Abstract—In this paper, the scheduling management of delay-constrained medical packet transmissions in Internet of Things (IoT) based healthcare networks is studied. Unlike most existing works in the literature, we focus on beyond wireless body area network (beyond-WBAN) communications, i.e., data transmissions between smart WBAN-gateways (e.g., smart phones) and the base station of remote medical centers. In our model, various medical packets are randomly aggregated at each gateway (which ordinarily stands for one patient), and their delay-constrained beyond-WBAN transmission requests are immediately reported to the network controller (i.e., base station) with different priority levels reflecting their medical importance. The base station schedules the uplink beyond-WBAN transmissions by forming a queueing system which addresses specific medical-grade quality of service (QoS) requirements, including the priority awareness and the delay constraints of medical packet transmissions. By taking into account the natural device intelligence of smart gateways in IoT-based networks, we design a truthful and efficient mechanism which can prevent gateways from strategically misreporting the priority levels of medical packets, while incentivizing the base station to manage the transmission scheduling according to the desired manner. Both theoretical and simulation results examine the feasibility of the proposed mechanism, and demonstrate its superiority over the counterparts.

Index Terms—IoT-based healthcare, beyond-WBAN, delay-constrained transmission scheduling, priority awareness, truthful mechanism design.

I. INTRODUCTION

With the rapid growth of aging population and the prevalence of chronic diseases, conventional healthcare systems have been facing an increasing burden of overload due to their low efficiency in providing timely and high-quality medical treatments, reducing financial costs in supporting hospital infrastructures, and improving the convenience of medical services. Meanwhile, recent advances in information and networking technologies have led to the emergence of Internet of Things (IoT) [1], [2], a key technology aiming to interconnect all objects in our daily life, including sensors, devices and systems. IoT goes beyond machine-to-machine (M2M) communications [3]–[9] and covers a variety of domains and applications, so that enabling IoT-based healthcare networks becomes a promising trend for future medical systems to support pervasive electronic-health (e-health) services, such as computer-assisted rehabilitations, ubiquitous health-care monitoring and emergency notifications.

Wireless body area network (WBAN) is the core network component in IoT-based healthcare systems [10]. Generally, a WBAN consists of a number of different biosensors that are deployed in, on or around a patient for continuously monitoring the health conditions. The sensed physiological signals are first collected at a gateway for data aggregation via intra-WBAN communications (i.e., from biosensors to the gateway) and then forwarded to remote medical centers for interpretation and analysis via beyond-WBAN communications (i.e., from gateways to remote medical centers). Gateways in IoT-based WBANs can be patients’ smart phones or any other smart devices, each of which ordinarily stands for one patient. In recent years, numerous applications and prototypes for IoT-based healthcare systems have been developed [11]. Research efforts in this area include the design of platforms [12], [13], the coordination of interoperability [14] and the protection of privacy and security [15]. However, among all these, technical problems related to medical data transmissions in IoT-based healthcare networks, especially in the IoT-based beyond-WBAN, have not been well studied [16].

In fact, there are a number of emerging issues that need to be taken into account in the beyond-WBAN transmission scheduling for IoT-based healthcare: i) it is expected that with an extensive use of IoT-based WBANs, a large amount of medical data traffic will be produced, leading to an imperative requirement for radio resource management with high utilization efficiency; ii) traditional wireless technologies, such as Wi-Fi and cellular networks, cannot be directly applied for supporting anywhere and anytime healthcare services because of their inherent restrictions in radio coverage [17] or network capacities [18]; iii) to guarantee medical-grade quality of service (QoS), it is necessary to offer an absolutely prioritized transmission order in the beyond-WBAN [19]–[21], i.e., more emergent medical packets should always be scheduled for transmissions prior to the ones with less criticality; and last but not least iv) since the effectiveness of each piece of medical information lasts only for a limited period of time [22], it is required that a successful beyond-WBAN packet transmission can be done within a certain delay limit. In other words, medical packet transmissions in the beyond-WBAN should be delay-constrained.

Furthermore, it is worth noting that, to follow the medical-grade priority rule, the beyond-WBAN transmission scheduling should be based on the priority information of medical packets, which is private and has to be reported by associated gateways. However, it is widely known that gateways in
IoT-based WBANs are smart devices (e.g., smart phones) with high device intelligence [10], [16], [23], so that they may strategically misreport higher or lower priorities of their medical packets to potentially manipulate the beyond-WBAN transmission scheduling for benefiting themselves. If this happens, conventional prioritized scheduling methods cannot work properly because the packet priority information may be highly inaccurate. As a result, the medical-grade QoS can never be guaranteed, which will introduces serious consequences in healthcare services. Thus, for maintaining an efficient and robust IoT-based healthcare systems, all potential untruthful behaviors from smart gateways in the beyond-WBAN have to be prevented. In addition, since the network management consumes both computation and communication resources, the network controller in the beyond-WBAN should be compensated with enough incentives (e.g., monetary remuneration) to encourage its participation.

