Received February 10, 2020, accepted March 2, 2020, date of publication April 6, 2020, date of current version April 21, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2986132*

Software-Defined Networking-Enabled Heterogeneous Wireless Networks and Applications Convergence

HAMADA ALSHAER, (Senior Member, IEEE), AND HARALD HAAS, (Fellow, IEEE)



School of Engineering, Institute for Digital Communications, The University of Edinburgh, Edinburgh EH9 3JL, U.K.

Corresponding author: Hamada Alshaer (h.alshaer@ed.ac.uk)

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) (TOUCAN Project) under

Grant EP/L020009/1. The work of Harald Haas was supported in part by the EPSRC through the Established Career Fellowship under Grant EP/R007101/1, in part by The Wolfson Foundation, and in part by the Royal Society.

 **ABSTRACT** A software-de ned networking (SDN) architecture is capable of integrating all radio frequencyand optical wireless small cell networks (e.g. fth generation (5G), long-term evolution (LTE) femtocell, wireless delity (WiFi), light delity (LiFi)) in one network domain. This paper considers a SDN-enabled heterogeneous network (HetNet) comprised of LiFi, LTE femtocell and WiFi access points (APs). The HetNet control plane maintains the state of the network topology and wireless resources, which can support the development of intelligent service provisioning and ef cient data communications in x generation (x G) wireless networks. The SDN applications use the network state to provide services in the data plane. However, when the state of network and wireless resources constantly changes, the SDN applications cannot provide reliable and guaranteed services to the wireless user equipments. This paper develops a queuing theoretic framework, which provides a performance evaluation for the SDN-enabled HetNet and applications convergence. A traf c engineering (TE) scheme is developed to support dynamic agnostic downlink ows routing to APs and differentiated granular services across the HetNet. Network and user centric policies are developed to make applications aware of network resource availability on the northbound and southbound interfaces of a SDN controller. Numerical models are introduced to study the impact of the computation and communication resources of northbound and southbound interfaces on the SDN-enabled HetNet scalability and the quality-of-service (QoS) guarantee of applications. Also, simulation scenarios are conducted to evaluate the performance of the TE scheme in provisioning effective and reliable services for subscribers.



 **INDEX TERMS** SDN controller, traf c engineering, heterogeneous wireless networks, QoS, LiFi, VLC,WiFi, LTE, software agents, 5G.



**I. INTRODUCTION**

Visible light communication (VLC) systems and light delity (LiFi) attocellular networks have been technologi-cally enhanced to support high data rate point-to-point (p2p) and multiuser wireless communications [1]. Radio and opti-cal wireless access points (APs) can coexist and operate in a small cell network without causing interference to each other, as shown in Fig. [1.](#page2) Next generation small cell networks are expected to have more APs and utilize 200x more spec-trum than the fourth generation (4G), which can support the

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

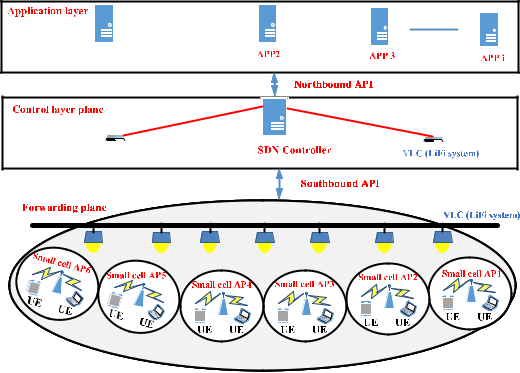
unprecedented growth in traf c volume, service diversity and bandwidth-intensive applications [2], [3].

With increasing small cell densi cation, the distances between users and APs are reduced. More wireless links become available for users to meet their quality-of-service (QoS) requirements [4]. A wireless user equipment (UE) can bene t from the wide service coverage of radio fre-quency (RF) wireless systems, and enjoy secure high-data rate communications in LiFi attocellular networks. Traf-c and user of oading in an integrated long-term evolu-tion (LTE) femtocell/wireless delity (WiFi)/LiFi network is viewed as a signi cant progress towards a true integration of RF and optical wireless technologies. The recent

|  |  |  |  |
| --- | --- | --- | --- |
| 66672 | This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/ | VOLUME 8, 2020 |  |
|  |  |

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

**FIGURE 1.** A SDN architecture for HetNet and applications convergence.



3rd generation partnership project (3GPP) 5G standard report [5] explains that the non-stand alone (NSA) archi-tecture enables the fth generation (5G) radio access network (AN) and its new radio (NR) interface to be used in conjunction with the existing LTE infrastructure packet core network. This makes the NR technology available with-out network replacement, while enabling the provided 4G services to enjoy the capacities offered by the 5G New Radio [5] [7].

However, the current 3GPP wireless networks integration options are not exible to enable ef cient multi-radio connec-tivity [8], yet alone the integration of radio and optical wire-less technologies. To this end, any non-3GPP access network like LiFi or WiFi should have a termination point at the LTE packet data network gateway (P-GW) to enable UEs route their traf c through an integrated LTE/WiFi/LiFi HetNet, following the mobile internet protocol (IP) principles [9]. The existing 3GPP architecture [8] does not solve the problem of ef cient multi-radio and optical wireless networking, because every time the data path is switched from LTE to WiFi or LiFi and vice-versa, some packets are lost, while simultaneous usage is not possible [9].

A software-de ned networking (SDN)-enabled heteroge-neous wireless network (HetNet) is comprised of LiFi, WiFi and LTE APs, which should be designed to support diverse services with different data traf c pro les. These can be provided by operators and/or trusted third parties. Smart 5G-enabled UEs, which have multiple air interfaces, can be set by preference to communicate through different wire-less networks. For example, a UE may be set to down-load music through the WiFi interface. A UE streams live high de nition (HD) videos and establishes secure wire-less communications through the LiFi interface. A UE with high mobility makes on-the-move voice over internet pro-tocol (VoIP) calls through the LTE interface. In HetNets, resource availability, users mobility, traf c volume, connec-tion duration and service requirements change constantly. To this end, they should be planned, and resources be dimen-sioned to support autonomous and adaptive reliable services

VOLUME 8, 2020

provisioning to UEs and service providers (SPs) offering vir-tual networks [10]. They should also support intelligent traf c routing in the user plane to cope with the UE mobility during active communications. A centralized SDN architecture pro-vides a platform that can meet these requirements [11], [12].

The SDN architecture decouples the control plane and the data plane of the HetNet, as shown in Fig. [1](#page2) [13]. The net-work events (e.g. AP failure and services disruption), traf c statistics and payload analysis are presented in information reports that describe the network state. The SDN controller collects it through the southbound and advertises it peri-odically through the northbound application programming interfaces (APIs) to the application plane [11]. The SDN applications manage, through the controller, OpenFlow (OF) rules in the different APs to support the offered services in the data plane. A tting traf c engineering (TE) scheme can leverage the centralized view of the network state to support intelligent traf c ows routing and network function virtualization (NFV) modules [14]. For example, it can route traf c ows from failed, congested or underutilized APs to others, which can improve network resource ef ciency and support provisioning mission-critical applications [9].

The proposed SDN architecture provides an agnostic net-work platform which can embody the integration of radio and optical wireless technologies in line with the 5G stan-dardization roadmap [7], [8]. The SDN controller can be interconnected through its east and west interfaces to the 5G core network via a non-3GPP inter-working function (IWF) on the N3 interface (N3IWF). This N3IWF interfaces the 5G core network control plane (CP) and user plane (UP) functions via N2 and N3 interfaces, respectively [6] [8]. The N2 and N3 reference points are used to connect standalone non-3GPP accesses to 5G core network CP and UP func-tions, respectively [6] [8]. This enables the SDN-enabled HetNet to bene t from the full services supported in the rst phase of 5G. This paper extends the SDN concepts to manage, in an agnostic manner, LiFi, WiFi and LTE small cells interconnected through the SDN architecture shown in Fig. [1.](#page2) Note that the SDN control plane implementation for wireless networks is more complex than the wired optical networks [15]. They must account for physical constraints, including channel gain variations, bandwidth availability and granularity, heterogeneous multiuser access and ows routing recon guration speed.

This paper evaluates a queueing model that helps to better understand the dynamic of traf c volume on the north and south bound interfaces of the SDN controller and wireless APs in the HetNet data plane. We develop an analytical mathematical framework for modelling the proposed SDN architecture shown in Fig. [1.](#page2) We are motivated by developing an agnostic TE framework for a SDN-enabled HetNet, which supports autonomous traf c load routing, ef cient resource and service resilience provisioning. These help to support a exible and intelligent integration of the LTE/WiFi/LiFi Het-Nets. Analytical and simulation analysis have been developed to corroborate the mathematical analysis and the comparison

66673

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

of the proposed TE routing policies. The convergence of multiple heterogeneous wireless networks enable them to operate seamlessly under the SDN architecture, which is an important enabler for ef cient service provisioning.

The remainder of this paper is organized as follows. Section [II](#page3) presents the HetNet system model. Section [III](#page4) states the research problem and challenges in the network, service and application dimensions. Section [IV](#page4) summarizes the main research work in the literature regarding queuing theory for TE in HetNet, SDN-enabled switch and controller modelling. It also summarizes the contributions of this paper. Section [V](#page6) introduces mathematical queuing models for the different planes of SDN-enabled HetNet and discusses the analytical results. Section [VI](#page9) introduces the proposed TE framework modules. Section [VII](#page14) explains the simulation results. Section [VIII](#page18) introduces the key future research chal-lenges. Finally, Section [IX](#page19) concludes this paper and presents the main ndings.

**II. SDN-ENABLED HetNet SYSTEM MODEL**

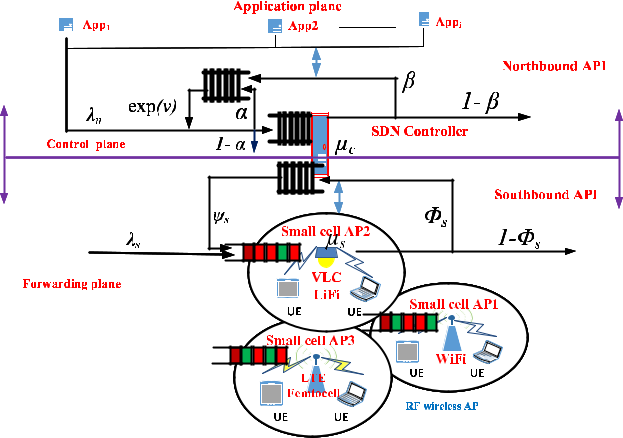
The SDN-enabled HetNet system model is composed of data (forwarding), control and application planes. A number of SDN-enabled heterogeneous LiFi, WiFi and LTE wireless APs are con gured in the data plane. They are managed by a SDN controller that runs in the control plane. An orthog-onal frequency-division multiple access (OFDMA) protocol allocates the available resources on the downlink channels of LiFi and LTE APs. A wireless resource is a slot that occupies a space in the time and frequency domains. It is the smallest resource unit which can be allocated to a UE. An average number of slots, *S*, are available for data transmission in the downlink part of each time frame, *TF* . The wireless channel states are represented by a number of modulation and coding schemes (MCS) *i*, 1 *i I* . A UE uses the MCS*i* to transmit a number of bits, b*i*, per slot. At each frame time-step, T*F* , a UE has the probability, p*i*, to use the MCS*i*.

The software switch, Open vSwitch [16], runs in the APs, which makes them OpenFlow (OF) enabled switches. They are connected to the controller through a tree topol-ogy via an OF-enabled switch, as shown in Fig. [1.](#page2) A set of northbound representational state transfer (REST) APIs enable communications between the application plane and the control plane, which support various applications. For example, users association and traf c engineering, security, access control, network resilience, etc., as shown in Fig. [2.](#page3) Similarly, a set of southbound open APIs (e.g., OF protocol) enables communications between the data plane and the con-trol plane, which supports the reporting of a global HetNet view (i.e. the state of resources in the data plane and network topology connectivity) and setting traf c forwarding rules in the routing tables of AP switches. Whenever AP switches receive new ows, it requests the controller to set forwarding rules in the switches that will handle them [17].

The downlink channel gains of wireless APs vary in time and space, which justify the assumption of the gen-eral (G) service distribution of the APs. Each small cell

66674

**FIGURE 2.** SDN-enabled HetNet architecture and applicationsconvergence modelling.



has a single AP to serve the UEs. Each AP switch has a buffer of size *K* packets. The arrivals of UEs and downlink ows follow a Markovian (M) Poisson distribution. Thus, an M/G/1/K queuing model [18] can be adopted to describe the networking operations of AP switches in the data plane, as shown in Fig. [2.](#page3) As the controller manages *N* switches in the data plane, an M/G/1/K/N queuing model can be adopted to describe the traf c packets handling operations on the southbound interface of the SDN controller. The traf c ows that arrive at the switch follow a Poisson process with arrival rate, *s*. The service times at the AP switches are assumed to

follow an exponential distribution with an average of 1 . The

*s*

service time at the SDN controller is assumed to follow an

exponential distribution with average rate, 1 . A traf c ow

*c*

with a pre-set rule in the AP switch is served and forwarded

with a probability, *s* D 1 *s*. However, if a traf c ow does not have a rule set in the AP switch, a packet-in is sent to the SDN controller, with a probability, *s*, to de ne its for-warding rule in the AP switch. The input traf c rate from the AP switch to the controller and vice-versa can be expressed statistically as, *s s*, and *s s s*, respectively. Note that *s* denotes the probability of the traf c without preassigned rules are successfully assigned rules by the controller.

Some of the SDN applications, which need to use the same resource advertised as available in the current network state, may try again to access it in the next round of network state advertisement. The retrial access happens due to a huge vol-ume of requests generated from the SDN applications on the northbound interface; or resource insuf ciency and unavail-ability in the data plane. This happens when the applica-tions require constant networks state update and simultaneous increase of network resources. To this end, the M/M/1 retrial queue with geometric loss and feedback [19] can be adopted to model the access of applications to the data plane through the northbound interface of the SDN controller. The arrival rate of applications requests sent through the northbound interface is denoted by *n*. The requests of applications which cannot be admitted at the controller join a retiral queue with a

VOLUME 8, 2020

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

retrial probability, , or dropped with a probability, D 1 . After a random amount of time, the requests in the retrial queue retry again following an exponential distribution with average rate 1*v* to access the network resources. An applica-tion that requires a single service leaves the SDN controller with a probability ( D 1 ). If it requires further services like an increase of bandwidth or other resources, it rejoins the retrial queue with a probability .

**III. PROBLEM STATEMENT**

Network function virtualization (NFV) orchestration, user access and traf c engineering applications provide services across the HetNet to support the QoS and quality-of-experience (QoE) requirements of users. These run appli-cations on UEs that request services from external endpoints like an edge computing or a cloud. The developed network access and traf c engineering application manages users association to APs and resource allocation to user appli-cations in an ef cient-manner. It takes into consideration user priority, HetNet resource granularity and allocation ef - ciency. When the number of deployed APs and active users increases in the network, it becomes more challenging for the control plane to apply fast forwarding rules and support concurrent applications access to the available resources in the data plane. Note that every service traf c ow requires proper forwarding rules to be set up in the APs that serve the users requested it. For example, suppose that there are user association and traf c engineering, and user localiza-tion (tracking) applications running in the SDN applica-tion plane. The former requires information regarding the resources availability to engineer users association to APs and resource allocation to users applications, whereas the latter requires information regarding the users in the network. How-ever, both applications compete for the available network resources they need to transport their traf c back and forth among servers, SDN controller and users. During the next state collection time period, the SDN controller advertises on the northbound interface the network state in terms of the available bandwidth on the downlink channel of AP*i*. Based on the received network state, both applications make attempts to request the controller to con gure forwarding rules in the AP*i* to use the same available bandwidth resource. However, this may no longer be enough to support both of them. In this case, one of the applications needs to retry again during the next state collection time period to access the same resource in the forwarding plane. To this end, the constant changes in the global state of the HetNet make it challenging for the SDN applications to provide consistent and ubiquitous services to their UEs in the data plane. A research challenge lies in developing a mathematical model that can evaluate the consolidation of the control plane in a centralized SDN controller and its impact on the network performance and scalability. Moreover, the HetNet has APs with different transmission ranges, service coverages and data rates. The UEs have various connection preference trade-offs among the bandwidth, latency, throughput and seamless connectivity.

