

# Software Defined Networking Enabled Wireless Network Virtualization: Challenges and Solutions

Ning Zhang, Peng Yang, Shan Zhang, Dajiang Chen, Weihua Zhuang, Ben Liang, and Xuemin (Sherman) Shen

## ABSTRACT

Next generation (5G) wireless networks are expected to support massive data and accommodate a wide range of services/use cases with distinct requirements in a cost-effective, flexible, and agile manner. As a promising solution, WNV, or network slicing, enables multiple virtual networks to share a common infrastructure on demand, and to be customized for different services/use cases. This article focuses on network-wide resource allocation for realizing WNV. Specifically, the motivations, the enabling platforms, and the benefits of WNV, are first reviewed. Then, resource allocation for WNV and the technical challenges are discussed. Afterward, an SDN enabled resource allocation framework is proposed to facilitate WNV, including the key procedures and the corresponding modeling approaches. A case study is provided as an example of resource allocation in WNV. Finally, some open research topics essential to WNV are discussed.

## INTRODUCTION

The wireless communication network has achieved laudable success in providing users with seamless connectivity. However, new challenges are arising with the ever increasing connected devices, tremendous data traffic growth, and the growing number of services/use cases [1]. For instance, it is predicted that the number of connected devices will reach 50 billion by 2020, while traffic will continuously skyrocket, resulting in a 1000-fold increase, compared with that in 2010 [2]. In addition, a wide range of services and use cases will emerge with diverse service requirements in terms of data rate, reliability, latency, security, and so on. The Next Generation Mobile Network (NGMN) alliance envisions 25 representative use cases, which are grouped into eight families such as higher user mobility, massive Internet of Things (IoT), extreme real-time communications, and ultra-reliable communications [3].

The next generation (5G) wireless networks is expected to effectively accommodate the massive connectivity, the traffic surge, and the great diversity in services/use cases. Consequently, to boost network capacity, mobile network operators (MNOs) need to densely deploy network infrastructure such as small cell base stations, and add more radio spectrum [4]. As a matter of fact, MNOs have long suffered from the experience of over-provisioning to deal with peak demands, which results in very low resource utilization

and extremely high capital expenditure (CAPEX). Resource sharing among MNOs is a potential option to reduce costs and improve resource utilization. For instance, 3GPP has specified two types of sharing with respect to the elements in the core network and the radio access network (RAN) [5]. On the other hand, the variety of services/use cases should also be supported efficiently. The current one-type-fits-all networks cannot achieve this goal due to poor scalability, limited adaptability, and inflexibility. Those facts translate into the pressing need for a new solution to support tremendous traffic and a multitude of services/use cases in a cost-effective, flexible, and agile manner.

Wireless network virtualization (WNV) or network slicing<sup>1</sup> has great potential to meet the above needs efficiently. It slices the wireless physical infrastructure<sup>2</sup> to create multiple virtual networks sharing a common resource pool [6–8]. A slice or virtual network is a combination of network resources required to support a service/use case. Virtual networks are logically isolated from each other, and any changes in a virtual network will not affect the operation of others. Furthermore, a virtual network can be customized according to the specific requirements of a service/use case, in terms of network topology, control policy, security mechanisms, and so on. With WNV, the following benefits can be achieved:

- Resource utilization can be greatly improved through efficient sharing by multiple virtual networks.
- Use cases or services with diverse requirements can be better supported since the virtual networks can be customized.
- It is cost-effective and flexible to roll out new services.<sup>3</sup>

From a technical perspective, WNV partitions and assigns a set of resources to different virtual networks, where the resources can be used in an isolated, disjunctive, or shared manner [9]. Therefore, the essence of WNV is network-wide resource allocation for different virtual networks. However, it is a great challenge considering the following facts:

- Network-wide resource allocation should satisfy the end-to-end service requirements of virtual networks on a common infrastructure resource pool.
- Traffic and network status are highly dynamic.
- The physical infrastructure exhibits multi-dimensional heterogeneity in devices spanning from access to core segments, and in net-

Ning Zhang, Shan Zhang, Weihua Zhuang, and Xuemin (Sherman) Shen are with the University of Waterloo. Peng Yang is with Huazhong University of Science and Technology.

Dajiang Chen is with the University of Electronic Science and Technology of China.

Ben Liang is with the University of Toronto.

<sup>1</sup> The phrases “wireless network virtualization” and “network slicing” are used interchangeably throughout this article.

<sup>2</sup> Heterogenous resources from different sources can be put in a resource pool for sharing.

<sup>3</sup> A new service or technique can be introduced and tested in a virtual network, without affecting the running services in other virtual networks.

work resources (e.g., bandwidth, computing, and storage).

In this article, we mainly focus on resource allocation for WNV to address the aforementioned challenges. A software defined networking (SDN) enabled resource allocation framework is proposed to support multiple virtual networks flexibly and dynamically. The procedure for resource allocation and the corresponding modeling approaches are provided. Specifically, based on the service description, the resource requirements are determined for each virtual network such that their Quality of Service (QoS) can be guaranteed. Then, based on the requirements from virtual networks and the physical resource constraints, sets of resources are dispatched to different virtual networks. The resources dispatched can be dynamically adjusted to adapt to the variations in the service requirement and in the network status. A case study is provided to illustrate resource allocation for WNV. Finally, open research topics are discussed.

