RT-TelSurg: Real Time Telesurgery Using SDN, Fog, and Cloud as Infrastructures

SHAHRZAD SEDAGHAT AND **AMIR HOSSEIN JAHANGIR**

Department of Computer Engineering, Sharif University of Technology, Tehran 14588-89694, Iran

Corresponding author: Amir Hossein Jahangir (jahangir@sharif.ir)

ABSTRACT This paper proposes a novel and efficient real time network architecture, named RT-TelSurg, for one of the most appealing tactile Internet applications, i.e., Telesurgery. In telesurgery, the patient's vital signs and status and the required robotic commands during the surgery should be received on time. Otherwise, the life of the patient or the safety of the operation is endangered. Hence, transmitted packets should meet their respective relative deadlines. Software-defined networking is a relatively new architecture for computer and telecommunications networks in which the network control plane is separated from the data plane. One way to achieve real time telesurgery is to employ cloud and fog networks using SDN as infrastructure. By using a real time cloud controller in the SDN as the core and a fog controller on the edge of the network on the master (physician) side, one may satisfy the acceptable level of the timing constraints of the telesurgery. Accordingly, based on the presented architecture, we develop the statistical performance model of our proposed approach, RT-TelSurg. This statistical queuing theory model is designed for critical and non-critical states of the network. Based on these states, network resources are allocated to telesurgery data in both modes, and a computational bottleneck is detected. RT-TelSurg is evaluated according to real time and surgery efficiency parameters. The results show that the average deadline hit ratio is 98.2% in different conditions, which is quite acceptable for telesurgery applications.

INDEX TERMS Real time, telesurgery, SDN, cloud, fog, statistical modeling, timing scheduling algorithm.

I. INTRODUCTION

Telesurgery (also known as remote surgery) is a technique for performing surgery online on a patient by a physician through an electronic channel without being physically present at his/her location. It can revolutionize traditional healthcare by allowing the use of expertise and practice of specialized physicians worldwide at any time [1]–[3].

The high prevalence of coronavirus (Covid-19) is a clear example of the importance of telemedicine/remote surgery. As health protocols force people to keep a minimum distance to prevent infection, which is very difficult in a small, enclosed operating room, telesurgery can protect surgeons, anesthesiologists, and other operating room staff from being in direct contact with affected patients.

Software defined network (SDN) is a novel network structure in which the control unit of the communication network is separated from the packet forwarding unit. Traditional network algorithms cannot adapt quickly to changing network conditions and are therefore not suitable for meeting real time requirements. SDN provides a way to implement this flexibility. Because these network controllers have a centralized view of the network, they can respond faster and adapt to changing network conditions. They also make network management and monitoring easier [4].

In packet switching networks, a packet flow (or network flow) is a sequence of packets sent from a source to a destination. In real time (RT) networks, time-sensitive flows that miss their deadlines are either considered useless (as in "firm" RT systems) or less important or "gracefully degraded" (as in "soft" RT systems). Packet flows that should reach their destinations at a given deadline are called RT flows as opposed to non-RT flows. In ordinary Internet networks, "meeting the deadline" (also called deadline hit) does not matter and is not considered. However, there are applications in which the commitment to perform an action or send a message before a deadline is essential. In our study, we divide network applications into three categories: 1- Those related to telesurgery and requiring high deadline hit ratio, 2- Those not related to telesurgery but requiring the

respect of time constraints, and 3- Those which are not subject to such constraints (Non-RT). The latter case is generally called the "best-effort" policy. RT network applications are sensitive to time constraints, and their deadlines are either firm (i.e., a packet that misses its deadline has to be discarded) or soft (i.e., the "importance" or the "utility" of a late packet degrades gracefully to zero after its deadline). Hard RT applications (i.e., those which cause severe harm or trouble to humans and the environment if they miss their deadlines) are rarely implemented over the Internet due to the "best-effort" and non-RT policy used in it. RT characteristic is the ability to perform an assigned task correctly before the expiration of its deadline. In a network context, the RT forwarding algorithm is referred to as any forwarding method that improves packets' delivery before their relative deadlines. Generally, the purpose of RT forwarding is to route packet streams on time, and this paper uses several techniques to accomplish this, including statistical modeling, our novel scheduling algorithm LE2EDF (Largest End-to-End Delay First), and bandwidth reservation. We propose an architecture with a simple algorithm whose implementation causes minimum delay, high deadline hit ratio, and relatively low loss of packets in the network.

Telesurgery is one of the most recognized and obvious RT applications that depend on an acceptable deadline hit ratio, end-to-end delay, and jitter. In telesurgery, the end-to-end delay should be less than the human body reflex time [3], otherwise, there may be a disruption in the telesurgical operation. It seems that the main obstacle to the implementation of telesurgery is not its high cost, but its dependency on network delay. It is a low deadline hit ratio, or a non-deterministic behavior of the network will cause significant problems. Telesurgery has master and slave parts: the surgeon plays the role of the master actor, and the patient is in the slave's role. The communication between them is established through the Internet, taking into account the real time requirements. One of the most fundamental topics in the RT network design is the scheduling policy. It can ensure the quality of service or experience.

In terms of network infrastructure architecture, RT-TelSurg, may generally be viewed as a combination of three main technologies: SDN, cloud, and fog/edge. The communication infrastructure presented in this paper is entirely consistent with the SDN architecture in which cloud controllers are designed to forward traffic of non-telesurgery data, and the fog controller, which is located close to the master, controls telesurgery data traffic. SDNs have significant capabilities to achieve this goal. Due to the strict real time requirements compliance for telesurgical procedures, these flows should be distinguished from other ones, and network resources should be allocated in compliance with the flows' priority. We have tried to provide flexible forwarding according to the network status to increase the deadline hit ratio and reduce the end-to-end jitter.

To increase the deadline hit ratio, improve network efficiency, provide high bandwidth to users, and reduce network congestion, the network devices should be located as close as possible to the end-user equipment. This purpose can be attained using fog sub-networks in which computation is performed at most one hop away from the end-user equipment [5]. Hence, we use fog computing for this potential application.

In the following, using Touch Surgery simulator and with medical students' participation, the performance of RT-TelSurg is measured by comparing several RT parameters and the effectiveness of surgery. Finally, implementation results show the suitability of RT-TelSurg for the mentioned properties.

A. OBJECTIVE

This paper aims to present and investigate an efficient solution to meet RT traffic requirements, particularly improving deadline hit ratio in telesurgery applications based on fiber-optic infrastructure by using SDN, cloud, and fog technologies in different network conditions. It is evaluated in terms of several criteria, with some physicians and surgeon medical students' consultation and cooperation. One of the main objectives of this research is that RT-TelSurg is presented in a way that considers a high priority for RT telesurgery flows and is also fair to avoid the starvation of lower priority RT and non-RT flows.

B. INNOVATION

In the present research, the most important innovation is to provide a new effective packet forwarding solution so that RT packets in telesurgery applications meet their RT requirements using some techniques such as exploiting the reserved bandwidth based on network status and the proposed novel scheduling algorithm, LE2EDF. Also, a statistical model for SDN packet forwarding specific to the proposed infrastructure is presented. Based on our gap analysis, this research is one of the few studies that has explicitly addressed the deadline hit ratio (See Table 2). On the other hand, the proposed scheduling algorithm has been customized to meet RT requirements in the SDN because the traditional algorithms for SDN are not appropriate, effective, and applicable. We show that the RT-TelSurg has provided an acceptable end-to-end delay and deadline hit ratio because in the implementation phase (V. implementation and analysis of result), we have compared two of the studies that are close to our proposed solution with RT-TelSurg and the results are proof of this claim. In addition, in the mathematical model presented in this paper, the average number of packets in each queue is calculated. This allows system designers to estimate the hardware requirements for the queue. As Internet applications have different RT requirements, we have a tradeoff solution that allocates as many resources as needed to each application while maintaining the average deadline hit ratio acceptable. Using objective (scoring-based surgical simulator) and subjective (scoring-based surgeons who participated in the test) evaluation of the telesurgery network infrastructure is another innovation of this research.

C. CONTRIBUTION

This study's main contribution is the ability to deal with the short acceptable end-to-end delay and sufficient deadline hit ratio in telesurgery. The other contributions of the paper can be summarized as follows:

- This study presents a scheduling algorithm that works effectively for the telesurgery system.

- The proposed method can be used in public Internet infrastructure, and not using a dedicated network will be affordable.

- According to Table 2, a small number of previous studies have explicitly addressed the deadline hit ratio issue. This study has expressed the importance of this basic parameter in real time application.

- We have used a feedback queue model that is fully compatible with the telesurgery system. So far, similar statistical modeling in telesurgery applications has not been observed in previous studies. Most of the previous research, such as [6], [7], and [8], has used optimization methods for analytical modeling. However, since the implementation of these algorithms is generally time-consuming, we have tried to use more appropriate methods for analytical modeling in this study.

- The presented analytical model is general and flexible that can be customized according to the network's characteristics.

- The utilization of statistical modeling to determine the network status and the usage of the reserved routes and a new scheduling algorithm for a well-known combination of SDN, Cloud, and Fog infrastructure for telesurgery application are among the contributions of this study.

The remainder of the paper is organized as follows. In Section II, the methodology is described, while in Section III, the related concepts and prior studies are presented. Section IV outlines the proposed strategy called RT-TelSurg. The emulation results are provided and analyzed in Section V. Finally, Section VI concludes the paper and suggests possible future research.

II. METHODOLOGY

- For the telesurgery application, our proposed architecture has used the latest and well-known infrastructures as SDN, Could, and Fog like [1], [3], [8], [9], and [10].