VOLUME 8, 2020

The centralized controller collects the channel state infor-mation, QoS requirements of users and load state of APs. A research challenge lies in developing user association to APs and agnostic traf c ow routing algorithms. These should run on any AP, independent of its underlying wireless technology, while meeting the service requirements of user applications.

**IV. RELATED WORK**

The operations underlying the OF switch, SDN controller and their interconnections have been investigated by researchers to develop mathematical models that can evaluate the perfor-mance and scalability of SDN-enabled networks and services provisioning. The following subsections discuss the main developed models for the SDN switch, controller, architecture and software-de ned wireless networks.

**A. SWITCH AND CONTROLLER MODELLING**

An OF switch should realise fast look up tables that match ows to forwarding instructions. The SDN controller should have enough processing capacity to handle all the messages coming from the OF switches on the southbound interface and the applications running on the northbound interface. The controller handles a large number of packets per second just to monitor the network state. This is not just a burden on the controller, but it also increases the load on the network control plane. A software Open vSwitch [16] can run with different hardware and operating systems, which support scalable OF operations in network switches. An OF switch is implemented on a networking eld-programmable gate array (NetFPGA) platform [20]. It is used to conduct perfor-mance analysis on forwarded traf c ows. An OF switch [21] is deployed on a Linux platform to evaluate the performance of multiple forwarding switches and compare switching and routing ows at layers 2 and 3, respectively. A modular and parametrised implementation of a hardware-based OF switch is implemented on three different platforms: NetF-PGA, FPGA ML605 and DE4 models. The performance evaluation of these platforms shows that the OF switch design can be implemented across them with minor performance variations [22].

In [23], the authors used a network calculus-based approach [24] to derive upper bounds for the buffer length of a single OF switch and controller. The upper bounds for the service busy period of the switch and controller were derived to calculate the worst delay of packet processing. However, the derived bounds are deterministic and apply only to an OF switch con gured in xed networks. They cannot be used to derive an end-to-end delay bound for traf c ows traversing multiple OF switches. In [17], the authors modelled an OF switch and controller as an M/M/1 and an M/M/1/S queuing systems, respectively, where *S* denotes the controller queue length on the southbound interface. The developed model considers a bidirectional feedback between the switch and controller.

66675

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

In [25], the authors proposed an approach that identi es the elephant ows based on some SDN information (e.g. ows, packets or bytes counters) on the different ports of network switches. This research study is useful to understand the relationship between the collected traf c statistics and the OF rules that need to be con gured in the OF switches to handle long and short traf c ows. However, the proposed approach may not support a large-scale data centre. Note that a single OF rule can be con gured in an OF switch to handle an aggregate of traf c ows shaped through a leaky bucket controller (LBC) at the network edges to receive the same forwarding services. In [26], the authors developed a network calculus-based admission control mechanism that ensures enough allocation of bandwidth and buffer space for complied traf c ows. These are serviced by the switches along a path set up by a SDN controller. However, the buffer and bandwidth allocation are based on a loose bound, which cannot accurately quantify the bandwidth for the different ow aggregates. In [27], the authors considered a Jackson network composed of some OF switches. They evaluated the performance of traf c in the data plane under nite and in nite buffer size scenarios at the switches. The packets, which are sent by the controller to install OF rules in the switch, cannot be sent back to it again. In [17], the authors developed an approximate model for the operations underly-ing the southbound interface. It does not distinguish between the external traf c and the packets sent by the controller to the switch. The packet-in requests for setting rules in the switches are modelled as a batch of ows, with size X, arriving at the controller queue, which is modelled as an M[*X* ]/M/1 queuing model [27]. This work applies mostly to xed networks. It also does not consider link dynamics in the forwarding plane. Most of the above developed SDN models consider a single switch in the data plane intercon-nected with a controller, which require further analysis to investigate the support of real-time applications. Also, none of the developed models [20] [22] provides a comprehen-sive and generic framework for evaluating the performance of SDN-enabled wireless networks. The developed queuing model helps to evaluate the impact of packet arrival rate and ow size of external traf c on the performance of the controller and switch. However, more research work is still needed to understand the performance of a SDN controller that manages multiple AP switches in a wireless network.

1. **SOFTWARE-DEFINED WIRELESS NETWORKS ARCHITECTURE**

SDN architectures and solutions are increasingly adopted in the context of wireless networks to enable network programmability and support mobile operators to manage resource allocation and service provisioning to end users. In [13] the authors proposed a software de ned radio access network (SoftRAN) architecture that encapsulates all the base stations in a radio access network (RAN) represented as a virtualized big base station. This centralizes the control and management of RAN. The paper only explains the bene ts

66676

of this architecture to support applications and use-cases in wireless networks. In [28] the authors discuss the capa-bilities of 3GPP network management and evolved-UMTS (Universal Mobile Telecommunications Service) Terrestrial Radio Access Network (E-UTRAN) architectures to sup-port self-organizing wireless networks, automation, mobility and load balancing. This is an overview paper, which only explains self-organizing networks (SON) 3GPP standardiza-tion activities. It discusses the 3GPP architecture, without explaining how this can support small cell wireless networks integration.

A SDN architecture can be used to interconnect data centres to support traf c engineering functionalities in ultra-dense networks (UDNs). The virtual cell technology has been widely used to improve the user experience of UDNs. By adopting the virtual cell technology, a new type of UDN, termed user-centric UDNs, could be developed [29]. The main idea of user-centric UDNs is to cluster APs dynamically to provide better services for end-users. Based on the virtual cells in UDNs, a new type of beam forming technology named the balanced beam forming algorithm was developed to optimize the network capacity. In [30] the authors summa-rize the main challenges in dense user-centric cloud random access networks (C-RAN). They advocate the application of dense user-centric C-RAN approaches to UDNs using some cloud computing techniques.

In [31] the authors developed an analytic model for evalu-ating the queueing delays and channel access times at nodes in 802.11 based WiFi networks. Each node is modelled as a discrete time G/G/1 queuing system model, which is used to derive closed form solutions to obtain the values of the delay and queue length. In [9] the authors proposed a high-level architecture for tight integration of WiFi-LTE systems. The idea is to integrate WiFi into a viable 3GPP architecture. A Markov chain model framework was developed to analyse the performance of the proposed LTE-WiFi UDNs. It pro-vides some insights regarding the gained bene ts in terms of average data rate per user and service coverage in the considered area. In [3] the authors provide a comparison study for different AP densities with various bandwidth in an UDN. The paper [3] proposes an approach for increasing the average user capacity in an UDN to 1 Gbps.

The support of SDN for control and forwarding isolation for traf c ows and the centralized network management provides a platform for realizing traf c engineering based on traf c measurement and management [32]. The integra-tion of SDN in legacy wired networks can improve traf c engineering [33]. In [34] the authors proposed a dynamic cell speci c uplink/downlink (UL/DL) frame recon guration mechanism, which can customize the TD(time division)-LTE frame ratio to ef ciently meet the QoS requirements of UEs. This research study only considers the local state of frame and traf c load in cells, without considering the network state. In [35] the authors demonstrated the capability of a SDN management algorithm to re-direct traf c ows to mobile femtocell, RF AP or LTE eNodes. The obtained results show

VOLUME 8, 2020

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

some improvement in the user experience, which is quanti ed in the increase of their service satisfactions, throughput and the reduction of waiting times. The algorithm also improves cellular resource utilization through of oading traf c ows to WiFi and mobile femtocells. It ensures load balancing based on ow admission control to each supported traf c technology.

The above related work handled the research problems from different angles, though more analysis is required to investigate services provisioning performance and applica-tions support in the network. The related studies are focused on some entities in the proposed SDN architecture, whereas a robust SDN solution should consider the different SDN planes. The previous developed approaches propose solutions that cannot run in an agnostic manner, independent of the wireless technologies underlying the different APs in the SDN-enabled HetNet.

**C. CONTRIBUTIONS**

The contributions of this paper can be described as three-fold. The rst uses queuing theory principles and models to develop a mathematical framework for modelling the inter-planes of the SDN-enabled HetNet and applications convergence. The second fold introduces a TE scheme that runs at the network and medium access layers. Routing poli-cies are developed to support TE and load balancing on the network layer and multiple services on the medium access layer. The proposed TE scheme handles the dynamics of the SDN-enabled HetNet through enabling the AP switches to dynamically serve ows per frame basis. The SDN controller exposes appropriate feedback information for applications to improve the QoS and QoE of their UEs. The third fold develops network and service provisioning automation mech-anisms, which can support service resilience, dynamic APs selection and resource allocation for UEs.

**V. INTER-PLANE INTERFACES MODELLING**

The SDN controller has horizontal (i.e. east, west) and verti-cal (i.e. northbound, southbound) interfaces, though, in this paper, we will focus on modelling the vertical interfaces. A buffer on the northbound interface keeps and passes the requests generated from the SDN applications to access the data plane. Likewise, a buffer on the southbound interface passes the packet-in rules and incoming packets to the con-troller and switches, respectively, as shown in Fig. [2.](#page3) The fol-lowing subsections introduce mathematical queuing models for the northbound and southbound interfaces, and data plane switches.

**A. NORTHBOUND INTERFACE MODELLING**

The SDN applications view the state of resources in the Het-Net data plane through the northbound interface of the control plane. They offer services via the centralized SDN controller by requesting to set up their forwarding rules in the data plane switches through a southbound protocol (e.g. Open-Flow). The applications may renege on accessing the data

VOLUME 8, 2020

plane, because of resource unavailability or controller pro-cessing capacity constraint. This reneging process in uences the performance of network services provided throughout the HetNet data plane.

Based on the HetNet system model, the SDN applica-tions generate requests following a Poisson process with arrival rate, *n*. The M/M/1 queuing model describes the packets buffering and processing at the northbound inter-face of the controller. We adopt the M/M/1 retrial queue-ing system model with geometric loss and feedback [19] to model the requests processing for data plane access through the northbound interface, as shown in Fig. [2.](#page3) The state of the SDN-enabled HetNet is described by a pair ( (*t*); *N* (*t*)), where (*t*) denotes the number of busy controllers and *N* (*t*) denotes the number of requests in the retrial buffer at time *t*.

* stochastic process ( (*t*); *N* (*t*)) V *t* 0) is formed as a time-homogeneous Markov process with a state space ( ; *N* ) as a limiting variable of ( (*t*); *N* (*t*)). When the controller receives a small number of requests from the network appli-cations to access the data plane, the average queue length of the controller is given as follows [19]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *E*(*N* V D0) | | | | | | | |  |  | C |  |  | *c* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| D *C* |  | | ( | |  | *c* C *v* | | | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | | |  |  | *v*) | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *F* | | | | | | ( | | *c* C | | *v* | | C | |  |  |  | *c* I | | | | |  |  |  |  | *c* C *v* C | | | | | | | |  |  |  |  |  |  |  | *c* | | | | I *vc* ! | | | | |  |  |
|  |  |  |  | | | | |  |  | 2*v*) | |  |  |  | | |  |  |  |  |  |  | | ( | |  |  |  | *v*) | | | | |  | | | | | | |  |  |  |  |  | |  |  |  |  |  |
|  |  |  | |  |  |  |  |  |  | | |  | |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | *F* | | | | |  |  |  | *c* C *v* | | | C | |  |  |  |  | *c* I | | | | ( | | | |  | *c* C *v* C | | | | | | |  |  |  |  |  |  |  |  | *c* | | | I *vc* ! | | | | |  |  |
|  |  |  |  |  |  | ( | | |  |  | *v*) |  |  |  | | | |  |  |  |  |  |  | |  | |  |  |  | *v*) | | | |  | | | | | | | |  |  |  |  | |  |  |  |  |  |
|  |  |  | |  | |  |  | | | | | | | |  |  | | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | *c*( *c* C *v*) C *c c* | | | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | *v*( |  | | ( *c* C *v*) C *c* | | | | | | | | |  | ) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | *c* I *vc* ! | | | | |  |  |
| *F* | | | | | |  |  | *c* C | | *v* | | C | |  |  |  | *c* I | | | | | ( | | | | *c* C *v* C | | | | | | | | | |  |  |  |  |  |  |  |  |  | ; |  |
|  |  |  | ( | | | | |  |  | 2*v*) | |  |  |  | | |  |  |  |  |  |  | |  | |  |  |  | 2*v*) | | | | |  |  |  | | | | | | | |  |  | | | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

where *F*(*a*I *b*I *w*) is the Kummer’s function; and *C* is a nor-

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| malizing constant, given by [19]: | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |
|  | D | |  |  |  | *c* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *C* |  |  | *c* C *c* | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | |  | | |  |  |  |  | I |  |  | *V* | | | I *V* | | | | ! |  |
|  |  |  |  |  |  | C *V* C | | | |  |  |  |  |
|  |  |  |  |  | ( *c* *V* ) | | | | | |  |  |  |  |  |  |  |  |  |  |  | 1 |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | *c* | | *c* C *c* | | | | | *c* | | | |  |
|  |  | *F* | | | : (1) |  |
|  |  |  | | | | | |  |  |  | |  |  | |  |  |  |  |  |

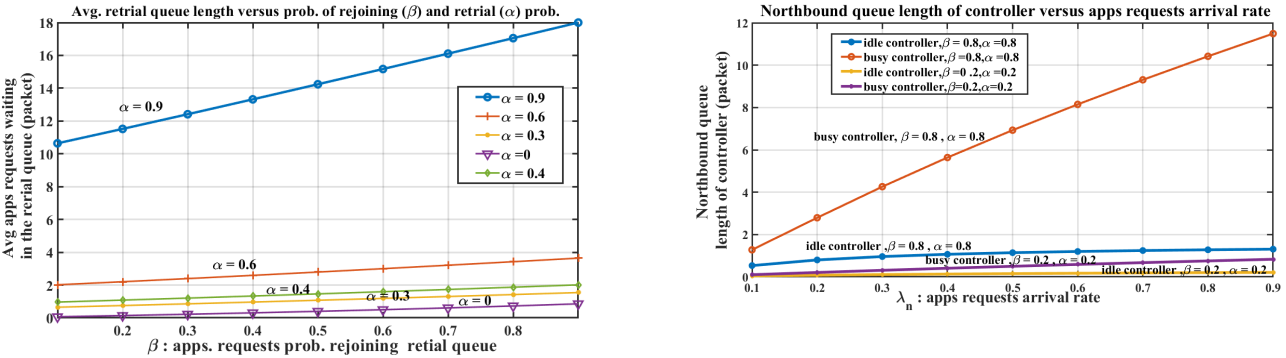
However, when the number of requests generated from the SDN applications increases, the controller may become busy on its northbound and southbound interfaces. In this case, the average queue length of the northbound interface is expressed as follows [19]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *E*(*N* V D1) | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| D | *C* |  | *c*( *c* C *v*) C *c c* | | | | | | | | | |  | | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | ( |  |  | *v*( *c* C *v*)C *c* | | |  | *v*) | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | |  |  |  |  |  | *c* I *vc* ! | | | | | |  |  |
|  | *F* | | ( | | | | *c* C *v* C |  |  |  | *c* I | | | | | *c* C *v* C | | |  |  | : |  |
|  |  |  |  | | | | 2*v*) |  | | |  | | |  | ( | 2*v*) | | |  | |  | | | | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

(2)

