

# NFV and Blockchain Enabled 5G for Ultra-Reliable and Low-Latency Communications in Industry: Architecture and Performance Evaluation

Haojun Huang, Wang Miao, Geyong Min, Jialin Tian, and Atif Alamri

**Abstract**—5G networks are expected to provide cost-efficient, reliable and flexible services for industrial productions and applications potentially, by introducing emerging network technologies like blockchain and Network Functions Virtualization (NFV), which virtualizes network functions and runs them on standard infrastructure rather than customized hardware. However, how to deal with the emerging security challenges and fulfil the requirement of Ultra-Reliable and Low-Latency Communications (URLLC) has not been fully resolved. In this paper, we present an NFV-enabled 5G paradigm for industry with the guarantee of URLLC through service chain acceleration and dynamic blockchain-based spectrum resource sharing among a variety of industry applications running in NFV-based equipment. First, we elaborate the benefits and shortcomings of NFV for industry, by executing an industry application experiment in virtualized and non-virtualized Data Center Networks (DCNs). Then, we illustrate an NFV-enabled 5G paradigm for URLLC in detail, with a special focus on the service chain acceleration and spectrum sharing built on NFV, blockchain, Software-Defined Networking (SDN) and Mobile Edge Computing (MEC). Finally, we establish a mathematical model to study the worst-case transmission latency of NFV-enabled 5G with the input of the bursty traffic. The proposed model can be exploited to support the plan, management and optimization of NFV-enabled 5G URLLC systems for industry.

**Index Terms**—Network Functions Virtualization, 5G, URLLC, Analytical Modelling, Industry.

## I. INTRODUCTION

THE traditional mobile networks, such as 3G and 4G, have become an important information platform for industrial communications and applications, but suffered from their inherent common limitations, such as complicated management and manual configuration [1], [2]. Specifically, the network hardware infrastructure, for example, servers, switches, and routers, are traditionally developed by network equipment manufacturers in the industrial Cyber-Physical Systems (CPS). Each network vendor requires to design dedicated firmware and related software to operate purpose-built industrial hardware in a proprietary and closed manner. This slows down the

progress of innovations in industry and thus leads to additional expenditures in network management and operation whenever new industrial services, technologies or hardware are to be deployed within the existing industrial networks. Moreover, the updating and upgrading of communication modules in industrial systems will limit the overall development and cause the additional Capital Expenditure (CAPEX) and Operating Expense (OPEX) with the increased industrial scale and requirements of Ultra-Reliable and Low-Latency Communications (URLLC) in industrial automation environments [3].

One efficient solution to overcome these limitations is to design the promising 5G networks, which are characterized by ultra-high traffic volume density, ultra-high connection density, and ultra-high mobility [1], [4], by introducing emerging technologies, like Network Functions Virtualization (NFV) [5], [6], [7]. The concept of NFV originated from service providers looking to increase the agility and flexibility of deploying new network services to support ever-increasing customer requirements. The base idea of NFV is to virtualize entire classes of network functions or middle-boxes into building blocks that may connect, or chain together, to create network services, and run them on standard infrastructure rather than customized hardware in software, known as Virtualized Network Functions (VNFs). Examples of VNFs include virtualized load balancers, firewalls, intrusion detection devices and WAN accelerators [8], [9]. The initial commercial applications show that NFV can provide numerous benefits for mobile networks, ranging from scalable programmability to reduced CAPEX and OPEX, and will greatly promote industrial automation [10], [11].

Because of its potential for machine-centric industry applications and services, NFV-enabled 5G has attracted considerable research efforts from the academia and industry. A large number of companies and standardization groups are pushing the research and development of NFV paradigm to improve cost efficiency, flexibility, and performance guarantees of 5G [2], [5], [6], [7]. The significant nimbleness and scalability of NFV in 5G can be found in [5], [6], [7]. In [5], an integrated NFV/Software-Defined Radio (SDR)/Software-Defined Networking (SDN) architecture for 4G/5G networks has been developed, with the aim of both CAPEX and OPEX reductions and openness promotion. Moreover, a novel 5G NFV-based experimental platform for URLLC implementation in Tactile Internet [2] has been proposed, which adopts SEMIoTICS architecture to reach industrial KPIs with SDN/NFV technologies in the control loop, and thus provides useful insights for profitable automation processes industry applications.

H. Huang and J. Tian are with Department of Communication Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (E-mail: hhj0704@hotmail.com and jialin.tian@hotmail.com).

W. Miao and G. Min are with Department of Computer Science, University of Exeter, Exeter, EX4 4QF, UK (E-mail: wang.miao@exeter.ac.uk and g.min@exeter.ac.uk).

A. Alamri is with the Research Chair of Pervasive and Mobile Computing, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia (E-mail: atif@ksu.edu.sa).

Manuscript received June 29, 2020; revised Sep. 11, 2020 (Corresponding author: Wang Miao).

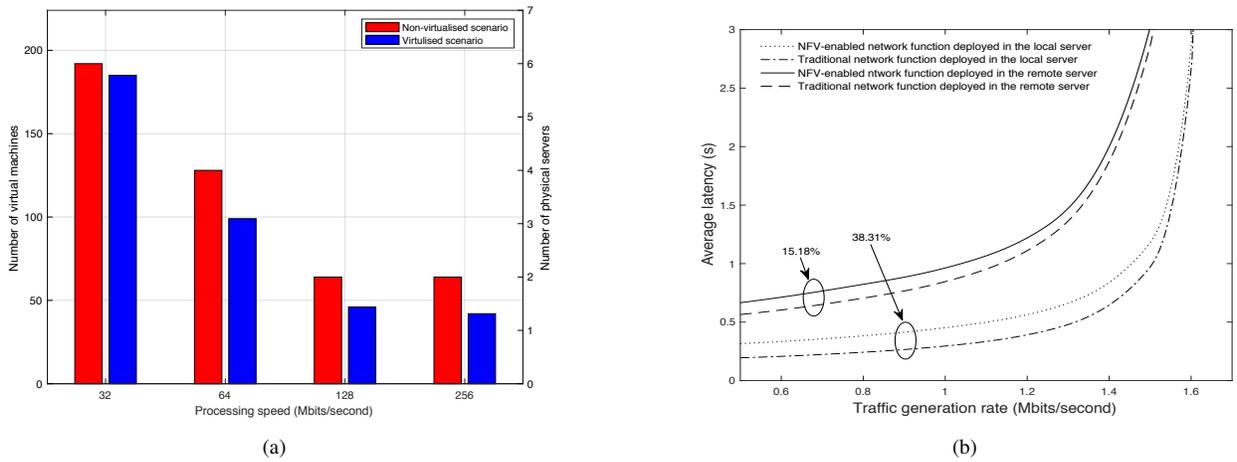


Fig. 1. NFV benefits and shortcomings in expenditure and performance for industry in quantitative manners. Both CAPEX and OPEX reductions have been achieved in (a), while only suboptimal performance guarantee compare to that of purpose-built industrial hardware implementations in (b).