- Real time applications have several design requirements. In this study, and generally in computer networks, relative deadlines are of interest. The guarantee of on-time delivery in a communication system for sporadic messages incurs a high resource utilization cost. Moreover, such a guarantee may not be required for the bulk of real time traffic. Therefore, algorithms and protocols for real time communications are typically designed to tolerate some deadline misses, provided that the miss ratio is less than an acceptable threshold. Real time networks should not be confused with fast networks. To explain their relationship, we can point out that if the end-to-end delay is longer than the relative deadline, the packet has not met its deadline and maybe useless. Hence, in this study, the average deadline hit ratio is defined as the ratio of the ''number of received packets in the slave with the relative deadline less than end-to-end delay'' to the ''total number of sent packets by the master unit.'' Scheduling is one of the most important and key issues in the design of real time networks. It can ensure the quality of service or experience. Except for some hard real time networks that use pre-determined tasks with offline scheduling, most real time networks that transfer the flow of Internet traffic, as discussed in this paper, use scheduling algorithms in the form of online and priority-driven scheduling [11]–[14]. To summarize, Table 1 presents a categorization of time-sensitive network protocols and their properties.

Our proposed approach is introduced at the network layer. To determine the critical and non-critical status, the average sojourn time of telesurgery packets, which is the result of queue theory analysis, has been used.

In the meantime, other techniques have also been implemented in networks for giving priority to fast or time-dependent flows. For instance, the Differentiated Services (DiffServ) protocol is a class-based mechanism for managing traffic composed of several predefined classes. Network flows are mapped to these classes. This mechanism is based on the classification and marking of each class of flows or packets. Intermediate network devices route and forward packets differently, according to their respective classes. Our proposed solution is very similar to the Diffserv mechanism. Still, the latter is well suited for traditional distributed network structure and step-by-step routing. It is not for SDN network architecture, which is centralized and has a global view of network conditions. On the other hand, the Diffserv mechanism only provides the quality of service for each traffic group in a domain and not for end-to-end communications [15]. This research aims to improve the deadline hit ratio, end-to-end delay, and end-to-end jitter of telesurgery flows in SDN centralized platform, and other existing approaches do not guarantee such time constraints. Hence, we use a different paradigm to deal with this issue.

- We choose Touch Surgery simulator [16] to implement and evaluate the proposed approach since it is one of the most widely used simulators in academic research studies [17]–[20]. It has a free trial version, covers a diverse range of surgery types with various difficulty levels, lets the physicians give a user assessment score, and is also accepted by the medical community. An adequate number of medical students (forty participants) have participated in this evaluation, and the results are validated accordingly and compared to previous research.

III. BASIC CONCEPTS AND RELATED WORK

A. TELESURGERY SYSTEM

Telesurgery works as a master-slave system. The following describes each component of the system.

1) MASTER SITE

At the master site, a surgeon can pass the input commands to the controlled robot at the slave site, such as impedance of

TABLE 1. Time-sensitive network protocols and their specifications.

Real Time Applications	Data Link Layer	- Categorized often as hard real time - Consider absolute deadline - Provide deadline guarantee - Consider the worst response time - Tasks are assumed to be either periodic, non-periodic, or sporadic in scheduling algorithms. - Some important protocols are Generic Object Oriented Substation Events (GOOSE), Sampled Values (SV), Time-Sensitive Networking (TSN), Network Time Protocol (NTP), and Precision Time Protocol (PTP), and they are destined mainly for industrial applications.
	Network Layer	- Categorized often as frim/soft real time - Consider relative deadline - Consider average response time - Tolerate some deadline misses which is below the acceptable threshold - Do not guarantee the deadline for all transmissions - One of the important protocols is DiffServ that is destined mainly for communication networks for Internet applications. - The transmission of packets on the Internet has almost all type of flows and patterns, so one must use a hybrid scheduling algorithm for improving deadline hit ratio
	Transport Layer	- Some important protocols are RSVP (reserve resources for a real time flow), SCTP (handle multiple simultaneous streams and multiplexed streams), and DCCP (serve a general congestion-control mechanism for time-sensitive applications)
	Application Layer	- Generally used for transmitting multimedia (e.g., streaming video) - Use codecs to encode the payload data - Tolerate some packet loss - Some important protocols are: Real time Transport Protocol (RTP), Real time Transport Control Protocol (RTCP), Reliable Data Transfer (RDT), Real Time Streaming Protocol (RTSP), Real time Publish-Subscribe Protocol (RTPS), Secure Real time Transport (SRTP), Real Time Messaging Protocol (RTMP), Zimmermann Real Time Transport Protocol (ZRTP), and Simple Network Time Protocol (SNTP)

the robotic arm, velocity, task description, force, and position. The capability to perform sensitive procedures relies exceedingly on feedbacks. In robotic surgery, surgeons require to be mighty to comprehend the value of strength being applied without intermediaries touching the surgical gadgets. This organization is well known for force-feedback. The master site receives feedback data like medical evaluation images, video, audio, tactile sensing, proprioceptive data, text, and the commands it is executing from the slave site [3], [21], [22].

2) SLAVE SITE
It is composed of teleoperators like haptic devices, surgical robots, 3D cameras, actuators, and sensors that utilize the edict of a human operator. The haptic data is sent and received to/from the master site as feedback.

3) NETWORK COMPONENT
Using an exclusive communication network media for implementing telesurgery is very expensive. The Internet can be considered as the first alternative to set up telesurgery as a generally valuable and feasible solution. However, using the Internet has its problems and challenges including the level of QoS (Quality of Service)/QoE (Quality of Experience) and the guarantee to meet packet deadlines. Like any RT application, telesurgery is strongly affected by end-to-end network delay, jitter, and packet loss. Though telesurgery is an RT application, it should tolerate some packet loss to accomplish its goal. Suppose the average deadline hit ratio, end-to-end delay, and jitter are greater than the standard

desired ones. In that case, various malfunctions could happen like poor quality audio/video, delayed response, oscillations, and instability. Hence, telesurgery requires a very high deadline hit ratio, availability, reliability, security, and extreme constraints for end-to-end delay, throughput, and safety.

The best physical layer infrastructure to meet these needs is fiber optics, and the mentioned expectations are entirely consistent with the features of the SDN infrastructure. Effectual network handling is essential to provide good qualified communication for telesurgery applications. Network administrators can readily manipulate the whole of the switches through the central SDN controller. So, it is no longer challenging to configure various vendor's devices, and there are no worries about meeting our expectations for the network's flexibility and scalability. Recently, many research studies have discussed SDN capabilities for use in emerging technologies, such as [23]–[28]. Such studies have examined various network factors. Among all these features, the deadline hit ratio and end-to-end delay are the important issues that this paper addresses. In [29], providing real time requirements in SDN networks (especially hitting the deadlines) is not addressed at all, while the main challenge of our research is to present a solution to provide real time SDN-based networks for the use of telesurgery application. In [30], the authors propose a method based on spectral clustering for SDN traffic classification. They design the flow table representation and extraction procedure. The combination of machine learning methods and SDN is presented in [31]. Some machine learning methods are used for traffic classification based

on the applications in a network with SDN architecture. IHSF [32] has used three approaches, an algorithm to place SDN switches and legacy ones together, a regression-based algorithm for predicting the reliability of legacy links, and a deep deterministic policy gradient optimization algorithm that determines forwarding routes for time-sensitive IoT flows. This study uses deep learning techniques to predict link reliability, which in turn increases network response time. The authors do not address the concept of hitting the deadline, and their solution does not especially consider the telesurgery requirements.

The common point seen in most previous research studies is that they all point to solving the real time problem, but they try to speed up tasks and not meet the deadline.

In [30] and [31], practical methods for classifying data streams are presented, but their analytical techniques are time-consuming, which makes the application of these methods impossible for timely applications such as telesurgery. Instead, our proposed method for flow separation is straightforward and, of course, fast, which makes it ideal for use in telesurgery applications with strict real time constraints.

Generally, a large amount of data is sent to the cloud zone platform through the network to be further processed, stored, integrated, and available at any location or routed. In this research, the cloud's mission has been the last case. It can be customized to the requirement of applications [26]. Considering that the cloud zone requires a long time to handle a massive amount of data, it is essential to select useful real time routing [25].

Fog/edge computing is a new technology considered as a platform at the edge of the network, providing computing, storage, and networking services between smart objects and traditional cloud computing data centers with salient characteristics of low delay, real time interaction, and many others [1]. It moves the computing and storage resources to the edge of the network, close geographically to the user where data is generated. Using fog beside the cloud is a well-known and effective solution. This approach can decrease the amount of data sent to the cloud and reduce the end-to-end delay. In addition to all of this, fog computing integrated with SDN can significantly improve the user quality of experience.

B. PREVIOUS RESEARCH ON TELESURGERY

Initially, telesurgery was performed between France (Strasbourg) and the United States (New York City) in 2001 using a dedicated fiber optic infrastructure. The surveys [3], [21], [22], [33], and [34] provide a comprehensive overview of telesurgery with essential details and their changes over time. In the following, several important papers in this field are introduced.

Some studies have examined the feasibility of telesurgery. For example, in [2], the authors have examined the feasibility of telesurgery through a public network according to QoS requirements by a simulator. In [3], the authors have presented an architecture for telesurgery using a traditional network and the 5th generation mobile network (5G).

They have performed a remote heart surgery using the proposed infrastructure and, in this way, they have made their proposed solution feasible.

