IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 13, 2020, accepted October 9, 2020, date of publication October 15, 2020, date of current version October 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3031306

Using 5G Network Slicing and Non-Orthogonal Multiple Access to Transmit Medical Data in a Mobile Hospital System

PARFAIT IFEDE TEBE[1,2], (Member, IEEE), KWADWO NTIAMOAH-SARPONG[2], (Graduate Student Member, IEEE), WENHONG TIAN[1], (Senior Member, IEEE), JIAN LI[2], (Member, IEEE), YONGJUN HUANG[2], (Member, IEEE), AND GUANGJUN WEN[2], (Senior Member, IEEE)

[1]School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Center for RFIC and System Technology, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Jian Li (lj001@uestc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Project 61701082, Project 61701116, Project 61601093, Project 61971113, and Project 61901095; in part by the National Key Research and Development Program under Project 2018YFB1802102 and Project 2018AAA0103203; in part by the Guangdong Provincial Research and Development Plan in Key Areas under Project 2019B010141001 and Project 2019B010142001; in part by the Sichuan Provincial Science and Technology Planning Program under Project 2019YFG0418, Project 2019YFG0120, and Project 2020YFG0039; in part by the Ministry of Education—China Mobile Fund Program under Project MCM20180104; in part by the Yibin Science and Technology Program—Key Projects under Project 2018ZSF001 and Project 2019GY001; and in part by the Fundamental Research Funds for the Central Universities under Project ZYGX2019Z022.

ABSTRACT In this work, we propose a novel approach combining 5G network slicing and non-orthogonal multiple access (NOMA) to transmit medical data in a mobile hospital system. We consider both the uplink and downlink of a 5G cellular network with an ambulance bus located at a remote site for data transmission in the uplink scenario and a hospital unit as the receiving site in the downlink scenario. We propose and model a NOMA slicing system where the medical data are categorized and assigned to two different slices based on 5G services. That is, 4K video from patients is assigned to an enhanced mobile broadband (eMBB) NOMA slice in both uplink and downlink, and all other medical data are assigned to an ultra-reliable and low latency communication (uRLLC) NOMA slice also in both uplink and downlink. Based on the system model and principles of NOMA, we formulate and use a joint power allocation optimization technique under users' minimum rate requirements and transmission power constraints, and successive interference cancellation (SIC) to maximize the medical data throughput as well as the system sum-throughput in each slice in both uplink and downlink. Our results show that, with the optimal power allocation technique, high throughput can be achieved for the 4K video and other medical data in the eMBB NOMA slice and uRLLC NOMA slice, respectively, but other users transmitting and receiving ordinary data in the slices will see their throughput decrease. Hence, in the interest of fairness for all users, we use truncated channel inversion power allocation in the downlink to prevent the decrease of the throughput of those users regardless of their channel conditions.

INDEX TERMS 5G cellular network, network slicing, non-orthogonal multiple access (NOMA), power allocation, throughput maximization.

I. INTRODUCTION

With the rapid development in mobile communication, big data, artificial intelligence, and other technologies in recent years, mobile hospitals can carry more advanced medical equipment and apply more intelligent means of diagnosis and treatment, such as online ultra-high-definition video (UHDV) expert diagnosis and treatment, critical emergency nursing, artificial intelligence imaging diagnosis and clinical examination and analysis of big data [1]. However, traditional mobile hospitals relying on the fourth-generation (4G) communication system can no longer support such advanced medical

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

methods, which strictly restricts the development of mobile hospitals. First, the transmission rate of the traditional 4G communication system is rather low, not supporting UHDV (4K and 8K) real-time transmission required for online therapy. Second, 4G technology does not support network slicing and cannot solve the business application requirements of mobile hospitals for medium-high speed, low delay, high reliability, and large capacity connection [1]–[3]. The development and gradual maturity of 5G mobile communication technology offers an effective technical approach for high-quality medical data transmission in mobile hospitals. Indeed, with its broadband, high speed, low delay, self-adaptation, and large capacity access capability [4], 5G has the potentials to serve as an effective communication technology approach for the development of mobile hospitals. It is, therefore, necessary and urgent to develop industry terminals and systems integrating 5G communication and mobile medical technology.

Unlike 4G, 5G systems are expected to be more versatile with a high variety of requirements on network functionalities and performances and are then intended to support a wider range of applications and business models [4]–[6]. 5G network services have been classified into three categories by the International Telecommunication Union (ITU): Enhanced Mobile Broadband (eMBB), Ultra-reliable and Low-latency Communications (uRLLC), and Massive Machine Type Communications (mMTC). eMBB focuses on high bandwidth requirement services; uRLLC focuses on latency-sensitive services, and mMTC focuses on services with high connection density requirements [5], [7]. However, those services cannot always be achieved through a common network setting. For instance, network optimization for low latency and high reliability could come at the expense of reduced spectral efficiency. In this context, network slicing has become a key enabler to enhance 5G systems by supporting their diversified services.

Network slicing is a network architecture designed for deployment and operation of multiple virtualized and logical networks over a common network infrastructure. Each logical network is called slice and can be tailored to fulfill the needs of specific applications [8]–[10]. Combined with 5G, network slicing can therefore allow simultaneous support of the eMBB, uRLLC, and mMTC services. Network slicing has recently received a lot of attention in research. In [8], a comprehensive survey of 5G network slicing with its driving forces is provided, and some open issues and challenges toward the practicability of the concept in 5G networks are also discussed. Another comprehensive survey on network slicing and its architecture is provided by [9] with further research directions. In [11], a network slicing implementation in the 5G communication system is detailed, where the deployment of vertical and horizontal slicing is done over the air interface, the radio access network (RAN), and the core network (CN). In [12], network slicing is analyzed in multi-cell RAN to support resource splitting among various slices, and the authors have proposed and compared four types of slicing approaches. RAN requirements for network slicing implementation are thoroughly covered by [13] where the impacts of network slicing on various aspects of the design and functions of 5G RAN have been explained. The authors in [14] discussed how network slicing can be deployed in heterogeneous cloud RAN (H-CRAN) to improve the system throughput. In [15], a comprehensive discussion on network slicing using both software defined network (SDN) and network function virtualization (NFV) is provided. Some challenges and future research directions in 5G network slicing have also been provided in the work. The authors in [16] presented an overview of the various cases and requirements of network slicing. Details of slicing solutions for the RAN, CN, and transport network (TN) of the 5G system have been provided, and some research directions have also been listed in their work. In [17], a communication-theoretic model involving non-orthogonal sharing of RAN resources in uplink network slicing is studied with the three types of heterogeneous services such as eMBB, uRLLC, and mMTC, and the authors refer the approach to as heterogeneous non-orthogonal multiple access (H-NOMA). They compared the results to the case where orthogonal sharing of the RAN resources is done and called H-OMA. Similar to [17], the authors in [18] also considered both H-NOMA and H-OMA network slicing. However, they considered only eMBB and uRLLC heterogeneous services and used max-matching diversity (MMD) algorithm to improve the achievable rate of both services. Authors in [19] defined network slicing in RAN, TN, and CN domains as an end-to-end network slicing system, and proposed the system architecture and components to demonstrate its capability with hardware and software. A novel RAN slicing architecture with slices providing services by choosing an appropriate functional split option and its placement is provided in [20]. SDN and NFV are used in the work to achieve large scalability and high flexibility. In [21], the optimal utility in terms of high throughput and low delay 5G slices is studied, and to maximize the utility, the authors have proposed a joint virtual resource optimization algorithm. In both slices, their results show a tradeoff between delay and throughput with satisfied QoS requirements.

Based on the aforementioned restrictions in the development of mobile hospital systems, and the potentials of network slicing, this article proposes a 5G network slicing approach to transmit medical data from a remote place to a hospital unit. Moreover, for efficient utilization of channel resources to achieve high data throughput, non-orthogonal multiple access (NOMA) is used within each slice of our designed system. The main contributions of the paper are summarized as follows:

- We considered a 5G cellular network to transmit data in our mobile hospital system. Both uplink and downlink are considered, with an ambulance bus located at the remote site for data transmission in the uplink scenario, and a hospital unit as the receiving site in the downlink scenario. We proposed and modeled a network slicing system where the medical data to be transmitted are

categorized and assigned to two different slices based on 5G services. That is, 4K video from patients intended to be transmitted from the ambulance to the hospital is assigned to an eMBB NOMA slice, and all other medical data are assigned to an uRLLC NOMA slice.

- Based on the principles of NOMA, we formulated and used a joint power allocation optimization technique to maximize the medical data throughput as well as the system sum-throughput in each slice in both uplink and downlink. The optimization problem is formulated under users' minimum rate requirements and transmission power constraints, and successive interference cancellation (SIC). We showed that the optimal power allocation can help to achieve high throughput for the 4K video and other medical data in the eMBB NOMA slice and uRLLC NOMA slice, respectively; but this to the detriment of other ordinary users in the slices who will see their data rate decrease.
- Finally, in the interest of fairness for all users, we used truncated channel inversion power allocation in the downlink to prevent the decrease of the ordinary users' throughputs.

The rest of the paper is organized as follows. Section II provides an overview of the fundamentals of network slicing and NOMA technologies. Our proposed system model is described in section III. Section IV provides the system throughputs as the performance metrics with the optimal power allocations used. The system performance metrics are evaluated numerically in section V, and section VI concludes the paper.

II. FUNDAMENTALS OF NETWORK SLICING AND NOMA

In this section, we provide an overview of 5G network slicing and non-orthogonal multiple access (NOMA) technologies.

A. NETWORK SLICING

As a key foundation to support diversified 5G services, network slicing is a network architecture designed for deployment and operation of multiple virtualized and logical networks over a common network infrastructure [8]–[10]. 5G network slicing technology divides a physical network into a plurality of virtual logical networks (slices) through virtualization, and each virtual network corresponds to a different application scenario. According to different business requirements, different network resources are provided, and different application scenarios are connected to the network in the most appropriate way, to improve the utilization rate of network resources. The three business scenarios defined by 5G are enhanced mobile broadband (eMBB) characterized by a large amount of bandwidth to support high data rate such as ultra-high definition video stream; ultra-reliable and low-latency communications (uRLLC) characterized by high reliability and ultra-low latency services; massive machine-type communications (mMTC) characterized by services with high connection density requirements provided by a massive number of Internet of Things (IoT) devices [16]–[20].

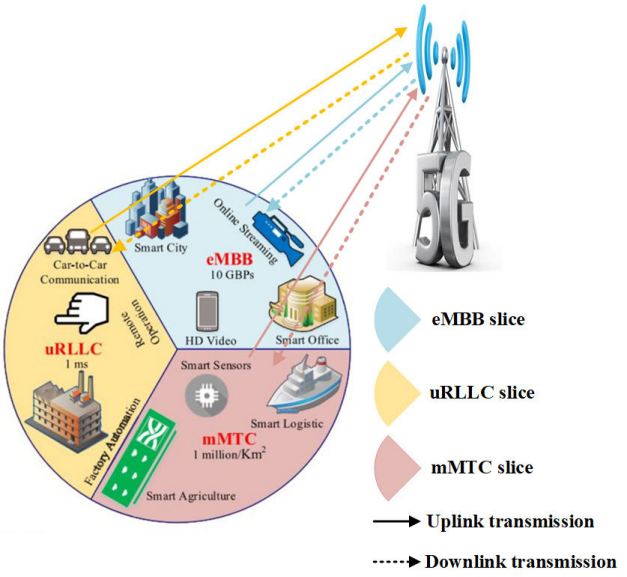

FIGURE 1. Illustration of 5G network slicing.

Network slicing then allows operators to partition the networks in a structured, elastic scalable, and automated manner so that each slice is dedicated to one of the three categories of 5G services, and it is then customized and dimensioned to best serve the needs of specific applications. An illustration of 5G network slicing with the three categories of services is depicted in Fig. 1.