The remainder of this article is organized as follows. The motivations and enabling platforms of WNV are presented in the following section. The resource allocation for WNV and the technical challenges are then discussed. Following that the resource allocation framework for WNV is presented, followed by a case study. Open research topics are then discussed, followed by the conclusion.

## WIRELESS NETWORK VIRTUALIZATION: MOTIVATIONS AND CONCEPTIONS

### MOTIVATIONS

Traditionally, MNOs invest and operate the entire network infrastructure on their own. To scale up, this business model incurs heavy costs and increases operational complexity, leading to unsustainable and inflexible operation. In the upcoming 5G era, this model will be severely challenged due to tremendous connectivity and massive data. MNOs have been constantly investing to upgrade the network and increase network capacity, while their profit remains almost stagnant, mainly because the network operates as a pipeline. Moreover, 5G networks are expected to accommodate great diversity of services/use cases efficiently. Those services/use cases have disparate requirements in terms of latency, reliability, data rate, and so on. Traditional methods for augmenting network capacity and service provisioning cannot work well. A cost-effective, flexible, and agile solution is required to meet the needs of users and provide value-added service.

### WHAT IS WNV

WNV or network slicing can create multiple virtual networks sharing the same physical infrastructure, which helps lower CAPEX. Furthermore, it can provide value-added services (e.g., provide Networking as a Service), by enabling virtual network customization for different applications or use cases. As shown in Fig. 1, WNV enables multiple logical, self-contained virtual networks on a common infrastructure spanning access and core segments. Different sets of resources are provisioned to create virtual networks on demand for services/use case driven networking to their

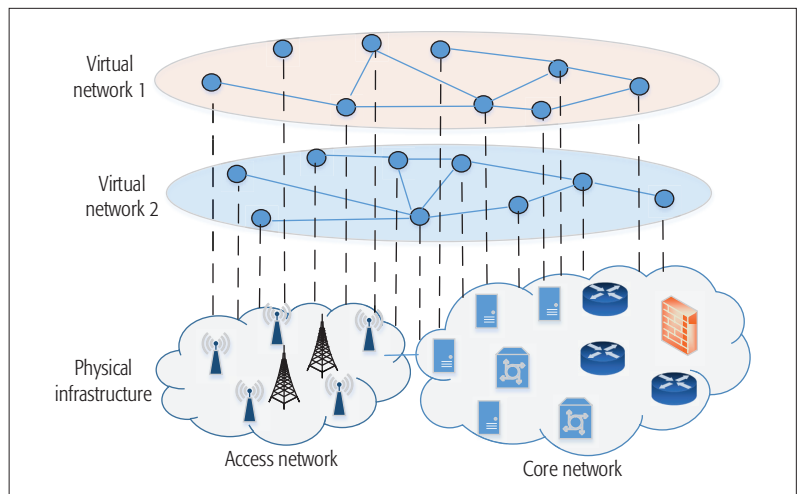


FIGURE 1. Wireless network virtualization.

respective users. Moreover, the physical resources in the access and core network can be further virtualized, e.g., through network function virtualization (NFV) [10]. With NFV, appropriate network functions can be efficiently chained such that service-oriented functions/policies can be enforced on the traffic from a specific service/application, giving rise to a service function chain (SFC) [11]. The resulting SFC can be considered as virtual network, with virtual functions chained in an appropriate way.

With WNV, a virtual network can be customized for particular services/use cases. Different from virtual private networks or virtual local area networks, which mainly focus on a particular protocol layer, WNV allows virtual networks to be completely customized. Virtual network customization can be performed on the network topology, management policies, security mechanisms, resource management strategies, protocol stack, and so on. To create a virtual network, the following three main steps can be performed. First, the virtual network topology is customized, which consists of virtual nodes and virtual links and their corresponding capacities, based on the service description (e.g., user distribution, service request statistic, and end-to-end QoS requirements). Then, the virtual network is mapped to the physical substrate network, which is also referred to as network embedding/mapping. Last, the service-oriented protocol stack can be tailored to better fit the services/use cases [12]. Additionally, sub-virtual networks can be created on a virtual network to support multiple services simultaneously.

### ADVANTAGES

WNV can achieve many benefits, as follows.

**Resource Sharing:** With WNV, the physical infrastructure resources can be shared by multiple virtual networks. Collections of resources are dispatched to multiple virtual networks on demand, helping to avoid high CAPEX caused by the traditional over-provisioning method. In addition, it has the potential to significantly improve resource utilization.

**QoS Provisioning:** Through WNV, virtual networks can be customized for specific services/use cases. Therefore, WNV can better accommodate different use cases with diverse service require-