Although the fore-mentioned promising investigations have been made in the area of NFV-enabled 5G systems, there are still two challenges that hinder the adoption of the existing NFV-enabled 5G to support the emerging URLLC industry applications. Firstly, smart manufacturing exploits the promising Internet of Things (IoT), Artificial Intelligence (AI) and network technologies to revolutionise the operation and management of industry services. While, with the increased reliance of factory operations on complex and actionable data flows, comes the substantial risk of cyber-attacks on the physical systems, which brings the new challenges of security and reliability in smart manufacturing. Secondly, compared with the traditional non-virtualised approaches, NFV brings the benefits of flexible resource orchestration and utilisation, while suffers from the transmission performance degradation in term of the latency, making it challenging adaptable to support delay-sensitive industry applications.

To bridge these gaps, this paper proposes a novel NFV-enabled 5G system for industry through exploiting the latest development of blockchain, SDN and Mobile Edge Computing (MEC). Specifically, due to the inherent feature of distributively providing immutable data record, storage and sharing across networks in a peer-to-peer manner [12], [13], [14], [15], we exploit blockchain to manage the wireless spectrum sharing and resource utilisation of the entire NFV-enabled 5G system to ensure the security during resource allocation and service provisioning. To improve the programmability and scalability, SDN is used to facilitate the design and implementation of NFV service chains in the underlying 5G. Furthermore, the MEC is utilized to improve the agility of the service provisioning for the proposed system architecture, where the VNF nodes are deployed in the MEC servers in the vicinity of smart factories to reduce the data transmission latency and provides the computation offloading for resource-constraint industry devices to conduct blockchain operations. In addition, to shed the light on the transmission capability of the NFV-enabled 5G, we study the performance of NFV services in the proposed system architecture in a quantitative manner. The main contributions of this paper are summarized as follows:

- We elaborate the benefits and shortcomings of NFV

for industry in a quantitative manner by executing an experiment in virtualized and non-virtualized industrial Data Center Networks (DCNs).

- We illustrate the promising NFV-enabled 5G architecture for industry to support URLLC services through service chain acceleration and dynamic blockchain-based spectrum and resource sharing among a variety of industry applications, buildings and appliances.
- We establish a mathematical model to study the capability of NFV service provisioning in 5G system with respect to the worst case transmission latency and reliability guarantee. The proposed model can be exploited to support the design and management of NFV service chains in 5G industry applications.

The organization of the paper is as follows: Section II introduces both NFV benefits and shortcomings in industry. Section III elaborates the NFV-enabled 5G architecture for URLLC, along with its potential applications and innovations for industry. Section IV then analyzes the performance with respect to the end-to-end (E2E) latency bound for NFV-enabled 5G for industry in a quantitative manner, followed by the conclusion in Section V.

## II. NFV BENEFITS AND SHORTCOMINGS IN INDUSTRY

NFV, with its inherent decoupling of network functions from the underlying infrastructure, offers a great management and control of the network through programming. This combined feature would bring the potential benefits of the decreased expenditure, improved industrial production efficiency, and encouraged innovation in network architecture and operations. For instance, a given industrial service, in the form of service function chain, can be decomposed into a set of VNFs, which could then be implemented in software running on one or more industry standard servers. Furthermore, the VNFs may then be relocated and instantiated at different network locations without requirements to purchase and install new hardware.

In addition, the potential benefits of NFV are evidenced by the fact that NFV offers a convenient environment for experimentations of new techniques and encourages new net-

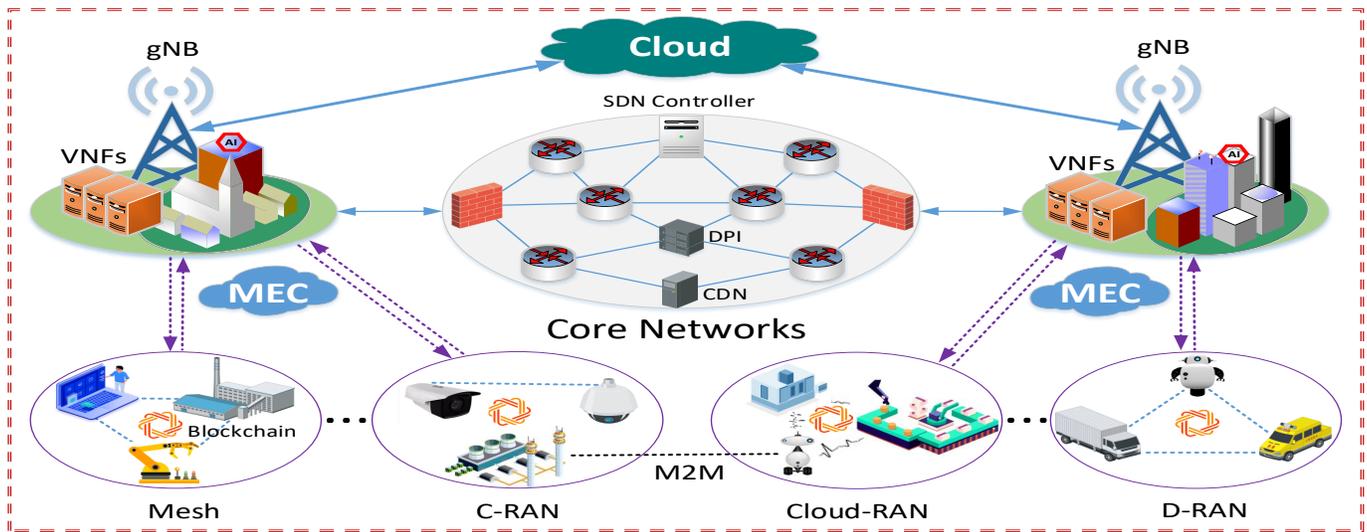


Fig. 2. NFV-enabled 5G paradigm for URLLC in which both the mobile-network base station and packet core networks are virtualized through NFV, and blockchain has been introduced to share spectrums and resource among industrial devices.

work designs, attributed to its network programmability and the ability to define isolated virtual networks via software.

To estimate the benefits of NFV for industry in a quantitative manner, Fig. 1(a) illustrates two scenarios in which more than 15 network services are implemented in virtualized and non-virtualized DCNs for industry applications like energy data storage and e-commerce transactions, respectively. There are 16 workstation servers, each of which has been virtualized into 32 virtual machines, to conduct experiments. As shown in Fig. 1(a), when the processing speed is 64 Mbits/second, the number of the equivalent VMs consumed by the non-NFV service deployment is around 128, which is approximate 28% of the total VM virtualised. And the number of VM occupied by the VNF-based service scenario is 99, which is around 78% of the equivalent VMs for non-NFV based service deployment, thus achieving an approximate 22% reduction of resource consumption. In addition, when the processing speeds are 128 and 256Mbit/second, the NFV-based service deployment can obtain more than 29% and 33% less resource consumption compared with the non-NFV-based service deployment, respectively. However, from Fig. 1(b), it is observed that NFV just achieves suboptimal performance with the ever-increasing traffic in the network. For example, the average transmission delays of virtualized DCNs are 15.18% and 38.31% more than that of non-virtualized DCNs due to the additional delay in the virtualization layer, respectively.