With network slicing 5G mobile networks can be designed to efficiently embrace a plethora of services with different service level requirements (SLR). The realization of this service-oriented view of the network leverages the concepts of software-defined networking (SDN) and network function virtualization (NFV) that allow the implementation of flexible and scalable network slices on top of a common network infrastructure [10]. A 5G network slice spans across multiple parts of the 5G network consisting of 5G terminals, 5G RAN, and 5G CN [10], [22]. The 5G network slicing can also be deployed across multiple operators where each mobile network operator providing mobile services has its own public land mobile network (PLMN) [6].

B. NON-ORTHOGONAL MULTIPLE ACCESS (NOMA)

Non-orthogonal multiple access (NOMA) is a technique whereby multiple users are multiplexed into the transmission power domain and then non-orthogonally scheduled for transmission on the same spectrum resources. The technique is based on channel gain differences to allow users to share the same time and frequency resources to enhance their connectivity, their throughput, and the spectral efficiency while some degree of multiple access interference is allowed at receivers [23]–[26]. NOMA has, therefore, emerged as another solution for 5G networks, and has been proposed as a candidate to meet the requirements for the network spectrum efficiency. NOMA schemes are generally classified into power-domain multiplexing and code-domain

multiplexing. In the power-domain multiplexing, different power coefficients are allocated to different users based on their channel conditions, for a high system performance achievement. More specifically, the users' information signals are superimposed at the transmitter side and successive interference cancellation (SIC) is applied at the receiver side to decode the signals one by one until the desired user' signal is obtained. In the code-domain multiplexing, different codes are allocated to different users who are multiplexed over the same time-frequency resources [24]–[26]. Clustering is an important method in NOMA which consists of pairing users with diverse channel responses together so that they can be mapped onto orthogonal frequency resources to avoid intercluster interference. It has the advantage of maximizing users' throughput gain and capacity and simplifying the SIC at the receiving end [24]. NOMA technique has different modes of operation in uplink and downlink.

1) UPLINK NOMA

In an uplink NOMA cluster, signals are transmitted to a single base station (BS) on the same channel. The transmission is non-orthogonal, and is done independently by each user with its maximum power or controlled transmit power. SIC and signal decoding are performed at the BS to distinguish the received signals channel gains. The BS decodes the signal for the user with the highest channel gain first since that signal is likely the strongest among the received ones. Then the BS decodes the signal for the next highest channel gain user, and so forth [23], [24], [27].

An example of an uplink NOMA cluster with m users is depicted in Fig. 2 [23],[27]. As shown by the figure, $h_1, h_2, \ldots, h_{m-1}$ and h_m are the channel gains of $UE_1, UE_2, \ldots, UE_{m-1}$ and UE_m, respectively. We assume that $h_1 > h_2 > \ldots > h_{m-1} > h_m$. Also, $x_1, x_2, \ldots, x_{m-1}$ and x_m are the desired messages of $UE_1, UE_2, \ldots, UE_{m-1}$ and UE_m, respectively, while w_0 is the additive white Gaussian noise at the BS. Based on the principle of operation described above, the signal of user UE_1 will be decoded first, then the signal of UE_2, and so on. Thus, the achievable data rate of UE_1 will depend on the interferences from $UE_2, \ldots, UE_{m-1}$ and UE_m. The achievable data rate of UE_2 will depend on the interferences from$UE_3, \ldots, UE_{m-1}$ and UE_m, and so on.

2) DOWNLINK NOMA

In the downlink, the SIC is assumed to be performed successively by each user until the user' signal is recovered. Users are allocated power coefficients in an inversely proportional manner according to their channel conditions. That is, low power is used to transmit signals to users with high channel gain, and high power is used to transmit signals to users with low channel gain. Hence, the user with the highest channel gain will perform SIC to suppresses all interfering signals, whereas the user with the lowest channel gain cannot cancel any interference, but considers the signals from other users as noise, and recovers its signal without performing any SIC [23], [24], [27].

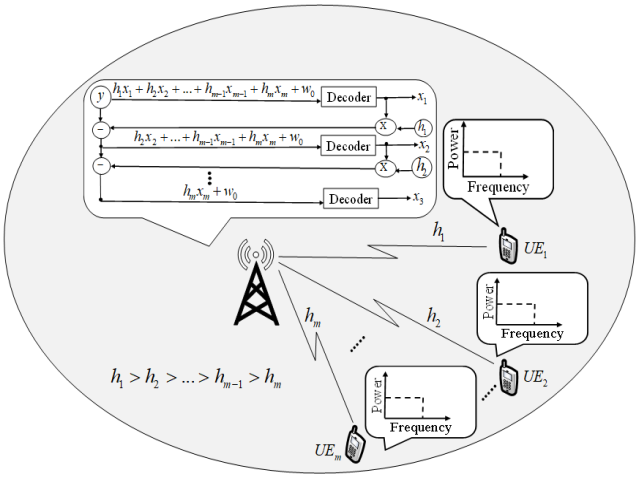

FIGURE 2. Illustration of an uplink NOMA cluster with m users.

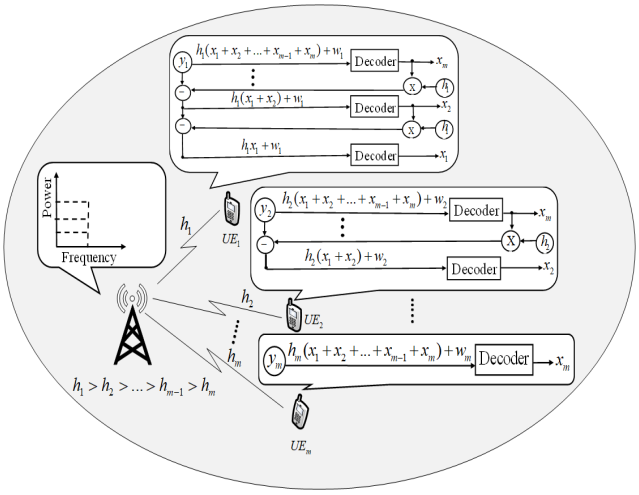

FIGURE 3. Illustration of a downlink NOMA cluster with m users.

An example of a downlink NOMA cluster with m users is depicted in Fig. 3 [23]. Similar to the uplink, $h_1, h_2, \ldots, h_{m-1}$ and h_m are the channel gains of $UE_1, UE_2, \ldots, UE_{m-1}$ and UE_m, respectively. We also assume that $h_1 > h_2 > \ldots > h_{m-1} > h_m$; $w_1, w_2, \ldots, w_{m-1}$ and $w_m$are the respective additive white Gaussian noise. Based on the principle of operation in the downlink described above, SIC can be performed by UE_1 to cancel interference from $UE_2, \ldots, UE_{m-1}$ and UE_m. UE_2 can only cancel interference from $UE_3, \ldots, UE_{m-1}$ and UE_m, and so on. UE_m will experience interference from $UE_1, UE_2, \ldots,$ and UE_{m-1} but cannot cancel any of them [23], [27].

III. PROPOSED SYSTEM MODEL

Our mobile hospital system consists of an ambulance bus transmitting patients' medical data from a remote site to a hospital unit, using a 5G network. The whole system is depicted in Fig.4.

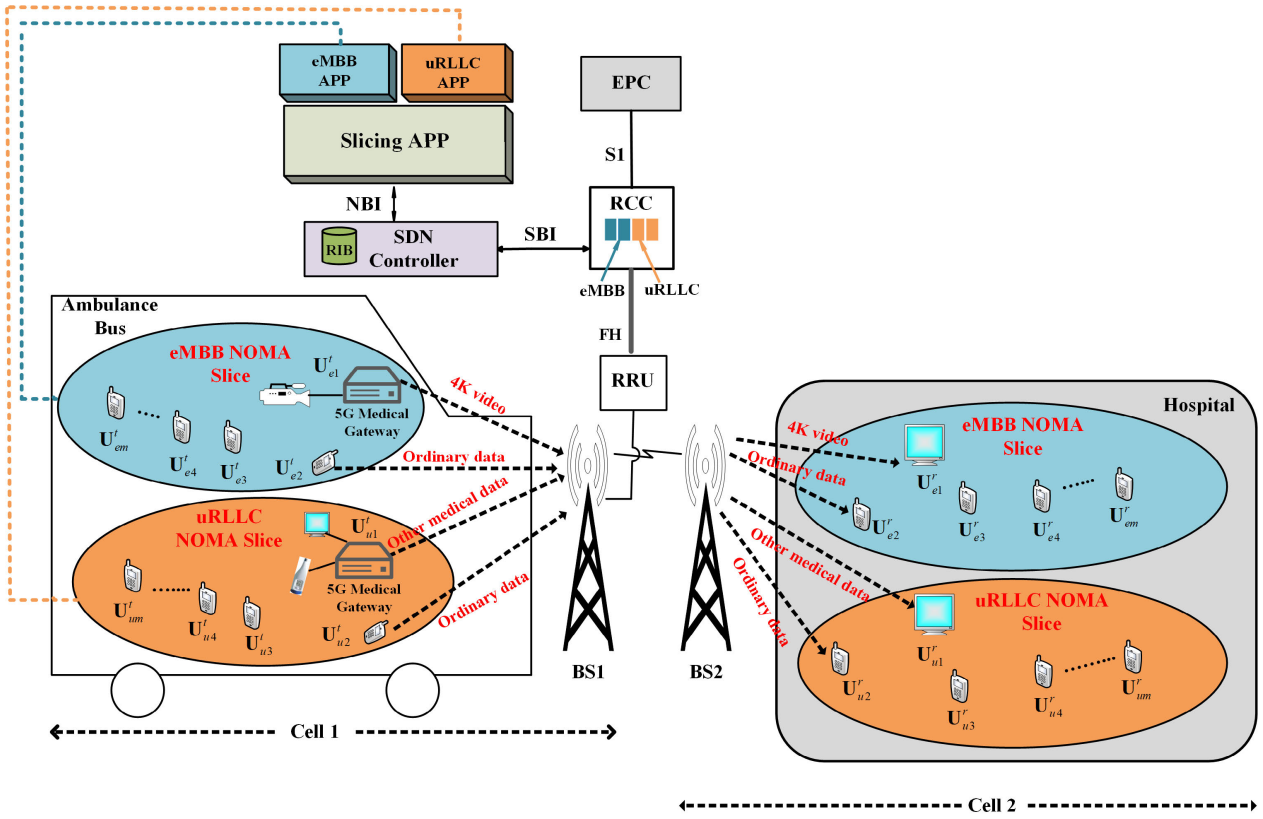

FIGURE 4. Schematic diagram of the proposed mobile hospital system.

We consider a 5G cellular network with two cells for simplicity (Cell 1 and Cell 2). Cell 1 and Cell 2 are served by BS1 and BS2, respectively. Two cells are considered because of the eventual long distance between the remote site and the hospital. The medical data include patients' recorded videos in 4K format and other data such as body temperature, blood pressure, blood oxygen saturation, electrocardiogram (ECG), etc. We assume that the ambulance bus does not move and is within Cell 1 which serves for the uplink scenario to transmit the medical data to BS1. BS1 is responsible for forwarding the data to the hospital via BS2 of Cell 2 containing the hospital unit. The data transmission from BS2 to the hospital is then considered as the downlink scenario of our system. As shown by the figure, we apply network slicing at the radio access network (RAN) in each cell for the data transmission. Additionally, for better network resources utilization and higher throughput, non-orthogonal multiple access (NOMA) is also applied where each slice is modeled as a NOMA cluster. The proposed slicing technique is described in the following subsection.

