, video streaming, file transmission, and best effort (web browsing, social networks, *etc.*). The proportions among these services are based on real statistics and predictions in [28–30]. The four service types constitute, respectively, 8%, 46%, 11%, and 35% of the traffic [30]. Because the current mobile network follows the pipelined method, it does not distinguish these different services. Our architecture is capable of differentiating these services. As voice services possess the highest priority, the bandwidth requirements of voice services are guaranteed first. Although video streaming services call for large bandwidth, video buffering allows the bandwidth fluctuation to a certain extent. As a result, the bandwidth allocation of the video flows is dynamically scheduled to meet the QoS and QoE of the whole network, as long as the practical transmit time is guaranteed to be not more than the video length. The best effort services have a lower priority. Considering that each best effort flow always occupies very little traffic duration time, the required bandwidth can be met as well. Finally, as file transmission services have the lowest priority, the bandwidth of the file transmission services is adjusted dynamically depending on the current network state. Figure 6(a) shows that the proposed architecture accepts 27% more services than the current network does, while Figure 6(b) confirms that the proposed architecture outperforms the current network in the user satisfaction by

Table I. Simulation parameters.

	Parameter	Parameter setting	Value
Base station	$N_A N_B^2/2$	$N_A = 10, N_B = 8$	320
Aggregation switch	N_B^2	$N_B = 8$	64
Core switch	$N_B^2/2$	$N_B = 8$	64
Voice (realtime services)	percent	———-	8%
Video (stream services)	percent	———-	46%
File (data transmission)	percent	———-	11%
Best effort and others	percent	———-	35%



(a) Accepted flows (b) Satisfied flows (c) Overhead cost

Figure 6. Performance comparison between proposed architecture and current networks.

22% on average, which is reflected by the timeliness rate, as the users are sensitive to the completion times of most services including voice, online video, and file transmission. With the growth of flow number, the leading edge of the architecture is increasingly enlarged. Moreover, Figure 6(c) depicts the overhead cost of our architecture. We can observe that the total overhead caused by SDN controlling is very small, and naturally, the overhead cost increases as the total flow number keeps growing.

6. CONCLUSIONS

As the mobile Internet has become an irreversible trend, future mobile network services will be ever multifarious and proliferate dramatically. The future mobile network can no longer rely on the traditional pipeline method, and it must directly orient towards the services and is capable of configuring the network to meet the end-to-end QoS and QoE. In this article, we have proposed an end-to-end service-oriented 5G architecture based on a software defining approach, which addresses several key technical challenges facing the future mobile network. The architecture introduces a logically centralized control plane consisting of the CN controller, RAN controller, and service coordinator. The two controllers set the behaviors of the CN and RAN, respectively, while the service coordinator deals with the cooperative events. The architecture is able to efficiently guarantee the end-to-end QoS and QoE. Because the key technologies of 5G are still open, we believe our flexible software-defining architecture is capable of supporting the 5G standards that eventually emerge and efficiently satisfy the requirements of mobile demands. In our future work, the detailed control strategy is one of the most important studies. Furthermore, efficiently guaranteeing the mobility feature and designing the open APIs also remain as future work.

ACKNOWLEDGEMENTS

This work is supported by National Basic Research Program of China (973 Program Grant No. 2013CB329105), National Nature Science Foundation of China (Grants No. 61171065, No. 61021001, No. 61133015, No. 61271279, and No. 61201157), National High Technology Research and Development Program (Grants No. 2013AA010601, No. 2013AA010605, and No. 2014AA01A707), Program for Changjiang Scholars and Innovative Research Team in University, and the Fundamental Research Foundation of Northwestern Polytechnical University (Grant No. JCY20130131).

REFERENCES

1. Kennedy M, Ksentini A, Hadjadj-Aoul Y, Muntean G. Adaptive energy optimization in multimediacentric wireless devices: a survey. *IEEE Communications Surveys Tutorials* 2013; **15**(2):768–786.
2. Marsch P, Raaf B, Szufarska A, Mogensen P, Guan H, Farber M, Redana S, Pedersen K, Kolding T. Future mobile communication networks: challenges in the design and operation. *IEEE Vehicular Technology Magazine* 2012; **7**(1):16–23.
3. Khanjari SA, Arafeh B, Day K, Alzeidi N. Bandwidth borrowing-based QoS approach for adaptive call admission control in multiclass traffic wireless cellular networks. *International Journal of Communication Systems* 2013; **26**(7):811–831.
4. Zhang J, Ansari N. On assuring end-to-end QoE in next generation networks: challenges and a possible solution. *IEEE Communications Magazine* 2011; **49**(7):185–191.