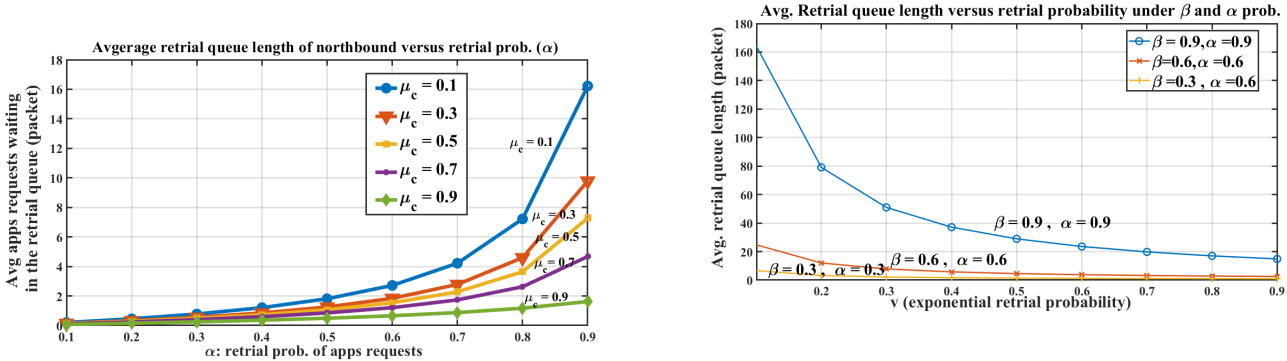
66677

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence



**FIGURE 5.** Average queue length of northbound controller versus **n**.

**FIGURE 3.** Average retrial queue length versus .



**FIGURE 6.** Average retrial queue length versus **v**.

**FIGURE 4.** Average retrial queue length versus .

When the SDN controller runs beyond its capacity, the net-work is down or resources in the data plane are overbooked, the excess requests generated from the SDN applications are kept in the retrial queue or are dropped. In this case, the aver-age length of the retrial queue is expressed as follows [19]:

*E*(*N*)

D *E*(*N* V D0)C*E*(*N* V D1)

( *c* C *v*) C *c*

* *C*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | *v* | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | *vc* ! | | | | | |  |  |
| *F* | ( | | | *c* | C *v* C |  | *c* I | | | | | | | | *c* C *v* C | | | | | |  |  |  |  | *c* I | | | |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  | |  | 2*v*) |  |  |  |  |  |  | | ( | |  |  |  | *v*) | | |  | | | |  |  |  | | | | | | | |  |  |  |  |
|  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *F* | ( | | | | *c* C *v* C |  |  | *c* I | | | | | | |  | *c* C *v* C | | | | |  |  |  |  |  | *c* I | | |  |  |  | *vc* ! | | | | | : |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | | | | *v*) |  | |  | | | |  | | ( | |  |  | *v*) | | |  | | | | |  | | | | | | | | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | |  | |  | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | (3) | | | | | | | | |  |

Analytical models have been developed in MATLAB, which evaluate [(1), (2)](#page6) and [(3)](#page7). They give some indications regarding the relationship among the applications requests, queue length of the northbound interface, retrial queue length and controller processing rate. Based on [(3),](#page7) the request retrial probability, , in uences the retrial queue length more than the rejoining probability, , as shown in Fig. [3.](#page7) This is attributed to the fact that not all the rejoining requests wait in the retrial queue length. The applications, which are allocated resources in the previous round, are quickly served by the controller. Based on [(3),](#page7) the controller service

66678

rate, *c*, directly in uences the retrial queue length. this even grows more rapidly with the increase of the requests retrial probability, , as shown in Fig. [4.](#page7) This also happens when the controller is busy or requested resources in the network data plane become unavailable during the access time of applications. This emphasises the importance of the controller processing rate and requested resource availability for applications to provide reliable services in the HetNet data plane. Based on [(1),](#page6) the retrial and rejoining probabilities signi cantly impact the queue length of the northbound inter-face, as shown in Fig. [5.](#page7) When the exponential probability, *v*, of sending requests to the controller increases, it is obvious to see that the retrial queue length decreases. However, based on [(2)](#page6) and [(3),](#page7) the queue length of the northbound interface still depends on the controller processing rate, retrial and rejoining probabilities, as shown in Fig. [6.](#page7) When the rejoining and retrial requests decrease, the average retrial queue length decreases as well and vice-versa. This case indicates that the controller can manage the allocation of resources that are requested by the different SDN applications. New ow rules can also be set in the APs based on the retrial and rejoining probabilities range.

**B. SOUTHBOUND INTERFACE MODELLING**

As mentioned before, the SDN applications can manage the data plane traf c through requesting the controller to con-gure their OF rules in the APs to deliver network services. Thus, the APs and controller may become a bottleneck for provisioning services in the data plane, where traf c ows

VOLUME 8, 2020

And *c* denotes the controller throughput in terms of the number of processed in-packet requests per unit of time, and

VOLUME 8, 2020

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

wait for a longer time in the queue of APs to set up their OF rules. The queuing model, M/G/1/K/N, which captures the operations of the SDN controller on the southbound interface, is proposed to investigate the impact of the buffer size, number of switches and traf c ow rate on the network performance. It is used to evaluate the queue length and response time of the controller and AP switches under traf c load.

The external traf c load rate, *s*, and the traf c load rate of SDN applications, *n*, represent the total traf c load rate at the southbound interface of controller, which also follows a Poisson distribution with a traf c arrival rate, expressed as: *c* D *n* C *s s*. The traf c load intensity at the controller is expressed as: *c* D *c* . When a new ow arrives at an AP that

*c*

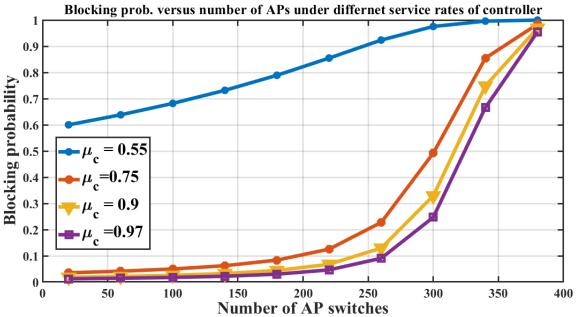
does not have a correspondent forwarding rule, an in-packet request is sent to the controller to de ne the forwarding rule. However, if the in-packet request arrives, where the controller has already *K* packets in its buffer, the *K* C 1 packet is dropped. A stochastic relationship between the distribution of the controller queue at an arbitrary time and an arrival time of an in-packet rule request is used to describe a Markov process [18]. This is used to derive the relationship formulas for the blocking probability, throughput and response time. Let *PBc* denotes the probability that an arriving in-packet rule request is blocked by the controller. It (*PBc* ) can be expressed as follows [36]:

|  |  |  |  |
| --- | --- | --- | --- |
| *c*(*N K* )*PK* |  | (4) |  |
| *PBc* D (1 *Po*) C *c*(*N K* )*PK* | ; |  |

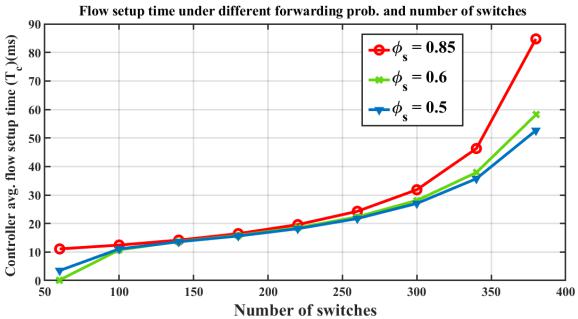
where *PK* denotes the probability that there are *K* in-packet requests in the controller queue at an arbitrary time. And the probability *Po* denotes the queue of the controller is empty at an arbitrary time, given by [37]:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| given by [18]: | |  |  |  |  |
| 1 | | *Po* |  | (9) |  |
| *c* D |  |  | ; |  |
|  | *c* |  |

where *c* denotes the average service time of controller.



**FIGURE 7.** Southbound controller blocking probability versus number ofswitches.



**FIGURE 8.** Controller average flow set up time versus number switches.

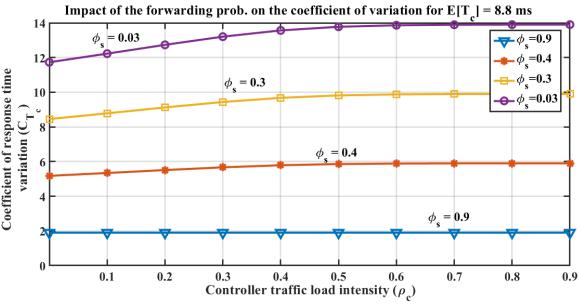
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | *c* | | | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | As a result, based on [(4),](#page8) the probability that a traf c ow |  |
| *Po* D | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | : |  |  | (5) | be blocked by the controller increases when the number of |  |
| *c*2((1Cp | | | |  |  |  | *s*2 | | | | p | | |  |  |  | C*k*)=(2Cp | | | | |  |  | *s*2 | |  | p | |  |  | )) | |  |  |  |  | AP switches increases. But, it decreases, when the controller |  |
| *c* | | | *c* | | | *c* | | *c* | |  | 1 |  |  |
| And *PK* is expressed as follows [37]: | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | service rate increases, as shown in Fig. [7.](#page8) The controller |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | service rate in uences the blocking probability of new traf c |  |
|  |  | ((1Cp | |  | *s*2 p | | | | | | | |  | | C*K* )=(2Cp | | | | | | | |  | *s*2 | | |  | p |  |  | ))( | | | | |  | 1) | | |  |  | ows, that arrive at the AP switches, without pre-assigned |  |
|  | *c* | *c* | | *c* | *c* | | *c* |  |  |  |
| *PK* D |  | *c* | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ; | (6) | forwarding rules. To this end, the capacity of SDN controller |  |
|  | *c*2((1Cp | | | |  | |  | *s*2 | | | | | p | |  | |  | C*K* )=(2Cp | | | | | | |  | | *s*2 |  | p | | |  | | )) | |  | |  | in terms of processing rate and buffer space should be care- |  |
|  | *c* | | | *c* | | | *c* | | *c* | | 1 | |  |  |
| where *s*2 | 1=2 denotes the squared coef cient variation | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | fully dimensioned to support both the SDN applications and |  |
| AP switches. Based on (8), the ow set up time for traf c |  |
| of the service process. *Lc* denotes the average number of | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | ows, that arrive at the APs without pre-assigned forwarding |  |
| in-packet requests in the controller queue at an arbitrary time, | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | rules, increases in terms of the in-packet rule requests sent to |  |
| which is expressed as follows [37]: | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | the controller, as shown in Fig. [8.](#page8) This becomes more obvious |  |
|  | 1 | | | | | | | | | | | | |  |  |  |  | *Po* | | | | (*N* | | | | |  | *K*)*PK*: | | | | | | | |  |  |  |  |  | (7) | when the number of AP switches exceeds a critical value. For |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | *Lc* D *N* | | | | | | | | |  |  | |  | | | |  | | |  |  |  |  |  |  |  | example, when the number of switches exceeds 200, the ow |  |
|  |  |  |  | | *c* | | | | | | | |  |  |  |  |  |  |  |
| And *Tc* denotes the average response time from the arrival to | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | set up time and blocking probability start to grow rapidly, |  |
| as shown in Fig. [8](#page8) and Fig. [7.](#page8) This means that the controller |  |
| service time completion, which is expressed as follows [18]: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | needs an effective mechanism that can proactively assign ow |  |
|  | 1 | | | | | | | | | |  |  |  |  |  |  |  | (*N* | | | | *K* ) *cPK* | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  | rules in the AP switches for traf c ows. This signi cantly |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | (8) |  |  |
|  |  |  | *Tc* D | | | | |  | | | |  | |  | | | |  | | | |  |  | | | |  |  |  | : |  |  |  |  |  |  |  |  |  |  | reduces the controller set up time, which shortens the sojourn |  |
|  |  |  | *c* | | | | 1 | | | | | | | |  | *Po* | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

time of traf c packets arrive at the AP switches. The average sojourn time for packets that arrive at an AP, in the HetNet, is de ned as the expected amount of time to wait before they

66679

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

**FIGURE 9.** Controller response time variation versus traffic load.



are handled by the AP. This mainly includes the controller ow set up time. Based on [(8),](#page8) the coef cient of response time variation, C*Tc* , for the sojourn times is dependent on the traf c ows, that arrive without pre-assigned ow-rules, at the southbound interface of the SDN controller, as shown in Fig. [9.](#page9) The coef cient of response time variation (*CTc* ) is de ned as the ratio of the standard deviation, , to the average of controller service (response) time (E[*Tc*]), namely: *CTc* D *E*[*Tc*]. It is calculated for an average controller servicetime value of E[*Tc*] D 8:8 ms. When the traf c load, arriving without pre-assigned ow-rules, increases at the AP switches, the coef cient of response time variation decreases. With a smaller *s*, less traf c packets are subject to the delay imposed by the controller and therefore the deviation from the average value for these packets is much higher, as shown in Fig. [9.](#page9) The more traf c ows require ow-rules assign-ment in the APs by the controller, the more traf c ows have a longer sojourn time caused by the controller service time and queueing waiting time. The controller response time in uences the total packet sojourn time more than the AP switches [17].

There is not a clear approach to identify the amount of resources (buffer space and channel bandwidth) required by a ow to guarantee its performance requirements, except for a peak bandwidth assignment. Here, the idea is to dynamically allocate a suf cient capacity to traf c ows served by the AP switches. These provide bandwidth based on the wire-less channel conditions. Thus, the M/G/1/K queuing system model provides a good approximation for buffering packets at the southbound buffer and the users access to the wireless channels of AP switches in the data plane.

An embedded Markov chain of the M/G/1/K queue is developed, in which the embedded points are selected imme-diately after each packet transmission completion. A switch state is de ned at a particular time instant as the number of packets in its queue at that time. Then, the considered switch has *K* C 2 states from state 0 to state *K* C 1, including *K* waiting packets in the queue and one packet being transmit-ted. However, when observing the switch immediately after each packet transmission completion, the utilized embedded Markov chain has only *K* C 1 states from state 0 to state

* C 1 [36].

66680

The packet blocking probability, *PBs* , is the probability that there are *K* packets waiting in the buffer of the AP switch, which is given by [37]:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (1 | | *s*) *sK* | |  |  |
| *PBs* D *P*[*N* D *K* ]D |  |  | ; | (10) |  |
| (1 | *sK* C1) |  |

where *s* denotes the traf c load intensity, expressed as,

*s* D *s* . The switch throughput, *s*, represents the number

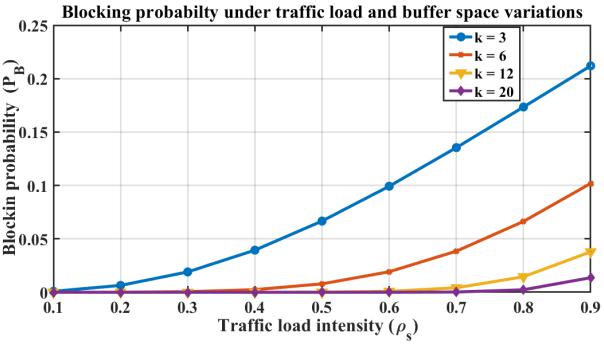
*s*

of packets that are completely served when the AP switches reach an equilibrium point, given as follows [37]:

|  |  |
| --- | --- |
| *s* D *s*(1 *PBs* ): | (11) |

The response time of packets served in the AP switches is given as follows [37]:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *T* | *s* D | *s* C *sK* C1(1 C *K* *sK* ) | : | (12) |  |
|  |  |
|  | *s*(1 *sK* C1)(1 *s*) | |  |  |



**FIGURE 10.** AP switch blocking probability versus traffic load intensity.

As a result, based on [(10),](#page9) the blocking probability at the AP switches increases in terms of the traf c load intensity and decreases when the buffer size gets larger, as shown in Fig. [10.](#page9) This re ects the importance of ensuring a proper buffer size at the different AP switches, because it can control the number of admitted downlink traf c ows. Based on [(12),](#page9) the AP switches continue to serve ows with minimal response time, irrespective of the buffer size until the traf c load intensity exceeds a speci c value, as shown in Fig. [11.](#page10) For example, after the load intensity exceeds 0:6, the impact of buffer size starts to become clearer on the switch response time, as shown in Fig. [11.](#page10) The throughput of AP switches has an impact on the QoE of UEs. When the QoS offered to the UEs is improved, this enhances the offered service. For example, based on [(11),](#page9) despite the increase in traf c load intensity, the buffer size remains a key in enhancing the satisfaction of users about the received throughput, as shown in Fig. [12.](#page10) This is re ected in the data rates of services offered to the UEs.