A. PROPOSED RAN SLICING TECHNIQUE

Our proposed technique involves two types of slices sharing the same cloud RAN within each cell of the network: enhanced mobile broadband (eMBB) slice dedicated for the 4K video transmission in the uplink, and its reception in the downlink, and ultra-reliable and low latency communications (uRLLC) slice dedicated for the other medical data transmission in the uplink and their reception in the downlink. The main objective is to efficiently share the bandwidth resources among the slice segments to achieve higher transmission speed and lower latency for the medical data. Each slice is then designed as a NOMA cluster where each user can use the whole frequency resource blocks allocated to its slice, according to NOMA principles of operation explained in Section II. We, therefore, jointly apply the concepts of network slicing and NOMA to achieve our objective, and then use the terms eMBB NOMA slice and uRLLC NOMA slice throughout the work. However, to conform with the slices isolation principle [8], [9], the frequency resource blocks of the eMBB NOMA slice are orthogonal to those of the uRLLC NOMA slice in each cell. That is, a user from a NOMA slice cannot use frequency resource blocks allocated to another slice.

As shown by the figure, in the uplink scenario, a 4K camera is used in the eMBB NOMA slice. Such a camera records patients' health conditions from the remote place of the mobile hospital system. Via a 5G medical gateway, the recorded 4K video is transmitted to the hospital unit through the 5G C-RAN of both base stations. Similarly, some medical equipment are used in the uRLLC NOMA slice to measure other medical data of the patients. The measured data are also transmitted to the hospital unit via another 5G medical gateway connected to them. In both cases, the data are periodically collected by the 5G medical gateways to

establish a connection with the C-RAN for data delivery to the cloud. For the flexibility of the radio resources allocation to the different slices, a software defined networking (SDN) is used in the C-RAN network [28]. The SDN consists of a slicing application "Slicing APP" and a logical centralized controller. The controller is in charge of the slicing process in real time. The slicing application is in charge of the slicing process by interacting with the controller via the northbound interface (NBI). It consists of two different modules. The first module refers to as "eMBB APP" and stores the eMBB related traffic information including 4K video streaming from patients. The second module refers to "uRLLC APP" and stores the uRLLC related traffic information including other medical data. The traffic profile of each data category can be obtained by a database called RAN information base (RIB) which is within the SDN controller. The number of resources to be allocated to each slice is identified by the "slicing APP" via each module. This is done by taking into account each slice's traffic profile. The base station of the cell (BS1) is linked to the remote radio unit (RRU), which is also linked to the radio cloud center (RCC) via a fronthaul (FH). A south-bound interface (SBI) is used to connect the RCC to the SDN controller. The RCC is connected to the evolved packet core (EPC) of the network via a switch S1. In the downlink, BS2 forwards the data to the UEs in their respective slices. That is, the 4K video data is forwarded to the appropriate medical equipment belonging to the eMBB NOMA slice at the hospital unit, and the other medical data are also forwarded to the appropriate medical equipment belonging to the uRLLC NOMA slice.

B. NETWORK MODEL AND ASSUMPTIONS

A macro base station is considered in each cell with M UEs in each NOMA slice. We denote by B each cell's available bandwidth which is divided into frequency resource blocks. The bandwidth of each resource block is denoted by B_r and the total number of frequency resource blocks in each cell is then given by $\Omega = B/B_r$. Moreover, we represent the number of resource blocks allocated to each NOMA slice by ω, where $1 \leq \omega \leq \Omega$.

For simplicity of analysis, we will consider only two UEs in each slice in both uplink and downlink in the work. More specifically, in the uplink, we will consider only the 5G medical gateway transmitting the 4K video and one ordinary user in the eMBB NOMA slice. That is, we consider U^t_{e1} and U^t_{e2}. In the uRLLC NOMA slice, we will consider the 5G medical gateway transmitting the other medical data (U^t_{u1}) and one ordinary user (U^t_{u2}). Similarly, in the downlink, we will consider only the medical equipment receiving the 4K video (U^r_{e1}) and one ordinary user (U^r_{e2}) in the eMBB NOMA slice; the medical equipment receiving the other medical data (U^r_{u1}) and one ordinary user (U^r_{u2}) in the uRLLC NOMA slice. We, therefore, assume $M = 2$ in each slice and in the rest of the work we will adopt $K = M = 2$. Moreover, we assume that BS1 and BS2 have the same characteristics; B, B_r, Ω and ω have the same values in both uplink and downlink scenarios.

We denote P_t the maximum transmission power budget of each user; P_{BS} the maximum transmission power budget of each base station and P_{DS} the maximum transmission power budget per downlink NOMA slice. We represent by η_k the normalized channel gain between the k^{th} user equipment and the base station. In both uplink and downlink, we define a variable $\beta_{k,l}$ as follows:

$$\beta_{k,l} = \begin{cases} 1, & \text{if a user } k \text{ belongs to NOMA slice } l \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $l = 1, 2$ and $k = 1, 2, 3, 4$.

IV. ACHIEVABLE THROUGHPUT WITH OPTIMAL POWER ALLOCATIONS

In this section, we provide the achievable throughput for each user equipment in both uplink and downlink, based on the various considerations and assumptions presented in the previous sections. We also provide the optimal power allocations used to maximize the users' throughput and the system sum-throughput in both uplink and downlink.

A. UPLINK SCENARIO

1) ACHIEVABLE THROUGHPUT

As stated earlier, we consider a total of four UEs from the two slices to transmit data to BS1 in our uplink scenario. Those users include U^t_{e1} and U^t_{e2} in the eMBB NOMA slice; U^t_{u1} and U^t_{u2} in the uRLLC NOMA slice. The achievable throughput for any user equipment $UE_k, \forall k = 1, 2$ from any of the two NOMA slices can be expressed as [23]

$$R^t_k = \omega B_r \log_2 \left(1 + \frac{P_k \eta_k}{\sum_{l=k+1}^{2} P_l \eta_l + \omega} \right) \tag{2}$$

where $\sum_{l=k+1}^{2} P_l \eta_l$ is the inter-user interference at the k^{th} user level; P_k is the transmission power of UE_k with $P_k \leq P_t$; $\eta_k = h_k / N_0 B_r$ with h_k being the user's channel gain and N_0 the additive noise.

2) OPTIMAL POWER ALLOCATION FOR USERS' THROUGHPUT MAXIMIZATION

Based on the principle of uplink NOMA described in section II, to achieve high video and other medical data throughputs in our uplink scenario, we assume that, in the eMBB NOMA slice, the user equipment transmitting 4K video (U^t_{e1}) has higher channel gain than the other ordinary user equipment (U^t_{e2}). That is, $h_{e1} > h_{e2}$. Similarly, in the uRLLC NOMA slice, the user equipment transmitting other medical data (U^t_{u1}) is assumed to have higher channel gain than the other ordinary user equipment (U^t_{u2}). That is, $h_{u1} > h_{u2}$. In such cases, U^t_{e1} and U^t_{u1} will not interfere with U^t_{e2}

and U_{u2}^t, respectively, because the interference will be canceled by SIC. Therefore, U_{e1}^t and U_{u1}^t can transmit with their maximum power for higher 4K video throughput and other medical data throughput in their respective NOMA slice.

Let's denote P_{e1} and P_{e2} the transmission power of users U_{e1}^t and U_{e2}^t, respectively; P_{u1} and P_{u2} the transmission power of users U_{u1}^t and U_{u2}^t, respectively. For efficient SIC at BS1, the power allocations must satisfy the following requirements [23].

$$P_{e1}\eta_{e1} - P_{e2}\eta_{e2} \geq P_{tol} \tag{3}$$

$$P_{u1}\eta_{u1} - P_{u2}\eta_{u2} \geq P_{tol} \tag{4}$$

where P_{tol} is the minimum power difference required to distinguish between the signal to be decoded and the remaining non-decoded signal; η_{e1} and η_{e2} are the normalized channel gains between user U_{e1}^t and BS1 and between user U_{e2}^t and BS1, respectively; η_{u1} and η_{u2} are the normalized channel gains between user U_{u1}^t and BS1 and between user U_{u2}^t and BS1, respectively. Equation (3) represents the necessary condition to efficiently decode the signal of U_{e1}^t prior the decoding of the signal of U_{e2}^t in the eMBB NOMA slice. Similarly, (4) represents the necessary condition to efficiently decode the signal of U_{u1}^t prior the decoding of the signal of U_{u2}^t in the uRLLC NOMA slice. It infers from the above explanations that optimal power allocation is necessary to maximize users' throughput in our work. Because the power constraints are necessary for efficient SIC. We, therefore apply optimal power allocation to maximize the achievable throughputs of the 4K video and other medical data in the eMBB NOMA slice and uRLLC NOMA slice, respectively.

Let's denote $R_{e1}^{t\,\min}$ and $R_{u1}^{t\,\min}$ the minimum rate requirements for user U_{e1}^t in the eMBB NOMA slice and user U_{u1}^t in the uRLLC NOMA slice, respectively, with $R_{e1}^{t\,\min} > 0$ and $R_{u1}^{t\,\min} > 0$.

For the eMBB NOMA slice, we can express the optimal power allocation problem as:

$$\begin{aligned}
&\max_{P}\ \omega B_r\left[\log_2\left(1+\frac{P_{e1}\eta_{e1}}{P_{e2}\eta_{e2}+\omega}\right)\right.\\
&\qquad\left.+\log_2\left(1+\frac{P_{e2}\eta_{e2}}{P_{e1}\eta_{e1}+\omega}\right)\right]\\
&\text{Subject to: } C_1:\ P_{e1}\leq P_t\\
&\qquad C_2:\ \omega B_r\log_2\left(1+\frac{P_{e1}\eta_{e1}}{P_{e2}\eta_{e2}+\omega}\right)\geq R_{e1}^{t\,\min}\\
&\qquad C_3:\ P_{e1}\eta_{e1}-P_{e2}\eta_{e2}\geq P_{tol}
\end{aligned} \tag{5}$$

For the uRLLC NOMA slice, we can express the optimal power allocation problem as:

$$\begin{aligned}
&\max_{P}\ \omega B_r\left[\log_2\left(1+\frac{P_{u1}\eta_{u1}}{P_{u2}\eta_{u2}+\omega}\right)\right.\\
&\qquad\left.+\log_2\left(1+\frac{P_{u2}\eta_{u2}}{P_{u1}\eta_{u1}+\omega}\right)\right]\\
&\text{Subject to: } C_1':\ P_{u1}\leq P_t\\
&\qquad C_2':\ \omega B_r\log_2\left(1+\frac{P_{u1}\eta_{u1}}{P_{u2}\eta_{u2}+\omega}\right)\geq R_{u1}^{t\,\min}\\
&\qquad C_3':\ P_{u1}\eta_{u1}-P_{u2}\eta_{u2}\geq P_{tol}
\end{aligned} \tag{6}$$

Note that the optimization problems (5) and (6) are convex under the constraints $C_1 - C_3$ and $C_1' - C_3'$, respectively. Moreover, even though the main objective is to maximize the achievable throughput for U_{e1}^t in the eMBB NOMA slice and the achievable throughput for U_{u1}^t in the uRLLC NOMA slice, the main function of each problem involves the sum rate of all the users of the slice. This is because optimal power allocation is applied to be available for the whole of each slice. Constraints C_1 and C_1' are users U_{e1}^t and U_{u1}^t transmit power constraints, respectively. C_2 and C_2' are the minimum rate requirement constraints for users U_{e1}^t and U_{u1}^t, respectively. C_3 and C_3' are the SIC constraints for each NOMA slice, respectively.