Technically, designing a mechanism which satisfies all above requirements is challenging:

a. To explicitly analyze the complicated relationships between the QoS of delay-constrained medical packet transmissions and the prioritized beyond-WBAN scheduling mechanism, a packet-level queueing modeling is required. However, the corresponding queueing analyses are very difficult.

b. Since gateways in the beyond-WBAN are ordinarily intelligent and potentially selfish, the designed mechanism should be able to force or induce all smart WBAN-gateways to behave truthfully by reporting actual priorities of their medical packets.

c. In order to facilitate the wide employment of IoT-based healthcare systems, the revenue of the network controller should be maximized (for compensating potential costs in implementation and management) while guaranteeing all other desired properties in data transmissions (including priority awareness and delay-constrained requirements).

In our previous works [24], [25], mechanism design technique was first employed for dealing with the beyond-WBAN transmission scheduling. However, both of them imposed a strong assumption that there was no stringent delay limit for medical packet transmissions, so that the timeliness feature of IoT-based healthcare services was ignored. This largely simplified the corresponding scheduling analyses. Moreover, the analytical frameworks in [24], [25] relied on either the heavy traffic approximation or the assumption of larger-scale settings, which may not always be compatible with any practical IoT-based healthcare networks.

To tackle aforementioned difficulties and address limitations in our previous works, in this paper, we redesign a novel priority-aware truthful mechanism for scheduling delay-constrained medical packet transmissions in IoT-based healthcare networks. In the considered system model, multi-class medical packets generated by biosensors arrive randomly at each gateway via intra-WBAN communications. Upon receiving a medical packet, the associated gateway will immediately declare a beyond-WBAN transmission request to the base station (BS) which is further connected to remote medical centers via Internet. All medical packets are temporarily stored in gateways’ buffers until they have been successfully transmitted to the BS or dropped by gateways due to excessive delays (i.e., waiting longer than their required delay limits). Different medical packets may experience different beyond-WBAN transmission time and are constrained by different delay requirements. As the network controller, the BS manages beyond-WBAN medical packet transmissions with the guarantee of medical-grade QoS. The packet-level operation of the beyond-WBAN transmission scheduling is then formulated as a multi-class delay-constrained multi-server priority queueing system. Based on this model, an efficient mechanism is proposed which can ensure that all smart WBAN-gateways will truthfully report their packet priority levels and can incentivize the BS to manage the transmission scheduling system by maximizing its revenue. Note that the focus of this paper is on the scheduling management of medical packet transmissions in the beyond-WBAN, and thus the details of intra-WBAN communications are omitted. Interested readers can refer to [26]–[28] for intra-WBAN designs.

The main contributions of this paper are in the following.

- The management of beyond-WBAN transmissions for IoT-based healthcare networks is modeled as a multi-class delay-constrained multi-server priority queueing system.
- With the consideration of smart gateways’ strategic behaviors, a mechanism design problem for beyond-WBAN transmission scheduling is formulated.
- As a key prerequisite, the performance of the considered delay-constrained queueing scheduling system is extensively analyzed.
- Based on derived queueing outcome and observed characteristics, a truthful mechanism for scheduling medical packet transmissions with delay constraints is proposed.
- Theoretical and simulation results show that our proposed mechanism can meet all design requirements, and can achieve a superior performance compared to counterparts.

The rest of this paper is organized as follows: Section II reviews the related literature and explains the novelties of this work. Section III describes the considered system model and shows the problem formulation. In Section IV, a truthful mechanism for scheduling medical packet transmissions with delay constraints in the IoT-based beyond-WBAN is proposed and analyzed. Simulation results are presented in Section V, followed by conclusions in Section VI.

II. RELATED WORK

As a promising paradigm for future medical systems, IoT-based healthcare applications and services have attracted more and more interests from both academia and industry. For example, Yang et al. in [29] developed an intelligent IoT-based platform for in-home healthcare with the aim of improving patients’ experience and convenience. Kwon et al. in [30] introduced the implementation of IoT-based healthcare systems for industrial applications. Lu et al. in [15] constructed a secure and privacy-preserving opportunistic computing framework for minimizing the privacy disclosure and preventing malicious attacks in mobile healthcare. Recently, WBANs, as the most important supporting network components, have
also drawn a lot of research attentions. In [26], Liu et al. proposed a novel time division multiple access protocol for minimizing the energy consumption of biosensors in WBANs. In [27], Su et al. presented a cross-layer based battery-aware transmission scheme for WBANs by jointly considering time-varying fading channels and packet queuing characteristics. In [28], Argyriou et al. studied optimal data forwarding from biosensors to gateways in the presence of body shadowing. However, all these works restricted their emphases on intra-WBAN communications.