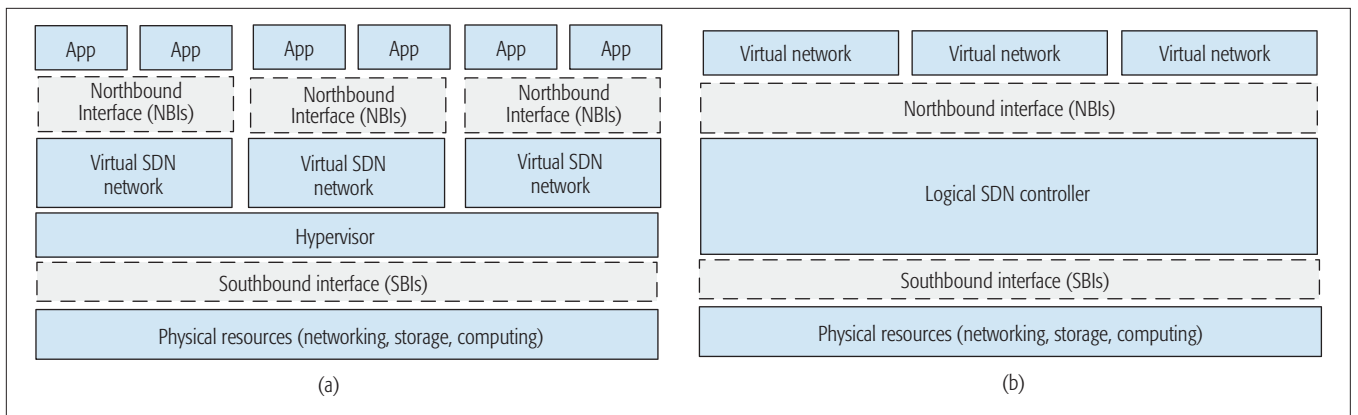


FIGURE 2. Wireless network virtualization.

| Schemes       | Scheduling based  | Software defined networking platform based                             |   |
|---------------|---|--|---|
|               |   | Hypervisor based   | Pure SDN based                                      |
| Features      | Exploit existing schedulers or their extensions             | Rely on SDN platform; add an extra hypervisor layer                    | Rely on the three layers of SDN platform            |
| Advantages    | Less modification to the existing system; easy to implement | Fully support virtual network customization; simplified SDN controller | Fully support virtual network customization         |
| Disadvantages | Limited customization; less flexibility                     | Require new networking platform; more layers                           | Require new networking platform; complex controller |

TABLE 1. A comparison of different schemes.

ments, compared with the conventional one-type-fits-all networking approach.

**Flexible Management:** Since virtual networks are isolated from each other, each virtual network can have its own control and management policies independently. Any change in a virtual network will not disturb the services in other virtual networks.

**Service Innovation:** With WNV, an isolated virtual network can be created to deploy new services or techniques, which avoids disturbing other running services. New techniques/products can be verified in production networks before widespread deployment.

## WIRELESS NETWORK VIRTUALIZATION: ESSENCE AND ENABLERS

### ESSENCE OF WNV

The essence of WNV is network-wide resource allocation to support disparate service requirements of virtual networks. Specifically, a collection of resources is allocated by the infrastructure provider (InP) to constitute a slice (or a virtual network), spanning the access network and core network. For conventional resource allocation, resources are directly allocated to users from the operators. In contrast, in the context of WNV, groups of resources are first allocated to different virtual networks to satisfy the service requirements, considering their user distribution, service request statistics, average end-to-end service requirements, and so on. Then, within each virtual network (VN), based on the instantaneous user requests (service request instances), resources of a virtual network are allocated to users to meet their requests. The former and the latter correspond to inter-VN and intra-VN

resource allocation, respectively. Intra-VN resource allocation is similar to conventional resource allocation, and this article mainly focuses on inter-VN resource allocation. Different from conventional resource allocation, inter-VN resource allocation is performed in a network-wide manner for multiple virtual networks to satisfy their end-to-end service requirements. Moreover, the resource allocation among virtual networks should be dynamically adjusted according to the spatial and temporal variations in traffic demands.<sup>4</sup>

### ENABLING SCHEMES

To perform network-wide resource allocation for multiple virtual networks, the enabling schemes can be mainly classified into two categories: scheduling based [6, 13] and software defined paradigm based [9, 14]. The former slices resources for different virtual networks by exploiting the existing schedulers or their extensions, for example, the scheduling functions at base stations or routers. The latter relies on the new network paradigm of software defined networking (SDN) to allocate resources for multiple virtual networks, through centralized control and open interfaces. Scheduling based approaches require less modification to the existing system, which is fast to implement. However, they cannot fully support virtual network customization. Thanks to the features of SDN separation of the control plane and data plane, logically centralized control, and network programmability, MNV can be performed in a flexible and efficient manner. Moreover, virtual network customization can be fully supported, where each virtual network can have its own SDN controller to enforce different control and management policies.

<sup>4</sup> Although static slicing can guarantee isolation, it results in low spectrum utilization.