5. Yu M, Rexford J, Freedman MJ, Wang J. Scalable flow-based networking with DIFANE. *ACM SIGCOMM Computer Communication Review* 2010; **41**(4):351–362.
6. Kim H, Feamster N. Improving network management with software defined networking. *IEEE Communications Magazine* 2013; **51**(2):114–119.
7. Jiang Y, Lan J, Wang Z, Deng Y. Embedding and reconfiguration algorithms for service aggregation in network virtualization. *International Journal of Communication Systems*, posted on 2014. DOI: 10.1002/dac.2794, (to appear in print).
8. Chen JL, Kao SP. *QoS Mechanism for Virtualized Wireless Networks with Software – Defined Networking*, 2014.
9. Chen JL, Liu SW, Wu SL, Chen MC. Cross-layer and cognitive QoS management system for next-generation networking. *International Journal of Communication Systems* 2011; **24**(9):1150–1162.
10. Liao J, Qi Q, Li T, Cao Y, Zhu X, Wang J. An optimized QoS scheme for IMS-NEMO in heterogeneous networks. *International Journal of Communication Systems* 2012; **25**(2):185–204.
11. Kumar S, Cifuentes D, Gollakota S, Katab D. Bringing cross-layer MIMO to today's wireless LANs. *ACM SIGCOMM Computer Communication Review* 2013; **43**(4):387–398.
12. Bansal M, Mehlman J, Katti S, Levis P. OpenRadio: a programmable wireless dataplane. *Proceedings of Workshop on Hot Topics in Software Defined Networking*, Helsinki, Finland, 2012; 109–114.
13. Gudipati A, Perry D, Li LE, Katti S. Softran: software defined radio access network. *Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking, HotSDN*, Hong Kong, China, 2013; 25–30.
14. Yang M, Li Y, Jin D, Su L, Ma S, Zeng L. Openran: a software-defined ran architecture via virtualization. *ACM SIGCOMM Computer Communication Review* 2013; **43**(4):549–550.
15. Pentikousis K, Wang Y, Hu W. Mobileflow: toward software-defined mobile networks. *IEEE Communications Magazine* 2013; **51**(7):44–53.
16. ETSI. Network Functions Virtualisation (NFV) White Paper. *SDN and OpenFlow World Congress*. ETSI. Network Function Virtualization: Network Operator Perspectives on Industry Progress: Frankfurt, Germany, 2013. (Available from: portal.etsi.org/NFV/NFV_White_Paper2.pdf) [accessed on 2 January 2014].
17. Jin X, Li LE, Vanbever L, Rexford J. SoftCell: scalable and flexible cellular core network architecture. *Proceedings of Conference on Emerging Networking EXperiments and Technologies, CoNext*, Santa Barbara, CA, 2013; 163–174.
18. Gebert S, Hock D, Zinner T, Tran-Gia P, Hoffmann M, Jarschel M, Schmidt ED, Braun RP, Banse C, Köpsel A. Demonstrating the optimal placement of virtualized cellular network functions in case of large crowd events. *Proceedings of the 2014 ACM Conference on SIGCOMM*, Chicago, USA, 2014; 359–360.
19. Basta A, Kellerer W, Hoffmann M, Morper HJ, Hoffmann K. Applying NFV and SDN to LTE mobile core gateways, the functions placement problem. *Proceedings of the 4th Workshop on All Things Cellular: Operations, Applications, & Challenges*, Chicago, USA, 2014; 33–38.
20. McKeown N, Anderson T, Balakrishnan H, Parulkar G, Peterson L, Rexford J, Shenker S, Turner J. Openflow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review* 2008; **38**(2):69–74.
21. Lara A, Kolasani A, Ramamurthy B. Network innovation using openflow: a survey. *IEEE Communications Surveys Tutorials* 2013; **99**:1–20.
22. Ofcom. Sitefinder: mobile phone base station database. (Available from: http://www.sitefinder.ofcom.org.uk/) [accessed on 15 January 2014].
23. Zhang Y, Long H, Peng Y, Vasilakos AV, Wang W. QoE and energy efficiency aware resource allocation for OFDM systems in group mobility environments. *International Journal of Communication Systems*, posted on 2013. DOI: 10.1002/dac.2555, (to appear in print).
24. An C, Xie R, Ji H, Li Y. Pricing and power control for energy-efficient radio resource management in cognitive femtocell networks. *International Journal of Communication Systems*, posted on 2013. DOI: 10.1002/dac.2700, (to appear in print).
25. Zhang DJ, Lee CN, Hwang RH. An energy-saving algorithm for cloud resource management using a Kalman filter. *International Journal of Communication Systems*, posted on 2013. DOI: 10.1002/dac.2599, (to appear in print).
26. Howard M. Using carrier Ethernet to backhaul LTE. *Infonetics Research White Paper*, 2011.
27. Nadiv R, Naveh T. Wireless backhaul topologies: analyzing backhaul topology strategies. *Ceragon White Paper*, 2010.
28. Sivaraman V, Moors T, Gharakheili HH, Ong D, Matthews J, Russell C. Virtualizing the access network via open APIs. *Proceedings of Conference on Emerging Networking EXperiments and Technologies, CoNext*, Santa Barbara, CA, 2013; 31–42.
29. Briggs P, Chundury R, Olsson J. Carrier Ethernet for mobile backhaul. *IEEE Communications Magazine* 2010; **48**(10):94–100.
30. Ericsson. *Ericsson Mobility Report: On the Pulse of the Networked Society* 2013. (Avialable from: http://www.ericsson.com/res/docs/2013/ericsson-mobility-report-june-2013.pdf) [accessed on 10 December 2013].