A major requirement in scheduling transmissions of multi-class medical packets with different criticality is the priority awareness [31]. Lee et al. in [19] designed an efficient prioritized medium access protocol based on IEEE 802.11e for medical wireless local area networks. Rezvani et al. in [32] developed a context-aware adaptive resource allocation algorithm for WBANs, in which data traffics were prioritized according to medical situations and channel conditions. Phunchongharn et al. in [33] introduced an innovative wireless access scheme for healthcare networks which could avoid electromagnetic interference and provide two classes of priorities. Besides these, delay constraints for medical packet transmissions, as another key feature of IoT-based healthcare services, have been discussed in [34], [35]. Specifically, Liu et al. in [34] proposed an energy-efficient transmission rate adaption policy for delay-constrained real-time healthcare monitoring, and Kim et al. in [35] analyzed the coexistence of ZigBee-based WBAN and WiFi with the guarantee of medical-grade delay requirements. However, none of these works investigated impacts of device intelligence on medical packet transmission scheduling.

Mechanism design technique has been widely studied in various wireless applications for preventing untruthful behaviors from intelligent and selfish users [36]–[40]. For instance, Xu et al. in [36] presented a collision-resistant double auction mechanism to determine the optimal relay assignment for maximizing the overall network throughput. Yi et al. in [37] formulated a two-stage spectrum sharing mechanism for cognitive radio networks with uncertain primary spectrum usages. Since queuing modeling is an effective tool in describing the dynamic management of wireless transmissions [41], mechanism design integrating queuing scheduling has been recently recognized in the literature [42], [43] as a natural and powerful way to explore the performance of dynamic systems with private information from individuals. However, all these mechanisms cannot be directly applied because their considered models did not fit the settings of IoT-based healthcare systems, and all of them ordinarily ignored medical-grade QoS requirements on transmission scheduling.

The most related works are [24] and [25], in which two different priority-aware truthful mechanisms were respectively developed for the beyond-WBAN management with homogeneous and heterogeneous packet transmission time. Though both of them aimed to address delay-sensitive issues in healthcare monitoring, they ignored the impact of delay constraints on transmission scheduling. Moreover, the mechanism in [24] was devised based on the heavy traffic approximation, and the one in [25] was shown to be efficient under large-scale settings only. All these limitations will be relaxed in this paper.

In IoT-based healthcare monitoring, the remote collection of patients’ physiological information relies on a multi-tier network architecture, called WBAN, as illustrated in Fig. 1. In this paper, with the specific focus on medical packet transmissions from patients to the medical center, we consider a cellular-like beyond-WBAN communication model [24], [25] consisting of a single BS and $K$ gateways (each of which represents one patient). The BS is responsible to manage the scheduling of medical packet transmissions from all gateways on $N$ homogeneous and orthogonal channels that are dedicated for healthcare services. Each gateway aggregates a variety of medical packets (e.g., EEG, ECG and EMG) generated by its associated biosensors through intra-WBAN communications (which has been standardized in IEEE std. 802.15.6 [44]), and then forward them to the BS through beyond-WBAN transmissions. According to the existing standard [44], medical packets are categorized into a finite set of priority levels, denoted by $\mathcal{L} = \{0, 1, \ldots, L\}$, where 0 and 1 represent the lowest and the highest priority levels, respectively.

In IoT-based healthcare networks, each WBAN consists of up to 256 heterogeneous biosensors deployed on a patient [44], and it is expected that with the developments in lightweight sensors and the low-power transmission technologies, this number may even increase for fulfilling more comprehensive and accurate healthcare monitoring [10]. Thus, at each gateway $k, \forall k \in \{1, 2, \ldots, K\}$, the aggregate arrival of medical packets collected from a large number of independent biosensors can be well approximated as a Poisson process with an average rate $\lambda_k$ [17], [27]. However, our proposed mechanism can also be applied to scenarios where packet arrivals are more generally distributed. Besides, with a long-term health condition tracking on patients, it is reasonable to assume that there is a known distribution $P_k = (P_{k,0}, P_{k,1}, \ldots, P_{k,L})$ on the medical packet arrival from different priority levels at each gateway $k, \forall k \in \{1, 2, \ldots, K\}$, where $P_{k,\ell}$ indicates the probability

III. System Model and Problem Description

In this section, the system model for medical packet transmissions in IoT-based healthcare networks is first described. Then, the problem of designing a truthful mechanism for priority-aware transmission scheduling of medical packets with delay constraints is formulated.