This means that NFV-enabled 5G networks will significantly reduce the CAPEX and OPEX and provide flexible network services for industry with NFV as a potential enabler, but just achieve suboptimal performance compare with that of purpose-built industrial hardware implementations. It is worth noting that VNF can generally assure performance guarantees of VNFs operations, such as maximal pack failure rate, longest delivery delay, and tolerable packet loss [7], [16]. Therefore, it is requisite to introduce the emerging technologies into 5G, enabling it to be more intelligent, flexible and powerful in industry.

### III. NFV-ENABLED 5G ARCHITECTURE FOR URLLC IN INDUSTRY

#### A. NFV-Enabled 5G Architecture

The NFV-enabled 5G paradigm for URLLC in industry can be illustrated in Fig. 2, which consists of a series of base stations, mobile core networks, and the emerging network paradigms and technologies like NFV, blockchain, SDN, MEC, and cloud. The base stations and mobile core networks will be partially or fully virtualized through NFV. In such an architecture, many network equipment types have been consolidated onto industry standard high volume servers, routers, and storage devices located in data centers, network nodes, and even end user premises. The underlying physical infrastructure and physical radio resources among manufacturing shops, power stations, and manufactories, as shown in Fig. 2, have been abstracted and sliced into virtual cellular network resources holding certain corresponding functionalities, and shared by multiple parties with each other without interfering, through resource slicing and isolation [17], [18], [19]. These enable the sharing of not only the infrastructure, but also the content, among different service providers [18]. The network functions in this architecture are implemented in a software package running in virtual machines, MEC and even remote cloud. The open standard APIs will be available for various applications, enabling the establishment of new mechanisms to deploy and operate network and infrastructure services. These can be rapidly provisioned and released with minimal efforts, and enable the standard network appliance to migrate from one hardware to another [12].

In order to realize URLLC services with high security for industry, there are three considerations in our proposed NFV-enabled 5G paradigm. The main idea of these considerations is to improve NFV service chains in core networks and share spectrum resources in base stations, built on blockchain, SDN and EMC, for a variety of industry applications running in NNFV-based equipment, described as follows.

First, the dynamic blockchain-based spectrum sharing of microwave and millimetre waves among industrial machines will be exploited. This can ensure that a variety of radio waves used in 2G, 3G 4G and 5G will be assembled to send more signal carriers in downlinks and uplinks, enabling ultra-high bandwidth and ultra-high reliable communications for industrial manufacturing and applications. Once a machine hopes to communicate with another, it first signs a contract with the base station for spectrum sharing, then the authorized available spectrum can be allocated to them. The spectrum sharing will be allowed in different machines and base stations. In order to improve the agility of the service provisioning for the proposed architecture, the blockchain operations will be migrated to MEC and cloud computing of 5G to improve the security of the system management to fight against the potential malicious attacks. The VNF nodes will be deployed in the MEC servers in the vicinity of smart factory to reduce the data transmission latency.

Second, two kinds of VNF service chain orchestrations, referring to the sequential NFV chain and the parallelized NFV chain, are designed in the proposed NFV-enabled 5G architecture to boost NFV-based service performance. In the sequential NFV deployment, the VNFs are connected in series to provide network services, where each VNF stands for a specific network function and multiple VNFs are linked together in a strict order for realising a specific network service. In the parallelized NFV deployment, multiple VNFs, for example, firewall and intrusion detection system, are parallel executed, which significantly improves NFV performance, such as reduced processing latency and increased system throughput, for diversified industrial manufacturing and applications.

Finally, to enhance the intelligence of networks and support the programmability of networks for industry applications, NFV-enabled 5G paradigm will integrate with the emerging network paradigms and technologies such as SDN [20], cloud computing, MEC, AI, and data-driven thought, to address the existing issues related to URLLC and their derivations [5], [6]. The NFV-based network functions will be executed in SDN controllers, which are logically regarded as the smart brain of the whole network by collecting network-wide topology information and introducing data-driven information plane. For example, the NFV-enabled 5G in Fig. 2 can be managed and controlled with SDN controllers, thereby further simplifying network management and making intelligent decisions for information transmission with global network information in diversified industry applications. In addition, MEC will be exploited in the proposed architecture to provide the computation and storage resources for blockchain processing and NFV deployment. For instance, the computation-intensive computation tasks can be offloaded from the resource-constrained industrial devices to MEC servers to reduce the latency related to blockchain operation. Meanwhile, the VNFs can be installed in the MEC to enable the data collected from industrial sensors to be processed locally, strengthening the data privacy and alleviating the workload in the core network and reducing the data processing latency. Furthermore, NFV-enabled 5G will coexist with legacy technologies like 4G and even 3G to support URLLC services for industry. This means

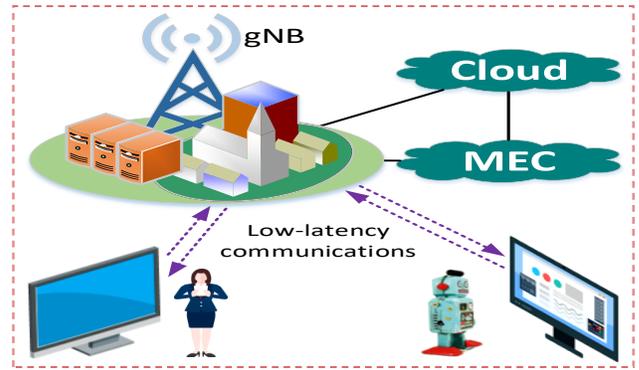


Fig. 3. Low-latency communications among cloud-based robots, cloud and MEC in industry.

that the upgraded industry standard communication devices need to be compatible with the traditional facilities, and will communicate with others via 3G and 4G when 5G cannot be available due to its short-range communications.

### B. Industry Applications and Service Innovations

NFV-enabled 5G paradigm has unleashed great power and potentials in various industry applications, and provided extraordinary opportunities for industrial service innovations. There will be a number of potential industry applications and service innovations related to our paradigm, listed as follows:

(1)**Industrial automation control.** Being one basic application of industrial manufacturing, automation control takes the closed-loop control system as its core components, where each sensor measures the real-time data for the controller to set the actuators, within a few milliseconds. This will benefit from our NFV-enabled 5G paradigm, which will reduce the communication latency to support the URLLC services for automation control.