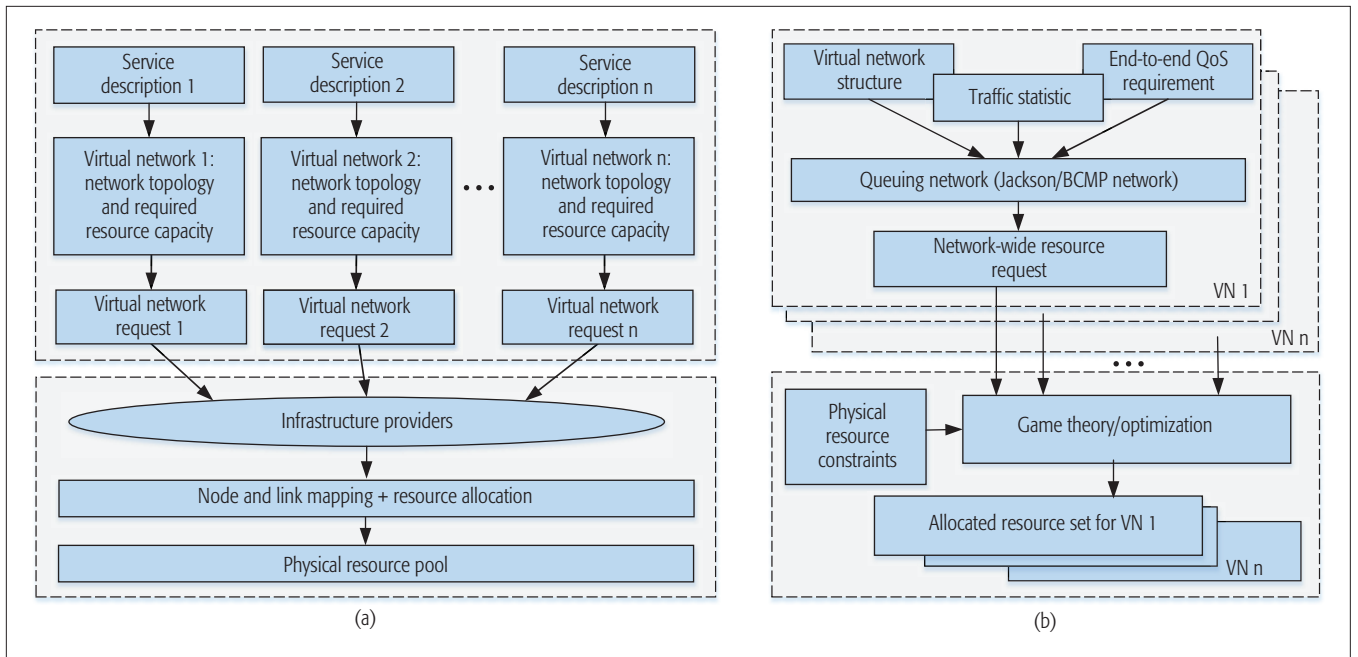


FIGURE 3. The resource allocation framework for WNV.

SDN based schemes can be further divided into hypervisor based [14] and pure SDN controller based [9], as shown in Fig. 2. For the former, an extra layer (i.e., hypervisor plane) is added to the system, above the southbound interface of the SDN layered architecture, as shown in Fig. 2a. A hypervisor introduced at network components helps create multiple virtual entities, for example, multiple virtual base stations and gateways. A network-level hypervisor coordinates local hypervisors to perform network-wide resource scheduling. For implementation, Flowvisor, and its extensions including VerTIGO and OpenVirteX, can be employed [15]. For the latter, with the global view of the network, the logically centralized SDN controller manages and allocates all the underlying physical resources for different virtual networks, as shown in Fig. 2b. Network virtualization can be regarded as an application on the top of the SDN control plane. Compared with the previous method, this approach requires the SDN controller to be more powerful to dynamically orchestrate virtual networks with diverse requirements. A comparison of those schemes is given in Table 1.

### CHALLENGES

Resource allocation for WNV is of significance yet very challenging. The main challenges lie in the following aspects:

- Interference between adjacent nodes in different virtual networks, due to the broadcast channels.
- Network-wide resource allocation to meet the requirement of virtual networks.
- Support of multiple virtual networks with diverse end-to-end service requirements concurrently and efficiently.
- Multi-dimensional heterogeneity in devices spanning from access to core networks, in radio access technologies (RAT), and in network resources (such as bandwidth, computing resources, and storage).

- Adaption to the changes in virtual network requests, variations in physical network status, and time-varying traffic distribution in virtual networks.

### FRAMEWORK FOR WNV

In this section, we propose a resource allocation framework for WNV, including the procedure, modeling, and approaches. It is illustrated in Fig. 3.

#### PROCEDURE FOR WNV

As shown in Fig. 3a, for each virtual network, a suitable virtual network topology is first established (in network planning) for the given service description. The service description may include users' distribution, statistics of service requests, and end-to-end service requirements. The virtual network topology determines the virtual nodes with different functions and their corresponding connections (i.e., virtual links). This process is similar to network planning in wireless networks, for example, base station placement in the service area and core network design. Similarly, based on service requirements, the SFC selects different virtual functions and connects them in an appropriate manner, which can also result in a virtual network. Based on the virtual network topology, the required capacity of virtual nodes and links are determined in order to satisfy the end-to-end service requirement (such as in terms of average delay and jitter). With the network-wide resource capacity requirements and the virtual network topologies, virtual network requests can be formed. Based on those requests from multiple virtual networks and the physical resource constraints, the virtual nodes are mapped to the physical nodes, the virtual links are mapped to the physical paths, and resources at individual physical entities are allocated for each virtual network.<sup>5</sup>

#### RESOURCE ALLOCATION FRAMEWORK

For the given virtual network topology and physical substrate network, resource allocation mainly includes two phases: network-wide resource

<sup>5</sup> The ways for mapping include one-to-one, one-to-many, and many-to-many.