A. Network Model

In IoT-based healthcare monitoring, the remote collection of patients’ physiological information relies on a multi-tier network architecture, called WBAN, as illustrated in Fig. 1. In this paper, with the specific focus on medical packet transmissions from patients to the medical center, we consider a cellular-like beyond-WBAN communication model [24], [25] consisting of a single BS and $K$ gateways (each of which represents one patient). The BS is responsible to manage the scheduling of medical packet transmissions from all gateways on $N$ homogeneous and orthogonal channels that are dedicated for healthcare services. Each gateway aggregates a variety of medical packets (e.g., EEG, ECG and EMG) generated by its associated biosensors through intra-WBAN communications (which has been standardized in IEEE std. 802.15.6 [44]), and then forward them to the BS through beyond-WBAN transmissions. According to the existing standard [44], medical packets are categorized into a finite set of priority levels, denoted by $\mathcal{L} = \{0, 1, \ldots, L\}$, where 0 and 1 represent the lowest and the highest priority levels, respectively.

In IoT-based healthcare networks, each WBAN consists of up to 256 heterogeneous biosensors deployed on a patient [44], and it is expected that with the developments in lightweight sensors and the low-power transmission technologies, this number may even increase for fulfilling more comprehensive and accurate healthcare monitoring [10]. Thus, at each gateway $k, \forall k \in \{1, 2, \ldots, K\}$, the aggregate arrival of medical packets collected from a large number of independent biosensors can be well approximated as a Poisson process with an average rate $\lambda_k$ [17], [27]. However, our proposed mechanism can also be applied to scenarios where packet arrivals are more generally distributed. Besides, with a long-term health condition tracking on patients, it is reasonable to assume that there is a known distribution $P_k = (P_{k,0}, P_{k,1}, \ldots, P_{k,L})$ on the medical packet arrival from different priority levels at each gateway $k, \forall k \in \{1, 2, \ldots, K\}$, where $P_{k,\ell}$ indicates the probability...
that an arrived medical packet at gateway $k$ is in priority level $\ell, \forall \ell \in \mathcal{L}$. Obviously, $\sum_{\ell=0}^{L} P_{k,\ell} = 1$. Thus, the average arrival rate of medical packets in priority level $\ell$ at gateway $k$ can be calculated as $\lambda_k P_{k,\ell}$, $\forall k \in \{1, 2, \ldots, K\}, \forall \ell \in \mathcal{L}$.

To join the beyond-WBAN for transmitting any medical packet to the BS, each gateway is required to immediately declare a transmission request along with the corresponding packet priority when it receives a packet from the intra-WBAN. All medical packets that have not been scheduled for beyond-WBAN transmissions are temporarily stored in gateways’ buffers. In this paper, we do not consider buffer overflow\(^1\). However, it is worth noting that since medical packet transmissions are delay-constrained (i.e., there is a stringent delay requirement for each medical packet transmission, as shown in Table I [22]), a medical packet will be dropped by the gateway (and will no longer be transmitted) whenever it has been waiting longer than its required delay limit. Since delay limits of various medical packets may be different, we can define a generic random variable $D$ to describe the packets’ delay limits observed by the BS (while the realizations of packets’ delay limits are allowed to be different). Note that these delay limits may lead to potential packet loss in the beyond-WBAN scheduling system.

As the network controller, the BS schedules the uplink transmissions of medical packets from all gateways in the beyond-WBAN. To guarantee medical-grade QoS, i.e., medical packets with higher criticality (in higher priority levels) should always be delivered/reported before the others with less emergency (in lower priority levels) [19], the BS will determine the beyond-WBAN transmission order purely based on medical packets’ priorities, and independent of the identities of gateways. Thus, the BS can treat all transmission requests from a single virtual gateway, as depicted in Fig. 2, which consists of $L+1$ different arrival processes with respect to the total $L+1$ priority levels in $\mathcal{L}$. Because of the independency among gateways, their aggregate arrivals at the virtual gateway are still Poisson distributed, and the average rates are

\[ \Lambda_\ell = \sum_{k=1}^{K} \lambda_k P_{k,\ell}, \quad \forall \ell \in \mathcal{L}. \]

(1)

Considering the diversities in term of packet sizes and achievable signal-to-noise ratios (SNRs) at different gateways, medical packets may experience different beyond-WBAN transmission time. From the view of the network controller (i.e., the BS), the transmission time of medical packets on a beyond-WBAN channel can thus be represented by a generic random variable $T$. Then, we can formulate the operation of the beyond-WBAN transmission scheduling as a multi-class delay-constrained multi-server priority queueing system with $L+1$ Poisson-distributed packet arrivals corresponding to $L+1$ priority levels, different service time (i.e., transmission time), heterogeneous delay limits, and $N$ servers (i.e., channels). Fig. 2 shows the formulated queueing model of medical packet transmissions in the beyond-WBAN. Note that, following the similar discussions as in [17], [25], the overhead caused by control signalling is ignored since it is negligible compared to regular medical packet transmissions.

For convenience, Table II lists some important notations used in this paper.