Several papers have addressed the issue of increasing the telesurgery QoS. Some of them have tried to reduce the delay, and some have tried to use a flow control mechanism. For example, the authors of [35] have performed an experimental adjustment with various simulated transmission delays to investigate the effect of video transmission delay and surgical performance. In [36], the authors address some time-sensitive issues, such as activating a high-frequency surgical device to propagate force. They use the time-sensitive standard (IEEE 802.1). A combination of Service-oriented Device Connectivity (SDC) and Time-sensitive Networking (TSN) based standards enables semantic collaboration and real time communication. They propose a new way to automatically configure the TSN network using the self-description of SDC-based medical devices. It should be noted that these types of research are limited to devices and equipment within an operating room and have not been used in Internet network environment.

Two studies that address the congestion control issue of telesurgery networks are explained. Network Adaptive Flow Control Algorithm for Haptic data (NAFCAH) [37] is a network-compatible flow control algorithm for tactile data. It aims to control network congestion to improve the quality of telehaptic data, such as telesurgery. In the case of network congestion, it gradually reduces the transmission speed. However, NAFCAH uses RTT (Round Trip Time) to monitor congestion, which is not accurate for measuring estimates. Another application layer that operates over the congestion control mechanism for telehaptic applications is DPM (Dynamic Packetization Module) [38]. The proposed approach monitors network status to measure the end-to-end delay of telehaptic packets. The DPM then adjusts the amount of telehaptic data based on network status feedback. The authors have performed some simulations, and their results show that DPM meets the QoS requirements of telehaptic applications even in very dense network traffic. Although DPM complies with QoS requirements, it must modify end-user applications, and all hosts must be synchronized using the Network Time Protocol (NTP).

In the following, some recent papers relevant to our research (especially those who have used the SDN architecture), are presented. In these papers, techniques similar to our ideas in this research have been used, Such as SDN, fog, scheduling algorithm, allocating segregated resources.

In [1], the physician and the master control terminal were placed on one side, and on the other side, the patient and his controllable desk were present at the edge of the network. The fog computing node provided a slice of the network based on communication needs for 5G users using SDN and Network functions virtualization (NFV) technologies. They also used AI techniques to predict the physician's following behavior unique to her/him. In [8], using SDN and 5G, the network was configured as if each stream was sent through a dedicated

fiber, and this configuration was done quickly to support most dynamic applications. Finally, the researcher evaluated the proposed solution for telesurgery.

The authors of [6] have designed a hierarchical framework based on the priority of network flows. Then Some scheduling algorithms are presented according to different preferences with various characteristics. They have also developed a smooth scheduling strategy based on user experience.

In [7], the authors have proposed a mechanism for optimizing the network resources, which dynamically creates network slices for users based on tactile cyber-physical systems. Their solution consists of two main components. One is the clustering algorithm to determine the sections and their specifications to support the Transmission Control Protocol (TCP) stream, and the second is the use of programmable switches P4 (Programming Protocol-independent Packet Processors) [39]. SDN can provide and change these switches. They also use the SDN controller to pre-calculate the path of the specified sections.

The authors of [10] have isolated users' haptic data from other data and provided diverse, resources. Such a separation scheme provides resource availability for applications and allows resource personalization for them. In [40], the haptic Internet system based on the three-layer MEC (Multi-access Edge Computing), SDN, and NFV has been proposed relative to the structure of 5G. The proposed design includes a user device, RAN (Radio Access Network), cloud units, access switches, OpenFlow switches, Middleboxes, and SDN controllers. Finally, the proposed solution is compared to the evolved traditional mobile network [41]. In [42], the authors have minimized end-to-end delay for telesurgical applications using the Constraint satisfaction problem (CSP) and Ant Colony Optimization (ACO) models. They have used SDN as an infrastructure that is aware of all the network links and resources.

The combination of Type-2 Fuzzy System (T2FS) and Enhanced Cuckoo Optimization Algorithm (E-COA) [43] is an SDN-based communication model proposed to achieve optimal and reliable paths for telesurgery applications. In this paper, the delay has been considered a limitation of the CSP problem, which should be solved by finding minimum cost routes. To provide reliability, two separate paths with no common link were considered simultaneously. In this way, the first optimal path is set as the default path. It is a path destined to compensate for any possible failure. In [44], the authors have proposed SDN-based network architecture for shared internet for telesurgery, used optimization algorithms for traffic management, and simulated and evaluated their solution. Their research used TCP protocol to transmit surgical instructions and force feedback data and the User Datagram Protocol (UDP) protocol to send audio/video data.

In [9], the authors have proposed integrating network coding and SDN to meet the need for very low haptic internet delay. The authors claim that the widespread use of a flexible network encoding mechanism such as random linear network coding (RLNC) across the network can improve delay performance and reduce the required packet retransmission frequency. Finally, the authors have implemented a software router with network coding capability.

In Table 2, the studies, as mentioned above, are compared in terms of their respective distinguishing features. These studies are compared based on the techniques used in this area. In Table 2, the expression "Optimization algorithms" refers to whether different optimization techniques are used in the mentioned articles.

TABLE 2. Comparison of previous studies and RT-TelSurg.

Research	SDN- based infrastructure	Cloud-based infrastructure	Fog-based infrastructure	Priority-based approach	Optimization Techniques	Mathematical Modeling	Scheduling Algorithm	Reducing Delay	Reducing end-to-end Jitter	Consider the concept of a deadline
[7]	✓	✓	×	✓	✓	✓	✓	✓	×	×
[35]	×	×	×	×	×	×	×	✓	×	×
[37]	×	×	×	✓	×	✓	×	✓	×	×
[38]	×	×	×	✓	×	✓	✓	✓	✓	×
[36]	✓	×	×	✓	×	✓	✓	✓	×	×
[3]	✓	✓	✓	×	×	×	×	✓	×	×
[9]	✓	✓	×	×	×	✓	×	✓	×	×
[1]	✓	✓	×	×	✓	×	×	✓	×	×
[8]	✓	✓	✓	×	✓	✓	✓	✓	✓	×
[6]	×	×	✓	✓	✓	✓	✓	✓	×	×
[44]	✓	×	×	×	✓	✓	×	✓	×	×
[10]	✓	✓	✓	✓	×	×	✓	×	×	×
[40]	✓	✓	✓	×	×	×	✓	✓	×	×
[38]	×	×	×	✓	×	✓	×	✓	×	✓
[42]	✓	×	×	×	✓	✓	×	✓	×	×
[43]	✓	×	×	×	✓	✓	×	✓	×	×
RT-TelSurg	✓	✓	✓	✓	×	✓	✓	✓	✓	✓

IV. THE PROPOSED APPROACH- RT-TelSurg

This section explains how to send data in the proposed architecture, RT-TelSurg.

Packets are routed from the sender to the exterior of the local network via fiber optics are divided into two general categories: telesurgery packets and non-telesurgery ones. The proposed network structure treats these two categories differently.

A brief overview of SDN orchestration, cloud, and fog architecture is presented in the following. Then a statistical model is given.

A. PROPOSED RT-TELSURG ARCHITECTURE

In Fig. 1, a general schema of the proposed architecture based on SDN, cloud, and fog is presented.

Although 5G is one of the communication infrastructure candidates for telesurgery implementation, the use of fiber optics also provides high-quality communication, which makes it a good choice for use in-network surgical network infrastructure. Very high bandwidth, low packet loss rate
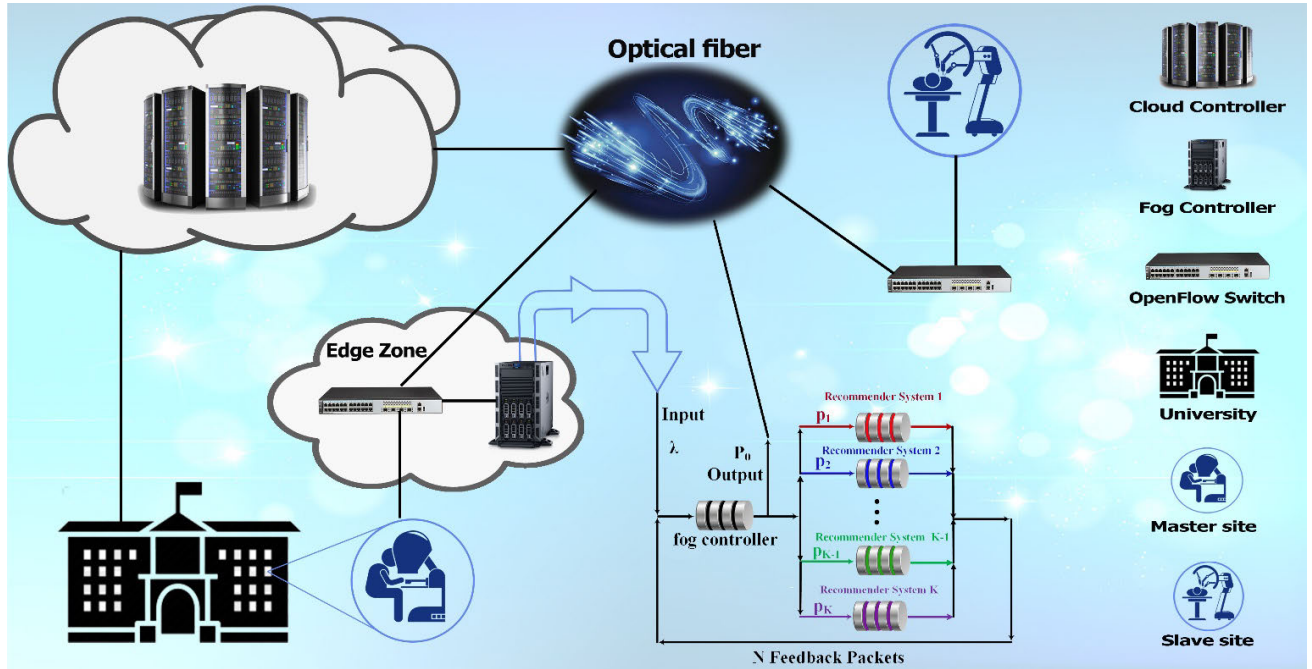
FIGURE 1. The general outline of the proposed architecture based SDN, cloud, and fog.

compared to other physical layer communication channels in the Open Systems Interconnection (OSI) model, a very satisfactory level of security, as well as the highest transmission speed (i.e., the speed of light) are some of the advantages of fiber optics compared to other communication media. Due to these valuable features, the 5G network also uses fiber optics in its lowest level infrastructure, including its inter Base Transceiver Station (BTS) communications.