To solve the optimization problems, we will use the Lagrange function. We will solve only (5) since it's similar to (6), and the solutions will be applied to both problems. The Lagrange function of problem (5) can be expressed as

$$\begin{aligned}
&L(P,\lambda,\alpha,\mu)\\
&=\omega B_r\left[\log_2\left(1+\frac{P_{e1}\eta_{e1}}{P_{e2}\eta_{e2}+\omega}\right)+\log_2\left(1+\frac{P_{e2}\eta_{e2}}{P_{e1}\eta_{e1}+\omega}\right)\right]\\
&\quad+\lambda_1\left(P_t-P_{e1}\right)+\lambda_2\left(P_t-P_{e2}\right)+\alpha_1\\
&\quad\times\left(P_{e1}\eta_{e1}-\phi_1 P_{e2}\eta_{e2}-\phi_1\omega\right)\\
&\quad+\alpha_2\left(P_{e2}\eta_{e2}-\phi_2 P_{e1}\eta_{e1}-\phi_2\omega\right)\\
&\quad+\mu_1\left(P_{e1}\eta_{e1}-P_{e2}\eta_{e2}-P_{tol}\right)\\
&\quad+\mu_2\left(P_{e1}\eta_{e1}-P_{e2}\eta_{e2}-P_{tol}\right)
\end{aligned} \tag{7}$$

where $\phi_k = \left(2^{\frac{R_{ek}^{t\,\min}}{\omega B_r}}-1\right)$; λ_k, α_k, and μ_k are the Lagrange multipliers with $k = 1, 2$. In general, the Karush-Kuhn-Tucker (KKT) conditions are the first-order necessary conditions for a solution to an optimization problem involving Lagrange function, and the combinations of the Lagrange multipliers need to satisfy the conditions. To obtain the KKT conditions in our problem we take the derivatives of the Lagrange function with respect to $P_{e1}, P_{e2}, \lambda_1, \lambda_2, \alpha_1, \alpha_2, \mu_1$ and μ_2; and next, to find the optimal powers for the users we set the derivatives equal to zero. However, we will take the derivatives only with respect to $P_{e1}, \lambda_1, \alpha_1$ and μ_1 because we are only interested in maximizing the throughput of U_{e1}^t. The KKT conditions can then be written as

$$\begin{aligned}
\frac{\partial L}{\partial P_{e1}^*} &= \omega B_r\times\left[\left(\frac{\eta_{e1}}{\ln 2\left(1+\frac{P_{e1}\eta_{e1}}{P_{e2}\eta_{e2}+\omega}\right)\left(P_{e2}\eta_{e2}+\omega\right)}\right)\right.\\
&\quad\left.-\left(\frac{P_{e2}\eta_{e2}\eta_{e1}}{\ln 2\left(1+\frac{P_{e2}\eta_{e2}}{P_{e1}\eta_{e1}+\omega}\right)\left(P_{e1}\eta_{e1}+\omega\right)^2}\right)\right]\\
&\quad-\lambda_1+\alpha_1\eta_{e1}+\mu_1\eta_{e1}\geq 0,\quad \text{if } P_{e1}^*\geq 0
\end{aligned} \tag{8}$$

Proof: See **Appendix A.**

$$\frac{\partial L}{\partial \lambda_1^*} = P_t - P_{e1}\geq 0,\quad \text{if } \lambda_1^*\geq 0 \tag{9}$$

$$\frac{\partial L}{\partial \alpha_1^*} = P_{e1}\eta_{e1}-\phi_1 P_{e2}\eta_{e2}-\phi_1\omega\geq 0,\quad \text{if } \alpha_1^*\geq 0 \tag{10}$$

$$\frac{\partial L}{\partial \mu_1^*} = P_{e1}\eta_{e1}-P_{e2}\eta_{e2}-P_{tol}\geq 0,\quad \text{if } \mu_1^*\geq 0 \tag{11}$$

TABLE 1. Closed-form solutions for optimal power allocation in uplink.

NOMA Slice	Optimal transmission power	Necessary conditions
eMBB	$P_{e1} = P_t$ $P_{e2} = \frac{P_t \eta_{e1}}{\phi_1 \eta_{e2}} - \frac{\omega}{\eta_{e2}}$	$P_{e1}\eta_{e1} - P_{e2}\eta_{e2} - P_{tol} > 0$ $P_{e2} < P_t$
	$P_{e1} = P_t$ $P_{e2} = \frac{P_t \eta_{e1}}{\eta_{e2}} - \frac{P_{tol}}{\eta_{e2}}$	$P_{e1}\eta_{e1} - \phi_1 P_{e2}\eta_{e2} - \phi_1 \omega > 0$ $P_{e2} < P_t$
uRLLC	$P_{u1} = P_t$ $P_{u2} = \frac{P_t \eta_{u1}}{\phi_1 \eta_{u2}} - \frac{\omega}{\eta_{u2}}$	$P_{u1}\eta_{u1} - P_{u2}\eta_{u2} - P_{tol} > 0$ $P_{u2} < P_t$
	$P_{u1} = P_t$ $P_{u2} = \frac{P_t \eta_{u1}}{\eta_{u2}} - \frac{P_{tol}}{\eta_{u2}}$	$P_{u1}\eta_{u1} - \phi_1 P_{u2}\eta_{u2} - \phi_1 \omega > 0$ $P_{u2} < P_t$

The KKT condition (8) still involves the Lagrange multipliers whereas conditions (9), (10), and (11) do not. Therefore, (9), (10) and (11) are necessary and sufficient conditions for the solution of our problem. Following the guidelines in [23], we set the derivatives equal to zero, then we combine (9) and (10) on one hand, combine (9) and (11) on the other hand to get closed-form solutions for the problem with the necessary conditions. The solutions are provided in Table 1, and as mentioned earlier, solutions for problem (6) are similar to those for problem (5) and are also provided in the table.

Remark 1: To maximize the 4K video and the other medical data throughputs, the power control needs to be applied only on the ordinary users, and users transmitting the 4K video and other medical data will transmit with their maximum power. In other words, users U^t_{e1} and U^t_{u1} will achieve high throughput 4K video and other medical data with all the transmit power allocated to them, in eMBB NOMA slice and uRLLC NOMA slice, respectively. And the power control will be applied to users U^t_{e2} and U^t_{u2} to transmit other ordinary data in eMBB NOMA slice and uRLLC NOMA slice, respectively. However, applying the power control to the ordinary users may affect their throughputs due to their lower channel conditions. A solution to correct such a situation is then provided and presented at the end of the next subsection.

B. DOWNLINK SCENARIO

1) ACHIEVABLE THROUGHPUT

In the downlink scenario, we also consider a total of four UEs from the two slices (two in each) to receive the data signals transmitted by users in the uplink. As shown by Fig.4, those users are represented by U^r_{e1} and U^r_{e2} in the eMBB NOMA slice; and by U^r_{u1} and U^r_{u2} in the uRLLC NOMA slice. The achievable throughput for any user equipment $UE_k, \forall k = 1, 2$ from any of the two NOMA slices can be expressed as [23]

$$R^r_k = \omega B_r \log_2 \left(1 + \frac{P_k \eta_k}{\sum_{l=1}^{k-1} P_k \eta_l + \omega} \right) \tag{12}$$

where $\sum_{l=1}^{k-1} P_k \eta_l$ is the inter-user interference at the k^{th} user level.

2) OPTIMAL POWER ALLOCATION FOR USERS' THROUGHPUT MAXIMIZATION

The main objective in our downlink scenario is to receive patients 4K videos and other medical data with high achievable throughput at the hospital unit. Based on the principle of downlink NOMA described in section II, to achieve the objective, we assume that, in the eMBB NOMA slice, the user equipment receiving the 4K videos (U^r_{e1}) has a higher channel gain than the one receiving ordinary data (U^r_{e2}). That is $h_{e1} > h_{e2}$. Similarly, in the uRLLC NOMA slice, the user equipment receiving other medical data (U^r_{u1}) has higher channel gain than the one receiving ordinary data (U^r_{u2}). That is $h_{u1} > h_{u2}$. This is because interferences from users with the lowest channel gain can be suppressed by users with the highest channel gain by using SIC. But the users with the lowest channel gain cannot suppress any interference [23]. The transmission power for the receiving user in each NOMA slice then needs to be selected properly with different levels to perform SIC.

Let's denote P_{e1} and P_{e2} the transmission power of U^r_{e1} and U^r_{e2}, respectively in the eMBB NOMA slice; P_{u1} and P_{u2} the transmission power of U^r_{u1} and U^r_{u2}, respectively in the uRLLC NOMA slice. For efficient SIC at U^r_{e1} and U^r_{u1} receivers, the power allocations must satisfy the following requirements [23].

$$P_{e2}\eta_{e1} - P_{e1}\eta_{e1} \geq P_{tol} \tag{13}$$

$$P_{u2}\eta_{u1} - P_{u1}\eta_{u1} \geq P_{tol} \tag{14}$$

with $P_{e1} + P_{e2} \leq P_{DS}$ and $P_{u1} + P_{u2} \leq P_{DS}$; η_{e1} and η_{e2} are the normalized channel gains between BS2 and user U^r_{e1} and between BS2 and user U^r_{e2}, respectively; η_{u1} and $\eta_{u2}$are the normalized channel gains between BS2 and user U^r_{u1} and between BS2 and user U^r_{u2}, respectively. Equation (13) represents the power allocation requirement to cancel interference from U^r_{e2} at U^r_{e1} receiver in the eMBB NOMA slice. Similarly, Equation (14) represents the power allocation requirement to cancel interference from U^r_{u2} at U^r_{u1} receiver in the uRLLC NOMA slice.

It is evident from (13) and (14) that the transmit power for the user with lower channel gain must be higher than that for the user with higher channel gain in each downlink NOMA slice. That is, $P_{e2} > P_{e1}$ and $P_{u2} > P_{u1}$. This is because $P_{tol} > 0$. We will, therefore, also apply optimal power allocation in the downlink to maximize the achievable throughputs for the 4K video and the other medical data.

The optimization problem will follow the same procedure and principle as in the uplink.

Let's denote $R_{e1}^{r\,\min}$ and $R_{u1}^{r\,\min}$ the minimum rate requirements for U_{e1}^{r} in the eMBB NOMA slice and U_{u1}^{r} in the uRLLC NOMA slice, respectively, with $R_{e1}^{r\,\min} > 0$ and $R_{u1}^{r\,\min} > 0$.

For the eMBB NOMA slice, we can express the optimal power allocation problem as:

$$\begin{aligned}
&\max_{P}\ \omega B_r\left[\log_2\left(1+\frac{P_{e1}\eta_{e1}}{P_{e1}\eta_{e2}+\omega}\right)\right.\\
&\qquad\left.+\log_2\left(1+\frac{P_{e2}\eta_{e2}}{P_{e2}\eta_{e1}+\omega}\right)\right]\\
&\text{Subject to: } C_1:\ P_{e1}+P_{e2}\le P_{DS}\\
&\qquad C_2:\ \omega B_r\log_2\left(1+\frac{P_{e1}\eta_{e1}}{P_{e1}\eta_{e2}+\omega}\right)\ge R_{e1}^{r\,\min}\\
&\qquad C_3:\ P_{e2}\eta_{e1}-P_{e1}\eta_{e1}\ge P_{tol}
\end{aligned}\tag{15}$$

For the uRLLC NOMA slice, we can express the optimal power allocation problem as:

$$\begin{aligned}
&\max_{P}\ \omega B_r\left[\log_2\left(1+\frac{P_{u1}\eta_{u1}}{P_{u1}\eta_{u2}+\omega}\right)\right.\\
&\qquad\left.+\log_2\left(1+\frac{P_{u2}\eta_{u2}}{P_{u2}\eta_{u1}+\omega}\right)\right]\\
&\text{Subject to: } C_1':\ P_{u1}+P_{u2}\le P_{DS}\\
&\qquad C_2':\ \omega B_r\log_2\left(1+\frac{P_{u1}\eta_{u1}}{P_{u1}\eta_{u2}+\omega}\right)\ge R_{u1}^{r\,\min}\\
&\qquad C_3':\ P_{u2}\eta_{u1}-P_{u1}\eta_{u1}\ge P_{tol}
\end{aligned}\tag{16}$$

Note that the optimization problems (15) and (16) are also convex under the constraints $C_1 - C_3$ and $C_1' - C_3'$, respectively. Constraints C_1 and C_1' are the total power constraints in eMBB NOMA slice and uRLLC NOMA slice, respectively. Constraints C_2 and C_2' are the minimum rate requirement constraints for users U_{e1}^{r} and U_{u1}^{r}, respectively. Constraints C_3 and C_3' are the SIC constraints for each NOMA slice, respectively.