### B. Problem Formulation

Naturally, patients will benefit from pervasive healthcare monitoring. Such benefit implies a utility gain at each gateway for successfully transmitting a medical packet to the BS in the beyond-WBAN. To characterize this, define $v_\ell$ as the valuation for the successful beyond-WBAN transmission of a medical packet in priority level $\ell, \forall \ell \in \mathcal{L}$. Intuitively, packets with higher priorities have higher valuations. Thus, we must have

\[ v_0 < v_1 < v_2 < \ldots < v_L. \]

(2)

Denote $\mathcal{V} = \{v_0, v_1, \ldots, v_L\}$. In practice, $\mathcal{V}$ can be pre-determined by medical specialists, and hence can be consid-
ered as a common knowledge to all gateways and the BS in the network. For explanation purpose in later analyses, we let 
\[ v_{t+1} - v_t = \delta, \forall \ell \in \{0, 1, \ldots, L - 1\}, \] 
where \( \delta > 0 \) is a pre-defined system parameter.

On the contrary, for each medical packet \( i \), its priority level \( \ell_i \in \mathcal{L} \) is a private information that is only available to the associated gateway, while unknown to other gateways and the BS. According to the network model, upon receiving medical packet \( i \) with priority \( \ell_i \), the associated gateway will immediately declare a beyond-WBAN transmission request to the BS by reporting the priority level of this packet. However, as an intelligent and rational entity, a smart gateway may strategically report \( \ell'_i \neq \ell_i \) if and only if it can benefit more from such behavior. By taking into account all gains and costs of a medical packet transmission in the beyond-WBAN, the net utility obtained by the gateway from transmitting packet \( i \) with actual priority \( \ell_i \) but reporting \( \ell'_i \) can be defined as

\[ U_i(\ell'_i|\ell_i) = v_{\ell_i}(1 - x(\ell'_i)) - \pi(\ell'_i), \] \hspace{1cm} (3)

where \( x(\ell'_i) \in \{0, 1\} \) is the indicator of packet loss (i.e., \( x(\ell'_i) = 1 \) means that packet \( i \) is dropped due to the over-limit waiting delay, and \( x(\ell'_i) = 0 \) otherwise); \( v_{\ell_i} \) and \( \pi(\ell'_i) \) are the valuation of successful packet transmission and the charge by the BS for beyond-WBAN channel service, respectively. Since the BS is unaware of gateways’ private information about the actual priorities of their data packets, it is intuitive that the beyond-WBAN transmission scheduling outcomes (i.e., \( x(\ell'_i) \) and \( \pi(\ell'_i) \)) are based on the reported priority level \( \ell'_i \).

Obviously, as an essential requirement to guarantee medical-grade QoS (i.e., the proper execution of the absolutely prioritized transmission scheduling), the designed mechanism should be able to induce all gateways to truthfully report the actual priority levels of their medical packets. Note that (3) is an ex-post utility function because the packet loss indicator \( x(\ell'_i) \) depends on the instantaneous queueing performance of the system, which is unknown in advance. Thus, a smart gateway will consider to potentially misreport the priority of a packet only for maximizing its expected utility (according to the packet loss probabilities of the queueing system). To prevent such misreport, we introduce the following truthfulness condition:

\[ \ell_i = \arg \max_{0 \leq \ell'_i \leq L} \{E[U_i(\ell'_i|\ell_i)]\}, \text{ for any packet } i. \hspace{1cm} (4) \]

The above equation indicates that the expected utility of transmitting packet \( i \), i.e., \( E[U_i(\ell'_i|\ell_i)] \), is always maximized when the gateway behaves truthfully by reporting \( \ell'_i = \ell_i \). With the utility function (3), \( E[U_i(\ell'_i|\ell_i)] \) can be expressed as

\[ E[U_i(\ell'_i|\ell_i)] = v_{\ell_i}(1 - Q(\ell'_i)) - \pi(\ell'_i), \] \hspace{1cm} (5)

where \( Q(\ell'_i) \) indicates the packet loss probability given priority level \( \ell'_i \). Substituting (5) into (4), we can rewrite the truthfulness condition in a general form as

\[ v_{\ell_i}(1 - Q(\ell'_i)) - \pi(\ell'_i) \geq v_{\ell_i}(1 - Q(\ell_i')) - \pi(\ell_i'), \forall \ell, \ell' \in \mathcal{L}. \hspace{1cm} (6) \]

In addition, to encourage medical packet transmissions in the beyond-WBAN, the designed mechanism should also ensure individual rationality, i.e., non-negative expected utility for transmitting any packet that is reported truthfully:

\[ E[U(\ell)|\ell] = v_{\ell}(1 - Q(\ell)) - \pi(\ell) \geq 0, \forall \ell \in \mathcal{L}. \hspace{1cm} (7) \]