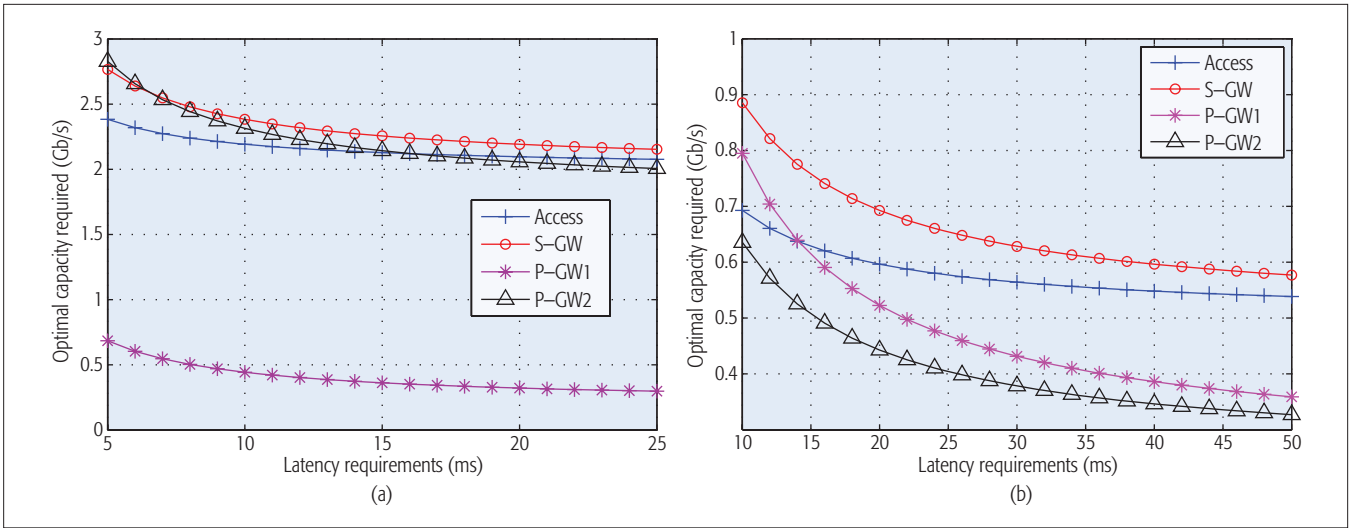


FIGURE 4. Optimal resource allocations of two applications versus latency requirement.

request determination and resource dispatch for multiple virtual networks. The former determines the resources required for each virtual network to satisfy the QoS requirements, given the virtual network topology, whereas the latter assigns resource sets to different virtual networks, based on their requests and the physical resource constraints. To model the resource allocation procedure, potential approaches include optimization and game theory under different scenarios, as shown in Fig. 3b.

For the optimization based approach, the overall objective can be maximization of resource utilization defined by the number of virtual networks concurrently supported. In this case, each virtual network determines the resource capacity needed, and the InP maximizes the total virtual networks supported. Specifically, each virtual network first determines the optimal resource capacity required to satisfy the average end-to-end performance requirement. Taking delay requirement as an example, queuing networks such as the Jackson network or the Baskett-Chandy-Muntz-Palacios (BCMP) network can be employed to analyze the average end-to-end delay performance, given the virtual network topology, the traffic arrival rate, and the capacity of different network components. Then, optimization techniques can be employed to minimize the capacity required at different network components (i.e., the service rate  $\mu_i$ , where  $i$  is the index of the entities) subject to the QoS requirement. With virtual network topology and the optimal capacity requirement, virtual network resource requests are formed. Based on those requests, virtual networks are embedded into the physical substrate network through suitable mapping and resource allocation strategies, to maximize the number of concurrent virtual networks. The mapping problem may be formulated as integer programming and solved accordingly. Node and link mapping can be solved in an uncoordinated or coordinated fashion, considering the trade-off between performance and complexity. Similarly, mapping and resource allocation can also be solved separately or jointly.

When utilizing the resources from the InP, certain costs will be incurred for virtual networks.

Therefore, virtual networks attempt to satisfy QoS requirements with minimal costs (e.g. leasing costs), while the InP aims to maximize its revenue from the virtual networks. In such a case, game theoretical approaches can be employed for the resource allocation procedure. Specifically, based on the queuing network, virtual networks determine the optimal resource capacity required to minimize costs, that is,  $\sum_i p_i \mu_i$ , where  $p_i$  is the price for resource type  $i$  set by the InP. For the InP, with those requests from different virtual networks, it maximizes revenue by selecting suitable price vectors and mapping strategies. Contrary to the previous case, virtual networks have no fixed resource requests and can adjust the capacity requests to balance the achievable performance and cost. Additionally, an auction can also be considered, where the multiple virtual networks compete for resources from the InP.

## TWO TIME-SCALE OPERATIONS

Considering different characteristics of inter-VN and intra-VN operation, resource allocation for WNV is carried out in a two time-scale fashion. In small time scale, intra-VN resource allocation is dynamically performed to adapt to instantaneous user requests, while inter-VN resource allocation remains unchanged. In large time scale, inter-VN resource allocation is adjusted dynamically. Dynamic inter-VN resource allocation is necessary because of:

- Service requirement changes in existing virtual networks, due to the variation in traffic demand and service requests (e.g., various load statistics in different hours).
- New virtual network requests.

For the first case, the resource allocation for different virtual networks will be adjusted by the InP, in order to adapt to the time-varying traffic statistics. For the second case, the InP not only adjusts resource allocation among existing virtual networks, but also allocates resources for the newly coming virtual networks. An admission control policy is required such that the new virtual networks can only be admitted when both their requests and the service requirements of the existing virtual networks are satisfied.