(2)**Logistics tracking.** There are a number of stakeholders in current industrial logistics, including staff and high-priced goods in transit. The expensive expenditures of tracing them slow down the progress of industrial logistics. Fortunately, this will be resolved with NFV-enabled 5G paradigm, which will accurately and securely locate all staff and high-priced goods anytime anywhere at lower OPEX and CAPEX, main due to its URLLC and blockchain in tracking.

(3)**Cloud-based robots.** There will be ever-increasing cloud-based robots, who own the ability of self-organization and friendly cooperation in flexible production, in smart manufacturing. The cloud-based robots will control the real-time computing of manufacturing through migrating the massive computing into the cloud and MEC via mobile networks, as illustrated in Fig. 3. The NFV-enabled 5G paradigm will provide URLLC services to fulfill this requirement, enabling flexible industrial manufacturing in future.

(4)**Industrial augmented reality.** Augmented reality (AR) has played a key role in the monitor and production process, including manual assembly guidance and remote expert support. The AR equipment will obtain the necessary real-time information, for example, production environment and equipment data, from the cloud through wireless networks. In

order to ensure content synchronization between AR cameras and screens, the two-way transmission latencies should be no more than 10ms. The NFV-enabled 5G paradigm has the potential to provide URLLC-guaranteed services to promote smart manufacturing.

#### IV. PERFORMANCE ANALYTICAL MODEL

Faced by the strict transmission performance requirements of the industry applications, a new NFV-enabled 5G architecture was presented in the previous section, which jointly integrates NFV, SDN, and blockchain to improve the reliability, security and programmability of industrial service provisioning. Meanwhile, as shown in Fig. 1(b), the multi-layer virtualisation design of NFV-based 5G would lead to performance loss compared with the non-NFV systems with respect to the transmission latency. To theoretically analyse the performance loss of NFV-enabled 5G, we quantitatively study the worst case transmission latency of NFV-enabled 5G systems in the proposed architecture in this section. Different from the traditional queuing theory, which provides the average network performance, e.g., average latency, Stochastic Network Calculus (SNC) is used in this paper to obtain the reliability that a certain latency requirement can be met with given network resource allocation. Meanwhile, the work in [21] has revealed that the traffic generated by the blockchain system exhibits a high degree of the burstiness, which warrants an SNC-based analytical model capable of capturing the traffic burstiness. In this regard, a two-state Markov Modulated Poisson Process (MMPP) is exploited to model the traffic of blockchain-based architecture proposed in the previous section. There are two states in an MMPP process, “1” and “0”, respectively. The traffic rates when the system is in these two states, are denoted as  $\lambda_1$  and  $\lambda_0$ , respectively. The rates that the system transmits from/to state “1” to/from “0” are  $\varphi_1$  and  $\varphi_0$ , respectively. The number of VNFs is set to be  $n$ . In addition, the underlying 5G infrastructure is shared by multiple services and therefore it is necessary to capture the resource computation in the performance evaluation. In this regard, assume there are  $m_i$  traffic sources generated by other NFV chains at the  $i$ th VNF, which compete for the resources with the traffic generated by the interesting NFV chains. For each competing or cross traffic, the service rate is set to be  $\lambda_{cf}$ . The key notations used in the derivation of the performance model are summarised in TABLE I.

##### A. Modelling the Upper Latency Bound

In the SNC, the arrival process is denoted by the cumulative traffic,  $A_r(t)$ , which defines the amount of the packets entering the nodes of an NFV chain before the end of time  $t$ . Similarly, we use  $D_e(t)$  to represent the total number of the packets departing from NFV nodes before the time  $t$ . In addition, the total numbers of the packets entering and departing from NFV nodes between the time  $t$  and  $s$ , are represented by  $A_r(t, s)$  and  $D_e(t, s)$ , respectively. Given  $A_r(t)$  and  $D_e(t)$ , the service process, denoted by  $S(t)$ , can be calculated by the following equation,

TABLE I  
NOTATIONS

Parameter	Description
$A_r(t)/D_e(t)$	Accumulative traffic arrival/departure
$B_u(t)$	Accumulative traffic at the system
$w_{e2e}/W_{e2e}(t)$	E2E latency bound/at time $t$
$\epsilon/\theta$	Violation probability/Free parameter
$\rho$	Effective processing rate
$b$	Burstiness of the traffic process
$M_S/M_{A_r}$	MGF of an service/arrival process
$P(w_{e2e})$	Probability that a latency requirement is met
$Hr/\lambda_{cf}$	Unreliability entropy/Competing traffic rate
$n/m_i$	Number of VNFs/competing traffic
$\mu$	Service rate of the VNF node
$\varphi_1/\varphi_0$	Transmission rate from/to state 1 to/from 0
$\lambda_1/\lambda_0$	Traffic arrival rate in the state 1/0

$$D_e(t) \geq A_r \otimes S(t). \quad (1)$$

where ‘ $\otimes$ ’ stands for the min plus convolution operation in SNC and it is defined by  $A_r \otimes S(t) = \min_{\tau \in [0, t]} \{A_r(\tau) + S(\tau, t)\}$ . The accumulative traffic remaining the system is defined as the system backlog, ‘ $B_u(t)$ ’, which is obtained by  $A_r(t) - D_e(t)$ . Coupled with Eq. (1), the backlog is calculated by  $B_u(t) \leq \max_{\tau \in [0, t]} \{A_r(\tau, t) - S(\tau, t)\}$ . Similar to min plus, SNC also creates a new convolution operation, ‘ $\oslash$ ’ to calculate a max plus of two functions,  $A_r \oslash S(t) = \max_{\tau \in [0, t]} \{A_r(\tau, t) - S(\tau, t)\}$ . Then the backlog can be expressed as  $B_u(t) \leq A \oslash S(t)$ . For practical NFV deployments in the proposed system architecture, a single physical server will host multiple VNF nodes from different NFV chains. It is critically important to prevent performance impacts among different service chains. In this regard, let  $A_{th}(\tau)$  denote the accumulative traffic generated by the interesting NFV chain, and  $A_{cp}(\tau)$  the accumulative traffic from NFV chains that compete for the physical server or resources. Let  $W_{e2e}(t)$  represent the transmission latency of the NFV system at the time  $t$ . In the SNC, the value of  $W_{e2e}(t)$  is calculated by minimising the staying time. If packets arriving at the NFV system before the end of time  $t$  depart from the system before the end of the time  $t + v$ ,  $W_{e2e}(t) = v$ . Let  $w_{e2e}$  denote the worst case E2E latency bound for all  $W_{e2e}(t)$ . The objective of this study is to calculate the probability,  $1 - \epsilon$ , that the required latency bound  $w$  is not violated, which can be expressed as,

$$P(W_{e2e}(t) < w_{e2e}) > 1 - \epsilon. \quad (2)$$

It can be readily observed that the value of SNC for NFV performance analysis is that it can obtain the probability of  $\epsilon$  with given  $w$ , vice versa. In addition, given the cumulative arrival and departure, the E2E worst-case latency at time  $t$  can be calculated by

$$W_{e2e}(t) \leq \min_{w_{e2e} \geq 0} \{\max [A_r(\tau, t) < D_e(\tau, t + w_{e2e})]\}. \quad (3)$$