Meanwhile, the BS aims to maximize its revenue gained from beyond-WBAN transmissions of all medical packets. If packets’ priority levels are reported truthfully, the expected revenue of the BS can be calculated as

\[ R = \sum_{\ell=0}^{L} \Lambda_{\ell} \cdot \pi(\ell), \hspace{1cm} (8) \]

where \( \Lambda_{\ell} \pi(\ell) \) is the average service charge on beyond-WBAN transmissions of medical packets in priority level \( \ell, \forall \ell \in \mathcal{L} \). It is worth noting that the packet loss probability, \( Q(\ell), \forall \ell \in \mathcal{L} \), is not necessary to be included in (8). This is because \( \pi(\ell), \forall \ell \in \mathcal{L} \), is actually a function of \( Q(\ell), \forall \ell \in \mathcal{L} \) (which will be discussed in Section IV-B), so that the definition of \( R \) already implies the expected revenue charged from successful beyond-WBAN transmissions. As the network controller, the BS determines the scheduling discipline \( \zeta \) and the pricing rule \( \pi = [\pi(0), \ldots, \pi(L)] \) for maximizing \( R \) subject to required system constraints.

In summary, the problem of designing a truthful mechanism for scheduling delay-constrained medical packet transmissions in IoT-based healthcare networks can be formulated as

\[ [\zeta, \pi] = \arg \max_{L} \sum_{\ell=0}^{L} \Lambda_{\ell} \cdot \pi(\ell), \hspace{1cm} (9) \]

s.t., \( v_{\ell} Q(\ell) + \pi(\ell) \geq v_{\ell'} Q(\ell') + \pi(\ell'), \forall \ell, \ell' \in \mathcal{L}, \hspace{1cm} (10) \]

\[ v_{\ell}(1 - Q(\ell)) - \pi(\ell) \geq 0, \forall \ell \in \mathcal{L}, \hspace{1cm} (11) \]

\[ Q(\ell) \geq O_{\ell}(\zeta), \forall \ell \geq \ell' \text{ and } \forall \ell, \ell' \in \mathcal{L}, \hspace{1cm} (12) \]

\[ (Q(0), \ldots, Q(L)) \in S(A, T, D, \zeta), \hspace{1cm} (13) \]

where constraint (10) is derived from the truthfulness condition (6); constraint (11) imposes the requirement of individual rationality; constraint (12) indicates that the scheduling discipline \( \zeta \) should lead to an absolutely prioritized transmission order (to fulfill the required medical-grade QoS), i.e., medical packets in a higher priority level should always be granted with a higher transmission priority \( (O_{\ell}(\zeta)) \); and constraint (13) states that the packet loss probabilities are obtained from the beyond-WBAN queueing system, denoted by \( S \). Obviously, solving this problem directly is very challenging because i) \( Q(\ell), \forall \ell \in \mathcal{L} \), is an endogenous factor of the priority-aware delay-constrained queueing system \( S \) (depending on the queueing discipline \( \zeta \) and the delay limits of medical packets \( D \)); ii) \( \pi \) is not a simple decision vector but an undetermined pricing function highly relying on the resulting queueing dynamics; and iii) exhaustive searching will result in an immeasurable computational complexity. Therefore, in the following, we will propose a novel approach to construct the mechanism, i.e., \([\zeta, \pi]\), which can meet all aforementioned requirements.

IV. A TRUTHFUL MECHANISM FOR DELAY-CONSTRAINED TRANSMISSION SCHEDULING

In this section, a truthful mechanism for scheduling medical packet transmissions with delay constraints (named as
TMDC) in IoT-based healthcare networks is designed. We first study the performance of the considered priority-aware delay-constrained queueing system. Then, the characteristics of the mechanism are investigated. Based on these, an efficient pricing rule is devised and the desired properties of TMDC are analyzed.

A. Analysis of the Queueing Scheduling System

As explained in Section III-B, a key in designing TMDC is the explicit analysis of the formulated queueing scheduling system $S(\Lambda, T, D, \zeta)$ and its corresponding performance $Q(\ell), \forall \ell \in L$. Note that the absolutely prioritized transmission requirement (12) will naturally result in a non-preemptive priority queueing discipline as the realization of $\zeta$, where arriving medical packets in higher priority levels will always be put in front of the medical packets with lower priorities that are waiting in the buffer. Besides, the non-preemptive priority queueing discipline can also ensure that medical packets that are currently under beyond-WBAN transmissions will never be interrupted so that the continuity and the completeness of medical information can be maintained.