In this research, two efficient commercial surgery products named, Sina (Sina Robotics & Medical Innovators Co. [45], [46], and [47]) and Parsiss (Parsiss Surgical Navigation Co. [48], [49], and [50]), have been used. These products include some recommender systems (such as removing unwanted vibrations or accidental movements of the surgeon's hand, providing a 3D map of the surgical site) that the surgeon can choose during surgery, and each of them can be helpful. They are placed in the master side fog zone. These units only help surgeons make better decisions and improve their performance, especially when tired due to work pressure, lack of concentration, or critical operation room environment. Hence, they are not a substitute for surgeons. The different parts of the proposed structure are shown in Fig. 1, which and the function of each part is described below.

RT-TelSurg behaves differently with the telesurgery packets and the other ones. The telesurgery packets will be sent to the cloud zone for routing, as described in Section IV.D. RT-TelSurg in fog zone (as described in Section IV.c) can be divided into RT-TelSurg two parts. The first one, which is performed in the fog controller/switch, uses our proposed LE2EDF scheduling algorithm that determines which packet from the queue has to be selected for routing/forwarding. The second part is about deciding under which conditions and

how long the reserved bandwidth can be used (as described in Section IV.C.2). The specified reservation paths are part of the bandwidth of all links used to forward extremely urgent real time telesurgery packets before their relative deadlines in critical mode. The rest of the communication links' bandwidth is to be used for forwarding packets under all remaining conditions is called common bandwidth. Statistical modeling has been used for this decision (as described in Section IV.B).

The outline of RT-TelSurg is illustrated in Fig. 2. In the following subsection, we explain how queuing theory, especially networks of queues, can model and analyze the proposed architecture to determine the critical and non-critical states.

B. STATISTICAL MODELING OF RT-TELSURG IN THE FOG ZONE

Queueing theory is the primary statistical method and natural choice for analyzing network performance and estimating the delay, close to real value. The results of queuing theory to understand network behavior are very noteworthy. Researchers shall usually be eager to approximate some parameters like the average number of packets in the system, the average sojourn time (that is, the time a packet spends in the queue, and service for forwarding) per packet.

These parameters are calculated based on the average arrival rate and the average service rate. The following packets forwarding in the fog zone is modeled based on queue theory. In this way, Little's theorem coupled with Jackson's networks helps us a lot for analysis.

The architecture of the master fog zone is depicted in Fig. 3. The statistical modeling performed to identify the critical state of the network is based on this schema.
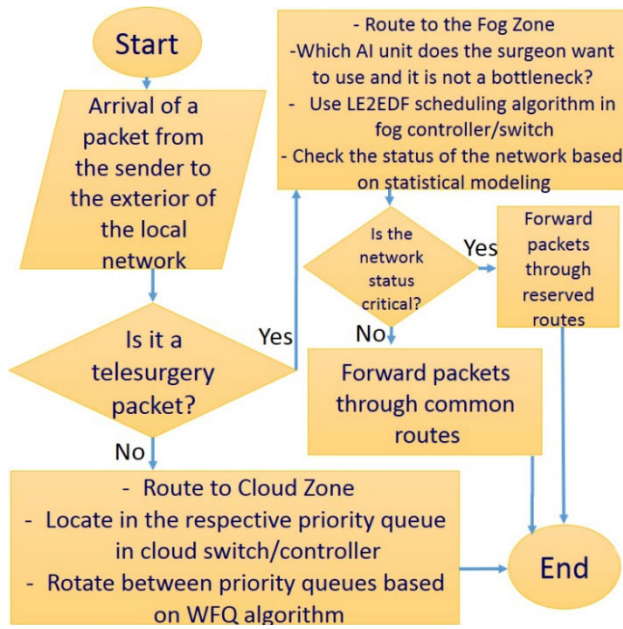
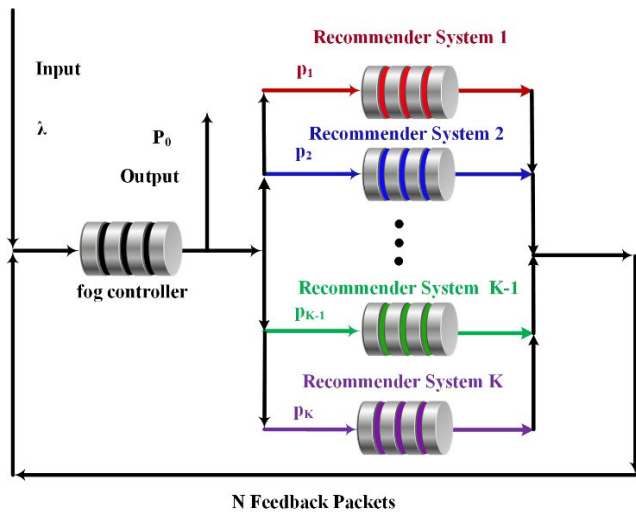**FIGURE 2.** RT-TelSurg overall scheme.

FIGURE 3. Proposed network architecture in the edge zone.

Another feature of the proposed approach is to specify a recommender system that is recognized as a bottleneck. One of the crucial concepts in the design of an RT system is identifying the bottleneck component queue. During the component workload growth, over time, the queues will be continually bustling and saturated.

According to the queuing theory, when the component queue is saturated, the queue length, packet waiting time, and dropped packet ratio sharply increase with the arrival of newer packets to the system unlimitedly. Therefore, the rate of packets meeting their deadlines decreases dramatically. So, identifying the bottleneck is very important to deal with the congestion problem. From a statistical point of view, a bottleneck component is characterized by ρ, the queue's efficiency, where ρ is very close to the 1.

1) RT-TELSURG STATISTICAL MODELING ASSUMPTIONS
In this section, the following two assumptions are considered for the model:

- The number of recommender systems in the fog is unknown, but there is always a fog controller.
- The number of input flows per fog controller is unknown.

2) RT-TELSURG STATISTICAL MODELING
Queuing networks that let input elements (packets) enter from outside the system and eventually leave it, and some packets that return to the system, are named open queues. Jackson open queue network is a numerable sequence of queues in which each input element visits each queue respectively. This queuing model is easy to evaluate and has been applying in many applications, mostly with accurate results. The symbols used in our statistical model are introduced in Table 3.

TABLE 3. The symbols used in statistical modeling.

Symbol	Description
K	Number of queue nodes
N_i	The number of packets in queue node i
$\overline{N_i}$	The average number of packets in queue node i
$\overline{N}$	The average number of packets in the system
λ_i	The average packet arrival rate of queue node i
λ	Average packet arrival rate into the system
μ_i	Average service rate in node i
$p_{i.d}$	The possibility that packet in queue node i leaves the network
T_i	The sojourn time of packet in node queue i
$\overline{T_i}$	The average sojourn time of packet in node queue i
$\overline{\overline{T}}$	The average sojourn time of packet in the system (fog zone)
$\rho_i = \dfrac{\lambda_i}{\mu_i}$	The efficiency of queue node i (average ratio of time when the processor queue is busy). If $\rho < 1$, the queue is stable
B	The index of the bottleneck queue node
CT	The threshold for the critical state of the telesurgery flows

Necessary conditions for an open Jackson network consisting of K nodes (the total number of nodes includes the fog controller and all recommender systems) are as follows:

1- Each node contains a FIFO queue that is infinite. Hence, the efficiency of all queues (i.e., throughput, utilization) is less than one.

2- Any arrival at a node follows a Poisson process with a rate equal to λ [51]–[53]. Therefore, the packet inter-arrival time follows the exponential distribution.

3- All service times in the queue of each node i are exponentially distributed with a service rate μ_i. The service time of each queue is independent of the other ones, and their service rates are different.

4- Upon leaving the queue from node i (fog controller), the packets enter the queue node j randomly with probability $p_{i.j}$ or departs the network with probability $p_{i.d}$.