Similar to the uplink case, we will use the Lagrange function to solve optimization problems (15) and (16), and we will solve only (15) whose results will be applied to (16). The Lagrange function of problem (15) can be then expressed as

$$\begin{aligned}
&L(P,\lambda,\alpha,\mu)\\
&=\omega B_r\left[\log_2\left(1+\frac{P_{e1}\eta_{e1}}{P_1\eta_{e2}+\omega}\right)+\log_2\left(1+\frac{P_{e2}\eta_{e2}}{P_{e2}\eta_{e1}+\omega}\right)\right]\\
&\quad+\lambda_1\left(P_{DS}-P_{e1}-P_{e2}\right)+\lambda_2\left(P_{DS}-P_{e1}-P_{e2}\right)\\
&\quad+\alpha_1\left[P_{e1}\eta_{e1}-\left(P_{e1}\eta_{e2}+\omega\right)\left(\varphi_1-1\right)\right]\\
&\quad+\alpha_2\left[P_{e2}\eta_{e2}-\left(P_{e2}\eta_{e1}+\omega\right)\left(\varphi_2-1\right)\right]\\
&\quad+\mu_1\left(P_{e2}\eta_{e1}-P_{e1}\eta_{e1}-P_{tol}\right)\\
&\quad+\mu_2\left(P_{e2}\eta_{e1}-P_{e1}\eta_{e1}-P_{tol}\right)
\end{aligned}\tag{17}$$

where $\varphi_k = 2^{\frac{R_k^{r\,\min}}{\omega B_r}}$; λ_k, α_k, and μ_k are the Lagrange multipliers with $k = 1, 2$.

TABLE 2. Closed-form solutions for optimal power allocation in downlink.

NOMA Slice	Optimal transmission power	Necessary conditions
eMBB	$P_{e1}=\frac{P_{DS}}{2}-\frac{P_{tol}}{2\eta_{e1}}$ $P_{e2}=\frac{P_{DS}}{2}+\frac{P_{tol}}{2\eta_{e1}}$	$P_{e1}\eta_{e1}+P_{e1}\eta_{e2}-\varphi_1\left(P_{e1}\eta_{e2}+\omega\right)+\omega>0$
uRLLC	$P_{u1}=\frac{P_{DS}}{2}-\frac{P_{tol}}{2\eta_{u1}}$ $P_{u2}=\frac{P_{DS}}{2}+\frac{P_{tol}}{2\eta_{u1}}$	$P_{u1}\eta_{u1}+P_{u1}\eta_{u2}-\varphi_1\left(P_{u1}\eta_{u2}+\omega\right)+\omega>0$

We, here too, take the derivatives of the Lagrange function only with respect to P_{e1}, λ_1, α_1 and μ_1 to write the KKT conditions as follows.

$$\begin{aligned}
\frac{\partial L}{\partial P_{e1}^{*}} &= \frac{\eta_{e1}\omega^2 B_r}{\ln 2\left(P_{e2}\eta_{e2}+\omega\right)^2\left(1+\frac{P_{e1}\eta_{e1}}{P_{e1}\eta_{e2}+\omega}\right)}\\
&\quad-\lambda_1+\alpha_1\left(\eta_{e1}-\eta_{e2}\varphi_1+\eta_{e2}\right)-\mu_1\eta_{e1}\ge 0\\
&\quad \textit{if } P_{e1}^{*}\ge 0
\end{aligned}\tag{18}$$

Proof: See **Appendix B.**

$$\frac{\partial L}{\partial \lambda_1^{*}} = P_{DS}-P_{e1}-P_{e2}\ge 0,\quad \textit{if } \lambda_1^{*}\ge 0 \tag{19}$$

$$\frac{\partial L}{\partial \alpha_1^{*}} = P_{e1}\eta_{e1}+P_{e1}\eta_{e2}-\varphi_1\left(P_{e1}\eta_{e2}+\omega\right)\ge 0,\quad \text{if } \alpha_1^{*}\ge 0 \tag{20}$$

$$\frac{\partial L}{\partial \mu_1^{*}} = P_{e2}\eta_{e1}-P_{e1}\eta_{e1}-P_{tol}\ge 0,\quad \text{if } \mu_1^{*}\ge 0 \tag{21}$$

Similar to the uplink case, the KKT condition (18) here too still involves the Lagrange multipliers, whereas conditions (19), (20), and (21) do not. Therefore, (19), (20) and (21) are necessary and sufficient conditions for the solution of our problem. Following the guidelines in [23], we set the derivatives equal to zero, combine (19) and (21) to get closed-form solutions to the problem with the necessary conditions. The solutions are provided in Table 2. The table also provides solutions to problem (16) which are found in a similar way as those of (15).

Remark 2: To maximize the throughputs of the received 4K video and other medical data in the downlink, the power control needs to be applied to all the UE receivers in each slice. However, the effect of the power control on the ordinary user's throughput from the uplink scenario will also be noticed on the throughput of the user equipment receiving ordinary data in the downlink, and may even make it worse.

This is because the power control is also applied to that ordinary user with lower channel conditions in the downlink. Hence, the final targeted results which are the downlink throughputs may be affected.

From the results of the optimal power allocation in both uplink and downlink of our proposed system, it infers that the UEs transmitting and receiving the medical data (4K

video and other data) will have favorable channel conditions to achieve high throughputs. This is because they will use their maximum power to transmit in the uplink, and in the downlink, they will receive a great portion of the total power budget. This goes against the intuition of fairness since users with lower channel conditions (lower channel gains) will suffer from low throughputs. Therefore, in the interest of fairness, we use truncated channel inversion (TCI) power control in the downlink to, at least, keep constant the throughputs of the users with lower channel conditions regardless of their channel gains. The technique is described in the next subsection.

3) TRUNCATED CHANNEL INVERSION POWER CONTROL

Channel inversion power control is a technique that aims for more fairness between channels with higher gain and channels with lower gain. In the technique, the transmitter compensates for the channel with lower gain to ensure a constant received data for its user [29]–[31].

Let's consider the channel inversion power control for user U^r_{e2} in the downlink of our system. The power allocation problem can be formulated as [30]

$$\breve{P} \triangleq \breve{P}_{e2} = \arg \max_{\{P_{e2} \geq 0\}} \min R^r_{e2} \quad \text{Subject to } P_{e1} + P_{e2} \leq P_{DS} \tag{22}$$

Following the guidelines in [29], [30], this problem can be solved by choosing the power of U^r_{e2} channel use inversely proportional to the channel power gain as

$$\breve{P}_{e2} = \frac{\tau}{\eta_{e2}} \tag{23}$$

where τ is a parameter related to the available total power for the slice as

$$P_{DS} = \tau \left(\frac{1}{\eta_{e1}} + \frac{1}{\eta_{e2}} \right) \tag{24}$$

This gives

$$\tau = \frac{\eta_{e1}\eta_{e2}P_{DS}}{\eta_{e1} + \eta_{e2}} \tag{25}$$

The achieved rate for U^r_{e2} is given by [29], [30]

$$C = \breve{R}^r_{e2} = \omega B_r \log(1 + \tau) \tag{26}$$

Substituting (25) into (26) gives

$$C = \breve{R}^r_{e2} = \omega B_r \log \left(1 + \frac{\eta_{e1}\eta_{e2}P_{DS}}{\eta_{e1} + \eta_{e2}} \right) \tag{27}$$

This is known as the zero-outage capacity and is expected to be constant whatever the user's channel condition [29]–[31]. However, this channel inversion power allocation method has a drawback. Indeed, if the channel is Rayleigh (like in our case) with more users served by the network with time the zero-outage capacity of the channel will be zero [29]. The solution is, therefore, to use a truncated channel inversion procedure where the idea is to compensate the fading only from a certain channel gain threshold η_0. The procedure is as follows.

If we denote $\breve{P}(\eta) = \tau/\eta$ as the power allocation, we have [29]

$$\breve{P}(\eta) = \begin{cases} 0, & \eta < \eta_0 \\ \dfrac{\tau}{\eta}, & \eta \geq \eta_0 \end{cases} \tag{28}$$

The channel is used only when the power is different from zero, that is when $\eta \geq \eta_0$. The user capacity is then given as [29]

$$C = \omega B_r \log(1 + \tau) P_r \{\eta \geq \eta_0\} \tag{29}$$

where $P_r\{\eta \geq \eta_0\}$ is the outage probability. This capacity is obtained by maximizing over all possible η_0 [29], and to get the minimum possible throughput for user U^r_{e2} in our case study we then consider $\eta = \eta_0$ to have

$$\breve{R}^r_{e2} = \omega B_r \log \left(1 + \frac{\eta_{e1}\eta_0 P_{DS}}{\eta_{e1} + \eta_0} \right) \tag{30}$$

Following the same procedure, we can get the minimum possible throughput for user U^r_{u2} in the uRLLC NOMA slice as

$$\breve{R}^r_{u2} = \omega B_r \log \left(1 + \frac{\eta_{u1}\eta_0 P_{DS}}{\eta_{u1} + \eta_0} \right) \tag{31}$$

V. NUMERICAL RESULTS

In this section, we provide analytical and simulated results to evaluate our proposed approach. We performed the simulations over 10000 channel realizations using Monte Carlo to evaluate the users' and system's throughputs under the proposed power allocation techniques. The total number of frames for the simulations is 100 and the length of each frame is 10. Besides, 2 units of resource blocks are allocated to each NOMA slice in both uplink and downlink in the simulations. For generality, we used some simulation parameters from [20], [23], [27], and the major ones are presented in Table 3. In all the simulations, we fixed the normalized channel gain of U^t_{e1}, U^t_{u1}, U^r_{e1}, and U^r_{u1} at 40 dB . We varied the normalized channel gain of U^t_{e2}, U^t_{u2}, U^r_{e2}, and U^r_{u2} in decreasing order from 40 dB to 5 dB [23]. Our simulation and analytical results almost coincide in all the cases, as shown in the figures.

A. PERFORMANCE EVALUATION IN UPLINK

In this subsection, we evaluate the performance of both NOMA slices in our uplink scenario. The performance includes the individual throughput of the user equipment transmitting the 4K video (U^t_{e1}) and the one transmitting ordinary data (U^t_{e2}) and their sum-throughput in the eMBB NOMA slice. It also includes the individual throughput of the user equipment transmitting the other medical data (U^t_{u1}) and the one transmitting ordinary data (U^t_{u2}) and their sum-throughput in the uRLLC NOMA slice.

TABLE 3. Simulations parameters.

Parameter	Value
System effective bandwidth	20 MHz
Bandwidth of a resource block, B_r	180 kHz
Number of available resource blocks	100
Uplink transmit power budget of each user, P_t	24 dBm
Downlink transmit power budget of each BS, P_{BS}	46 dBm
Downlink transmit power per NOMA slice, P_{DS}	23 dBm
Detection threshold at SIC receiver, P_{tol}	10 dBm

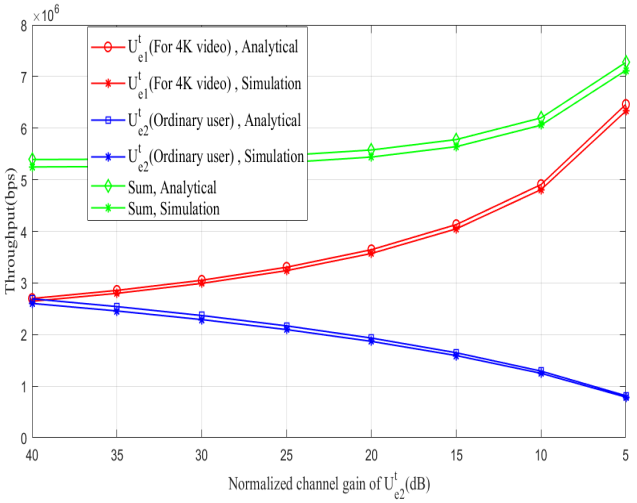

FIGURE 5. Throughput performance in uplink eMBB NOMA slice for $R_{e1}^{tmin} = 1$ Mbps and $\eta_{e1} = 40$ dB, using $P_{e1} = P_t$ and $P_{e2} = \frac{P_t \eta_{e1}}{\phi_1 \eta_{e2}} - \frac{\omega}{\eta_{e2}}$.