The methodology to obtain the upper bound latency is to firstly derive the traffic envelope for the traffic process,

$Ea$ , and service curve for service process,  $Es$ . Specifically,  $Ea$  defines the maximal amount of the traffic for a certain traffic arrival process. SNC describes this relation from the probability view. For instance, the work in [3] exploited the traffic envelope of  $Ea(t, \tau) = \rho(t - \tau) + b$  to characterise the traffic arrival,  $A_r(t)$ . SNC uses the following equation to calculate the traffic envelope of the traffic arrival,

$$P(A_r(t, \tau) < Ea(t - \tau)) \leq 1 - \epsilon(b). \quad (4)$$

By exploring the Chernoff bound, Eq. (4) could be transferred to  $P(A_r(t, \tau) < Ea(t, \tau)) \leq 1 - e^{\theta\sigma}e^{-\theta b}$ , where violation probability  $\epsilon(b)$  exponentially decays with the increase of  $b$  and is calculated by  $\epsilon(b) = \frac{e^{\theta\sigma}}{1 - e^{-\theta\sigma}}e^{-\theta b}$ . Similar to the traffic process, the accumulative service process can be characterised by the stochastic service curve,

$$P(S(t, \tau) > Es(t, \tau)) \leq 1 - \epsilon(b), \quad (5)$$

where  $Es(t, \tau) = \rho(t - \tau) - b$ . In addition, for a stable NFV-enabled 5G system, the traffic arrival rates should be smaller than the service rates. In the SNC, the traffic arrival rates are captured by the slope of the traffic envelop and the service rates are modelled by the slope of the service envelop. In this case, the slopes of the envelops in Eqs. (4) and (5) should satisfy the system stability requirement.

To obtain the worst case E2E latency, we need to calculate the Moment Generating Functions (MGF) for the traffic arrival and VNF service provisioning, which capture the effective traffic arriving and service capacity available in the system. According to [3], MGFs can be obtained by,

$$M_{A_r}(\theta, t) = E[e^{\theta A_r(t)}] \& M_S(-\theta, s) = E[e^{-\theta S(s)}]. \quad (6)$$

Consider there are two VNFs in a service chain, featured by  $M_{S1}(-\theta, s)$  and  $M_{S2}(-\theta, s)$ , then the work in [22] provides the approach to obtain an equivalent service process of two VNFs,  $M_{St}(-\theta, s)$ , expressed by,

$$M_{St}(-\theta, s) \leq \sum_{\tau=0}^t M_{S1}(-\theta, \tau) M_{S2}(-\theta, t - \tau). \quad (7)$$

Assume there are  $n$  VNFs in a service chain, the above approach could be used to iteratively derive the service capability of multiple VNFs in the form of MGF.

After obtaining the MGFs for the traffic arrival and the overall service capabilities of multiple VNFs, the worst case E2E latency experienced by a packet is given by,

$$w_{e2e} = \min_{\theta > 0} [\tau : \ln \epsilon - \frac{1}{\theta} \ln (\sum_{s=\tau}^{\infty} G(\tau, s, \theta)) > 0], \quad (8)$$

where  $G(\tau, s, \theta) = M_{A_r}(\theta, s - \tau) M_{St}(-\theta, s)$ . In this study, we assume the VNF nodes are processing the traffic based on the First-In-First-Out strategy. Then, the condition that Eq. (8) can obtain its minimal value is that the parameter,  $\tau$ , meets the constraint of

$$\ln \epsilon - \frac{1}{\theta} \ln (\sum_{s=\tau}^{\infty} G(\tau, s, \theta)) > 0. \quad (9)$$

Under the stability condition that the traffic arrival rate should be smaller than the service capability,  $\rho_{A_r}(\theta) \leq \rho_{St}(-\theta)$  and after applying the traffic envelop, service curve

and the MGFs of both the arrival process and service process in Eq. (9), the worst case latency bound,  $\tau$ , is obtained by,

$$\tau \geq \frac{\sigma(\theta)}{\rho_S(-\theta)} + \frac{n \ln \gamma - \ln \epsilon}{\theta \rho_S(-\theta)}, \quad (10)$$

where  $\gamma = 1 + \frac{1}{1 - e^{-\theta(\rho_S(-\theta) - \rho_{A_r}(\theta))}}$ . Eq. (10) captures the quantitative relationship between the reliability requirement,  $\epsilon$ , and the worst case latency,  $\tau$ , with the given system performance configuration and traffic arrival status.

## B. Modelling the Transmission Reliability

The worst case latency bound is derived in the previous subsection. For industry applications, 5G networks should not only provide the low latency for data transmission, but also need to offer the high reliability that the latency requirement can be satisfied at all time. To calculate the reliability that a certain latency can be met by the underlying NFV networks, we firstly rewritten Eq. (10) as follows,

$$\epsilon \geq e^{\sigma(\theta)\theta + n \ln \gamma - \theta \rho_S(-\theta)\tau}. \quad (11)$$

To guarantee the reliability of E2E transmission, multi-path transmission technology is used by 5G networks to split the traffic among multiple routings to improve the transmission reliability, where some VNFs in the NFV chains are distributively deployed to provide the network functions for each transmission routing. Furthermore, in the practical implementation of NFV-enabled 5G services, the NFV service chain can be deployed in both sequential and paralleled orders. In the former case, Eq. (11) could be directly used to obtain the violation probability for a certain service chain deployment strategy. While, for the latter case, multiple service paths would exist, and different service paths have different number of VNF hops and service resources. In this case, applying the proposed analytical model would lead to multiple violation probabilities to characterise the overall service reliability becomes a challenging issue for the proposed analytical model to be deployed in the parallel scenario. To solve this issue, we exploit the entropy of information theory to represent the uncertainty of traffic transmission for individual service path and use the principle of entropy addition to obtain the overall system uncertainty, which captures the degree of the traffic transmission reliability. Assume there are  $N$  paths for an NFV chain. Let  $P_i(w)$  denote the probability that the required latency could be met by the  $i$ th path. Then the transmission reliability is calculated from the following equation,

$$H_r = - \sum_i^N (1 - P_i(w)) \log_2 (1 - P_i(w)), \quad (12)$$

where  $H_r$  stands for the unreliability entropy of the E2E transmission. The larger value of the entropy, the lower transmission reliability that a certain system configuration can obtain, and vice versa. This metric refers to the degree of the unreliability for the multi-path transmission in 5G systems.