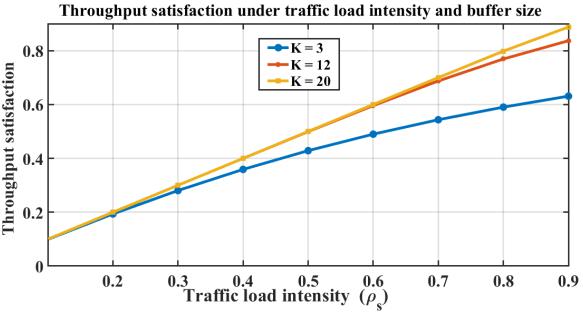
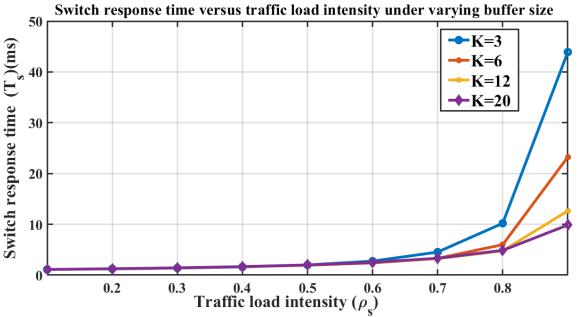
**VI. DYNAMIC TRAFFIC ENGINEERING SCHEME**

The available resources on the downlink channels of the APs can be viewed as virtual wireless paths that provide data connectivity to the UEs, as shown in Fig. [13](#page14) [38]. Four service classes are proposed to serve traf c ows generated from

VOLUME 8, 2020

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

**FIGURE 11.** AP switch response time versus traffic load intensity.



**FIGURE 12.** AP switch throughput versus traffic load intensity.

user applications, namely: Deterministic guaranteed service (DGS), delay-sensitive service (DSS), throughput sensitive service (TSS) and best effort service (BES). The DGS class is offered to traf c ows that have a constant bit-rate require-ment. The DSS class is offered to traf c ows that have a stringent delay requirement. The TSS class is offered to traf c ows that have a minimum throughput requirement but tolerate some delay bound violation. The BES class is offered to traf c ows that have no particular QoS guarantees. The software agents, that run in the AP switches, compute the sta-tistical information of downlink traf c ows and monitor the available bandwidth of wireless interfaces and queue length. They also determine the channel conditions of UEs and their MCS values. Local and generic network channel gain and information matrices are formed and updated periodically. The local matrix records the information of the UEs that can be only associated with a single AP. Whereas, the generic matrix records the information of the UEs that have a choice to be associated with one AP from a set of APs. The local matrix, 0***l*** , records the users’ tag, ow class, SNR, data rate of UEs and AP mark. The generic matrix, 0***g***, records the users’ tag, ow class, SNR, data rate of UEs and potential APs which can serve them. A downlink traf c ow is tagged with an identi er of the UE requesting it. A row in 0***l*** and 0***g*** has six elements, which are de ned in this order: user’s tag, AP mark, SNR value, data rate value, ow class and number. These information matrices are used by traf c routing and scheduling modules running under the proposed TE scheme. The traf c routing module routes ows to the different

VOLUME 8, 2020

**Algorithm 1** Traf c Engineering Scheme

**Input** : Local and generic matrices,0***l*** and0g

**Output**: 1. Network layer: Downlink ows routing to APs, 0n

**Output**: 2. Mac layer: traf c packets scheduling,0c

**Output**: 3. Network-Mac layer: Flow routing and traf cadaptation, 0nc

**Initialize**: Maximum number of users and ows perclass i. Number of slots and frame duration per AP switch, (*S*, *T*f);

* 0 ;

0n 0; 0c 0; 0nc 0 ;

**while** *true* **do**

The SDN controller receives updated 0***l*** and 0g ;

* size ( 0g ); y(1,1) gives the number of rows ; *z* size(0***l***); z(1,1) gives the number of rows ;0n(:,:) 0***l*** (:,:) ;

Routing on the networking layer; **for** *k*1*to y*(1;1) **do**

class 0g(k,5) ;

**if** class2DGS, apply smallest delay routingpolciy ;

**if** class2DSS, apply minimum network delayrouting polciy;

**if** class2TSS, apply maximum throughputpolicy;

**if** class2BES, apply probabilistic routingpolicy;

0n(z(1,1)+k,:) = 0g(k,:); **end**

0n;

Scheduling traf c on the MAC layer per AP ; **for** *v*1*to Tf* **do**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | *S*P *S*DGS; | | | |  |  |  |
| *S*rem | | *Si*; | |  |
| *S*DGS | |  | *i*DDGS20n | | |  |
|  |  |  |  |  |  |  |  |  |

**if** *S*rem>0;

slots are allocated to ows *j* 2 f DSS,TSSg proportional to their queue size;

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  | j | |  | | |  | | |  |  |  |  |  | *Qi* | | |  | k | ; |  |  |  |
| *Sj* | | | | |  | *Sj* C *S*rem | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |
|  | P | | *j*;*j*6D*Qij* | | | | |  |  |  |  |
| **if** | | | *S* | rem | | |  |  |  | ( | | *S* | DSS | | | | | | |  |  | TSS | | | | ) |  |  |  |  |  |
|  |  |  |  |  |  |  |  | | |  |  |  |  |  |  |  |  |  |  | C | | | *S* |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *S* | DSS | | | |  |  |  |  | min( | | | | | | | *S* | | | DSS; | | | | | *S* | | rem) ; | | | |  |  |  |  |
| *S* | TSS | | | |  |  |  | min( | | | | | | | *S* | | | TSS; ( | | | | | | | *S* | | rem | | |  | *S* | DSS)); |  |
| *S* | BES | | | |  |  |  |  | *S* | | rem | | | | | | | | |  | *S* | *j*;*j*2DSS;TSS; | | | | | | | | | | |  |
| 0c(j,:) | | | | | | |  |  |  | 0n(j,:) ; | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |
| **if** | | *S* | | *fi* | D 0; | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Change MSTR of this ow class or do reassocaition to another AP ;

0nc(i,:)= 0c(i,:) ;

**end**

* *t* C;

**end**

AP switches. The scheduling module forwards traf c packets to the medium access control of the selected APs, following the procedure explained in Algorithm [1.](#page10)

66681

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **A. TRAFFIC FLOWS ROUTING MODULE** | | | | | | | | | | | | | | | | | as follows [41]: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| The TE scheme uses this module to select an AP for each | | | | | | | | | | | | | | | | | *f*nd(*q*E*s*;E*s*)D | | |  |  |  |  |  |  |  |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| active UE that requests downlink traf c ows, as explained | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | | | |  |  |  |  |  |  |  |  | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | *n* | | | | 1;*qsj* | | | |  |  |  | *s* | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| in Algorithm [1.](#page10) This routes ows on the network layer and | | | | | | | | | | | | | | | | |  |  |  |  |  | *j* | |  |  |  |  | 0 | | |  |  |  | *j* | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| schedules their packets (frames) on the MAC layer based | | | | | | | | | | | | | | | | |  |  |  |  | P | |  | D | |  |  | *n* | | 6D | |  |  | 0 | | |  |  |  |  |  |  |  |  | *sk* | | | | |  |  |  |  |  |  |  |  |  | *f*nd(*qs* | | | | | | | | | ; *s*) | | |  |  |  |
| on their class and QoS requirements. The presence of UEs | | | | | | | | | | | | | | | | |  |  |  | C *k* | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | 1;*qs* | | |  |  |  |  |  |  |  |  | *n* | | | 1;*qsj* | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  | *j* |  |  |  |  |  |  |  | 0 *sj* | | | | | | | | |  |  |  |  |  |  |  |  | I |  |  |  |
| in the network are detected by the software agents, running | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  | D | |  | X*k* 6D | | | | | |  |  |  | P | | | |  | D | |  |  |  |  |  | 6D | | | |  |  |  |  |  |  |  |  |  |  |  |  | *qsk* | | | | | |  |  |  |
| in the APs, based on their active wireless interfaces and | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| service coverage. When the UEs are detected by a single | | | | | | | | | | | | | | | | |  |  |  | *f*nd(0; : : : ; *qsk* ; *sk* ; : : : ;0;0)D | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  | | | | | I | |  |  |  |
|  |  |  | *sk* | | | | |  |  |  |
| AP, they are associated with it and their traf c is classi ed | | | | | | | | | | | | | | | | |  |  |  | *f*ndD | | | | | | | |  | min | | |  |  |  | *fnd* (*q*E*s* | | | | | | | | | | | ; | | | |  |  |  |  | : |  |  |  |  |  |  |  |  |  |  |  |  |  | (14) | | | | |  |
| as local. Whereas, when the UEs are detected by multiple | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | *k*D1:::*n* | | | | | |  |  | E*s*) | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| APs, the module uses routing policies to select an AP that | | | | | | | | | | | | | | | | | Alternatively, this policy routes a ow to the AP *i* that has | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| can better serve them. It also classi es their traf c as generic | | | | | | | | | | | | | | | | |  |
| the shortest queue, which can be heuristically expressed as | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| and shares it among the selected APs. The UEs are cleared | | | | | | | | | | | | | | | | |  |
| follows [41]: | |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| from their APs in the local and generic matrices, when it | | | | | | | | | | | | | | | | |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| shows that their idle time exceeds or the SNR drops below | | | | | | | | | | | | | | | | |  | *f* |  | | (*q* | | ; | | | | | | ) | D *i* | | |  |  |  | min | | | | | | |  |  |  |  | *qsi* | | | C 1 | | | | | | |  | : |  |  |  |  |  |  |  |  |  | (15) | | | | |  |
| programmable preassigned thresholds for these parameters. | | | | | | | | | | | | | | | | |  | ld | |  |  |  |  |  |  |  |  |  |  |  |  | | | | |  |  |  |  |  |  |  |  |  |  |  |
|  |  | E*s* | |  |  |  | E*s* | | |  | D | | | 1;2;:::;*n* | | | | | | | | | |  |  | 2 | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| While we consider N AP switches, let n be a subset of APs | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | *si* | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| The maximum average throughput policy has a criterion | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| that detect and can serve a UE. | | | | | | | | | | | |  |  |  |  |  |  |
| Traf c routing policies are broadly classi ed into two | | | | | | | | | | | | | | | | | function, *f*mt(***q***E***s***, E***s***, ***f*** ), of the ow average traf c arrival rate, | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| types: deterministic and non-deterministic policies [39]. The | | | | | | | | | | | | | | | | | *f* , in addition to the average queue length vector, ***q***E***s***, and | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| former routes downlink traf c ows to APs based on criterion | | | | | | | | | | | | | | | | | service rate vector, E***s***, of APs. A ow is routed to an AP *i* | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| functions, whereas the latter routes traf c ows based on | | | | | | | | | | | | | | | | | that maximizes the average network throughput during the | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| probabilities, !*i*, assigned to APs, as shown in Fig. [13.](#page14) The | | | | | | | | | | | | | | | | | next inter-arrival time period. This policy also minimizes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| module runs a probabilistic policy, which will be discussed | | | | | | | | | | | | | | | | | the average network response time. The criterion function, | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| later along with three deterministic routing policies: smallest | | | | | | | | | | | | | | | | | *f*mt(***q***E***s***,E***s***,***f***), is expressed as follows [40]: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| delay, minimum network delay and maximum throughput. | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  | *n* |  |  |  |  |  |  | |  | *qsi* 1 | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Obviously, these policies mainly run on the network layer and | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  | X | | |  |  |  |  |  |  | X | | | | |  |  |  |  |  |  |  | *qsi* | | | | |  |  |  |  |  |  |  | *i* | | | | |  |  | )*k* |  |  |  |
| therefore use information related to this routing level. | | | | | | | | | | | | | | | | | *f*mt(*q*E*s*; E*s*; *f* ) D *i* | | | | | | | | | | | |  |  | *f* | | |  |  |  |  |  |  |  |  | (1 | | | |  |  |  | )( | | | |  |  | |  | |  |  |  |
| D | 1 |  |  |  | *k* | D | | | 1 |  |  | *k* | | | | | *f* | | | C | | | |  | *i* | |  |  |  |
| A deterministic smallest delay routing policy has a crite- | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | *f* | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| rion function, *f*sd(***q***E***s***, E***s***), of the average queue length vector, | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | *qsi* | | | | ln( | | | |  |  |  |  |  |  |  |  |  |  | ) | |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | C | | | | |  | | |  |  |  |  |  |  |  |  |  |  |  |
| E | | | | | | | | | E | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | *f* | | |  | *i* | | | | I | | | | |  | | | | | |  |
| ***qs*** , and service rate vector, | | | | | | | | | ***s***, of APs in the network. The | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| SDN controller advertises the state of network in terms of | | | | | | | | | | | | | | | | |  |  |  | *f*mtD | | | | | | | | max | | | |  |  |  |  | *f* | |  |  |  | (*q* | | | | ; | | | | | ; | | | |  | ) : | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |
| the queue length and service rate vectors, as follows: ***q***E***s*** D | | | | | | | | | | | | | | | | |  |  |  | *i*D1:::*n* | | | | |  | | | |  | mt | | |  |  | E*s* | |  |  | E*s* | | |  |  |  |  | *f* |  |  |  |  | |  |  |  |  |  |  |  |  | (16) | | | | |  |
| (*qs*1 ; *qs*2 ; : : : ; *qsn* ) and E***s*** D ( *s*1 ; *s*2 ; : : : ; *sn* ); 8*n N* , | | | | | | | | | | | | | | | | | Alternatively, this policy *f*mt(***q***E***s***, E***s***, ***f*** ) routes a ow to an | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| where q*si* denotes the number of packets waiting in the queue | | | | | | | | | | | | | | | | |  |
| and those under service in the AP switch i; and *si* | | | | | | | | | | | | | | | | denotes the | AP *i* that maximizes the expected number of ow service | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| completions before the next expected ow arrival. With this | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| service rate of the AP switch i. The policy, *fsd* ( ***q***E***s***, E***s***), routes | | | | | | | | | | | | | | | | |  |
| policy rule, the AP that satis es [41]: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| a downlink traf c ow to an AP *i* that has the smallest ratio of | | | | | | | | | | | | | | | | |  |  |  |  |  | *f* )D *i*D1:::*nf* | | | | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | |  |  |  |  | |  |  |  |  |
| the queue length to the service rate, expressed as follows [40]: | | | | | | | | | | | | | | | | | *f*mt(*q*E*s* | E*s* | |  |  |  | C *qsi* | | | | | | | C *si* | | | | | | | | | | *qsi* | | |  |  |  |  |
|  | | | | | | | | | | | | | | | | | | ; |  | ; | |  | | | | | | | max | | | | |  | | | | | | | *si* | | | | | | | | | | | | | | | | | | | | |  | | | | | | (17) | | | | |  |
| *f* | sd | (*q* | ; | ) | D *i* | | | |  | min | | *qsi* C1 | | | : | (13) | is selected. |  |  | | | | | | |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
|  |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
|  |  | | |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
|  | E*s* | E*s* | | D | 1;2;:::;*n* | | | | *si* | |  |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | The proposed TE scheme also runs a non-deterministic | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| If more than one AP has the same minimum queue length to | | | | | | | | | | | | | | | | | probabilistic policy [40] which aims to minimize the aver- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| age response time of all service ows in the network data | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| service rate ratio, the policy selects the AP *i* which has the | | | | | | | | | | | | | | | | | plane. The | UEs that | | | | | | | | | | can only | | | | | | | | | | | | |  | be | | | |  | served | | | | | | | | | | |  | under | | | | | | | | | a single | | | | |  |
| maximum service rate, | | | | | | *s* |  | . Since | | | *qsi* C1 | | provides the average | | | | AP generate local or dedicated traf c, | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  | | | |  |  |  |  | D ( | | | | |  | | | 1 | ;:::; | | | | *n*) | ; |  |
|  |  | |  |  |  |  |  |
|  |  |  |  |  |  |  | *i* |  |  |  | *si* | | |  |  |  |  | E*l* | | | |  |  |  |  |  |  |  |  |  |  |  |  |
| delay of the next admitted ow in the network, the smallest | | | | | | | | | | | | | | | | | 8*n N* . Whereas, the UEs that can be served under multiple | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| delay traf c routing policy provides an indication of how to | | | | | | | | | | | | | | | | | APs generate a generic traf c load, which is shared among | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| improve the response time of provisioned services. | | | | | | | | | | | | | | | | | the reachable APs based on splitting probabilities, !E*s* | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | D | |  |
| The minimum average network delay policy has a criterion | | | | | | | | | | | | | | | | | (!*s*1 ; : : : ; !*sn* ); | | | 8 *n N* , as shown in Fig [13.](#page14) When the | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| function, *f*nd(***q***E***s***, E***s***) , of the average queue length, ***q***E***s***, and | | | | | | | | | | | | | | | | | probability of ow routing to AP *i*, !*si* | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  | 6D1, the routing | | | | | | | | | | | | | | | | | | |  |
| service rates, E***s***, of APs in the network. It routes a downlink | | | | | | | | | | | | | | | | | decision is considered non-deterministic. When *si* | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  | C!*si* | | |  | |  |
| traf c ow to an AP *i* that minimizes the delay of all waiting | | | | | | | | | | | | | | | | | *si* , the AP *i* is considered unsaturated and the response time | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| and in service packets at the APs which can serve a UE, given | | | | | | | | | | | | | | | | | is calculated based on [(12)](#page9). | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 66682 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | VOLUME 8, 2020 | | | | | | | |  |

min*ni*D1 *si*

*si*

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

Let *si* and *si* be the local traf c arrival rate and load intensity at AP switch *i*; and *n* and *n* denotes the total generic traf c arrival rate and load intensity at *n* APs, respec-

tively. An AP *i* is assigned a generic traf c load based on,

*nrsi* !*si* , where *rsi* D ; and *n* D *n*(min*ni*D1 *si* ).