However, compared to existing studies on priority queues [45], analyzing the beyond-WBAN transmission scheduling system with $\zeta$ is still very difficult because i) each medical packet has a required delay limit so that the packet loss due to unsatisfied services has to be taken into account with the service process; and ii) there are multiple channels available for beyond-WBAN packet transmissions, which necessitates the investigation of a multi-server queueing model. To analyze this complicated queueing system so as to obtain the packet loss probabilities $Q(\ell), \forall \ell \in L$, in the following, we will construct an absorbing Markov chain to describe the serving and the dropping processes of the queue; and then calculate the packet transmission/loss probabilities by recursions. For explicit expressions, in this subsection, we illustrate the detailed analysis by assuming that $D$ and $T$ are both exponentially distributed random variables with means $1/\eta$ and $1/\mu$, respectively. However, the proposed TMDC is actually compatible with any random distributions\(^2\).

For notation simplicity, let us denote $\Lambda = \sum_{\ell=0}^L \Lambda_\ell$, $\rho_\ell = \frac{\Lambda_\ell}{N\mu}$ and $\rho = \sum_{\ell=0}^L \rho_\ell$. Furthermore, define $A_\ell^m(t)$ as the probability that all servers are busy (i.e., all $N$ channels are temporarily occupied) and there are $m$ medical packets with priorities higher than or equal to $\ell$ in the waiting buffer at time $t$, and $B_m(t)$ as the probability that there are totally $m$ medical packets (of all priorities) in the overall beyond-WBAN transmission scheduling system (including packets under transmissions and in buffers) at time $t$.

Then, by considering all possible events that may occur during a short interval $(t, t + \Delta t)$, we have

$$B_0(t + \Delta t) = B_0(t)(1 - \Delta \zeta t) + B_1(t)\mu \Delta t + o(t),$$

and

$$B_m(t + \Delta t) = B_{m - 1}(t)\Lambda \Delta t + B_m(t)(1 - \Lambda m \mu \Delta t) + B_{m + 1}(t)(m + 1) \mu \Delta t + o(\Delta t), \quad \forall m < N,$$

$$B_m(t + \Delta t) = B_{m + 1}(t)(N \mu + (m + 1 - N) \eta \Delta t) + B_m(t)(1 - (\Lambda + N \mu + (m - N) \eta \Delta t) + B_{m - 1}(t)\Lambda \Delta t + o(\Delta t), \quad \forall m \geq N;$$

and

$$A_0^0(t + \Delta t) = A_1^1(t)(N \mu + \eta \Delta t) + B_{N - 1}(t)\Lambda \Delta t + (A_0^0(t) - B_0(t))N \mu \Delta t + A_0^0(t) - (N \mu + \sum_{h=\ell}^{L} \Lambda_h) \Delta t + o(\Delta t),$$

$$A_m^0(t + \Delta t) = A_{m + 1}^1(t)(N \mu + (m + 1) \eta \Delta t) + A_m^1(t) - (N \mu + m \eta + \sum_{h=\ell}^{L} \Lambda_h \Delta t) + A_{m - 1}^1(t) - (N \mu + (m + 1) \eta) \Lambda t + o(\Delta t), \quad \forall m \geq 1.$$

By letting $\Delta t \to 0$, we can obtain the following two sets of differential equations:

$$\frac{dB_0(t)}{dt} = - \Lambda B_0(t) + \mu B_1(t),$$

$$\frac{dB_m(t)}{dt} = \Lambda B_{m - 1}(t) - (\Lambda + m \mu) B_m(t) + (m + 1) \mu B_{m + 1}(t), \quad \forall m < N,$$

$$\frac{dB_m(t)}{dt} = \Lambda B_{m - 1}(t) - (\Lambda + N \mu + (m - N) \eta) B_m(t) + (N \mu + (m + 1 - N) \eta) B_{m + 1}(t), \quad m \geq N;$$

and

$$\frac{dA_0^0(t)}{dt} = (N \mu + \eta) A_1^1(t) - \sum_{h=\ell}^{L} \Lambda_h A_0^0(t) + \Lambda B_{N - 1}(t) - N \mu B_N(t),$$

$$\frac{dA_m^0(t)}{dt} = \sum_{h=\ell}^{L} \Lambda_h A_{m - 1}(t) + (N \mu + (m + 1) \eta) A_{m + 1}^1(t) - (N \mu + m \eta + \sum_{h=\ell}^{L} \Lambda_h) A_m^1(t), \quad \forall m \geq 1.$$

In the steady state (i.e., as $t \to \infty$), we can expect that $A_m^0(t)/dt \to 0, dB_m(t)/dt \to 0, \forall m$. Defining that $A_m^0 = \lim_{t \to \infty} A_m^0(t)$ and $B_m = \lim_{t \to \infty} B_m(t), \forall m$, we can derive

$$\Lambda B_0 = \mu B_1,$$

$$\Lambda + m \mu) B_m = \Lambda B_{m - 1} + (m + 1) \mu B_{m + 1}, \quad \forall m < N,$$

$$\Lambda + (N \mu + (m - N) \eta) B_m = \Lambda B_{m - 1} + (N \mu + (m + 1 - N) \eta) B_{m + 1}, \quad \forall m \geq N;$$

and

$$N \mu B_m + \sum_{h=\ell}^{L} \Lambda_h A_0^0 = \Lambda B_{N - 1} + (N \mu + \eta) A_1^1,$$

$$(N \mu + m \eta + \sum_{h=\ell}^{L} \Lambda_h) A_m^0 = \sum_{h=\ell}^{L} \Lambda_h A_{m - 1}^0 + (N \mu + (m + 1) \eta) A_{m + 1}^0, \quad \forall m \geq 1.$$