## A CASE STUDY

In this section, a case study is presented to demonstrate how virtual networks determine the resources required to support the respective services. Specifically, we consider that an InP supports virtual networks with different applications and service level agreements (SLAs), and demonstrate how virtual networks respond to their application traffic, the SLAs, and the price of each resource set by the InP.

### NETWORK SETTING

Suppose that two virtual networks coexist, which contain the RAN and core network elements. We consider the following four-node topology: each virtual network consists of a radio access node, a serving gateway (S-GW) node, an administrative P-GW node, and a packet data gateway (P-GW) node, each with independent service rate  $\mu_i$  ( $i = 1, 2, 3, 4$ ), respectively. Assume data traffic coming to the radio access node follows a Poisson process with average arrival rate of  $\lambda$ , and the data packet size follows exponential distribution with mean 1M. The stream reaches the S-GW node after RAN processing. With probability  $q$ , the data packets are for administrative processing and leave the network through the administrative P-GW, whereas the rest leave through the service P-GW for Internet service. We consider the latency requirement (e.g., the time for a packet to pass through the RAN and core network) in the SLA, and present how to meet the latency criteria with minimum cost. Jackson network is employed to analyze the above problem. Jackson's theorem indicates that the numbers of queued packets  $N_i$  in all nodes are independent, and each node behaves as if its arrival stream  $\lambda_i$  were Poissonian. Furthermore, we have  $\lambda_1 = \lambda$ ,  $\lambda_2 = \lambda$ ,  $\lambda_3 = q\lambda$ , and  $\lambda_4 = (1 - q)\lambda$ , respectively.

### OPTIMAL CAPACITY REQUIRED

If latency is required to be less than  $T$  by the SLA, according to the Jackson theorem and Little's law, we have the following constraint:

$$\sum_{i=1}^4 \frac{\lambda_i}{\mu_i - \lambda_i} \leq T.$$

If the quoted unit price of service rates at each node is  $p_i$ , then the objective of each virtual network is to minimize the total cost  $\sum_{i=1}^4 p_i \mu_i$  under the preceding constraint. By applying the method of Karush-Kuhn-Tucker (KKT) multipliers, we can solve the optimization problem by minimizing

$$\sum_{i=1}^4 p_i \mu_i + \alpha \left( \sum_{i=1}^4 \frac{\lambda_i}{\mu_i - \lambda_i} - \lambda T \right).$$

After solving this dual problem, the optimal processing capacity at each node can be obtained.

We illustrate the optimal capacity under two different settings in Fig. 4. Suppose that the InP's price portfolio is  $\mathbf{p} = (p_1, p_2, p_3, p_4) = (0.8, 0.2, 0.05, 0.1)$ . Figure 4 shows the minimum resources required at each node under the scenario of high data volume uploading, such as live video streaming. The access arrival rate is 2000 Mb/s, and 90 percent of data are delivered to the service P-GW, while the remaining 10 percent are delivered to the administrative P-GW for management function such as data statistics. Figure 4b

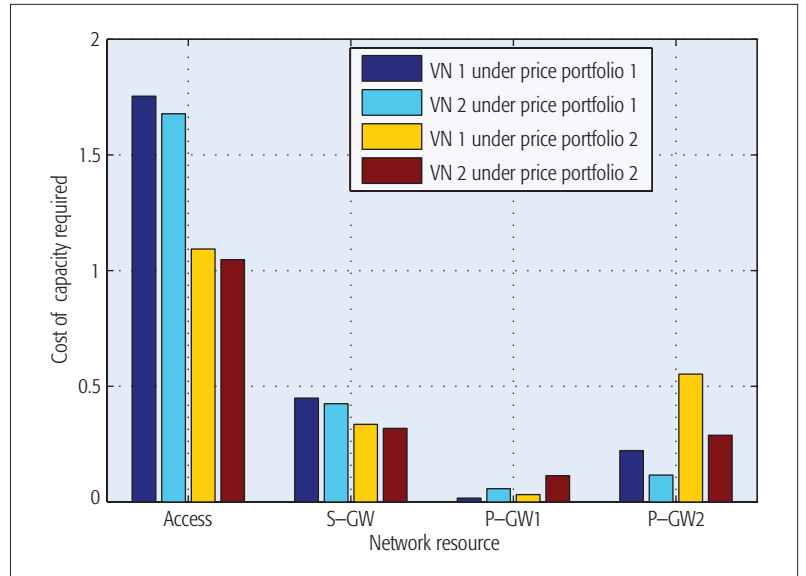


FIGURE 5. VN's cost on each node under different InP pricing Strategies. Price portfolio 1:  $\mathbf{p} = (p_1, p_2, p_3, p_4) = (0.8, 0.2, 0.05, 0.1)$ , Price portfolio 2:  $\mathbf{p} = (0.5, 0.15, 0.1, 0.15)$ .

shows the optimal resources required for low traffic applications such as network monitoring. The average data arrival rate is 50 Mb/s, and 50 percent of data are forwarded to the administrative P-GW. Figure 4 indicates that when the latency requirement loosens to a certain level, the optimal service rate for each node stagnates at an amount slightly larger than that required to process traffic with intensity  $\lambda_i$ . Otherwise, a backlog may accumulate at that node and the latency will run out of control.

Figure 5 further shows the cost on each node with different price tags for the above two scenarios, where the corresponding latency requirements are 10 ms and 20 ms, respectively. With this framework, the optimal capacity required can be obtained for the considered case.