The minimum average network response time optimization can be formulated as a non-linear programming problem with linear constraints, as follows [40]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| min | |  |  |  |  | 1 |  |  |  |  | *n* |  | *nrsi* | !*si* | C *si* | | |  |
|  |  |  |  |  |  | X | | | |  |
| *s t* |  | *si* |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 0 | |  | P |  |  |  |  |  | 11 ( *nrsi* !*si* | | | | C *si* ) | |  |
| !*s*1;!*s*2 | ;:::;!*sn* | | | *n* | C | | *jn*D1 | *sj i* | | D | |  |
|  |  |  |  |  |  | 1 |  |  |  |  |  |  |  |  |  |  |
| : : ! | |  |  |  | ; | *i* |  | *n* | |  |  |  |  |  |  |  |  |
|  |  | *n* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | X | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | *i*D1 | | !*si* D1 | | | |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | C *si* < 1; | | |  |  | 1 *i n*: | | | |  | (18) | |  |
|  | *nrsi* !*si* | | | | |  |  |  |  |

The solution of this optimization problem stated in [(18)](#page12) results in obtaining these optimal traf c load sharing prob-abilities: !*s* D [!*s*1 ; !*s*2 ; : : : ; !*sn* ]. These determine the proportions of generic traf c ows assigned to each AP that can serve the UEs.

An adaptive control algorithm uses the collected SDN information to maintain a similar level of controlled generic loads at *n* APs. A heuristic probabilistic load sharing algorithm is developed in [40] to solve this nonlinear pro-gramming problem. The proposed TE scheme receives the information regarding the local and generic matrices in addi-tion to the queue length and service rates of the APs every

seconds [42], [43], as explained in Algorithm [1.](#page10) These parameters are used in a function developed in MATLAB to solve the optimization problem [(18)](#page12). The splitting probabili-ties, !*si* , are kept constant during the measurement intervals; and they are updated upon the receipt of a new measurement every . The local traf c ows are rst associated with their APs. Then, the generic traf c ows are routed to the APs that can serve them, using the above routing policies, as explained in Algorithm [1.](#page10)

* 1. **GRANULAR RESOURCE ALLOCATION MODULE**

The scheduling module supports granular differentiated ser-vices on the MAC layer of LiFi and LTE APs. It allocates the wireless interface bandwidth to the UEs based on the number of bits that can be transmitted at their assigned MCS per frame. The smallest data unit allocation per resource block, *b*, differs in each wireless technology. Therefore, b*k* represents different values in LiFi and LTE. For LTE, the number of transmitted bits per resource block (RB) pair using MCS*i* is, *bi* D2 *R*re*b*re*i* , where *b*re*i* represents the number of bits transmitted per resource element using MCS*i* [44]. For LiFi, *bi* represents the number of transmitted bits per slot usingMCS*i*. Whereas, for WiFi, it represents the number of trans-mitted bits per frame (i.e. contention period).

A two-level hierarchical parametrized scheduler is deployed in the AP switches to adaptively allocate wireless resources. It can support various policies for sharing the AP

VOLUME 8, 2020

wireless interface and scheduling polices to provide multiple services. A traf c ow classi er module is integrated with the scheduler. It classi es packets arriving from the application and transport layers based on an identi er of their traf c class, (*f*id), and forwards them into a queue designated for their service class [38]. The upper interface medium access class scheduler distributes the available bandwidth among the supported services according to a criterion function of AP state and performance parameters. The lower interface medium access service scheduler supports light scheduling functionalities that serve separate queues designated for the supported service classes [38].

One of the QoS parameters considered for service classes is the maximum sustained traf c rate (MSTR) in bits per second (bps) [44], which is an upper bound for the user throughput. In our context, each service class can be associated with a different MSTR. The average download duration of a UE is,

* on D *QL* , where *Q* denotes the average number of active UEs;

and *L* denotes the average number of UEs that are served per unit of time. The average throughput of a UE can be calculated as, D *x*on , where *x*on denotes the average size of

* on

downlink data in bits during the ON period. A UE may receive a different amount of resources at each frame to achieve its guaranteed bit rate (GBR ( )), which varies with the MCS*i*. It is assigned a slightly greater bit rate than its GBR, called delivery bit rate (DBR ( )) to compensate for the data rate losses in the outage periods, given by [45]:

D ; (19)

* *p*0

where *p*0 denotes the wireless channel outage portability. A UE that achieves its (DBR) using MCS*i* needs *gi* slots, given as followsV

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *gi* D | *TF* | : | (20) |  |
| *bi* |  |

A UE with an outage probability, *p*0 D 1, is not allocated any slot (i.e. *g*0 D 0). The average number of slots per frame needed by *m* UEs belonging to a service class to obtain their DBR is given byV

|  |  |
| --- | --- |
| *I* |  |
| X |  |
| *g*(*m*)D *m pigi*: | (21) |

*i*D1

The hierarchical scheduler supports four service sched-ulers, which serve the proposed DGS, DSS,TSS and BES classes. There is a limit, *M*max, to the number of simultaneous DGS, DSS and TSS ows accepted in a small cell. The BES traf c ows can simultaneously transfer data, when there are enough resources remained after serving the higher priority services. The upper class scheduler dynamically shares the downlink wireless bandwidth of each AP among the sup-ported services based on their backlogged traf c and QoS requirements, as explained in Algorithm [1.](#page10) This improves the buffer utilization and wireless interface bandwidth of APs, while guaranteeing the requirements of services. The DGS scheduler assigns xed time slots for ows, which are

66683

where *t*off denotes the average OFF duration.

The steady-state probabilities, (*m*), can be expressed as follows [44], [45]:

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| negotiated in the initial service access phase and deducted | | | | | | | | | | | | | | | | | | | | (MLWDF) policy proposed in [47], [48]. It schedules UE *i* | | |  |
| from each downlink frame, as explained in Algorithm [1.](#page10) The | | | | | | | | | | | | | | | | | | | | such that [47], [48] |  |  |  |
| aggregate departure rates (*m*) per service class is expressed | | | | | | | | | | | | | | | | | | | |  | *i*[*Di*(*t*)] *i*(*t*) ; |  |  |
| as follows [44], [45]: | | | | | |  | | |  | | |  |  |  |  |  |  |  |  | *i* D *arg* max*i* | (29) |  |
| (*m*) D | |  |  |  | *S* | | | | | | |  |  |  | *m* | *R* | | | (22) |  |  |
|  |  |  |  | |  |  |  |  |  | |  |  |  |  | ; | where *Di*(*t*) is the head-of-the-line packet delay for UE *i* at | | |  |
|  |  |  |  | |  |  | |  |  |  |  |  | *x* | on |  |
| max(*mg* | | | | | | | | | | ; *S*) | | |  |
|  |  |
| where *R* denotes the MSTR of the service class offered to | | | | | | | | | | | | | | | | | | | | time *t*, *i*(*t*) is its instantaneous channel data rate at time *t* | | |  |
| and *i* and are arbitrary positive constants. It is recommend | | |  |
|  |  |  |  |  |  | |  |  | |  |  |  |  |  |  |  |  |  |  |  |
| *m* UEs. The rst part |  |  | *S* | represents the ratio of the | | | | | | | |  |
| max(*mg* | | | | | | ; |  | | ) |  | a value of [48] |  |  |  |
| *S* | |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| global departure rate achieved by the *m* concurrent transfers, | | | | | | | | | | |  |  |  | *ai* | |  |  |
| when there are *m* active UEs needing *m* | | | | | | | | | *g* | slots on average |  |  | *i* D | (30) |  |
| to obtain their *R*. The second part *m* | | | | | | *R* | | corresponds to the | | |  |  |  |  |  |
|  | *i* |  |
|  | *on* |  |  |
|  |  |  |  |  |  | *x* |  |  |  |  |  |  |  |  |  |
| rate at which any of the *m* active UEs completes its transfer, | | | | | | | | | | | for the UE *i*, where *ai* is a weight that may be based on delay | | | | | |  |
| assuming there are always enough available slots in the frame | | | | | | | | | | |  |
| to satisfy the *R*. The service class parameter intensity, , is | | | | | | | | | | | requirements, and |  | *i* is the user’s long-term average data rate. | | | |  |
|  |  |
| given by [44], [45]: | | | | | | | |  |  |  | This policy tries to balance weighted delays, and for the given | | | | | |  |
|  |  | *x* | | on | | | |  |  |  | choice of *i*, reduces to a channel-aware scheduling strategy | | | | | |  |
| D |  | (23) | | | when the users are otherwise equal. The DSS scheduler uses | | | | | |  |
|  |  |  |  | ; | | |  |
|  | | off*R* | |  |  |  |  |  | |  |
| *t* | | the exponential rule introduced in [47] and analysed in [49]. | | | | | |  |

It schedules the user *i* such that [49]

(*m*) D

where (0) D 1 given as followsV

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *M*W | | | *m* | | | | (0); | (24) | *i* Darg max*i* | | | | | | *i i*(*t*) exp( *a*1*iDi*(*t*p | (*aD*) | | ) ; | (31) |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ) | *aD* | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | *m* |  | *S* | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (*M m*)W *m*W5*i*D1 | | | |  |  |  |  |  | where *aD* D | | | | 1 | *i Di*(*t*), and *M* is the total number of users. | | | | | |  |
| max(*ig* | ;*S*) |  |  |  |  |
|  |  |  | *M* |  |
|  |  |  |  |  |  |  |  |  | This policy | | | tends to equalize the weighted delays when the | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  | P |  |  |  |  |  |  |

. The average number of active UEs is

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | differences are large, and falls to the Proportional Fair (PF) | |  |
|  |  |  |  | policy when the differences are small. A PF algorithm serves | |  |
|  |  | *M* |  | each UE *i* at peak channel conditions. At each scheduling | |  |
|  |  | X | (25) |  | V |  |
| *Q* D | | *m* (*m*): | time slot, the BES scheduler policy selects the UE *i* with the | |  |
|  |  | *m*D1 |  | highest priority value which is calculated as follows |  |  |

The average number of UEs, *L*, that complete their transfer per unit of time is given byV

*M*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *i* Darg max*i* | *i*(*t*) | : | (32) |  |
| *i* |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | (*m*) (*m*): | | | | | | (26) | The class scheduler aims to allocate a suf cient number |  |
|  |  |  |  |  |  |  |  | *L* D | | | | | |  |
|  |  |  |  |  |  |  | of slots from the frames to each service so that its active |  |
|  |  |  |  |  |  |  |  |  | *m*D1 | | | | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | X | | | | |  |  |  |  |  |  |  | UEs together can achieve *R*, as in Algorithm [1.](#page10) If a UE that |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| The average throughput *X* obtained by each active UE is | | | | | | | | | | | | | | | | | | | | |  |
| receives the BES is in an outage state, it will experience |  |
| given as followsV | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | a temporarily degraded throughput. If at any given time, |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | *M* | | | | | |  | the total number of available slots is not enough to satisfy *R* of |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | *x*on | | | | | *x*on | | | | | *m*D1 (*m*) (*m*) | | | | | : | (27) |  |  |
|  | *X* | all users, excluding those in outage, given a BES class, they |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | D *t*on | | | | | | | D | | | | |  |
|  |  | *M* | | | | | |  |  |  |
|  |  | P*m*D1 *m* (*m*) | | | | | |  | will all experience equally degraded throughput. The service |  |
| The average utilization | | | | | | | | | | |  | |  | Pof the frame is de ned as the | | | | | | | scheduler policies can be dynamically selected or updated for |  |
| *U* | |  |
| weighted sum of the ratios between the average number of | | | | | | | | | | | | | | | | | | | | | each service class to best meet its QoS requirements. The |  |
| slots needed by the *m* UEs to reach their *R* and the average | | | | | | | | | | | | | | | | | | | | | class scheduler policies can also be dynamically updated to |  |
| number of slots they obtainV | | | | | | | | | | | | | |  |  |  |  |  |  |  | ef ciently support the QoS requirements of supported service |  |
|  |  |  |  |  |  |  | classes. After the class scheduler policy allocates the slots to |  |
|  |  |  |  |  |  |  |  |  | *M* | | | | |  |  |  |  |  |  |  | the DGS class, the remaining slots are allocated to DSS and |  |
|  |  |  |  |  |  |  |  |  | *mg* | | | | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | *U* D | | | | | |  |  |  | | | |  | |  | (*m*): | | (28) | TSS ows proportional to their reported backlogged traf c. |  |
|  |  |  |  |  | max(*mg* | | | | | | ; *S*) |  |
|  |  |  |  |  |  |  |  |  | *m*D1 | | | | |  |  |  |  |  |  |  | If DSS or TSS ows cannot be assigned slots at an AP, |  |
|  |  |  |  |  |  |  |  |  | X | | | | |  |  |  |  |  |  |  |  |  |
| A feasible Earliest Due Date (FEDD) scheduler policy [46] | | | | | | | | | | | | | | | | | | | | | the algorithm either changes the sustainable data rate of the |  |
| selects, at each slot per frame, a UE with the earliest deadline | | | | | | | | | | | | | | | | | | | | | traf c class or routes them to another AP that can provide the |  |
| among the UEs which have good channel gains. We recognize | | | | | | | | | | | | | | | | | | | | | requested service. This process is updated per SDN period, |  |
| that this policy is not always throughput optimal. The TSS | | | | | | | | | | | | | | | | | | | | | , and a number of frames, *T*f, to adapt the QoS requirements |  |
| scheduler uses the Modi ed Largest Weighted Delay First | | | | | | | | | | | | | | | | | | | | | of supported services by the scheduler module. |  |
| 66684 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | VOLUME 8, 2020 |  |

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

**C. NETWORK RESILIENCE**

When the performance level of the provided DGS, DSS or TSS is degraded below a predetermined threshold or an AP becomes unresponsive (i.e. fails), the TE scheme runs the restoration Algorithm [2](#page14) to maintain service provisioning to the affected UEs. In a centralized tree-type SDN-enabled network, the dynamic traf c routing Algorithm [2](#page14) running in the SDN controller dynamically employs the wide coverage WiFi or LTE capacity to support a single AP failure recovery. It does not place any spare network capacity for mitigating the impact of an AP failure to provide service continuity with an acceptable QoS level to UEs during the AP failure. The agents running in the APs periodically send the load of each traf c class in addition to the information matrices, 0***l*** , 0g, to the SDN controller. This uses Algorithm [1](#page10) to periodically compute and update the information matrix, 0*nc*. When an AP fault occurs, the SDN controller produces a new matrix 0nc-r, which is used to reroute traf c from the failed AP. Algorithm [2](#page14) uses a traf c re-routing policy that considers the tree-topology constraints, volume and QoS requirements of the failed ows. The SDN controller reroutes the affected traf c ows based on the reported information of the failed AP before a failure occurrence. The local traf c ows are either routed to the LTE or WiFi AP by using the probabilistic routing policy. Whereas the generic traf c ows are managed according to the other routing policies following Algorithm [1.](#page10)

**Algorithm 2** Service Provisioning Restoration From aSingle AP Failure

**Input** : Local and generic matrices,0***l*** ,0g,0nc

**Network fault event**: An AP becomes unresponsive orservice ows’ throughput or delay drops below a speci c threshold;

**Output**: Updated Network-MAC layer matrix:re-routing failed downlink ows to operational APs, 0nc-r

**for** *i* 1 *to N*f 2 0***l*** **do**

**if** *f*20***l*** , route local traf c ows to LTE or WiFibased on their traf c load. **if** *f* 2 0g, route local traf c ows baed on Algorithm [1.](#page10) Update 0nc based on the rerouted traf c ow and their correspoding

inforamation.