1) THROUGHPUTS IN EMBB NOMA SLICE

Fig. 5 and Fig. 6 depict the throughput of U_{e1}^T and U_{e2}^T, and their sum-throughput in the eMBB NOMA slice under the two solutions for the optimal power allocation provided in Table 1, respectively. The throughputs are plotted as functions of channel variations of the user with lower channel gain, that is, the channel gain of U_{e2}^t. The minimum data requirement for U_{e1}^t is fixed at 1 Mbps.

It can be seen from Fig. 5 that the 4K video throughput significantly increases as the channel gain of the ordinary user reduces. The ordinary user throughput, on the other hand, decreases at lower channel gain. This is because the power control is only applied to the ordinary user while the user transmitting the 4K video uses its maximum allocated transmit power. This approach then helps to achieve one of our main objectives which is to get a high achievable throughput for the 4K video, and as stated in section IV, it disadvantages the user with lower channel gain. However, the system

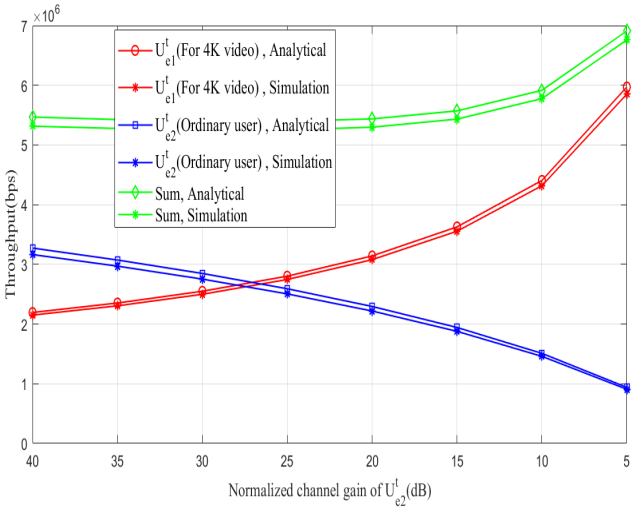

FIGURE 6. Throughput performance in uplink eMBB NOMA slice for $R_{e1}^{tmin} = 1$ Mbps and $\eta_{e1} = 40$ dB, using $P_{e1} = P_t$ and $P_{e2} = \frac{P_t \eta_{e1}}{\eta_{e2}} - \frac{P_{tol}}{\eta_{e2}}$.

sum-throughput also significantly increases despite the impact of the power control on the user with lower channel gain.

The results in Fig. 6 are similar to the ones in Fig. 5. That is, higher 4K video throughput and system sum-throughput are achieved at lower channel gain of the ordinary user whose throughput decreases. However, Fig. 6 shows that when both users have the same channel gain, the ordinary user has a better rate. Moreover, the maximum achievable sum-throughput and maximum achievable throughput of the user transmitting 4K video are lower in Fig. 6 compared to Fig. 5. These are due to the difference in the ordinary user's transmission power (P_{e2}) which is involved in the two different solutions of the power allocation problem (in Table 1). Indeed, the throughputs of both users depend on P_{e2} which is different in solution 1 and solution 2 used for Fig. 5 and Fig. 6, respectively.

2) THROUGHPUTS IN URLLC NOMA SLICE

Fig. 7 and Fig. 8 provide the sum-throughput and individual throughput of U_{u1}^t and U_{u2}^t in the uRLLC NOMA slice under the two solutions for the optimal power allocation provided in Table 1, respectively. The minimum data requirement for U_{u1}^t is fixed at 700 kbps, which is lower than that set for U_{e1}^t. This is because the other medical data transmitted by U_{u1}^t are mainly sent in form of texts and images and hence, require a lower size than a 4K video.

The results in both figures are similar to those of the eMBB NOMA slice. However, the users' individual throughput and system sum-throughput are lower in this slice compared to the eMBB NOMA slice, which is due to the difference between the minimum data requirements set for both slices.

B. PERFORMANCE EVALUATION IN DOWNLINK

Here, we evaluate the performance of both slices in our downlink scenario. The performance includes the individual

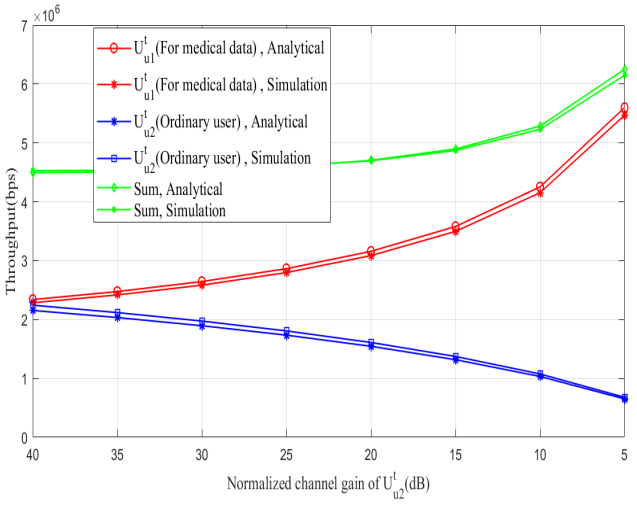

FIGURE 7. Throughput performance in uplink uRLLC NOMA slice for $R_{u1}^{tmin} = 700$ kbps and $\eta_{u1} = 40$ dB, using $P_{u1} = P_t$ and $P_{u2} = \frac{P_t \eta_{u1}}{\phi_1 \eta_{u2}} - \frac{\omega}{\eta_{u2}}$.

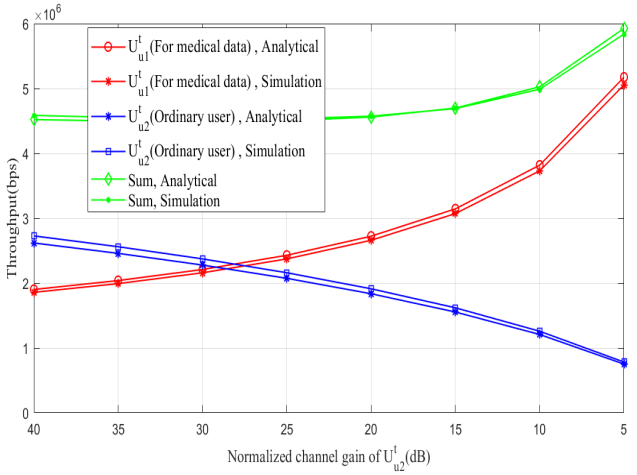

FIGURE 8. Throughput performance in uplink uRLLC NOMA slice for $R_{u1}^{tmin} = 700$ kbps and $\eta_{u1} = 40$ dB, using $P_{u1} = P_t$ and $P_{u2} = \frac{P_t \eta_{u1}}{\eta_{u2}} - \frac{P_{tol}}{\eta_{u2}}$.

throughput of the user equipment receiving the 4K video (U_{e1}^r) and the one receiving ordinary data (U_{e2}^r) and their sum-throughput in the eMBB NOMA slice. The performance also includes the individual throughput of the user equipment receiving the other medical data (U_{u1}^r) and the one receiving ordinary data (U_{u2}^r), and their sum-throughput in the uRLLC NOMA slice.

1) THROUGHPUTS IN EMBB NOMA SLICE

The sum-throughput and individual throughput of U_{e1}^r and U_{e2}^r in the eMBB NOMA slice under the optimal power allocation are depicted in Fig. 9. the results show that despite the power control applied to all users, higher 4K video throughput and sum-throughput can be effectively achieved at lower channel gain of the ordinary user. However, the throughputs are a little bit lower compared to those achieved in the uplink. This can be justified by some data loss due to the distance

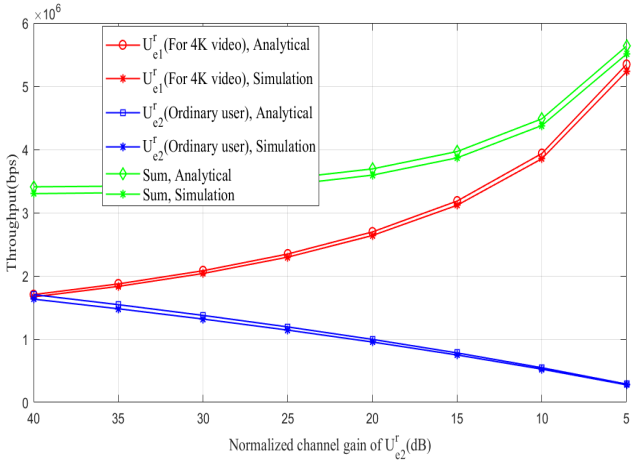

FIGURE 9. Throughput performance in downlink eMBB NOMA slice for $\eta_{e1} = 40$ dB, using $P_{e1} = \frac{P_{DS}}{2} - \frac{P_{tol}}{2\eta_{e1}}$ and $P_{e2} = \frac{P_{DS}}{2} + \frac{P_{tol}}{2\eta_{e1}}$.

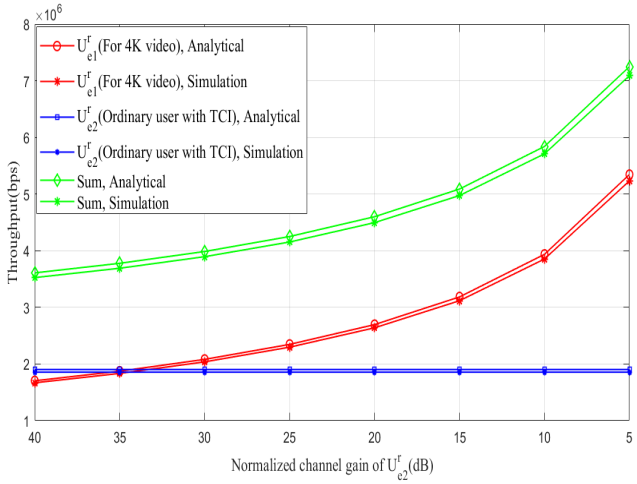

FIGURE 10. Throughput performance in downlink eMBB NOMA slice for $\eta_{e1} = 40$ dB, using $P_{e1} = \frac{P_{DS}}{2} - \frac{P_{tol}}{2\eta_{e1}}$ and applying TCI to U_{e2}^r.

between the remote site and the hospital unit, and then the signals' time difference of arrival. Moreover, the throughput of the ordinary user is still affected by this power control strategy. That is, the throughput decreases at lower channel gain.

Fig. 10 also provides the sum-throughput and individual throughput of U_{e1}^r and U_{e2}^r in the eMBB NOMA slice, but with truncated channel inversion (TCI) applied to U_{e2}^r. The results show that the TCI has been effectively useful for fairness for both users. Because it helped to keep the U_{e2}^r throughput constant regardless of its channel gains. Moreover, it is shown that the 4K video throughput and the sum-throughput have been improved with the TCI, compared to Fig. 9.