To define the threshold for the critical state of the telesurgery flows, we use the two concepts of average and variance of the sojourn time of packet in the system. One of the most important advantages of the variance is that it has the UMVUE (Uniformly Minimum Variance

Unbiased Estimator) feature. The term bias refers to several statistical issues that could be classified as measurement, sampling, and estimation bias. In measurement or sampling situations, bias is "the difference between a population mean of the measurements." Since the estimator for each random sample will have a different parameter value, it is expected that by repeated sampling, the estimator's average value will be approximately equal to the actual community parameter. Such an estimator is called an unbiased one. In fact, the existence of such a property separates parameter estimators into two classes of biased and unbiased estimators. Uniform means that this estimator has the least variance in the class of unbiased estimators for all points of the parametric space. This UMVUE estimator is unique [54].

$$N = N_1 + N_2 + \ldots + N_K.$$

$$\bar{N} = \bar{N}_1 + \bar{N}_2 + \ldots + \bar{N}_K = \sum_{i=1}^{K} \bar{N}_i.$$

$$\lambda = \sum_{i=1}^{K} \lambda_i p_{i.d}. \tag{1}$$

Equation (2) and (3) are driven, because of a geometric distribution [55] in this system;

$$\bar{N}_i = \frac{\rho_i}{1 - \rho_i}. \tag{2}$$

And

$$Var(N_i) = \frac{\rho_i}{(1 - \rho_i)^2}. \tag{3}$$

$$T_i = \frac{N_i}{\lambda_i}. \tag{4}$$

Using equation (2) and (3), we have:

$$\bar{T}_i = \frac{\bar{N}_i}{\lambda_i} = \frac{\rho_i}{\lambda_i(1-\rho_i)} = \frac{\rho_i}{1-\rho_i} \times \frac{1}{\rho_i \mu_i} = \frac{1}{\mu_i - \lambda_i}. \tag{5}$$

Because N_i is independent of the other one, we have:

$$E(\bar{N}) = E(\bar{N}_1) + \ldots E(\bar{N}_K) = \sum_{i=1}^{K} \frac{\rho_i}{1 - \rho_i}. \tag{6}$$

$$Var(\bar{N}) = Var(\bar{N}_1) + \ldots + Var(\bar{N}_K)$$
$$= \sum_{i=1}^{K} Var(\bar{N}_i) = \sum_{i=1}^{K} \frac{\rho_i}{(1-\rho_i)^2}. \tag{7}$$

Using equation (3):

$$Var(T_i) = Var\left(\frac{N_i}{\lambda_i}\right) = \frac{Var(N_i)}{\lambda_i^2} = \frac{\rho_i}{\lambda_i^2(1-\rho_i)^2}. \tag{8}$$

For modeling, a powerful and important, and still simple queueing theory formula, called Little's law [56], is used. Its importance is mainly due to its generality. It holds for roughly all queueing systems. According to Little's law, we have:

$$\bar{T} = \frac{\bar{N}}{\lambda} = \frac{1}{\lambda} \sum_{i=1}^{K} \frac{\lambda_i}{\mu_i - \lambda_i}. \tag{9}$$

Using equation (9), we have:

$$Var(\bar{T}) = Var\left(\frac{\bar{N}}{\lambda}\right) = \frac{Var(\bar{N})}{\lambda^2} = \frac{1}{\lambda}\left(\sum_{i=1}^{K} \frac{\rho_i}{(1 - \rho_i)^2}\right). \tag{10}$$

$$CT = \bar{T} + \sqrt{Var(\bar{T})}. \tag{11}$$

Using queuing theory, an efficient analysis is provided for finding bottleneck queues. So, the bottleneck queue node is b, where

$$0.9 \leq \rho_i \leq 1. \tag{12}$$

In this way, if the statistical properties of the recommender system(s) accord with the inequality (propositions 12), then the use of this unit is disabled in the master device as long as this condition is met.

C. ROUTING AND FORWARDING METHODS IN THE MASTER FOG ZONE

1) LARGEST END-TO-END DELAY FIRST (LE2EDF) SCHEDULING ALGORITHM

The proposed scheduling algorithm only applies to telesurgery packets. It is used in fog switch to select a packet from the queue for forwarding. By analyzing the traffic generated by the Touch Surgery simulator, it was found that the packets produced by this simulator at the transmission layer are either TCP or UDP packets. TCP packets transmit motion instruction (command) and feedback data, and UDP protocol packets transmit audio and video data. (It should be noted that some previous research studies like [40], had also used this classification).

One of the purposes of the proposed scheduling algorithm is to reduce end-to-end jitter. If the command packets reach the master sooner than the audio and video ones, the master must use complicated buffering techniques to reduce end-to-end jitter.

The fog switch maintains the latest end-to-end delays of the TCP and UDP protocol packets and, based on the latest information, decides which of the two categories has a higher priority to be selected from the fog switch queue for forwarding. In the fog switch, the slave system is continuously pinged. According to the last amount of the delay, which results from continuous ping, the priority is assigned to the category of the packet, i.e., TCP or UDP. In this way, an attempt has been made to reduce the end-to-end jitter.

2) PACKET FORWARDING STRATEGY

We change the packet forwarding strategy based on the network status. In the network, two types of conditions are foreseen for sending real time telesurgery flows: critical and non-critical.

Equation (11) is used as a threshold value to determine the telesurgery flows' critical state. Suppose the minimum end-to-end delay of the telesurgery flow is more than the threshold value given by Equation (11). In that case, the conditions of

the telesurgery flows are considered to be in a critical state. In this study, this assumption is made only as an example, and it can be modified or tuned according to real situations. This is intended to assure that the network functionality is satisfactory in different network conditions, and is tolerable in a predictable manner, even for undesirable deadline hit ratio.

By default, all packets in all flows, whether RT or non-RT, consume the common network bandwidth. Nevertheless, only real time packets of telesurgery flow in critical states are forwarded via specifically reserved paths.

When a critical telesurgery flow is detected, all its packets are forwarded by using the reserved bandwidth. This reservation lasts until the state of the packets becomes non-critical again.

It should be noted that the fog controller periodically and separately checks the status of each flow based on information received from recommender systems. To increase the flexibility of the network in a critical state and to react on time to the latest network status, the critical state checking period has been chosen in this paper equal to half of CT to determine the critical state in the mid-time of the chosen threshold. However, the period value is arbitrary and can be selected according to network conditions or the operation criticality.

D. ROUTING AND FORWARDING METHOD IN THE MASTER CLOUD ZONE

Non-telesurgery packets are forwarded to the cloud zone. Depending on the type of transport protocol, they are placed in priority queues as shown in Fig. 4. The Weighted Fair Queueing, WFQ, algorithm [57] is used to rotate between the controller/switch cloud queues, and each controller/switch queue applies the FIFO algorithm.
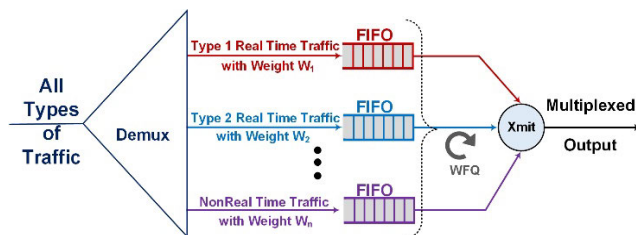

FIGURE 4. The operation process based on the priority of packets in the fog controller and the switch in RT-TelSurg.

The WFQ method is used to handle prioritized queues, whereas the Round Robin (RR) method [58] assumes equal priorities, which are not suitable for RT services. As in the WFQ method, the number of packets extracted from each queue is proportional to its weight. Hence, assigning higher weights to higher priority queues allows RT packets to exit queues sooner. In the model proposed in this section, the priority of each queue is determined in such a way that packets enter priority queues according to their respective priorities. In this algorithm, W_1, W_2, ..., W_N are weights assigned to each N ingress flow. Thus, each packet flow is served proportionally to its weight, and the data rate of each priority

queue, R_i, is obtained from formula 13. R represents the maximum total data rate for the channel.

$$R_i = \frac{RW_i}{(W_1 + W_2 + \ldots + W_N)} \qquad (13)$$

As shown in Table 1, RT applications are divided into four categories in terms of the OSI model. Data Link layer protocols only operate within the Local Area Network (LAN), and packets of these protocols do not access the Internet. In this study, non-telesurgery RT packets based on protocol layers are prioritized so that network-layer protocols, transport layer, and application layer are given greater priority, respectively. Other packets in the network that are not part of the mentioned protocols are considered non-RT packets and have the lowest priority.

The reason for this prioritization is explained below:
- Network layer protocol functionalities are point-to-point oriented, while the transport and application protocols operate end-to-end [59]. Therefore, network layer protocols naturally have a less relative deadline to perform their tasks than the other ones.
- Packets related to control protocols such as ICMP (Internet Control Message Protocol) and ECN (Explicit Congestion Notification) have a very high priority because they provide the communication path for sending the RT data. Actually, the forwarding path for RT packets is not determined until these packets are received by the destination after being sent.

V. IMPLEMENTATION AND ANALYSIS OF RESULTS

A. IMPLEMENTATION ENVIRONMENT

The conditions of the test environment are described in Table 4. It is necessary to mention:
- The Touch Surgery simulator is used to implement and test the proposed solution. It is configured to insert a delay in visual and command data packets between the master and slave sites. For the device (tablet) running the Touch Surgery simulator, a fixed IP is assigned whose traffic is routed to the fog zone, and other local network traffic is routed to the cloud zone.
- The PCAP Remote tool [60] is used to record the traffic of the Touch Surgery simulator, and after applying the changes related to the proposed solution described in Section IV-C, the Tcpreplay tool [61] is used to replay the traffic.
- Two groups of 20 medical students (40 students in total), one general medicine group and the other group comprised of some specialists, have participated voluntarily in this test.
- Some telesurgery exercises are provided in Touch Surgery simulator. Surgeons were initially allowed to select surgeries at three levels: simple, medium, and difficult, and they perform these simulated telesurgeries first with the default network settings (to get acquainted with the simulator) and then with the network changes according to what was described in Section IV. Finally, the effect of delay on surgeons' performance was measured. These people score "0" in their worst telesurgery experience after making changes to the network

TABLE 4. Implementation environment conditions.