**end**

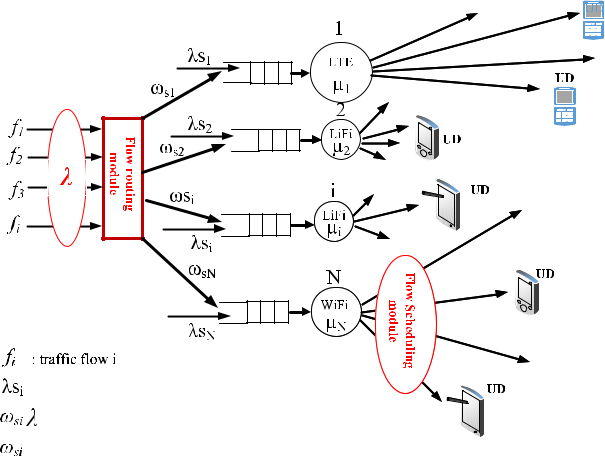
Remove the row corresponding to failed APs from 0nc; 0nc-r(:,:) 0nc(:,:)

**VII. SIMULATION AND RESULTS**

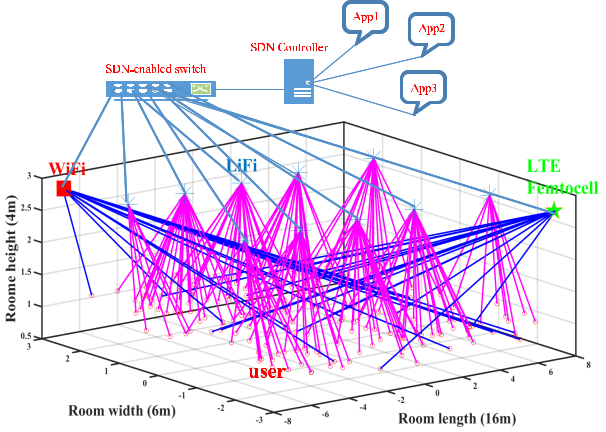
A discrete time simulation environment has been developed in MATLAB to evaluate the proposed TE scheme in the Het-Net. This is comprised of 10 LiFi APs, one LTE femtocell and another WiFi APs, which provide services to UEs uniformly distributed in a room of size 16 6 4 (*W L H* ), as shown in Fig. [14.](#page14) The user mobility pattern follows a random waypoint model [50], [51] with a uniform distribution low

VOLUME 8, 2020

**FIGURE 13.** A traffic engineering scheme for SDN-enabled HetNet.



movement speed between 0:1 and 0:4 m/s, which correspond to the users movement in the room. The downlink physical data rates of LiFi APs, the LTE femtocell AP and the WiFi AP are taken to be 45, 14 and 54 Mbps, respectively. The channel gains of LiFi APs are calculated based on the Lambertian model [52], which considers signal re ections from all the room walls. The main parameters of the LiFi attocellular network are summarized in Table 1 [53] and [54], [55] with a service coverage radius of 3 m and total number of OFDMA subcarriers on downlink channel, Nsc D 1024. The main parameters of the LTE femtocell system are summarized in Table 1 [56] and [57] with service coverage radius 20 m. The channel access parameters of the WiFi AP are summa-rized in Table 1 [58] and [31]. The interface queue length of all APs uses a drop tail policy; and it is set at 5000 packets. The maximum number of users that can be simultaneously connected to a LiFi AP, LTE femtocell or WiFi AP, is set to 8, 10 and 20, respectively. Similarly, the maximum number of each traf c connection class is set to communicate through the LiFi, LTE or WiFi AP switches.



**FIGURE 14.** SDN-enabled HetNet simulation environment.

A number of applications share the spectrum and buffer resources in the SDN-enabled HetNet, as shown in Fig. [14.](#page14)

66685

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

ON/OFF traf c sources generate DGS, DSS, TSS and BES traf c ows. An ON period is considered as a sequence of arrivals with less than *t* seconds between two OFF events. The duration of ON and OFF periods are exponentially dis-tributed. The traf c ows of DGS, DSS, TSS and BES are generated with data rate at 64 kbps, 1 5 Mbps, 12 18 Mbps, 0:5 6 Mbps, respectively. The guaranteed bit rate (GBR) of DGS, DSS and TSS is set to 64 kbps, 2 Mbps and 12 Mbps, respectively. A throttling policy schedules BES ows accord-ing to their MSTR. This is set to 1 Mbps, but it can be adjusted by the controller according to the service requirements of the other traf c classes and network state. The average downlink size of BES traf c, *x*on, is set to 2 Mbps. The packet size of DGS, DSS,TSS and BES ows is set to 200, 256, 512, 1512 bytes, respectively. The packet arrival rate, , of DGS, DSS, TSS and BES 90, 700, 350, 620 packet per second (pps), respectively. Each time a random number of UDs are active in the network, there are 30 % DGS UDs, 25 % DSS UDs, 25 % TSS UDs and 20 % BES UDs are distributed across the HetNet.

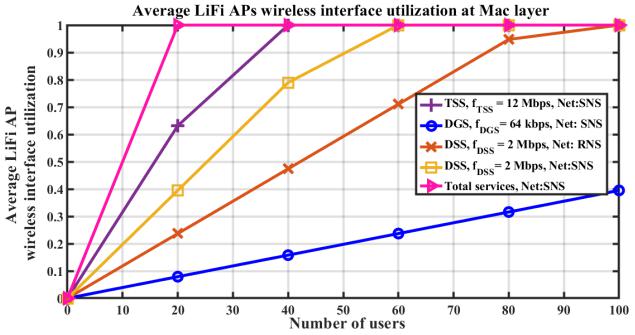
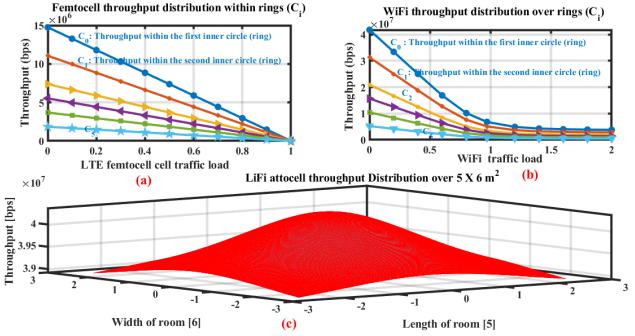
We have conducted a number of simulation scenarios to evaluate the impact of the proposed TE scheme on the service provided at the network, MAC and across network-MAC layers. The conducted scenarios are evaluated in two network settings: Random HetNet setting (RNS) and SDN-enabled HetNet setting (SNS). In the RNS, UEs receive services in a default HetNet that runs without the SDN support capabil-ities. For example, the UEs are associated to APs providing them with the best received signal strength (RSS). Whereas, in the SNS, the TE scheme supports service provisioning to UEs, which considers the network information feedback regarding the resource availability of APs, service class and target network performance.

The throughput distribution within LiFi, WiFi and LTE cells varies based on their path loss models, which changes in terms of the distance and transmission power. These de ne the virtual boundaries of circles (rings) forming the cells coverage. Within these cells, the throughput is considered uniform, but varies based on the number of UEs and traf c load, as shown in Fig. [15.](#page15) A cell capacity is de ned as the maximum traf c load that can meet its QoS requirements. The maximum average data arrival rate per UE is evaluated, given that all the UEs are uniformly distributed within the coverage area of each AP in the HetNet. The ow throughput and cell capacity in the different AP cells decrease, as the traf c load and radius increase. Deploying multiple opti-cal wireless technologies, such as LiFi APs, in an indoor environment is important for meeting the bandwidth and throughput requirements of UEs with heterogeneous wireless air interfaces.

The MAC scheduler shares the limited communication bandwidth of the AP wireless channel among the supported traf c services. The interface bandwidth of LiFi APs is shared in proportion to the number of UEs and generated traf c load per class, which is also in uenced by the network setting, as shown in Fig. [16.](#page15) Each evaluation point in this gure

66686

**FIGURE 15.** Throughput distribution under small cells: (a) LTE femtocell(b) WiFi cell (c) LiFi attocell.



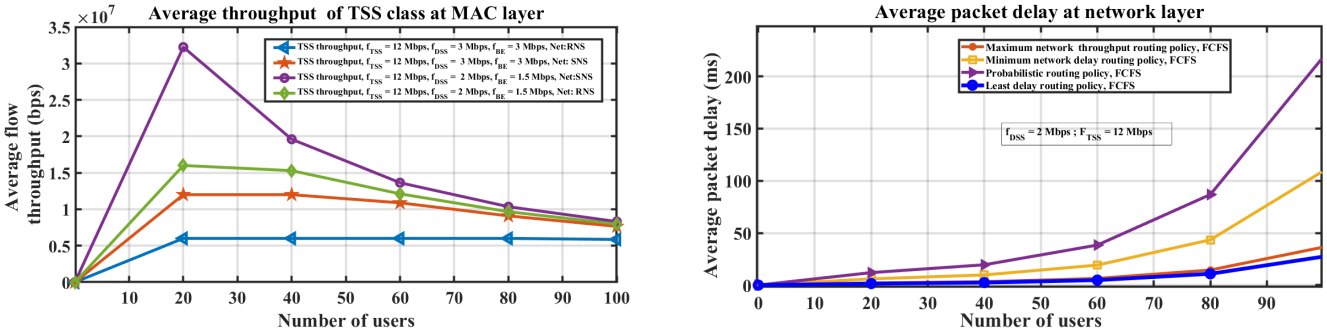
**FIGURE 16.** Average interface utilization of LiFi APs at MAC layer.

corresponds to the average utilization of service at each AP. The different scheduling functionalities can better manage the wireless interface of LiFi APs in SNS than RNS. This explains the admission of more DSS ows, which increases the utilization of LiFi AP interfaces in SNS. More DSS ows are dropped in RNS than SNS, which decreases the LiFi interface utilization, as shown in Fig. [16.](#page15) Furthermore, in RNS, some downlink traf c ows are blocked due to the unfairness of load distribution among the different APs in the HetNet.

The effective capacity is de ned as the maximum sustain-able constant data arrival rate by the channel process under some throughput or delay (QoS) constraints. The proposed TE scheme is evaluated under the minimum throughput and delay requirements of TSS and DSS traf c ows, respec-tively. The transmission of TSS real-time video streaming requires a large bandwidth and a minimum stable end-to-end delay. As a result, Fig. [17](#page16) shows the average ow through-put of TSS ows achieved under their minimum throughput requirements. We observe that the TSS ows are guaranteed a minimum bandwidth of 12 Mbps in SNS, which is dif cult to achieve in RNS, as shown in Fig. [17.](#page16) This is attributed to the fact that in SNS, traf c ows are balanced among the network APs; and wireless interfaces of APs are shared in proportion to traf c ows class and volume. In SNS, the TE scheme enhances the performances of HetNet by increas-ing the number of admitted traf c ows with minimum

VOLUME 8, 2020

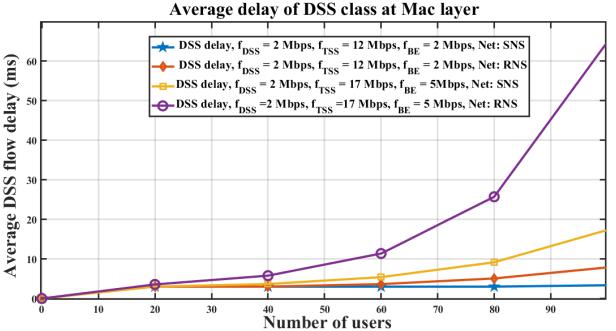
H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence



**FIGURE 17.** Average flow throughput in TSS class at MAC layer. **FIGURE 19.** Network response time for different traffic routing policies.

throughput requirement, which increase, in turn, the network throughput, as shown in Fig. [17.](#page16)

The DSS traf c packets experience a delay at the MAC layer of APs due to the resource contention and network congestion, which can be better maintained within strict performance bounds in SNS. The aggregate scheduler on the MAC layer ensures the resource allocation to the DSS traf c packets. Thus, they experience a shorter delay in SNS than RNS. The UEs are associated to APs based on service guarantee policies at the MAC layer, which can better exploit the LiFi and LTE/WiFi channels for data transmission in SNS than RNS, as shown in Fig. [18.](#page16) The average delay of DSS traf c grows exponentially, when the traf c volume of the other classes increases signi cantly, because the aggregate scheduler has to allocate some resources to the other traf c classes. Also, when the number of UEs that request the DSS increases signi cantly, their delay exceeds the upper delay bound requirement.



**FIGURE 18.** Average flow delay in DSS class at MAC layer.

The proposed TE scheme is evaluated on the network layer in two scenarios. The rst considers class-based traf c ows routing according to their minimum service require-ment and HetNet resource allocation ef ciency. The second scenario focuses on the HetNet resilience from a single AP failure, which considers re-routing the failed traf c ows. The network response time is de ned as the amount of time required for a packet to be routed on the network layer and scheduled at the MAC layer to reach the UEs in the downlink direction. Traf c packets are routed according to different

VOLUME 8, 2020

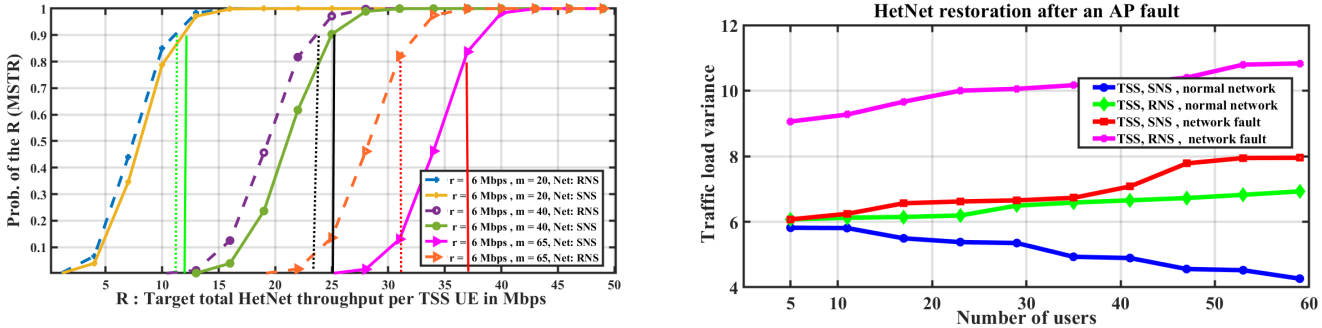
routing policies, which have unequal network response times, as shown in Fig. [19.](#page16) Obviously, packets delay increases as the number of users grows in the network. The smallest net-work delay and minimum throughput routing policies could maintain the delay within acceptable upper bounds, compared to the other policies, as shown in Fig. [19.](#page16) The minimum network delay routing policy requires more time to nd an AP that can minimize the whole network delay. Similarly, the probabilistic routing policy requires to classify the local and generic traf c, which require more time as ows grow in number and volume, making it dif cult to achieve a minimum network response time.