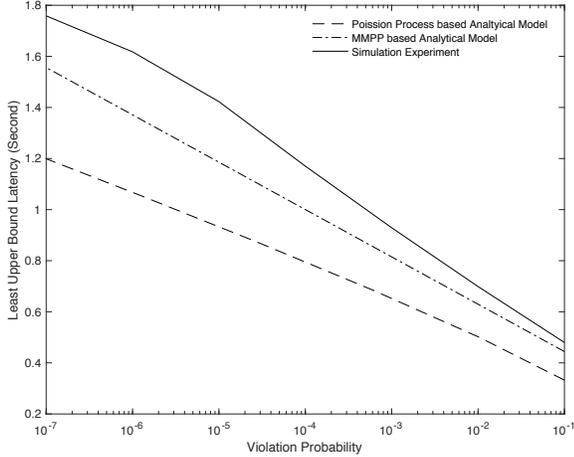


Fig. 4. Performance evaluation of the proposed MMPP-based analytical model, the Poisson based model and the simulation experiment by varying the violation probability.

### C. Performance Validation

In this section, we build an event-driven simulator based on OMNeT++ software to evaluate the performance of the proposed mathematical model. As described in Section III, there are multiple VNFs forming an NFV chain and multiple VNFs from different NFV chains would share the underlying physical devices. The traffic generated by the interesting VNFs follows an MMPP process and the traffic for the other VNFs sharing the same physical server follow the Poisson process. The MMPP process has two states (0 and 1). For the sake of illustration, the simulation parameters are configured as follows,  $\varphi_1=0.7$ ,  $\varphi_2=0.3$ ,  $\lambda_1=0$ ,  $\lambda_2=0.064$  Mbps and  $\mu_s=0.128$  Mbps. In addition, there are multiple NFV chains competing the computation resources of the physical servers and  $\lambda_{cp}=0.016$  Mbps. For the analysis of the E2E transmission reliability, the results of the classic queueing theory are also used to evaluate the accuracy of the proposed mathematical model, where the traffic arrival follows Poisson distribution. The classic queueing theory gave the worst-case latency as,  $\epsilon = \left[ \sum_{i=0}^{n-1} \frac{\mu_s ((1-\rho_s)w)^i}{i!} \right] e^{-\mu_s(1-\rho_s)w}$ , where  $\rho$  stands for the resource utilisation of the VNF node and calculated by  $\rho = \lambda/\mu_s$ . Therefore, apart from conducting the simulation experiments, the analysis results of the classical queueing theory with Poisson arrival traffic [23] are also used to predict the worst-case transmission latency of NFV-enabled 5G. The results predicted by the analytical models and collected from the simulation experiments are depicted in Fig. 4. It can be observed that both of the Poisson-based and MMPP-based analytical models achieve a similar trend of the upper bound latency with the increase of the violation probability. While, compared with the traditional queueing theory approach, the proposed analytical model obtains a more accurate performance prediction. This accuracy improvement is originated from the unique design of the proposed model, which can capture the bursty feature of the arrival traffic and resource competition among multiple NFV chains in the performance analysis. Furthermore, we can observe that there is still a gap between the simulation results from the OMNeT++ simulator and the analysis results of the proposed model, ranging from

9%-18%. This prediction inaccuracy is caused by the approximation processes during the mathematical derivation. While compared with the existing analytical work based on SNC approaches [3], the accuracy of 82%-91% is still acceptable for the performance analysis of the upper bound latency.

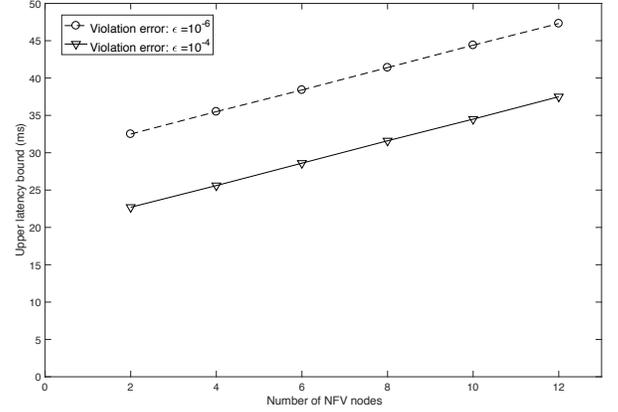


Fig. 5. The worst case latency by varying the number of VNF nodes with  $\mu_s=0.128$ Mbps,  $n = [2, 4, 6, 8, 10, 12]$ ,  $\epsilon = [10^{-6}, 10^{-4}]$ ,  $\varphi_1=0.7$ ,  $\varphi_2=0.3$ ,  $\lambda_1 = 0.064$ Mbps,  $m_i = 3$ , and  $\lambda_{cp_j}=0.001$ Mbps.

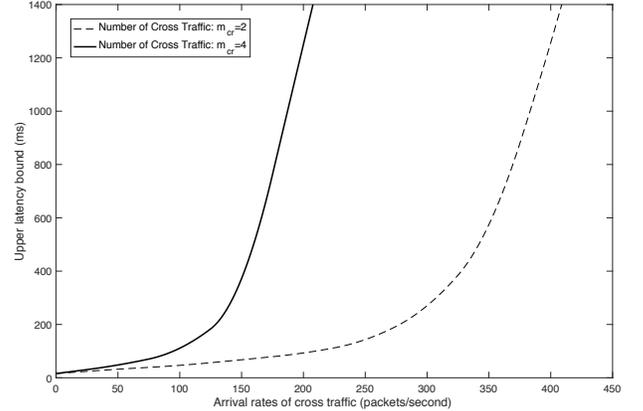


Fig. 6. The worst case latency by varying competing traffic rates with  $\mu_s=0.128$ Mbps,  $n = 2$ ,  $\epsilon = 10^{-6}$ ,  $\varphi_1=0.7$ ,  $\varphi_2=0.3$ ,  $\lambda_1 = 0.064$ Mbps and  $m_i = [2,4]$ .