Feature name	Value/Models
Cloud and Fog controller server at Master Site	HPE ProLiant DL380 Gen10 Server
Switch	HPE FlexFabric 12900E
Tablet	iPad Pro 2020 12.9
Cloud and Fog Controller at Master Site	OpenDayLight Helium SR1.1
Routing Algorithm in cloud/fog controller	OSPF
Communication protocol	OpenFlow1.3
Link data rate	32 Gb/s
Reserved bandwidth	25% of the bandwidth of each link
Average process time in AI8 and AI9	2-12 microsec.
Weights related to the WFQ algorithm { W_1, W_2, W_3, W_4 }	{1,2,3,4}
Number of iteration for each independent experiment	Ten times

compared to telesurgery without applying the changes. Their best experiences are scored ''5''.

- The quality of surgery was measured by several factors reported by Touch Surgery simulator. To evaluate the quality of the telesurgery, two methods have been used. First, the parameters were displayed by the program after the surgery completion (percentage of correct answer and completion time of surgery). Second, and the parameters that the surgeons themselves were asked to score about the quality of surgery.

- In this study, we transmit a sequence of medical tool images from the master side to the slave part. Each image has 2560×1600 pixels, each pixel being 64 bits. There are 90 screen images per second to transmit, so the required bandwidth is almost 24 Gbps. As shown in Table 4, we consider the bandwidth of the links greater than the required one. So, there will be no problem in transferring high-quality images and other data, including the location used by medical navigation systems and the other required data).

- Every second, each recommender system announces its ρ value to the master device. A recommender system whose condition is true in the inequality (propositions 12) is deactivated as long as it is in this condition.

- The master part is considered at the Information and Communication Technologies (ICT) site of Jahrom University of Medical Sciences. Throughout the test, physicians and all staff were present at the university, using the Internet to perform their daily and routine tasks. Therefore, the traffic of non-telesurgical data has been continuously changing during the test period. Nevertheless, we can be sure that the tested scenario is consistent with the reality of Internet traffic.

-At intervals equal to one-tenth of the smallest relative deadline, the alternative examinations listed in different modeling parts are redone.

- The priority of the non-telesurgery data in the test is as follows:

Priority 1: Diffserv, ICMP, and ECN protocols of non-telesurgery RT

*Priority 2:* RSVP, SCTP, and DCCP protocols of non-telesurgery RT

Priority 3: RTP, RTCP, RDT, RTSP, RTPS, SRTP, RTMP, ZRTP, and SNTP protocols of non-telesurgery RT

Priority 4: NRT packets

- To evaluate RT-TelSurg according to different metrics, two previous research methods closely related or similar to RT-TelSurg, have been selected for comparison. The solutions proposed in [7] and [8], which we implemented in detail for a fair comparison with RT-TelSurg. These papers have been selected for emulation and comparison with RT-TelSurg because, as can be seen in Table 2, they are very similar to our approach in terms of techniques.

B. ANALYSIS OF THE RESULTS

As shown in Fig. 5, and as expected, a lower packet rate in sending hosts leads to a better (i.e., greater) deadline hit ratio due to less congestion in the network. As can be seen, at different data rates, RT-TelSurg transmission performs better than other methods because of the proposed scheduling algorithm and the resource management technique used in critical network situations.
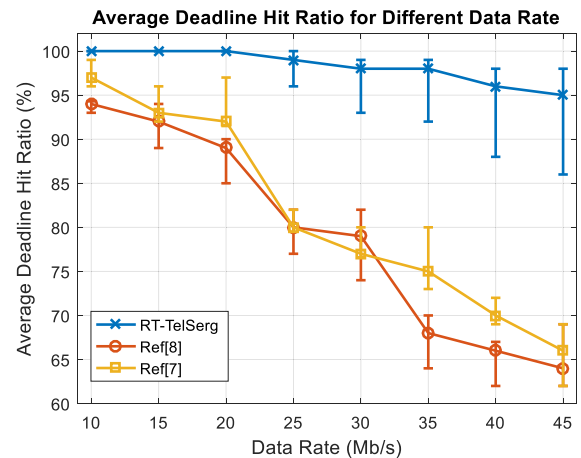
FIGURE 5. Average deadline hit ratio for different data rate.

It can be seen in Fig. 6 that RT-TelSurg can establish a good balance between the number of RT and N-RT packets that are discarded, and this shows the fairness applied to both types of flows. As shown in this graph, the packet drop ratio increases linearly initially but increases exponentially later as the queues approach the saturation point. RT-TelSurg could deliver a higher percentage of packets on time to the recipient, without discarding many N-RT packets. Note that in Fig. 6, the vertical scales for RT and N-RT flows are different, and their drop ratios reach a maximum of 0.16% and 3%, respectively. Hence, the average RT packet drop ratio obtained here is acceptable for telesurgery applications, as can be confirmed by [62], [21], [34], and [44].

As plotted in Fig. 7, the higher packet arrival rate to the switches leads to higher end-to-end delays. Due to the congestion, the waiting time for the packets in the switch and the response time to the inquiries of the switches from the fog/cloud controller increase. The advantage of the
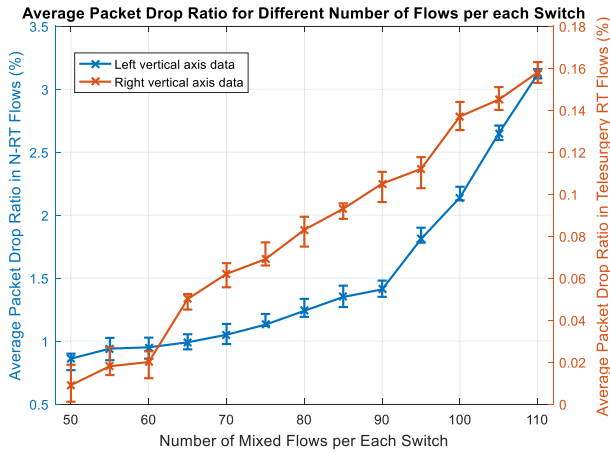
FIGURE 6. Average packet drop ratio for different number of flows per switch.
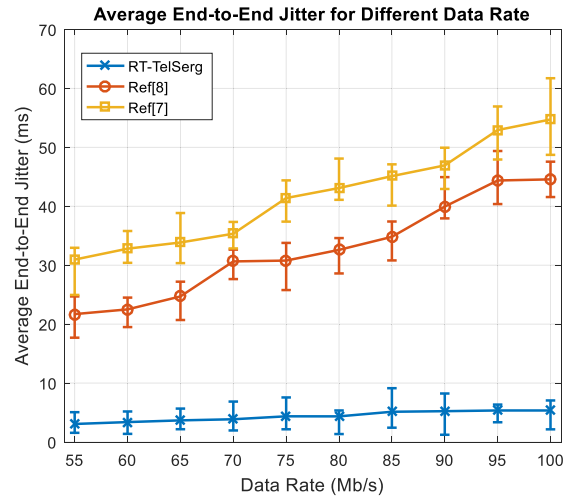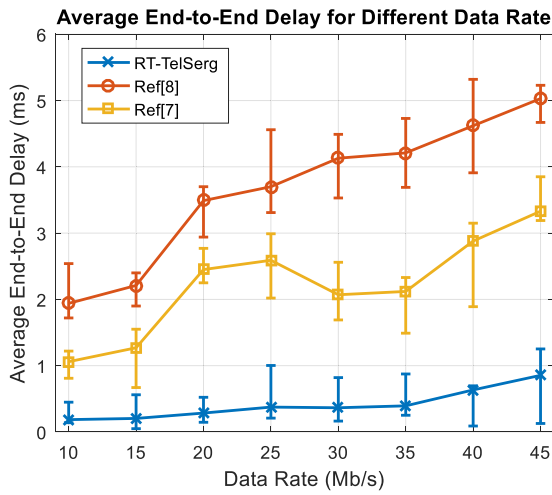

FIGURE 7. Average end-to-end delay for different data rate.

forwarding method used in RT-TelSurg is to manage the network resources according to its requirements. As a result, the congestion of the switch and the waiting time in the queue, and the end-to-end delay are reduced. As this diagram shows, RT-TelSurg gives better results than other methods. It should be noted that the telesurgery end-to-end delay obtained by the proposed method is acceptable since it is less than 1ms.

In Fig. 8, only telesurgery video flows are considered. As can be seen, RT-TelSurg has the least average end-to-end jitter value compared to other methods.

In the following, an examination is considered that evaluates the performance of the RT-TelSurge in the worst-case. In our opinion, the worst-case scenario for packets meeting their deadlines occurs when the links are all saturated. This means that the sending data rate in hosts is so high that all the link bandwidth is used. To get the worst-case scenario, two tools, SolarWinds WAN Killer Network [63] and Ostinato [64], are used simultaneously to generate traffic; both are among the best network traffic generator and simulator stress test gadgets. In the intended experiment, we have evaluated the average deadline hit ratio, based on the bandwidth

FIGURE 8. Average end-to-end jitter for different data rate.

usage percentage. As can be seen in Fig. 9, the RT-TelSurg has acceptable performance, even in the worst-case, and can handle this difficult situation gracefully, due to the use of the reserved bandwidth for telesurgery data in critical network situations occurred following the congestion of the links.
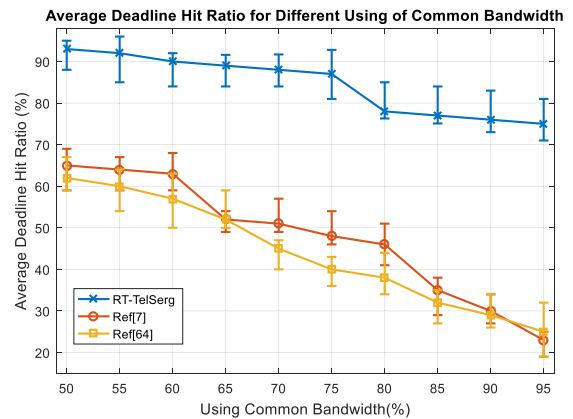

FIGURE 9. Average deadline hit ratio for different using of common bandwidth.