The proposed TE scheme is evaluated across the network-MAC layer in two scenarios. The rst discusses HetNet dimensioning according to the minimum data rate requirement and maximum target network throughput per active TSS UE. The second scenario discusses the capability of TE scheme to support reliable HetNet. Users transfer traf-c during the ON state at a minimum data rate, *r*, where the distribution of UEs, *m*, follows a binomial probability density function. The binomial cumulative distribution function is used to calculate the probability of the maximum target total HetNet throughput, *R* (MSTR), per TSS UE, considering different number of UEs in SNS and RNS con gurations. As a result, we observe that the TE scheme enables the HetNet to ef ciently allocate the bandwidth of each AP to maximize the target HetNet throughput, *R*, which supports the UEs receiving the TSS class. We notice that with a probability of 0:9, a total maximum sustainable throughput of 12 Mbps and 13 Mbps can support the TSS UEs in RNS and SNS con gurations, respectively, as shown in Fig. [20.](#page17) As the number of UEs increases, the target total throughput increases per TSS UE in SNS than RNS. For example, with a probability of 0:9, a maximum throughput of 32 Mbps and 38 Mbps are guaranteed per TSS UE in RNS and SNS, respectively. An extra 6 Mbps are added to the R (MSTR) of each active TSS UE in SNS than RNS, as shown in Fig. [20.](#page17)

The proposed TE scheme supports the HetNet to survive from a single AP failure, which enables it to autonomously handle the affected ows that were receiving services from a failed AP, as explained by Algorithm [2.](#page14) As mentioned before, an AP is identi ed as failed, when it can no longer provide

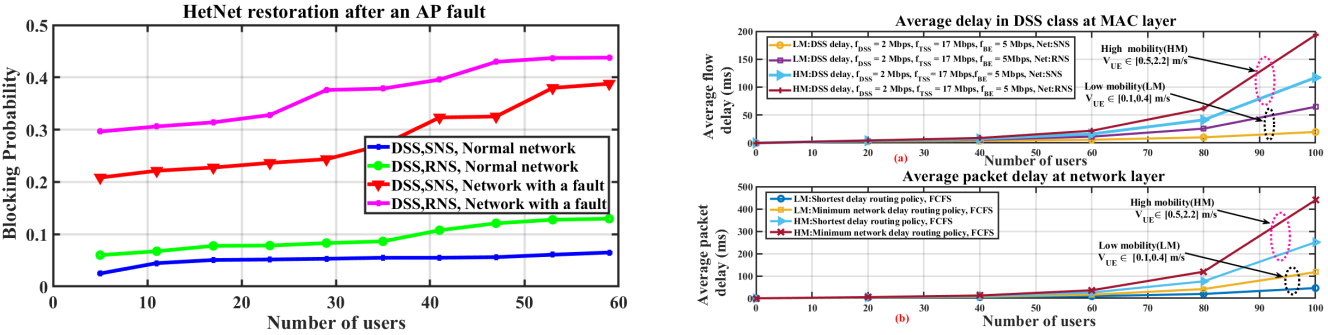
66687

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence



**FIGURE 20.** Target total HetNet throughput per TSS UE.

**FIGURE 22.** Traffic load variance of TSS flows after network fault.



|  |  |
| --- | --- |
| **FIGURE 21.** Blocking probability of DSS flows after network fault. | **FIGURE 23.** Mobility impact on average (a) DSS delay at MAC layer and |
|  | (b) packet delay at network layer. |

services to the users. While the affected ows from an AP failure can be rerouted to neighbouring LiFi APs, or wider coverage APs like WiFi or LTE AP, the network resources may not be suf cient to accommodate them. A computer sim-ulation is carried out, which involves the four neighbouring LiFi APs in the rst two columns, LTE and WiFi APs in the SDN-enabled HetNet topology shown in Fig. [14.](#page14) The focus is put on studying the performance of the affected DSS and TSS service ows from a single AP failure, subject to delay and throughput constraints, respectively. As discussed before, the DSS and TSS ows should receive their minimum e2e delay and throughput guarantees, respectively, to provide the requested services to the UEs. Therefore, some of these ows are blocked during an AP failure, because of being routed to an overloaded AP or their routing process takes a longer time. The blocking handover request probability is de ned as the number of dropped requests divided by the total affected ows that require association again with a new AP due to the failure. When the HetNet runs normally, without any AP failure, the DSS ows have a lower blocking probability in SNS than RNS, as shown in Fig. [21.](#page17) The probabilistic routing policy considers the affected ows as generic, which can be routed to any of the four APs or LTE/WiFi APs that can provide the requested services. Algorithm [2](#page14) introduces a SDN scheme to manage the affected ows, which reduces the number of blocked DSS ows in SNS, as shown in Fig. [21.](#page17)

The traf c load distribution varies at APs, because of network dynamics, which is expected to increase in case of AP failure. A traf c load variance metric is evaluated to show

66688

the spread of the traf c load at the APs involved in serving the ows before and after AP failures. The variance can be used to evaluate the standard deviation, which is a better measure for the traf c load variations at the APs. The standard devi-ation is the square root of the traf c load variance, which is

q

expressed as follows: *i*2 D *N*1 P*i*D1( *i i*)2; where *N* is the number of APs. In SNS, the load variance is smaller than in RNS, because the affected UDs are rerouted based on load balancing and scheduling policies. These keep the load of each AP around the average value measured during the AP failure.

While the previous results are obtained under low UE mobility, in the remaining part we investigate the impact of higher UE mobility on the performance of TE scheme. In essence, we compare the impact of low UE mobility to high mobility on traf c delay at MAC and network layers under the two network settings: RNS and SNS, separately, as shown in Fig. [23.](#page17) In this paper, a high UE mobility means that the UE’s minimum velocity vmin is set to 0:5 m/s, whereas the maximum velocity vmax varies over the range [0:5; 2:2] m/s. Initially, a UE is connected to a LiFi AP that can meet its service requirements in terms of delay and throughput. Otherwise, the TE scheme connects the UE to LTE or WiFi AP based on their traf c load. The UE is assumed to associate with only one AP at a time, and only vertical handover (VH) is performed from LiFi to LTE/WiFi and vice-versa.

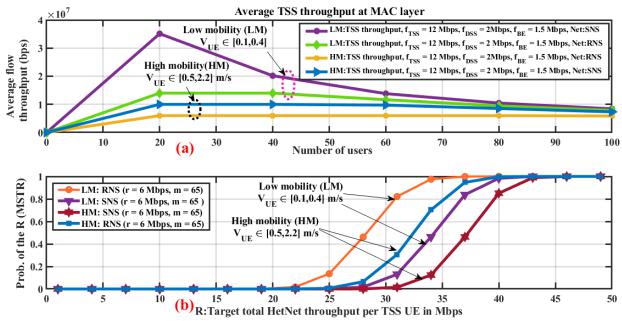
When the number of users is low, the TE scheme has shown an interesting performance by keeping the average

VOLUME 8, 2020

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

DSS delay and packet delay at the MAC and network layers almost the same under low and high mobility, as shown in Fig. [23](#page17) (a) and (b). However, when the number of active users started to exceed 40 in the network, the delay started to pick up under both low and high UE mobility. The gap between low and high mobility impact on the delay becomes obvious, when the number of active users exceeds 68. This can be considered as the maximum number of UEs that the SDN-enabled HetNet can offer services under low and high mobility scenarios, where we can clearly see that SNS alleviates the average DSS delay at MAC and network layers.

As a UE moves across the network under low and high mobility, it requires horizontal handovers (HO) between the LiFi APs and VH among the LiFi/WiFi/LTE APs. While the TE scheme executes both the HO and VH processes, traf-c ows and packets experience further delay and possible packet loss. They may also need to be connected again to a new AP to start receiving the requested traf c. This does not only increase the ows or packet delay at the MAC and network layers, but also decreases the network through-put at the MAC layer, as shown in [24](#page18) (a). Consequently, it decreases the target total HetNet throughput, *R*, per TSS UE, as shown in [24](#page18) (b). While the UEs move with high mobility under the LiFi APs, they may experience a series of return association to LiFi and LTE/WiFi RF APs, resulting in a ping-pong process. In real-time VH operations, the packet loss is caused due to the time waiting for the transmission and reception of signalling messages associated with network discovery and handover management procedures [59]. This mainly explains the higher delay and lower throughput of traf-c ows/packets under high mobility in Fig. [23](#page17) (a) and (b), and Fig. [24](#page18) (a) and (b). Also, these gures show that the ows/packets delay, throughput and network throughput are improved under low and high mobility in SNS than RNS. This shows that the TE scheme could alleviate the impact of high UE mobility on traf c, service and network performance.



**FIGURE 24.** Mobility impact on (a) average TSS throughput and (b) targetHetNet throughput per TSS UE.

**VIII. RESEARCH CHALLENGES**

The control and data planes of SDN architectures have limited available solutions to support reliable 5G cells and HetNet (i.e. WiFi/LTE/LiFi cells) integration in one network which

VOLUME 8, 2020

we call as 5GC network. The OpenFlow protocol [12] does not meet the requirements of 5GC wireless communica-tion and networking protocols [2]. The key research chal-lenges aim to support: (i) dynamic programmability in the wireless operations of APs through the open southbound interfaces, (ii) adaptable ef cient traf c, user and signalling management schemes in the control plane and (iii) appli-cations that can leverage the 5GC network brain (i.e. net-work control and information state) (iv) SDN for ultra dense 5GC network management and network-as-a-service. In the following sections, we discuss some open research challenges.

1. **SDN-ENABLED ULTRA-DENSE**

**5G**C **NETWORKS MODELLING**

The computing and processing capabilities of the SDN con-troller and OF switches constraint the potential performance of the control and data planes in terms of handling traf c ows volume and network size. Multiple SDN controllers are required to manage and control ultra-dense 5GC small cell network which provides services in often complex indoor environments (e.g. residential skyscrapers). While each con-troller manages its local network, it can be interconnected with others from its east and west interfaces or through com-plex hierarchical SDN architectures. In this context, the fol-lowing research challenges emerge to develop new solutions for enabling SDN in ultra dense 5GC network.

Analytical models are still required to express the capacity boundary conditions of large SDN-enabled ultra-dense 5GC network. Stochastic network calculus can be utilized for analytical modelling of multiple SDN enabled 5GC networks, though identifying the envelop of arrival and service curves at the different controllers and OF switches remains a very tedious work. Conse-quently, it is very challenging to develop closed forms for the end-to-end (e2e) packet delay upper bounds at the queues of controllers and switches.

The SDN applications can apply, through the SDN controller, ow control, routing, prioritization and QoS enforcement at the AP switches to offer services in the data plane. However, it is still not clear how an application running in one network can in uence the OF rules in other networks that are managed by dif-ferent controllers. More research work is needed to develop centralized mechanisms, which can support reli-able scalable applications and services provisioning in inter-5GC networks.

1. **TRAFFIC ANALYSIS IN SDN-ENABLED 5G**C **NETWORKS**

The different ports in the SDN architectures provide criti-cal information regarding the traf c ows and packets. The following research challenges require to analyse and process the SDN information and traf c characteristics to improve network performance and support SDN applications.

Correlating the analysis of SDN information with the traf c statistics on the ports of SDN architectures is

66689

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

one way to distinguish the elephant ows in each traf c data stream and group them based on a certain measure of similarity [25]. The volume and complexity of the traf c heterogeneity call for ef cient arti cial intelli-gence techniques to classify ows and detect on the y anomalies regarding network or information security. Besides, the computation of wireless resource allocation can be very challenging to meet the SDN time scale operations.

Real-time network capacity dimensioning through steer-ing traf c ows among APs is very challenging in 5GC networks. A research challenge lies in dynamically allo-cating the network capacity based on the current and future traf c while guaranteeing their e2e QoS require-ments. New approaches should be developed to support deterministic network models, which can use network calculus to compute the optimal paths for each ow through the priority queues of the different APs in the 5GC networks. A research challenge lies in developing a scheme that can guarantee e2e real-time QoS com-munication services across the SDN-enabled ultra dense 5GC networks.

5GC networks business models require clear workloads triggers for recon guring the whole network infrastruc-ture to meet their business (i.e service provisioning) requirements. However, the lack of programmable net-work control plane has so far hindered the realiza-tion of software-de ned 5GC networks. A research challenge lies in developing scheduling and automated resource allocation algorithms, which can consider the use of network state information and traf c characteristics to dynamically manage users to APs association.

1. **MOBILITY MODELLING IN SDN-ENABLED**

**5G**C **NETWORKS**

As a UE velocity increases, it spends shorter time to tra-verse the AP coverage. This increases the number of vertical handovers during a xed time interval. The frequent UE handover results in constantly changing OF rules in the net-work. The following research challenges require solutions to alleviate the potential impact of mobility on the performance of SDN-enabled 5GC networks.

Mobility robustness is a challenge, because moving UEs may switch rapidly among the 5GC cells. A UE transfer or traf c of oading policy from one LiFi AP to a WiFi AP depends on the AP selection and of oading policies, which, in turn, depends on the number of overlapped APs, the overlapping area size, and the user/network preference. A research challenge lies in developing a mobility manager module that can reacts fast and cope with the number of UEs and small cells in the 5GC networks. It should leverage the RSSI (Received Sig-nal Strength Indicator) and other parameters tracking primitives to trigger a handover process. The mobility

66690

manager module periodically checks if a better han-dover opportunity exists, even if the channel quality experience by the current APs offering services to the active UEs is still acceptable.

Mobility management functions can base their decisions on network status beyond local radio or optical wire-less channel quality at the cell site (e.g., energy, traf c, and interference awareness), while still providing mini-mal service interruptions during handovers. A research challenge lies in developing optimization techniques for optimal virtual network functions (VNFs) deployment. For example, operators can implement ef cient VNFs that of oad user traf c at the network edge, while dis-tributing traf c load among the different core nodes following load balancing policies. Small cells clustering can be done more ef ciently with network-supported decisions rather than terminal-supported decisions. One particular research challenge is where to place the VNF pool initially; that is, near the edge or near the core of the network. VNFs with real-time constraints are deployed near the edge and those with coordination requirements near the core. Although this split is intuitive, the deploy-ment scenario, where both requirements are present, is still unexplored.

**IX. CONCLUSION**

Analytical mathematical models have been developed and evaluated by using MATLAB, which clarify the relationship among (a) service requests rate of applications, (b) queue lengths of northbound and southbound interfaces (c) retrial queue length and (d) controller processing rate. The capacity of SDN controller in terms of processing rate and buffer space should be carefully dimensioned to support more SDN applications and AP switches in a reliable-manner. The SDN controller service rate can shorten the retrial queue length by quickly serving the applications that are already handled in a previous SDN time period. New ow-rules can be set in the APs based on the retrial and rejoining probabilities range, which can be incorporated in the con-troller’s brain to assign ow-rules in the AP switches in a proactive-manner. This signi cantly reduces the controller set up time and traf c ows blocking probability, which sup-port more SDN applications and services in the HetNet data plane.

A TE scheme has been developed, which runs on the net-work, MAC and across network-MAC layers. It has routing policies to route traf c ows to the different APs; and a two-level scheduler to transmit traf c packets on the MAC layer of APs according to their service class requirements and target HetNet performance. These functionalities, which run on the network and MAC layers, provide the SDN appli-cations with routing and scheduling algorithms and poli-cies to offer differentiated services and granular resource allocation in the HetNet data plane. Users can leverage the HetNet diversity to receive the data rate from the APs that

VOLUME 8, 2020

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

can provide them wireless coverage and the requested ser-vices. The simulation scenarios have been evaluated in two network settings, which demonstrate the impact of SDN on services provisioning performance and network autonomy in dynamically offering services to the UEs. A number of per-formance evaluation scenarios have been conducted, which demonstrate the capabilities of TE scheme to support multiple services, HetNet and applications convergence. The proposed TE scheme guarantees the minimum delay and throughput per service class in active and failed HetNet, while improving the target total HetNet network throughput per TSS UE. It also supports service provisioning restoration applications. The TE scheme algorithms ensure ef cient resource alloca-tion and seamless LTE/WiFi/LiFi cells integration. It can be deployed in any AP, independent of its underlying wireless technology.