2) THROUGHPUTS IN URLLC NOMA SLICE

Fig. 11 depicts the throughput of U_{u1}^r and U_{u2}^r and their sum-throughput in the uRLLC NOMA slice under the optimal power allocation. The results are similar to those of Fig. 9. That is, higher medical data throughput and sum-throughput

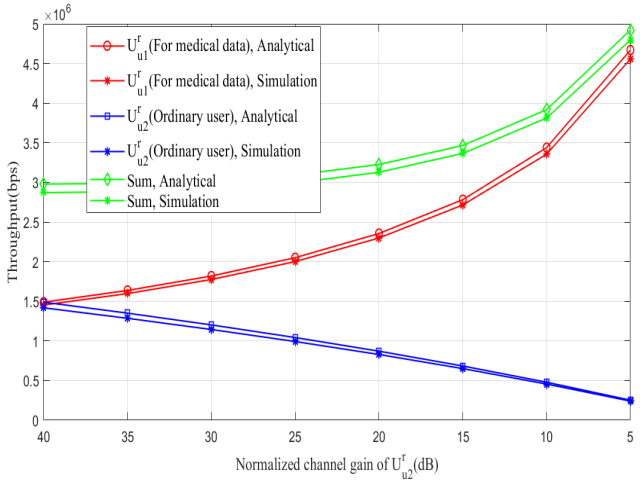

FIGURE 11. Throughput performance in downlink uRLLC NOMA slice for $\eta_{u1} = 40$ dB, using $P_{u1} = \frac{P_{DS}}{2} - \frac{P_{tol}}{2\eta_{u1}}$ and $P_{u2} = \frac{P_{DS}}{2} + \frac{P_{tol}}{2\eta_{u1}}$.

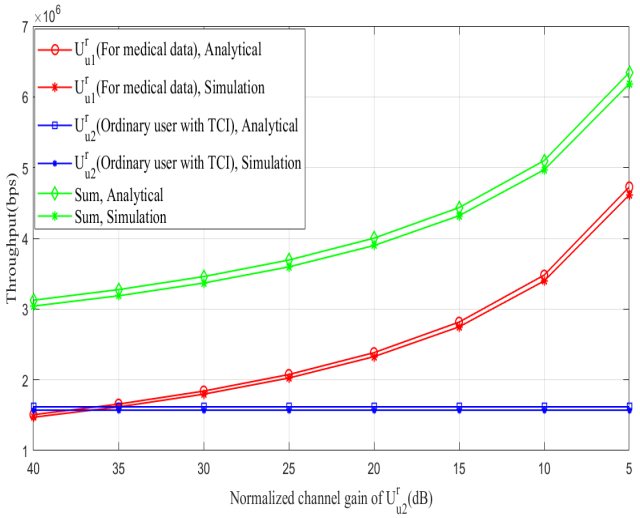

FIGURE 12. Throughput performance in downlink uRLLC NOMA slice for $\eta_{u1} = 40$ dB, using $P_{u1} = \frac{P_{DS}}{2} - \frac{P_{tol}}{2\eta_{u1}}$ and applying TCI to U^r_{u2}.

are achieved at lower channel gain of the ordinary user; the throughputs are a little bit lower compared to those achieved in the uplink, with the same reasons provided under the eMBB NOMA slice case. And the throughput of the ordinary user is also affected by the power control strategy.

The sum-throughput and individual throughput of U^r_{u1} and U^r_{u2} in the uRLLC NOMA slice are depicted in Fig. 12, but with truncated channel inversion (TCI) applied to U^r_{u2}. Similar to Fig. 10, the results show that the TCI has helped to keep the U^r_{u2} throughput constant regardless of its channel gains. Moreover, it is shown that the medical data throughput and the sum-throughput have been improved with the TCI, compared to Fig. 11.

VI. CONCLUSION

This work has proposed a 5G network slicing approach combined with non-orthogonal multiple access (NOMA) to transmit medical data from a remote place to a hospital unit in a mobile hospital system. Both the uplink and downlink of a 5G cellular network are considered with an ambulance bus located at a remote site for data transmission in the uplink scenario, and a hospital unit as the receiving site in the downlink scenario. The medical data include patients' recorded videos in 4K format and other data such as body temperature, blood pressure, blood oxygen saturation, electrocardiogram (ECG), etc. A NOMA slicing system is modeled where the medical data are categorized and assigned to two different slices based on 5G services. That is, 4K video from patients intended to be transmitted from the ambulance to the hospital is assigned to an enhanced mobile broadband (eMBB) NOMA slice in both uplink and downlink, and all other medical data are assigned to an ultra-reliable and low latency communication (uRLLC) NOMA slice also in both uplink and downlink. Moreover, we formulated and used a joint power allocation optimization technique for the medical data throughput and the system sum-throughput maximization in each slice in both uplink and downlink. The optimization problem was formulated under users' minimum rate requirements and transmission power constraints, and successive interference cancellation (SIC). Analytically and by simulations, we showed that, with the optimal power allocation technique, high throughput can be achieved for the 4K video and other medical data in the eMBB NOMA slice and uRLLC NOMA slice, respectively, but the throughput of the other users transmitting and receiving ordinary data in the slices will decrease. We, therefore, in the interest of fairness for all users, used truncated channel inversion power allocation in the downlink to prevent the decrease of the throughput of those users and keep it rather constant regardless of their channel conditions. In future work, we will focus on transmission latency minimization as well as more active UEs in each slice.

APPENDIX A PROOF OF EQUATION (8)

Given that the derivatives of (7) are to be taken only with respect to P_{e1}, λ_1, α_1 and μ_1, we have:

$$\begin{aligned}\frac{\partial L}{\partial P^*_{e1}} = \frac{\partial}{\partial P^*_{e1}}\Bigg[\omega B_r\Bigg[\log_2\left(1+\frac{P_{e1}\eta_{e1}}{P_{e2}\eta_{e2}+\omega}\right) \\ + \log_2\left(1+\frac{P_{e2}\eta_{e2}}{P_{e1}\eta_{e1}+\omega}\right)\Bigg]\Bigg] \\ + \frac{\partial}{\partial P^*_{e1}}\left[\lambda_1\left(P_t - P_{e1}\right)\right] + \frac{\partial}{\partial P^*_{e1}} \\ \times\left[\alpha_1\left(P_{e1}\eta_{e1} - \phi_1 P_{e2}\eta_{e2} - \phi_1\omega\right)\right] \\ + \frac{\partial}{\partial P^*_{e1}}\left[\mu_1\left(P_{e1}\eta_{e1} - P_{e2}\eta_{e2} - P_{tol}\right)\right] \qquad \text{(A.1)}\end{aligned}$$

Let's denote

$$\begin{aligned}A = \omega B_r\Bigg[\log_2\left(1+\frac{P_{e1}\eta_{e1}}{P_{e2}\eta_{e2}+\omega}\right) \\ + \log_2\left(1+\frac{P_{e2}\eta_{e2}}{P_{e1}\eta_{e1}+\omega}\right)\Bigg] \qquad \text{(A.2)} \\ B = \lambda_1\left(P_t - P_{e1}\right) \qquad \text{(A.3)} \\ C = \alpha_1\left(P_{e1}\eta_{e1} - \phi_1 P_{e2}\eta_{e2} - \phi_1\omega\right) \qquad \text{(A.4)} \\ D = \mu_1\left(P_{e1}\eta_{e1} - P_{e2}\eta_{e2} - P_{tol}\right) \qquad \text{(A.5)}\end{aligned}$$

Applying the addition rule of differentiation to (A.2), we have

$$\frac{\partial A}{\partial P_{e1}^*} = \omega B_r \left[\frac{\partial}{\partial P_{e1}^*} \left[\log_2 \left(1 + \frac{P_{e1}\eta_{e1}}{P_{e2}\eta_{e2} + \omega} \right) \right] + \frac{\partial}{\partial P_{e1}^*} \left[\log_2 \left(1 + \frac{P_{e2}\eta_{e2}}{P_{e1}\eta_{e1} + \omega} \right) \right] \right] \quad \text{(A.6)}$$

Applying the logarithmic rule of differentiation to (A.6), we have

$$\frac{\partial A}{\partial P_{e1}^*} = \omega B_r \left[\left[\frac{\frac{\eta_{e1}}{P_{e2}\eta_{e2}+\omega}}{\ln 2 \left(1 + \frac{P_{e1}\eta_{e1}}{P_{e2}\eta_{e2}+\omega} \right)} \right] + \left[\frac{\frac{-P_{e2}\eta_{e2}\eta_{e1}}{(P_{e1}\eta_{e1}+\omega)^2}}{\ln 2 \left(1 + \frac{P_{e2}\eta_{e2}}{P_{e1}\eta_{e1}+\omega} \right)} \right] \right] \quad \text{(A.7)}$$

It gives

$$\frac{\partial A}{\partial P_{e1}^*} = \omega B \left[\left(\frac{\eta_{e1}}{\ln 2 \left(1 + \frac{P_{e1}\eta_{e1}}{P_{e2}\eta_{e2}+\omega} \right) (P_{e2}\eta_{e2} + \omega)} \right) - \left(\frac{P_{e2}\eta_{e2}\eta_{e1}}{\ln 2 \left(1 + \frac{P_{e2}\eta_{e2}}{P_{e1}\eta_{e1}+\omega} \right) (P_{e1}\eta_{e1} + \omega)^2} \right) \right] \quad \text{(A.8)}$$

Moreover, we have

$$\frac{\partial B}{\partial P_{e1}^*} = -\lambda_1 \quad \text{(A.9)}$$

$$\frac{\partial C}{\partial P_{e1}^*} = \alpha_1 \eta_{e1} \quad \text{(A.10)}$$

$$\frac{\partial D}{\partial P_{e1}^*} = \mu_1 \eta_{e1} \quad \text{(A.11)}$$

Combining (A.8), (A.9), (A.10), and (A.11) gives Equation (8).

APPENDIX B PROOF OF EQUATION (18)

Given that the derivatives of (17) are to be taken only with respect to P_{e1}, λ_1, α_1 and μ_1, we have:

$$\begin{aligned} \frac{\partial L}{\partial P_{e1}^*} = & \frac{\partial}{\partial P_{e1}^*} \left[\omega B_r \left[\log_2 \left(1 + \frac{P_{e1}\eta_{e1}}{P_{e1}\eta_{e2} + \omega} \right) + \log_2 \left(1 + \frac{P_{e2}\eta_{e2}}{P_{e2}\eta_{e1} + \omega} \right) \right] \right] \\ & + \frac{\partial}{\partial P_{e1}^*} \left[\lambda_1 \left(P_{DS} - P_{e1} - P_{e2} \right) \right] \\ & + \frac{\partial}{\partial P_{e1}^*} \left[\alpha_1 \left(P_{e1}\eta_{e1} - \left(P_{e1}\eta_{e2} + \omega \right) \left(\varphi_1 - 1 \right) \right) \right] \\ & + \frac{\partial}{\partial P_{e1}^*} \left[\mu_1 \left(P_{e2}\eta_{e1} - P_{e1}\eta_{e1} - P_{tol} \right) \right] \end{aligned} \quad \text{(A.12)}$$

Let's denote

$$E = \omega B_r \left[\log_2 \left(1 + \frac{P_{e1}\eta_{e1}}{P_{e1}\eta_{e2} + \omega} \right) + \log_2 \left(1 + \frac{P_{e2}\eta_{e2}}{P_{e2}\eta_{e1} + \omega} \right) \right] \quad \text{(A.13)}$$

$$F = \lambda_1 \left(P_{DS} - P_{e1} - P_{e2} \right) \quad \text{(A.14)}$$

$$G = \alpha_1 \left(P_{e1}\eta_{e1} - \left(P_{e1}\eta_{e2} + \omega \right) \left(\varphi_1 - 1 \right) \right) \quad \text{(A.15)}$$

$$H = \mu_1 \left(P_{e2}\eta_{e1} - P_{e1}\eta_{e1} - P_{tol} \right) \quad \text{(A.16)}$$