### D. Performance Analysis

1) *The length of NFV service chains*: In the practical implementation of NFV services in 5G systems, a service chain will be composed of multiple VNF nodes, each of which will be responsible for a certain network function. During the lifecycle of the NFV service provisioning, the network service providers may need to increase the length of the NFV service chain to handle the unexpected situation, e.g., deep packet inspection to increase the security level, or remove a certain VNF node due to the completion of a certain network function. To provide the Service-Level-Agreement (SLA) guaranteed network service, the service provider needs to evaluate the impact of the adding or removing a VNF node on the worst-case transmission latency. Therefore, we exploit the proposed analytical model to predict the relationship between the length of the NFV

TABLE II

THE RELATIONSHIPS BETWEEN THE WORST CASE LATENCY (MS) AND THE VIOLATION PROBABILITY

Rate (Mbps)	The reliability requirement						
	$10^{-7}$	$10^{-6}$	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-2}$	$10^{-1}$
0.3	57.3	49.7	42.2	34.7	27.1	19.6	12.1
0.5	63.7	55.3	47.0	38.6	30.3	21.9	13.5
0.7	152.7	133.4	114.1	94.4	74.6	54.8	34.5

service chain and the worst case transmission latency in this subsection. The results of the worst-case transmission latency predicted by the proposed analytical model are depicted in Fig. 5. It can be seen that increasing the length of the NFV service chain would push up the worst case latency. This phenomenon keeps in line with the intuition of NFV service design. This is because the packet processing at each VNF would incur extra waiting time. The E2E transmission latency is the accumulative sum of the latency at each VNF. This reveals that if the strict latency is required for a certain industry application, the NFV chains should be designed with the least number of VNFs nodes. Following the design of NFV chains, the proposed analytical model could be used to revisit the NFV decision making and check whether the current design can satisfy the service requirement.

2) *Impact of the competing-traffic on the E2E transmission quality:* The underlying 5G networks will be shared by multiple NFV services. The traffic surge of an NFV service chain would impose the negative impacts on the other NFVs located at the same physical devices, referring as the resource competition. To provide the QoS-guaranteed 5G services for industry applications with strict delay requirements, we need to investigate the interaction between the resource competition of multiple NFV service chains and the worst case transmission latency, to ensure that the required latency can be met even with the serious resource competition. Towards this aim, the developed analytical model is used to obtain the upper bound latency with the increased arrival rate of the competing traffic. The results are shown in Fig. 6. Two scenarios are investigated with the number of competing traffic to be 2 and 4 respectively ( $m_i = [2,4]$ ). From Fig. 6, we can see that with the increase of the competing traffic, the degree of the resource competition becomes serious and the corresponding E2E latency is increased. In addition, through comparing the results of the solid and dashed lines, it can be observed that increasing the number of competing-traffic would also impose a negative impact on the E2E latency. Therefore, to guarantee the latency performance, 5G system should monitor the traffic conditions of each NFV chain to avoid performance degradation due to the surge of a certain NFV chain. In addition, the number of the NFV services sharing the same underlying physical devices should also be considered in the NFV orchestration and management to meet the strict latency requirement of industry applications.

3) *Relationship between the reliability requirements and the upper latency bound:* In this subsection, we firstly investigate the relationship between the worst case latency and violation probability in the case of the single-path transmission, where numerical analysis is conducted by varying the reliability

requirements and obtaining the worst case latency results. The system configurations are listed as follows,  $\mu_s=0.128$  Mbps,  $n = 2$ ,  $\epsilon = 10^{-1} - 10^{-7}$ ,  $\varphi_1=0.7$ ,  $\varphi_2=0.4$ ,  $\lambda_1 = [0.3, 0.5, 0.7]$  Mbps and  $m_i = 5$ ; the numerical results are demonstrated in TABLE II. We can observe that the higher reliability requirement (in the logarithmic form, e.g.  $10^{-7}$ ) would lead to an increased upper bound latency. This relationship can be proved from Eq. (11). If we impose log function on both sides of Eq. (11), it can be seen that the reliability requirement in the logarithmic form is bounded by a linear function with the variable of the latency.

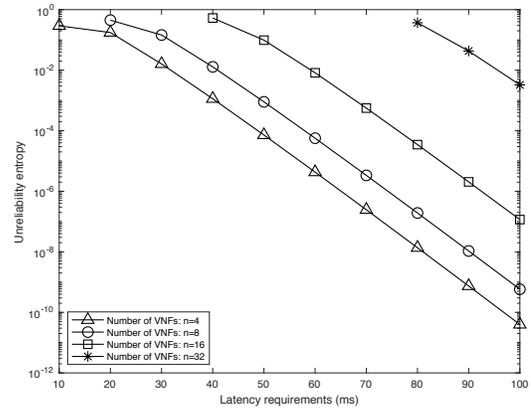


Fig. 7. Relationship between the transmission reliability and the latency requirements,  $\mu_s=0.128$  Mbps,  $n = [4,8,16,32]$ ,  $\varphi_1=0.5$ ,  $\varphi_2=0.1$ ,  $\lambda_1 = 0.4$  Mbps and  $m_i = 2$ .

Furthermore, we investigate the impact of the worst case latency requirement on the unreliability entropy of service provisioning in the case of the paralleled NFV design. The service reliability is calculated based on Eq. (12). The system configurations for each transmission path are set to be  $\mu_s=0.128$  Mbps,  $n = [4,8,16,32]$ ,  $\varphi_1=0.5$ ,  $\varphi_2=0.1$ ,  $\lambda_1 = 0.4$  Mbps and  $m_i = 2$ . Two transmission paths are considered in this scenario. Each equally shares the VNF traffic. The results predicted by the analytical model are depicted in Fig. 7. We can observe that when the latency requirement is relaxed from 10ms to 100ms, the transmission unreliability entropy is also reducing, which means the system becomes stable and the higher reliability could be obtained. On the other hand, when the strict latency requirement is posed, the unreliability entropy also increases, which aligns with the trend of the latency requirements. In this case, it can be concluded that the E2E transmission link may not be able to satisfy the predefined SLA requirements with respect to the transmission latency. In addition, when the length of the NFV chain is growing, for instance from 4 to 32 VNFs, under the same latency requirement, the unreliability entropy is also increasing, e.g. from  $10^{-10}$  for 4 VNFs to  $10^{-3}$  for 32 VNFs. This is because the traffic needs to visit more VNFs and stays more time in the network, making it challenging to satisfy the latency requirement. Therefore, to obtain low-latency and high reliability transmission, various aspects should be taken into consideration the design of NFV services, including the number of VNFs, the traffic conditions, the network resources as well as the service competition from other NFV services.

## V. CONCLUSIONS

NFV and blockchain have been regarded as the promising technologies for industry to realize URLLC services, integrated with the emerging network paradigms and technologies such as SDN and MEC, to address the current and emerging industrial application issues, and enable the design and deployment of 5G and beyond networks for industry. Both of them jointly present an extraordinary opportunity to rethink the current industrial networks. Therefore, we proposed an NFV-enabled 5G paradigm for industry with the service guarantee of URLLC, through the service chain acceleration and spectrum sharing built on NFV, blockchain, SDN and MEC. Furthermore, an analytical model was established to study the transmission capabilities of NFV-enabled 5G with respect to the worst-case latency bound and SLA violation probability with the input of the industrial bursty traffic.

## ACKNOWLEDGMENT

The authors are grateful to the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia for funding this work through the Vice Deanship of Scientific Research Chairs: Chair of Pervasive and Mobile Computing.