**REFERENCES**

1. H. Haas, *High-Speed Wireless Networking Using Visible Light*. Belling-ham, WA, USA: SPIE Newsroom, Apr. 2013.
2. 5GMF. *5G Mobile Communications Systems for 2020 and Beyond*. 5GMF White Paper V1.1, 2017. [Online]. Available: https://5gmf.jp/ en/whitepaper/5gmf-white-paper-1-1/
3. D. Lopez-Perez, M. Ding, H. Claussen, and A. H. Jafari, ‘‘Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deploy-ments,’’ *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2078 2101, Jun. 2015.
4. X. Ge, J. Ye, Y. Yang, and Q. Li, ‘‘User mobility evaluation for 5G small cell networks based on individual mobility model,’’ *IEEE J. Sel. Areas* *Commun.*, vol. 34, no. 3, pp. 528 541, Mar. 2016.
5. *Technical Speci cation Group Services and System Aspects*, document TR21.915 V15.0.0 (2019-09), 3GPP, Mar. 2019, pp. 1 112, vol. 1, no. 15.
6. *5G: A Technology Vision*, Huawei Technol., Huawei, Shenzhen, Apr. 2013.
7. E. Westerberg, ‘‘4G/5G RAN architecture: How a split can make the dif-ference,’’ Ericsson, Stockholm, Sweden, Tech. Rep. 6, Jul. 2016. [Online]. Available: https://www.ericsson.com/49ec87/assets/local/reports-papers/ ericsson-technology-review/docs/2016/etr-ran-architecture.pdf
8. *Technical Speci cation Group Services and System Aspects*, document TR21.915 V1.1.0 (2019-03), 3GPP, Mar. 2019, pp. 1 112, vol. 1, no. 15.
9. O. Galinina, A. Pyattaev, S. Andreev, M. Dohler, and Y. Koucheryavy, ‘‘5G multi-RAT LTE-WiFi ultra-dense small cells: Performance dynamics, architecture, and trends,’’ *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6,
   1. 1224 1240, Jun. 2015.
10. R. Abhishek, D. Tipper, and D. Medhi, ‘‘Network virtualization and sur-vivability of 5G networks: Framework, optimization model, and perfor-mance,’’ in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018,
    1. 1 6.
11. D. Kreutz, F. M. V. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, ‘‘Software-de ned networking: A compre-hensive survey,’’ *Proc. IEEE*, vol. 103, no. 1, pp. 14 76, Jan. 2015.
12. N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, ‘‘OpenFlow: Enabling innovation in campus networks,’’ *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, pp. 69 74, Mar. 2008.
13. A. Gudipati, D. Perry, L. E. Li, and S. Katti, ‘‘SoftRAN: Software de ned radio access network,’’ in *Proc. 2nd ACM SIGCOMM workshop Hot Topics* *Softw. De ned Netw. HotSDN*, New York, NY, USA, Dec. 2013, pp. 25 30.
14. Y.-D. Lin, P.-C. Lin, C.-H. Yeh, Y.-C. Wang, and Y.-C. Lai, ‘‘An extended SDN architecture for network function virtualization with a case study on intrusion prevention,’’ *IEEE Netw.*, vol. 29, no. 3, pp. 48 53, May 2015.
15. S. Gringeri, N. Bitar, and T. J. Xia, ‘‘Extending software de ned network principles to include optical transport,’’ *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 32 40, Mar. 2013.
16. *O. vSwitch*. Accessed: Apr. 9, 2020. [Online]. Available:https://www.openvswitch.org/

VOLUME 8, 2020

1. M. Jarschel, S. Oechsner, D. Schlosser, R. Pries, S. Goll, and P. Tran-Gia, ‘‘Modeling and performance evaluation of an open ow architecture,’’ in *Proc. 23rd Int. Teletraf c Congr. (ITC)*, San Francisco, CA, USA, Sep. 2011, pp. 1 7.
2. J. W. Cohen, *The Single Server Queue*, vol. 8. New York, NY, USA: North-Holland, 1982.
3. B. D. Choi, Y. C. Kim, and Y. W. Lee, ‘‘The M/M/c retrial queue with geometric loss and feedback,’’ *Comput. Math. with Appl.*, vol. 36, no. 6,
   1. 41 52, Sep. 1998.
4. J. Naous, D. Erickson, G. A. Covington, G. Appenzeller, and N. McKeown, ‘‘Implementing an OpenFlow switch on the NetFPGA platform,’’ in *Proc.* *4th ACM/IEEE Symp. Archit. for Netw. Commun. Syst. ANCS*, San Jose,CA, USA, Nov. 2008, pp. 1 5.
5. A. Bianco, R. Birke, L. Giraudo, and M. Palacin, ‘‘OpenFlow switching: Data plane performance,’’ in *Proc. IEEE Int. Conf. Commun.*, May 2010,
   1. 1 5.
6. A. Khan and N. Dave, ‘‘Enabling hardware exploration in software-de ned networking: A exible, portable OpenFlow switch,’’ in *Proc. IEEE 21st* *Annu. Int. IEEE Symp. Field-Program. Custom Comput. Mach.*, Seattle,WA, USA, Jun. 2013, pp. 145 148.
7. S. Azodolmolky, R. Nejabati, M. Pazouki, P. Wieder, R. Yahyapour, and D. Simeonidou, ‘‘An analytical model for software de ned networking: A network calculus-based approach,’’ in *Proc. IEEE Global Commun. Conf.* *(GLOBECOM)*, Dec. 2013, pp. 1397 1402.
8. R. L. Cruz, ‘‘A calculus for network delay. I. network elements in isola-tion,’’ *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 114 131, Jan. 1991.
9. L. Yang, B. Ng, and W. K. G. Seah, ‘‘Heavy hitter detection and identi - cation in software de ned networking,’’ in *Proc. 25th Int. Conf. Comput.* *Commun. Netw. (ICCCN)*, Aug. 2016, pp. 1 10.
10. J. Huang, Y. He, Q. Duan, Q. Yang, and W. Wang, ‘‘Admission control with ow aggregation for QoS provisioning in software-de ned network,’’ in *Proc. IEEE Global Commun. Conf.*, Atlanta, GA, USA, Dec. 2014,
    1. 1182 1186.
11. K. Mahmood, M. Jarschel, O. Østerbł, and A. Chilwan, ‘‘Modelling of OpenFlow-based software-de ned networks: The multiple node case,’’ *IET Netw.*, vol. 4, no. 5, pp. 278 284, Sep. 2015.
12. L. Jorguseski, A. Pais, F. Gunnarsson, A. Centonza, and C. Willcock, ‘‘Self-organizing networks in 3GPP: Standardization and future trends,’’ *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 28 34, Dec. 2014.
13. J. Kim, H.-W. Lee, and S. Chong, ‘‘Virtual cell beamforming in cooperative networks,’’ *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1126 1138, Jun. 2014.
14. S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, ‘‘User-centric ultra-dense net-works for 5G: Challenges, methodologies, and directions,’’ *IEEE Wireless* *Commun.*, vol. 23, no. 2, pp. 78 85, Apr. 2016.
15. O. Tickoo and B. Sikdar, ‘‘Modeling queueing and channel access delay in unsaturated IEEE 802.11 random access MAC based wireless networks,’’ *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 878 891, Aug. 2008.
16. Z. Shu, J. Wan, J. Lin, S. Wang, D. Li, S. Rho, and C. Yang, ‘‘Traf c engi-neering in software-de ned networking: Measurement and management,’’ *IEEE Access*, vol. 4, pp. 3246 3256, 2016.
17. M. T. Z. Win, Y. Ishibashi, and K. T. Mya, ‘‘QoS-aware traf c engineering in software de ned networks,’’ in *Proc. 25th Asia Paci c Conf. Commun.* *(APCC)*, Nov. 2019, pp. 171 176.
18. Y. Lin, Y. Gao, Y. Li, X. Zhang, and D. Yang, ‘‘QoS aware dynamic uplink-downlink recon guration algorithm in TD-LTE HetNet,’’ in *Proc.* *IEEE Globecom Workshops (GC Wkshps)*, Dec. 2013, pp. 708 713.
19. S. Namal, I. Ahmad, A. Gurtov, and M. Ylianttila, ‘‘SDN based inter-technology load balancing leveraged by ow admission control,’’ in *Proc. IEEE SDN Future Netw. Services (SDN4FNS)*, Trento, Italy, Nov. 2013, pp. 1 5.
20. L. Takacs, *Introduction to the Theory of Queues*. New York, NY, USA: Oxford Univ. Press, 1962.
21. J. M. Smith, ‘‘Optimal design and performance modelling of M/G/1/K queueing systems,’’ *Math. Comput. Model.*, vol. 39, nos. 9 10,
    1. 1049 1081, May 2004.
22. H. Alshaer and H. Haas, ‘‘Bidirectional LiFi attocell access point slicing scheme,’’ *IEEE Trans. Netw. Service Manage.*, vol. 15, no. 3, pp. 909 922, Sep. 2018.
23. S. Shenker and A. Weinrib, ‘‘The optimal control of heterogeneous queue-ing systems: A paradigm for load-sharing and routing,’’ *IEEE Trans.* *Comput.*, vol. 38, no. 12, pp. 1724 1735, Dec. 1989.

66691

H. Alshaer, H. Haas: SDN-Enabled Heterogeneous Wireless Networks and Applications Convergence

1. F. Bonomi and A. Kumar, ‘‘Adaptive optimal load balancing in a non-homogenous Multiserver system with a central job scheduler,’’ *IEEE* *Trans. Comput.*, vol. 39, no. 10, pp. 1232 1250, Oct. 1990.
2. S. T. Atan, ‘‘Solution methods for controlled queueing networks,’’ M.S. thesis, Iowa State Univ., Ames, IA, USA, Jun. 1997.
3. H. F. Chen, *Recursive Estimation Control Stochastic System*. New York, NY, USA: Wiley, 1985.
4. A. Kumar, ‘‘Adaptive load control of the central processor in a distributed system with a star topology,’’ *IEEE Trans. Comput.*, vol. 38, no. 11, pp. 1502 1512, Nov. 1989.
5. B. Baynat, ‘‘Analytical models for dimensioning of OFDMA-based cellu-lar networks carrying VoIP and best effort traf c,’’ *Int. J. Comput. Netw.*, vol. 4, no. 4, pp. 104 134, 2012.
6. S. Doirieux, B. Baynat, M. Maqbool, and M. Coupechoux, ‘‘An analytical model for WiMAX networks with multiple traf c pro les and throttling policy,’’ in *Proc. 7th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless* *Netw.*, Jun. 2009, pp. 1 8.
7. S. Shakkottai and R. Srikant, ‘‘Scheduling real-time traf c with deadlines over a wireless channel,’’ *Wireless Netw.*, vol. 8, no. 1, pp. 13 26, 2002.
8. M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, ‘‘Scheduling in a queuing system with asynchronously varying service rates,’’ *Probab. Eng. Informational Sci.*, vol. 18, no. 2, pp. 191 217, Apr. 2004.
9. M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, ‘‘Providing quality of service over a shared wireless link,’’ *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150 154, Nov. 2001.
10. S. Shakkottai and A. Stolyar, ‘‘Scheduling algorithms for a mixture of real-time and non-real time data in HDR,’’ *Teletraf c Sci. Eng.*, vol. 1, Jan. 2001, pp. 793 804.
11. E. Hyytiä and J. Virtamo, ‘‘Random waypoint mobility model in cellular networks,’’ *Wireless Netw.*, vol. 13, no. 2, pp. 177 188, Apr. 2007.
12. E. Hyytia, P. Lassila, and J. Virtamo, ‘‘Spatial node distribution of the random waypoint mobility model with applications,’’ *IEEE Trans. Mobile* *Comput.*, vol. 5, no. 6, pp. 680 694, Jun. 2006.
13. J. R. Barry, J. M. Kahn, W. J. Krause, E. A. Lee, and D. G. Messerschmitt, ‘‘Simulation of multipath impulse response for indoor wireless optical channels,’’ *IEEE J. Sel. Areas Commun.*, vol. 11, no. 3, pp. 367 379, Apr. 1993.
14. Z. Chen, N. Sera movski, and H. Haas, ‘‘Angle diversity for an indoor cellular visible light communication system,’’ in *Proc. IEEE 79th Veh.* *Technol. Conf. (VTC Spring)*, Seoul South Korea, May 2014, pp. 1 5.
15. H. Alshaer and H. Haas, ‘‘SDN-enabled Li-Fi/Wi-Fi wireless medium access technologies integration framework,’’ in *Proc. IEEE Conf. Stan-dards for Commun. Netw. (CSCN)*, Berlin, Germany, Oct. 2016, pp. 1 6.
16. D. Tsonev, S. Sinanovic, and H. Haas, ‘‘Practical MIMO capacity for indoor optical wireless communication with white LEDs,’’ in *Proc. IEEE* *77th Veh. Technol. Conf. (VTC Spring)*, Jun. 2013, pp. 1 5.
17. Y. L. Lee, J. Loo, T. C. Chuah, and A. A. El-Saleh, ‘‘Fair resource allocation with interference mitigation and resource reuse for LTE/LTE A femtocell networks,’’ *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8203 8217, Oct. 2016.
18. V. Chandrasekhar, J. Andrews, and A. Gatherer, ‘‘Femtocell networks: A survey,’’ *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59 67, Sep. 2008.
19. G. Bianchi, ‘‘Performance analysis of the IEEE 802.11 distributed coordi-nation function,’’ *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535 547, Mar. 2000.
20. M. A. A. Masri, A. B. Sesay, and A. O. Fapojuwo, ‘‘Session state aware handover procedure for VoIP sessions in heterogeneous wireless net-works,’’ in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Istanbul, Turkey, Apr. 2014, pp. 3011 3016.

66692

HAMADA ALSHAER (Senior Member, IEEE) received the Ph.D. degree in computer engi-neering and telecommunications from Pierre et Marie Curie University, France, in 2005. He is currently involved in software de ned wireless networking, security, and mobile networks vir-tualization research development with the LiFi Research and Development Centre, School of Engineering, The University of Edinburgh, U.K. He has researched for over 18 years with British



Telecom, Etisalat, INRIA, France, and Leads University, in academic and industrial research development. He has authored over 50 articles in various research topics in electrical and computer engineering, including information technology and systems, and authored a book entitled *Demanding Traf c* *Control and Management in Next Generation Networks: QoS Analysis, Network Simulation, Performance Modeling* (LAP LAMBERT AcademicPublishing, 2009). His research interests include cross-layer design for vehic-ular networks, software-de ned wireless networking, wireless networks vir-tualization, mobile communications security, optical communications and networking, wireless sensor networks, and intelligent transportation. He was a recipient of the 2009 Royal Academy of Engineering Travel Award and scholarships from UNRWA, French Government, and Pierre et Marie Curie University for academic distinctions. He is a Regular Reviewer of research grants submitted for prestigious funding bodies in Europe, Asia, and Middle East. He has served on the Technical Program Committee for various IEEE conferences, including the IEEE Intelligent Vehicles Symposium, the Vehic-ular Technology Conference, the IEEE GLOBECOM, ICC, and WCNC, and chaired some of their sessions.

HARALD HAAS (Fellow, IEEE) received the Ph.D. degree from The University of Edinburgh, in 2001. He is currently the Chair of mobile com-munications with The University of Edinburgh, the Initiator, the Co-Founder, and the Chief Sci-enti c Of cer of pureLiFi Ltd., and the Director of the LiFi Research and Development Centre, The University of Edinburgh. He has authored 500 conference papers and journal articles. His main research interests are in optical wireless com-



munications, hybrid optical wireless and RF communications, spatial mod-ulation, and interference coordination in wireless networks. He is a Fellow of the Royal Academy of Engineering. He was a recipient of the Prestigious Established Career Fellowship from the Engineering and Physical Sciences Research Council (EPSRC), U.K., in 2012 and 2017. In 2014, he was selected by EPSRC as one of ten Recognizing Inspirational Scientists and Engineers Leaders in U.K. He was a co-recipient of the EURASIP Best Paper Award for the *Journal on Wireless Communications and Networking*, in 2015, the Jack Neubauer Memorial Award of the IEEE Vehicular Technol-ogy Society, and recent best paper awards at VTC-Fall, 2013, VTC-Spring 2015, ICC 2016, ICC 2017, and ICC 2018. He received the Outstanding Achievement Award from the International Solid State Lighting Alliance, in 2016, and the James Evans Avant Garde Award of the IEEE Vehicular Technology Society, in 2019. He gave two TED Global talks, Wireless Data From Every Light Bulb and Forget Wi-Fi: Meet the New Li-Fi Internet, which together have been downloaded more than 5.5 million times. He is an Associate Editor of the IEEE JOURNAL OF LIGHTWAVE TECHNOLOGY.



VOLUME 8, 2020