Applying the addition rule of differentiation to (A.13), we have

$$\frac{\partial E}{\partial P_{e1}^*} = \omega B_r \left[\frac{\partial}{\partial P_{e1}^*} \left[\log_2 \left(1 + \frac{P_{e1}\eta_{e1}}{P_{e1}\eta_{e2} + \omega} \right) \right] + \frac{\partial}{\partial P_{e1}^*} \left[\log_2 \left(1 + \frac{P_{e2}\eta_{e2}}{P_{e2}\eta_{e1} + \omega} \right) \right] \right] \quad \text{(A.17)}$$

Applying the logarithmic rule of differentiation to (A.17), and after further derivations we have

$$\frac{\partial E}{\partial P_{e1}^*} = \omega B_r \left[\frac{\eta_{e1} \left(P_{e1}\eta_{e2} + \omega \right) - P_{e1}\eta_{e1}\eta_{e2}}{\left(P_{e1}\eta_{e2} + \omega \right)^2} \times \frac{1}{\ln 2 \left(1 + \frac{P_{e1}\eta_{e1}}{P_{e1}\eta_{e2}+\omega} \right)} \right] \quad \text{(A.18)}$$

It gives

$$\frac{\partial E}{\partial P_{e1}^*} = \frac{\eta_{e1}\omega^2 B_r}{\ln 2 \left(P_{e1}\eta_{e2} + \omega \right)^2 \left(1 + \frac{P_{e1}\eta_{e1}}{P_{e1}\eta_{e2}+\omega} \right)} \quad \text{(A.19)}$$

Moreover, we have

$$\frac{\partial F}{\partial P_{e1}^*} = -\lambda_1 \quad \text{(A.20)}$$

$$\frac{\partial G}{\partial P_{e1}^*} = \alpha_1 \left(\eta_{e1} - \eta_{e2}\varphi_1 + \eta_{e2} \right) \quad \text{(A.21)}$$

$$\frac{\partial H}{\partial P_{e1}^*} = -\mu_1 \eta_{e1} \quad \text{(A.22)}$$

Combining (A.19), (A.20), (A.21), and (A.22) gives Equation (18).

REFERENCES

[1] J. Favela, M. Rodriguez, A. Preciado, and V. M. Gonzalez, "Integrating context-aware public displays into a mobile hospital information system," *IEEE Trans. Inf. Technol. Biomed.*, vol. 8, no. 3, pp. 279–286, Sep. 2004.

[2] S. Saravanan, P. Harikrishna, and J. Vaideeswaran, "Big data exchange between ambulance bus to hospital network through Internet in telemedicine using computer communication network and 3G mobile antenna," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2015, pp. 1–7.

[3] G. Sahai, A. Goulart, W. Zhan, and R. Arnold, "Optimal selection of wireless channels for real-time communication in ambulances," in *Proc. IEEE Radio Wireless Symp.*, Jan. 2008, pp. 85–88.

[4] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G evolution: A view on 5G cellular technology beyond 3GPP release 15," *IEEE Access*, vol. 7, pp. 127639–127651, 2019.

[5] D. Marabissi, L. Mucchi, R. Fantacci, M. R. Spada, F. Massimiani, A. Fratini, G. Cau, J. Yunpeng, and L. Fedele, "A real case of implementation of the future 5G city," *Future Internet*, vol. 11, no. 1, pp. 1–16, Dec. 2018.

[6] R. Ferrus, O. Sallent, J. Perez-Romero, and R. Agusti, "On 5G radio access network slicing: Radio interface protocol features and configuration," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 184–192, May 2018.

[7] J. Rodriguez, "Drivers for 5G: The 'pervasive connected world,'" in *Fundamentals of 5G Mobile Networks*. Hoboken, NJ, USA: Wiley, 2015, pp. 1–25. [Online]. Available: https://www.wiley.com/en-us/Fundamentals+of+5G+Mobile+Networks-p-9781118867525

[8] S. Zhang, "An overview of network slicing for 5G," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 111–117, Jun. 2019.

[9] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 94–100, May 2017.
[10] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, "A comprehensive survey of RAN architectures toward 5G mobile communication system," *IEEE Access*, vol. 7, pp. 70371–70421, 2019.
[11] Q. Li, G. Wu, A. Papathanassiou, and U. Mukherjee, "An end-to-end network slicing framework for 5G wireless communication systems," 2016, *arXiv:1608.00572*. [Online]. Available: http://arxiv.org/abs/1608.00572
[12] O. Sallent, J. Perez-Romero, R. Ferrus, and R. Agusti, "On radio access network slicing from a radio resource management perspective," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 166–174, Oct. 2017.
[13] I. da Silva, G. Mildh, A. Kaloxylos, P. Spapis, E. Buracchini, A. Trogolo, G. Zimmermann, and N. Bayer, "Impact of network slicing on 5G radio access networks," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Athens, Greece, Jun. 2016, pp. 153–157.
[14] Y. L. Lee, J. Loo, and T. C. Chuah, "A new network slicing framework for multi-tenant heterogeneous cloud radio access networks," in *Proc. Int. Conf. Adv. Electr., Electron. Syst. Eng. (ICAEES)*, Putrajaya, Malaysia, Nov. 2016, pp. 414–420.
[15] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 80–87, May 2017.
[16] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2429–2453, 3rd Quart., 2018.
[17] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55765–55779, 2018.
[18] E. J. dos Santos, R. D. Souza, J. L. Rebelatto, and H. Alves, "Network slicing for URLLC and eMBB with max-matching diversity channel allocation," *IEEE Commun. Lett.*, vol. 24, no. 3, pp. 658–661, Mar. 2020.
[19] X. Li, R. Ni, J. Chen, Y. Lyu, Z. Rong, and R. Du, "End-to-end network slicing in radio access network, transport network and core network domains," *IEEE Access*, vol. 8, pp. 29525–29537, 2020.
[20] Y. Tsukamoto, R. K. Saha, S. Nanba, and K. Nishimura, "Experimental evaluation of RAN slicing architecture with flexibly located functional components of base station according to diverse 5G services," *IEEE Access*, vol. 7, pp. 76470–76479, 2019.
[21] G. Zhou, L. Zhao, K. Liang, G. Zheng, and L. Hanzo, "Utility analysis of radio access network slicing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1163–1167, Jan. 2020.
[22] A. Kaloxylos, "A survey and an analysis of network slicing in 5G networks," *IEEE Commun. Standards Mag.*, vol. 2, no. 1, pp. 60–65, Mar. 2018.
[23] M. Shipon Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.
[24] M. Aldababsa, M. Toka, S. Gökçeli, G. K. Kurt, and O. Kucur, "A tutorial on nonorthogonal multiple access for 5G and beyond," *Wireless Commun. Mob. Comp.*, vol. 2018, pp. 1–25, Jun. 2018.
[25] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.
[26] Y. Wang, B. Ren, S. Sun, S. Kang, and X. Yue, "Analysis of non-orthogonal multiple access for 5G," *China Commun.*, vol. 13, no. 2, pp. 52–66, Feb. 2016.
[27] K. Mehmood, M. T. Niaz, and H. S. Kim, "Dynamic fractional frequency reuse diversity design for intercell interference mitigation in nonorthogonal multiple access multicellular networks," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–18, Jul. 2018.
[28] S. Costanzo, I. Fajjari, N. Aitsaadi, and R. Langar, "Dynamic network slicing for 5G IoT and eMBB services: A new design with prototype and implementation results," in *Proc. 3rd Cloudification Internet Things (CIoT)*, Paris, France, Jul. 2018, pp. 1–7.
[29] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.
[30] H. Wymeersch and A. Eryilmaz, "Multiple access control in wireless networks," in *Transmission Techniques for Digital Communications*. Amsterdam, The Netherlands: Elsevier, Aug. 2016, ch. 12, Sec. 12.4.3, vol. 1, pp. 450–452.
[31] J. Hu, S. Yan, X. Zhou, F. Shu, and J. Li, "Covert wireless communications with channel inversion power control in Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12135–12149, Dec. 2019.

PARFAIT IFEDE TEBE (Member, IEEE) received the B.Sc. degree in telecommunication engineering from the Kwame Nkrumah University of Science and Technology (KNUST), Ghana, in 2011, and the M.Sc. and Ph.D. degrees in information and communication engineering from the University of Electronic Science and Technology of China (UESTC), China, in 2014 and 2019, respectively. He is currently a Postdoctoral Fellow with UESTC. His current research interests include wireless communication and signal processing with a particular focus on 5G technologies such as network slicing, non-orthogonal multiple access (NOMA), and massive MIMO.

KWADWO NTIAMOAH-SARPONG (Graduate Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from the Kwame Nkrumah University of Science and Technology (KNUST), Ghana, in 2005, and the M.Sc. degree in telecommunication management from the HAN University of Applied Sciences, The Netherlands, in 2011. He is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), China. His current research interests include wireless communication and signal processing with a particular focus on cooperative communications, non-orthogonal multiple access, resource management, and 5G.

WENHONG TIAN (Senior Member, IEEE) received the Ph.D. degree from the Computer Science Department, North Carolina State University, USA. He is currently a Full Professor with the University of Electronic Science and Technology of China. His research interests include dynamic resource scheduling algorithms and management in communication networks and cloud data centers, machine learning and deep learning algorithms for computer vision, and natural language processing. He has published about 70 journal and conference papers, and three English books in related areas. He is also a member of ACM and CCF.

JIAN LI (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in communication and information system from the University of Electronic Science and Technology of China, Chengdu, China, in 2007, 2010, and 2015, respectively. He was a Visiting Scholar with the Center for Computational Electromagnetics, Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Urbana, IL, USA, from 2016 to 2017. Since 2017, he has been an Associate Professor with the School of Communication and Information Engineering, University of Electronic Science and Technology of China. He has authored or coauthored over 90 papers in refereed journals and conferences. His current research interests include 5G, RFID, the IoT, passive communication systems, bioelectromagnetics, integrated circuits and systems, and electromagnetic metamaterials and its applications.

YONGJUN HUANG (Member, IEEE) received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), in June 2010 and December 2016, respectively. From September 2013 to September 2015, he was a Visiting Scholar of solid-state science and engineering, and mechanical engineering with Columbia University and the Visiting Project Scientist of electrical engineering with the University of California at Los Angeles (UCLA). He has been an Assistant Professor with UESTC, since February 2017. He has published more than 60 journal articles and more than 60 conference talks. His research interests include the chip-scale photonic crystal cavity opt mechanics, low-phase noise RF sources, and high-resolution force/field sensors; and the antennas, microwave passive components, electromagnetic metamaterials, and large scale MIMO systems. He was awarded as the Excellent Doctoral Dissertation Honor selected by the China Education Society of Electronics in 2017 and got the support by the National Postdoctoral Program for Innovative Talents.

GUANGJUN WEN (Senior Member, IEEE) received the B.Sc. and M.Eng. degrees from Chongqing University, Chongqing, China, in 1986 and 1992, respectively, and the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1998. From July 1986 to February 1995, he was a Lecturer with Chongqing University, China. He was a Postdoctoral Fellow/Associate Professor with UESTC, from July 1998 to May 2000, and a Postdoctoral Fellow with the Electronics and Telecommunication Research Institute, South Korea, from May 2000 to May 2001. He was a Research Fellow with Nanyang Technological University, Singapore, from May 2001 to September 2002. He has worked for VS Electronic Pte., Ltd., Singapore, and Sumitomo Electric Group, Yokohama, Japan, as a Senior RF Design Engineer, from September 2002 to August 2005. He was a Visiting Professor with the University of California at Los Angeles, from April 2015 to May 2015. Since January 2004, he has been a Professor with UESTC. He has authored or coauthored more than 300 journal articles and presented more than 150 conference papers. He holds more than 50 Chinese patents, two books, and two book chapters. His research interests include radio frequency integrated circuits and systems for various wireless communication systems, design of RFID tag and reader, circuit components and antennas design for the Internet of Things, wireless sensor networks, and wireless energy transmission systems.