To estimate the impact of deadline hit ratio, end-to-end delay, initial delay, end-to-end jitter, and packet loss ratio on the telesurgery efficiency, each criterion has been chosen in Table 5, according to the desired values mentioned in previous research works, such as [21], [33], [34], [44], [62], [65], and [6].

As shown in Table 5, the most important factors influencing the quality of telesurgery include deadline hit ratio, end-to-end jitter, end-to-end delay, packet loss ratio, and initial delay, respectively. To obtain the evaluation results shown in Table 5, we have considered the effect of each factor alone. This means that, for example, we change the network conditions described in Section IV in such a way that the "Deadline hit ratio" factor fluctuates in the range around its desired value, and in this case, the participants scoring the

TABLE 5. The impact of improving the quality of network communication on the efficiency of surgeons in telesurgery.

Factor	The average score of surgeons given by Touch Surgery Simulator (%)	Surgeons' evaluation score of network quality for telesurgery (0-5)
Deadline hit ratio	44	1
End-to-end delay	81	3
Initial delay	92	4
End-to-end jitter	46	2
Packet loss ratio	84	3

QoE of their telesurgeries. This process is repeated for other factors as well.

VI. CONCLUSION

Using dedicated communication media is very costly. Hence, we should look for means to overcome this issue and meet the real time requirements of telesurgery over the shared Internet. In this study, a method, named RT-TelSurg, has been proposed for RT forwarding of time-sensitive telesurgery data. It has been tailored to such requirements, using cloud and fog technologies and SDN for orchestrating them. In this study, statistical modeling has been presented and used to determine the critical and non-critical states of real time routing and the use of reserved routes, and the utilization of two efficient commercial surgery products named, Sina and Parsiss. Another important constituent of RT-TelSurg is the proposed scheduling algorithm, LE2EDF.

Using a surgical simulator and with the participation of some medical students, RT-TelSurg has been evaluated and it has proven to meet conveniently the real time requirements. It was shown the network factors that mainly affect the efficiency of telesurgery are the deadline hit ratio and end-to-end jitter. For further research, the design of reliable and fault-tolerant telesurgery networks, and distributed controller-based SDN could be considered.

ACKNOWLEDGMENT

REFERENCES

[1] Y. Miao, Y. Jiang, L. Peng, M. S. Hossain, and G. Muhammad, ''Telesurgery robot based on 5G tactile Internet,'' *Mobile Netw. Appl.*, vol. 23, no. 6, pp. 1645–1654, Sep. 2018.

[2] K. Kawamura, Y. Kobayashi, and M. G. Fujie, ''Development of real-time simulation for workload quantization in robotic tele-surgery,'' in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Kunming, China, Dec. 2006, pp. 1420–1425.

[3] R. Gupta, S. Tanwar, S. Tyagi, and N. Kumar, ''Tactile-Internet-based telesurgery system for healthcare 4.0: An architecture, research challenges, and future directions,'' *IEEE Netw.*, vol. 33, no. 6, pp. 22–29, Nov. 2019.

[4] M. Karakus and A. Durresi, ''Quality of service (QoS) in software defined networking (SDN): A survey,'' *J. Netw. Comput. Appl.*, vol. 80, pp. 200–218, Feb. 2017.

[5] K. Tiwari, S. Kumar, and R. K. Tiwari, ''FOG assisted healthcare architecture for pre-operative support to reduce latency,'' *Procedia Comput. Sci.*, vol. 167, pp. 1312–1324, Jan. 2020.

[6] L. Zhou, D. Wu, X. Wei, and J. Chen, ''Cross-modal stream scheduling for eHealth,'' *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 426–437, Feb. 2021.

[7] K. Polachan, B. Turkovic, T. V. Prabhakar, C. Singh, and F. A. Kuipers, ''Dynamic network slicing for the tactile Internet,'' in *Proc. ACM/IEEE 11th Int. Conf. Cyber-Phys. Syst. (ICCPS)*, Sydney, NSW, Australia, Apr. 2020, pp. 129–140.

[8] M. Szczerban, N. Benzaoui, J. Estaran, H. Mardoyan, A. Ouslimani, A. Kasbari, S. Bigo, and Y. Pointurier, ''Real-time control and management plane for edge-cloud deterministic and dynamic networks,'' *J. Opt. Commun. Netw.*, vol. 12, no. 11, pp. 312–323, Nov. 2020.

[9] D. Szabo, A. Gulyas, F. H. P. Fitzek, and D. E. Lucani, ''Towards the tactile Internet: Decreasing communication latency with network coding and software defined networking,'' in *Proc. 21th Eur. Wireless Conf.*, Budapest, Hungary, May 2015, pp. 1–6.

[10] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, ''5G-enabled tactile Internet,'' *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

[11] J. W. S. Liu, *Real-Time Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[12] G. C. Buttazzo, *Hard Real-Time Computing Systems*, 3rd ed. New York, NY, USA: Springer, 2011.

[13] F. Cottet, J. Delacroix, C. Kaiser, and Z. Mammeri, *Scheduling in Real-Time Systems*. Chichester, U.K.: Wiley, 2002.

[14] P. A. Laplante and S. J. Ovaska, *Real-Time Systems Design and Analysis*, 4th ed. Hoboken, NJ, USA: Wiley, 2012.

[15] J. Yan, H. Zhang, Q. Shuai, B. Liu, and X. Guo, ''HiQoS: An SDN-based multipath QoS solution,'' *China Commun.*, vol. 12, no. 5, pp. 123–133, May 2015.

[16] *Touch Surgery—Prepare for Surgery*. Accessed: Nov. 6, 2020. [Online]. Available: https://www.touchsurgery.com/

[17] N. E. James and A. R. Samsudin, ''Effectiveness of touch surgery in complementing technology-supported surgical anatomy teaching,'' in *Proc. ASGBI Surgical Congr. Coping With Complications*, Telford, U.K., vol. 106, Sep. 2019, p. 25.

[18] C. L. N. Tsang, J. Cao, K. Sugand, J. Chiu, and F. C. Pretorius, ''Face, content, construct validity and training effect of touch surgery as a surgical decision-making trainer for novices in open appendicectomy,'' *Int. J. Surg. Protocols*, vol. 22, pp. 19–23, Jan. 2020.

[19] S. Chidambaram, S. Erridge, D. Leff, and S. Purkayastha, ''A randomized controlled trial of skills transfer: From touch surgery to laparoscopic cholecystectomy,'' *J. Surg. Res.*, vol. 234, pp. 217–223, Feb. 2019.

[20] J. Tulipan, A. Miller, A. G. Park, J. T. Labrum, and A. M. Ilyas, ''Touch surgery: Analysis and assessment of validity of a hand surgery simulation 'app','' *HAND*, vol. 14, no. 3, pp. 311–316, May 2019.

[21] D. Van Den Berg, R. Glans, D. De Koning, F. A. Kuipers, J. Lugtenburg, K. Polachan, P. T. Venkata, C. Singh, B. Turkovic, and B. Van Wijk, ''Challenges in haptic communications over the tactile Internet,'' *IEEE Access*, vol. 5, pp. 23502–23518, Oct. 2017.

[22] M. Maier, M. Chowdhury, B. P. Rimal, and D. P. Van, ''The tactile Internet: Vision, recent progress, and open challenges,'' *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 138–145, May 2016.

[23] Z. Lv and W. Xiu, ''Interaction of edge-cloud computing based on SDN and NFV for next generation IoT,'' *IEEE Internet Things J.*, vol. 7, no. 7, pp. 5706–5712, Jul. 2020.

[24] S. Garg, S. Guo, V. Piuri, K.-K. R. Choo, and B. Raman, ''Guest editorial special issue on edge-cloud interplay based on SDN and NFV for next-generation IoT applications,'' *IEEE Internet Things J.*, vol. 7, no. 7, pp. 5690–5694, Jul. 2020.

[25] Z. Lv and W. Xiu, ''Software defined solutions for sensors in 6G/IoE,'' *Comput. Commun.*, vol. 153, pp. 42–47, Mar. 2020.

[26] L. Huo, D. Jiang, Z. Lv, and S. Singh, ''An intelligent optimization-based traffic information acquirement approach to software-defined networking,'' *Comput. Intell.*, vol. 36, no. 1, pp. 151–171, Nov. 2019.

[27] L. Huo, D. Jiang, and Z. Lv, ''A software-defined networks-based measurement method of network traffic for 6G technologies,'' *Trans. Emerg. Telecommun. Technol.*, no. e4172, pp. 1–13, Nov. 2020.

[28] Z. Lv, T. Li, and X. Luo, ''Introduction to the special section on network measurement for software-defined and information-centric networking,'' *Comput. Electr. Eng.*, vol. 70, pp. 1098–1101, Aug. 2018.

[29] W. Queiroz, M. A. M. Capretz, and M. Dantas, ''An approach for SDN traffic monitoring based on big data techniques,'' *J. Netw. Comput. Appl.*, vol. 131, pp. 28–39, Apr. 2019.