## REFERENCES

- [1] T. Koponen, S. Shenker, H. Balakrishnan, N. Feamster, I. Ganichev, A. Ghodsi *et al.*, "Architecting for innovation," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 3, pp. 24–36, 2011.
- [2] P.-V. Mekikis, K. Ramantas, A. Antonopoulos, E. Kartsakli, L. Sanabria-Russo, J. Serra, D. Pubill, and C. Verikoukis, "NFV-enabled experimental platform for 5G Tactile Internet support in industrial environments," *IEEE Trans. Ind. Inform.*, vol. 16, no. 3, pp. 1895–1903, 2019.
- [3] M. Fidler and A. Rizk, "A guide to the stochastic network calculus," *IEEE Commun. Surv. Tut.*, vol. 17, no. 1, pp. 92–105, 2014.
- [4] M. Dieye, W. Jaafar, H. Elbiaze, and R. Glietho, "Market driven multi-domain network service orchestration in 5G networks," *IEEE J. Sel. Areas Commun.*, 2020.
- [5] S. Sun, M. Kadoch, L. Gong, and B. Rong, "Integrating network function virtualization with SDR and SDN for 4G/5G networks," *IEEE Net. Mag.*, vol. 29, no. 3, pp. 54–59, 2015.
- [6] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [7] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "Network function virtualization in 5G," *IEEE Commun. Mag.*, vol. 54, no. 4, pp. 84–91, 2016.
- [8] W. Miao, G. Min, Y. Wu, H. Huang, Z. Zhao, H. Wang, and C. Luo, "Stochastic performance analysis of network function virtualization in future Internet," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 3, pp. 613–626, 2019.
- [9] X. Fu, F. R. Yu, J. Wang, Q. Qi, and J. Liao, "Dynamic service function chain embedding for NFV-enabled IoT: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 507–519, 2019.
- [10] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Commun. Surv. Tut.*, vol. 18, no. 1, pp. 236–262, 2015.
- [11] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surv. Tut.*, vol. 18, no. 3, pp. 1617–1655, 2016.
- [12] J. Wang, L. Wu, K.-K. R. Choo, and D. He, "Blockchain-based anonymous authentication with key management for smart grid edge computing infrastructure," *IEEE Trans. Ind. Inform.*, vol. 16, no. 3, pp. 1984–1992, 2019.
- [13] M. Zhaofeng, W. Xiaochang, D. K. Jain, H. Khan, G. Hongmin, and W. Zhen, "A blockchain-based trusted data management scheme in edge computing," *IEEE Trans. Ind. Inform.*, vol. 16, no. 3, pp. 2013–2021, 2019.
- [14] Z. Zhou, X. Chen, Y. Zhang, and S. Mumtaz, "Blockchain-empowered secure spectrum sharing for 5G heterogeneous networks," *IEEE Net. Mag.*, vol. 34, no. 1, pp. 24–31, 2020.
- [15] G. Praveen, V. Chamola, V. Hassija, and N. Kumar, "Blockchain for 5G: A prelude to future telecommunication," *IEEE Net. Mag.*, pp. 1–8, 2020.
- [16] M. Baggaa, T. Taleb, A. Laghrissi, A. Ksentini, and H. Flinck, "Coalitional game for the creation of efficient virtual core network slices in 5G mobile systems," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 469–484, 2018.
- [17] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Commun. Surv. Tut.*, vol. 17, no. 1, pp. 358–380, 2014.
- [18] X. Wang, C. Cavdar, L. Wang, M. Tornatore, H. S. Chung, H. H. Lee, S. M. Park, and B. Mukherjee, "Virtualized cloud radio access network for 5G transport," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 202–209, 2017.
- [19] J. Baranda, J. Manges-Bafalluy, I. Pascual, J. Nunez-Martinez, J. L. De la Cruz, R. Casellas, R. Vilalta, J. X. Salvat, and C. Turyagyenda, "Orchestration of end-to-end network services in the 5G-crosshaul multi-domain multi-technology transport network," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 184–191, 2018.
- [20] E. Hernandez-Valencia, S. Izzo, and B. Polonsky, "How will NFV/SDN transform service provider OPEX?" *IEEE Net. Mag.*, vol. 29, no. 3, pp. 60–67, 2015.
- [21] S. Geissler, T. Prantl, S. Lange, F. Wamser, and T. Hossfeld, "Discrete-time analysis of the blockchain distributed ledger technology," in *2019 31st International Teletraffic Congress (ITC 31)*, 2019, pp. 130–137.
- [22] M. Fidler, "WLC15-2: A network calculus approach to probabilistic quality of service analysis of fading channels," in *IEEE Globecom 2006*, 2006, pp. 1–6.
- [23] F. Ciucu, "Network calculus delay bounds in queueing networks with exact solutions," *Managing Traffic Performance in Converged Networks*, 2007.



networking.

**Haojun Huang** is an Associate Professor in the School of Electronic Information and Communications at Huazhong University of Science and Technology, China. He received his PhD degree in Communication and Information Engineering from the University of Electronic Science and Technology of China in 2012, and the BS degree in Computer Science from Wuhan University of Technology in 2005. His current research interests include Internet of Things, Network Function Virtualization, Software-Defined Networking, and Artificial Intelligence for



**Wang Miao** is currently a Postdoctoral Research Associate in the Department of Computer Science at the University of Exeter, United Kingdom. He received his PhD degree in Computer Science from the University of Exeter, United Kingdom in 2017. His research interests focus on Network Function Virtualization, Software-Defined Networking, Unmanned Aerial Networks, Wireless Communication Networks, Wireless Sensor Networks, and Edge Artificial Intelligence.



Computing, Ubiquitous Computing, Multimedia Systems, Modelling and Performance Engineering.

**Geyong Min** is a Professor of High Performance Computing and Networking in the Department of Computer Science within the College of Engineering, Mathematics and Physical Sciences at the University of Exeter, United Kingdom. He received the PhD degree in Computing Science from the University of Glasgow, United Kingdom, in 2003, and the BS degree in Computer Science from Huazhong University of Science and Technology, China, in 1995. His research interests include Computer Networks, Wireless Communications, Parallel and Distributed



**Jialin Tian** is currently a Master student in information and communication engineering at Huazhong University of Science and Technology, China. She received the BS degree in communication engineering from Northeastern University, China, in 2019. Her research interests include Network Function Virtualization and Blockchain.



**Atif Alamri** is currently a Full Professor in the Department of Software Engineering within College of Computer and Information Sciences (CCIS) at King Saud University (KSU), Riyadh, Saudi Arabia. He received the PhD degree in computer science from the School of Information Technology and Engineering, University of Ottawa, Canada, in 2010. His research interests include multimedia-assisted health systems, ambient intelligence, service oriented architecture, multimedia cloud, sensor cloud, Internet of Things, big data, mobile cloud, social networks, and recommender systems.