## RESEARCH ISSUES

In this section, some open research issues are discussed for the development of WNV.

**Automated Network Operation:** The SDN controller manages the network behavior, based on the network state obtained from information acquisition and analysis. Because of the scale and volume of information, big data techniques such as data mining and machine learning can be adopted to extract knowledge and insights for guiding the decisions of SDN controllers. In order to achieve automated network operation to efficiently support WNV with minimal manual efforts, the interplay between SDN and big data needs further investigation.

**Dynamic and Fast Network Update:** After the SDN controller makes decisions based on the current network state, the decision rules should be enforced in the network quickly and efficiently. However, due to distinct capabilities and loads on the network entities, the decision rule update times can vary, resulting in an inconsistency issue. Moreover, dependence on the rules further makes network updates more complex. Dynamic and fast network updates is of great importance to ensure the consistency and efficiency of WNV.

Based on the proposed framework, many research topics can be facilitated, such as interaction among different parties and dynamic network-wide resource allocation. To accelerate the pace of WNV development and to better support diverse services/use cases in the 5G era and beyond, great research efforts on WNV are expected.

**Virtual Network Customization:** Virtual networks can be customized to achieve service-oriented networking, taking into account the specific requirements of the services/use cases. Virtual network customization can be performed on virtual topology, functions of nodes, resource capacities, protocols, and so on. Optimal virtual network topology and resource requirements need to be determined to satisfy the end-to-end service requirements. Moreover, the network protocol can also be customized, such as the protocol stack, function modules of a layer, and parameters. In this line, great research efforts are needed to enable full network customization.

**Joint Mapping and Resource Allocation:** With the requests from virtual networks, the InPs need to map the virtual nodes/links to the physical nodes/paths, which is known to be a NP-hard problem. Moreover, the InPs also need to determine the network-wide resource allocation for virtual networks. Jointly performing mapping and resource allocation will bring more gains, but it is more complex and challenging. An efficient scheme that jointly performs mapping and resource allocation is needed.

**Network Security:** Along with the benefits, WNV also poses various security issues. Since the SDN controller is responsible for dispatching and managing resources to different VNOs, it is vulnerable to denial-of-service (DoS) attacks, which can paralyze all the virtual network operation. In addition, the SDN controllers can be compromised, whereby the attackers can disturb the normal operation of virtual networks. As an example, attackers can intentionally change the resource allocation policy so that the experience of users in certain VNOs can be severely degraded. Therefore, security issues should be thoroughly addressed before widespread deployment of WNVs.

## CONCLUSION

In this article, we have proposed a resource allocation framework to pave the road toward wireless network virtualization. The procedure for WNV and the corresponding modeling approaches introduced provide useful guidance and insights to study and realize WNV. Based on the proposed framework, many research topics can be facilitated, such as interaction among different parties and dynamic network-wide resource allocation. To accelerate the pace of WNV development and to better support diverse services/use cases in the 5G era and beyond, great research efforts on WNV are expected.

## ACKNOWLEDGMENT

This work has been supported by The Natural Sciences and Engineering Research Council of Canada (NSERC) and NSFC under Grant No. 61502085, and the China Postdoctoral Science Foundation funded project under Grants 2015M570775.

## REFERENCES

- [1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE JSAC*, vol. 32, no. 6, 2014, pp. 1065–82.
- [2] N. Zhang *et al.*, "Cloud Assisted HetNets toward 5G Wireless Networks," *IEEE Commun. Mag.*, vol. 53, no. 6, 2015, pp. 59–65.
- [3] A. Osseiran *et al.*, "Scenarios for 5G Mobile and Wireless Communications: The Vision of the METIS Project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, 2014.
- [4] N. Zhang *et al.*, "Beyond Coexistence: Traffic Steering in LTE Networks with Unlicensed Bands," *IEEE Wireless Commun.*, vol. 23, no. 6, 2016, pp. 40–46.
- [5] 3GPP, "Technical specification group services and system aspects; network sharing; architecture and functional description," TS 23.251 V11.5.0, 2015.
- [6] R. Kokku *et al.*, "NVS: A Substrate for Virtualizing Wireless Resources in Cellular Networks," *IEEE/ACM Trans. Networking*, vol. 20, no. 5, 2012, pp. 1333–46.
- [7] I. F. Akyildiz, S.-C. Lin, and P. Wang, "Wireless Software-Defined Networks (W-SDNs) and Network Function Virtualization (NFV) for 5G Cellular Systems: An Overview and Qualitative Evaluation," *Computer Networks*, vol. 93, 2015, pp. 66–79.
- [8] C. Liang and F. R. Yu, "Wireless Network Virtualization: A Survey, Some Research Issues and Challenges," *IEEE Commun. Surveys Tutorials*, vol. 17, no. 1, 2015, pp. 358–80.
- [9] ONF, "Applying SDN Architecture to 5G Slicing," TR-526, 2016.
- [10] J. Liu *et al.*, "Reliability Evaluation for NFV Deployment of Future Mobile Broadband Networks," *IEEE Wireless Commun.*, vol. 23, no. 3, 2016, pp. 90–96.
- [11] S. Mehraghdam, M. Keller, and H. Karl, "Specifying and Placing Chains of Virtual Network Functions," *Proc. IEEE CloudNet*, 2014, pp. 7–13.
- [12] M. Chiang *et al.*, "Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures," *Proc. IEEE*, vol. 95, no. 1, 2007, pp. 255–312.
- [13] A. Gudipati, L. E. Li, and S. Katti, "Radiovisor: A Slicing Plane for Radio Access Networks," *Proc. ACM Workshop Hot Topics Software Defined Networking*, 2014, pp. 237–38.
- [14] A. Blenk *et al.*, "Survey on Network Virtualization Hypervisors for Software Defined Networking," *IEEE Commun. Surveys Tutorials*, vol. 18, no. 1, 2016, pp. 655–85.
- [15] A. Al-Shabibi *et al.*, "OpenVirteX: Make Your virtual SDNs Programmable," *Proc. Third Workshop Hot Topics Software Defined Networking, ACM*, 2014, pp. 25–30.