[30] P. Xiao, N. Liu, Y. Li, Y. Lu, X.-J. Tang, H.-W. Wang, and M.-X. Li, "A traffic classification method with spectral clustering in SDN," in *Proc. 17th Int. Conf. Parallel Distrib. Comput., Appl. Technol. (PDCAT)*, Guangzhou, China, Dec. 2016, pp. 391–394.

[31] M. M. Raikar, M. S. Meena, M. M. Mulla, N. S. Shetti, and M. Karanandi, "Data traffic classification in software defined networks (SDN) using supervised-learning," *Procedia Comput. Sci.*, vol. 171, pp. 2750–2759, Jan. 2020.

[32] M. Ibrar, L. Wang, G.-M. Muntean, J. Chen, N. Shah, and A. Akbar, "IHSF: An intelligent solution for improved performance of reliable and time-sensitive flows in hybrid SDN-based FC IoT systems," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3130–3142, Mar. 2021.

[33] M. I. Majid, T. Aslam, A. M. Hashmi, D. Subzwari, and B. Siddiqui, "Tactile Internet and the remote surgeon," in *Employing Recent Technologies for Improved Digital Governance*. Hershey, PA, USA: IGI Global, Jan. 2020, pp. 290–316.

[34] G. Minopoulos, G. Kokkonis, K. E. Psannis, and Y. Ishibashi, "A survey on haptic data over 5G networks," *Int. J. Future Gener. Commun. Netw.*, vol. 12, no. 2, pp. 37–54, Jun. 2019.

[35] M. Perez, F. Quiaios, P. Andrivon, D. Husson, M. Dufaut, J. Felblinger, and J. Hubert, "Paradigms and experimental set-up for the determination of the acceptable delay in telesurgery," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Lyon, France, Aug. 2007, pp. 453–456.

[36] B. Rother, M. Kasparick, E. Schweisguth, F. Golatowski, and D. Timmermann, "Automatic configuration of a TSN network for SDC-based medical device networks," in *Proc. 16th IEEE Int. Conf. Factory Commun. Syst. (WFCS)*, Porto, Portugal, Apr. 2020, pp. 1–8.

[37] G. Kokkonis, K. E. Psannis, and M. Roumeliotis, "Network adaptive flow control algorithm for haptic data over the Internet-NAFCAH," in *Proc. Int. Conf. Genet. Evol. Comput.*, Yangon, Myanmar, vol. 388, Sep. 2015, pp. 93–102.

[38] V. Gokhale, J. Nair, and S. Chaudhuri, "Congestion control for network-aware telehaptic communication," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 13, no. 2, pp. 1–26, May 2017.

[39] Z. Zuo, C. Chang, Y. Zhang, R. He, X. Qin, and K. L. Yung, "P4Label: Packet forwarding control mechanism based on p4 for software-defined networking," *J. Ambient Intell. Humanized Comput.*, pp. 1–14, Jan. 2020.

[40] A. A. Ateya, A. Muthanna, I. Gudkova, A. Vybornova, and A. Koucheryavy, "Intelligent core network for tactile Internet system," in *Proc. Int. Conf. Future Netw. Distrib. Syst.*, New York, NY, USA, no. 22, Jul. 2017, pp. 1–6.

[41] A. Basta, W. Kellerer, M. Hoffmann, K. Hoffmann, and E. Schmidt, "A virtual SDN-enabled LTE EPC architecture: A case study for S-/P-gateways functions," in *Proc. IEEE SDN Future Netw. Services*, Trento, Italy, Nov. 2013, pp. 1–7.

[42] M. R. Parsaei, R. Mohammadi, and R. Javidan, "A new adaptive traffic engineering method for telesurgery using ACO algorithm over software defined networks," *Eur. Res. Telemed./La Recherche Europeenne Telemed.*, vol. 6, nos. 3–4, pp. 173–180, Nov. 2017.

[43] M. R. Parsaei, H. R. Boveiri, R. Javidan, and R. Khayami, "Telesurgery QoS improvement over SDN based on a type-2 fuzzy system and enhanced cuckoo optimization algorithm," *Int. J. Commun. Syst.*, vol. 33, no. 11, pp. 1–23, Apr. 2020.

[44] R. Mohammadi and R. Javidan, "On the feasibility of telesurgery over software defined networks," *Int. J. Intell. Robot. Appl.*, vol. 2, no. 3, pp. 339–350, May 2018.

[45] *Home—Sina*. Accessed: Feb. 10, 2021. [Online]. Available: https://sinamed.ir/

[46] A. Mirbagheri, F. Farahmand, S. Sarkar, A. Alamdar, M. Moradi, and E. Afshari, "The Sina robotic telesurgery system," in *Handbook of Robotic and Image-Guided Surgery*. New York, NY, USA: Elsevier, 2020, pp. 107–121.

[47] A. Alamdar, P. Samandi, S. Hanifeh, P. Kheradmand, A. Mirbagheri, F. Farahmand, and S. Sarkar, "Investigation of a hybrid kinematic calibration method for the 'Sina' surgical robot," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5276–5282, Oct. 2020.

[48] *Parseh Intelligent Surgical Systems Co-Home*. Accessed: Feb. 10, 2021. [Online]. Available: http://www.parsiss.com/

[49] J. Hasani-Bidgoli, S. M. S. Hosseini, A. Safavi-Naeini, H. Saberi, M. J. Sadeghi, S. Karimi-Fard, H. Salimi, A. Ahmadian, F. Farahmand, and S. Sarkar, "Experimental evaluation of PARSISS image guided surgery system," *Frontiers Biomed. Technol.*, vol. 1, no. 4, pp. 284–291, Dec. 2014.

[50] A. Nikouei, A. Seddighi, A. S. Seddighi, and A. Nikouei, "The results of image guided surgery using neuronavigation in resection of cerebral gliomas in eloquent cortical areas," *Arch. Phys. Med. Rehabil.*, vol. 97, no. 10, pp. e69–e70, Oct. 2016.

[51] A. S. Alfa, *Queueing Theory for Telecommunications, Discrete Time Modelling of a Single Node System*. New York, NY, USA: Springer, 2010.

[52] C.-H. Ng and S. Boon-Hee, *Queueing Modelling Fundamentals: With Applications in Communication Networks*. Hoboken, NJ, USA: Wiley, Jun. 2008.

[53] G. Bolch, S. Greiner, H. De Meer, and K. S. Trivedi, *Queueing Networks and Markov Chains: Modeling and Performance Evaluation With Computer Science Applications*. Hoboken, NJ, USA: Wiley, Mar. 2006.

[54] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York, NY, USA: Springer-Verlag, 1998.

[55] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*. Hoboken, NJ, USA: Wiley, 2011.

[56] A. O. Allen, *Probability, Statistics, and Queuing Theory With Computer Science Applications*. New York, NY, USA: Academic, Sep. 1990.

[57] A. Demers, S. Keshav, and S. Shenker, "Analysis and simulation of a fair queueing algorithm," in *Proc. Symp. Commun. Archit. Protocols (SIGCOMM)*, Austin, TX, USA, Aug. 1989, vol. 19, no. 4, pp. 1–12.

[58] R. A. Strehlow and S. E. Wright, *Standardizing Terminology for Better Communication: Practice, Applied Theory, and Results*. West Conshohocken, PA, USA: American Society for Testing and Materials, Apr. 1993.

[59] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach Featuring the Internet*. London, U.K.: Pearson, 2005.

[60] *PCAP Remote Tutorial—Andrey Egorov | Dev Blog & Portfolio*. Accessed: Nov. 6, 2020. [Online]. Available: https://egorovandreyrm.com/pcap-remote-tutorial/

[61] *Tcpreplay—Pcap Editing and Replaying Utilities*. Accessed: Nov. 6, 2020. [Online]. Available: https://tcpreplay.appneta.com/

[62] Q. Zhang, J. Liu, and G. Zhao, "Towards 5G enabled tactile robotic telesurgery," Mar. 2018, *arXiv:1803.03586*. [Online]. Available: http://arxiv.org/abs/1803.03586

[63] *Network Traffic Generator & Stress Test—WAN Killer | SolarWinds*. Accessed: Feb. 23, 2021. [Online]. Available: https://www.solarwinds.com/engineers-toolset/use-cases/traffic-generator-wan-killer

[64] *Ostinato Traffic Generator for Network Engineers*. Accessed: Feb. 23, 2021. [Online]. Available: https://ostinato.org/

[65] J. M. Llopis, J. Pieczerak, T. Janaszka, J. M. Llopis, J. Pieczerak, and T. Janaszka, "Minimizing latency of critical traffic through SDN," in *Proc. IEEE Int. Conf. Netw., Archit. Storage*, Long Beach, CA, USA, Aug. 2016, pp. 1–6.

SHAHRZAD SEDAGHAT received the B.Sc. and M.Sc. degrees in computer and information technology engineering from Yazd University, Iran, in 2008 and 2010, respectively. She is currently pursuing the Ph.D. degree in computer engineering with the Sharif University of Technology, Iran. In 2011, she joined the Department of Computer and Information Technology Engineering, Jahrom University. Her research interests include computer network and security, quality of service, and reliability modeling.

AMIR HOSSEIN JAHANGIR received the Ph.D. degree in industrial informatics from the Institut National des Sciences Appliquées, Toulouse, France, in 1989. Since then he has been with the Department of Computer Engineering, Sharif University of Technology, Tehran, Iran. During his career, he has served as the Head of the Department and the Head of Computing Center. He is currently an Associate Professor and the Director of Network Evaluation and Test Laboratory, an accredited recognized laboratory in the field of network equipment test. His research interests include comprise network equipment test and evaluation methodology, network security, and high-performance computer architecture.

• • •