## BIOGRAPHIES

NING ZHANG [S'12, M'16] earned the Ph.D degree from the University of Waterloo in 2015. He received his B.Sc. degree from Beijing Jiaotong University and the M.Sc. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2007 and 2010, respectively. From May 2015 to April 2016, he was a postdoc research fellow at the BCCR lab at the University of Waterloo. His current research interests include next generation wireless networks, software defined networking, green communication, and physical layer security.

PENG YANG received his B.Sc. degree from the Department of Electronics and Information Engineering, Huazhong University of Science and Technology (HUST), Wuhan, China, in 2013. Currently, he is pursuing his Ph.D. degree in the School of Electronic Information and Communications, HUST. From Sept. 2015, he has also been a visiting Ph.D. student in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His current research interests include next generation wireless networking, software defined networking, and fog computing.

SHAN ZHANG received her Ph.D. degree from the Department of Electronic Engineering, Tsinghua University, and a B.S. degree from the Department of Information from Beijing Institute Technology, Beijing, China, in 2016 and 2011, respectively. She is currently a post-doctoral fellow in the Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada. Her research interests include resource and traffic management for green communication, intelligent vehicular networking, and software defined networking. She received the Best Paper Award at the Asia-Pacific Conference on Communication in 2013.

DAJIANG CHEN [M'15] is currently a post doctoral fellow at the University of Waterloo, Waterloo, ON, Canada. He is also a post doctoral fellow in the School of Information and Software Engineering at the University of Electronic Science and Technology of China. He received the B.Sc. degree in 2005 and the M.Sc. degree in 2009 from Neijiang Normal University and

Sichuan University, respectively, and the Ph.D. degree in information and communication engineering from UESTC in 2014. His current research interests include information theory, channel coding, and their applications in wireless network security, wireless communications and other related areas.

WEIHUA ZHUANG [M'93, SM'01, F'08] has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, since 1993, where she is a professor and a tier I Canada Research Chair in Wireless Communication Networks. Her current research focuses on resource allocation and QoS provisioning in wireless networks, and on smart grid. She is a co-recipient of several best paper awards from IEEE conferences. She was the editor-in-chief of *IEEE Transactions on Vehicular Technology* (2007-2013), and the TPC Co-Chair of IEEE VTC Fall 2016. She is a Fellow of the IEEE, a Fellow of the Canadian Academy of Engineering, a Fellow of the Engineering Institute of Canada, and an elected member in the Board of Governors and VP Publications of the IEEE Vehicular Technology Society.

BEN LIANG [S'94, M'01, SM'06] received honors-simultaneous B.Sc. (valedictorian) and M.Sc. degrees in electrical engineering from Polytechnic University in Brooklyn, New York, in 1997, and the Ph.D. degree in electrical engineering with a computer science minor from Cornell University in Ithaca, New York, in 2001. In the 2001–2002 academic year, he was a visiting lecturer and postdoctoral research associate at Cornell University. He joined the Department of Electrical and Computer Engineering at the University of Toronto in 2002, where he is now a professor. His current research interests are in mobile communications and networked systems. He has served as an editor for *IEEE Transactions on Communications* and *IEEE Transactions on Wireless Communications*, and as an associate editor for Wiley's *Security and Communication Networks*, in addition to regularly serving on the organizational and technical committees of a number of conferences. He is a senior member of IEEE and a member of ACM and Tau Beta Pi.

XUEMIN (SHERMAN) SHEN [M'97, SM'02, F'09] received the B.Sc. (1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. He is a professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. He was the Associate Chair for Graduate Studies from 2004 to 2008. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. He served as the Technical Program Committee Chair/Co-Chair for IEEE Infocom'14, IEEE VTC'10 Fall, the Symposia Chair for IEEE ICC'10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, the Technical Program Committee Chair for IEEE Globecom'07, the General Co-Chair for Chinacom'07 and QShine'06, the Chair for IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He also serves/has served as the editor-in-chief of *IEEE Network*, *Peer-to-Peer Networking and Applications*, and *IET Communications*; a founding area editor for *IEEE Transactions on Wireless Communications*; an associate editor for *IEEE Transactions on Vehicular Technology*, *Computer Networks*, and *ACM/Wireless Networks*, among others; and as a guest editor for *IEEE JSAC*, *IEEE Wireless Communications*, *IEEE Communications Magazine*, and *ACM Mobile Networks and Applications*, among others. He received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004, 2007 and 2010 from the University of Waterloo, the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. He is a registered professional engineer of Ontario, Canada, an IEEE Fellow, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.