Solving equation set (14) – (17) yields

$$B_m = \begin{cases} 
B_0 \left( \frac{(N \mu)}{m!} \right)^{m - N} \quad m \leq N, \\
B_N \frac{N^{m-N}}{\Gamma(m-N+1)} \left( \frac{1}{\eta/N\mu} \right) \quad m > N.
\end{cases}$$

\(^2\)This is because the design of TMDC only requires $Q(\ell), \forall \ell \in L$, as an essential input in the pricing rule, which can always be obtained either analytically or numerically for all kinds of queueing models with different distributions.
Further applying the normalized condition $\sum_{m=0}^{\infty} B_m = 1$, we have
\[
\frac{1}{B_0} = \sum_{m=0}^{N-1} (N \rho)^m + \frac{(N \rho)^m}{N!} \sum_{m=0}^{\infty} \frac{\rho^m}{\prod_{i=0}^{m} (1 + i N \mu)},
\]
where $B_0$ shows the probability that the system is empty.

From (14) – (17), we are also able to derive $\Lambda B_{m-1} = m \mu B_m, \forall m \leq N$. Substituting this into equation set (18) – (19) and with some mathematical manipulations, we have
\[
\sum_{h=0}^{L} A_h A_{m-1} = (N \mu + m \eta) A_{m},
\]
which can be transformed as
\[
A_{m} = A_0^\ell \left( \sum_{h=0}^{L} \rho_h \right)^m / \prod_{i=1}^{m} (1 + i N \mu).
\]
Taking summations from $m = 0$ to $\infty$ on both sides of (23), we have
\[
\sum_{m=0}^{\infty} A_m = A_0^\ell \left( \sum_{h=0}^{L} \rho_h \right)^m / \prod_{i=1}^{m} (1 + i N \mu).
\]
Here, $\sum_{m=0}^{\infty} A_m$ implies the probability that all channels are busy, and thus can be calculated as
\[
\sum_{m=0}^{\infty} A_m = B_0 (N \rho)^N \sum_{m=0}^{\infty} \frac{\rho^m}{\prod_{i=1}^{m} (1 + i N \mu)}.
\]
Substituting (25) into (24), $A_0^\ell$ can be explicitly expressed as
\[
A_0^\ell = \sum_{m=N}^{\infty} B_m \prod_{m=0}^{\infty} \frac{(\sum_{h=0}^{L} \rho_h)^m}{\prod_{i=1}^{m} (1 + i N \mu)}.
\]
Clearly, with the help of (23) and (26), the explicit expression of $A_m$ can also be derived.

Now, define $H(\ell)$ as the probability that an arriving medical packet in priority level $\ell, \forall \ell \in \mathcal{L}$, will be transmitted in the beyond-WBAN, i.e.,
\[
H(\ell) = 1 - Q(\ell), \quad \forall \ell \in \mathcal{L},
\]
where $Q(\ell)$ is the packet loss probability. Furthermore, denote $H_m(\ell)$ as the probability that a medical packet with priority $\ell, \forall \ell \in \mathcal{L}$, which is currently waiting after $m$ medical packets with higher priorities, will be eventually transmitted. Obviously, the relationship between $H(\ell)$ and $H_m(\ell)$ can be established as
\[
H(\ell) = \sum_{m=0}^{N-1} B_m + \sum_{m=0}^{\infty} A_m H_m(\ell), \quad \forall \ell \in \mathcal{L},
\]
where $\sum_{m=0}^{N-1} B_m$ is the probability that there is at least one idle channel (or at most $N - 1$ medical packets in the system) when the considered medical packet with priority $\ell$ arrives (so that it can be transmitted immediately); and $\sum_{m=0}^{\infty} A_m$ is the probability that the considered medical packet will be scheduled for transmission even though the system is always busy. In (28), the only unknown variable is $H_m(\ell)$ while $B_m$ and $A_m$ have already been derived in (20) and (23), respectively. Thus, to obtain $H(\ell)$ so as to determine $Q(\ell)$, the remaining problem is to calculate $H_m(\ell)$, which is solved in the following by Markov chain analysis.

Here, we focus on a tagged medical packet in priority level $\ell$, and consider $m$ as the state which indicates that there are $m$ medical packets in front of the tagged one with priorities higher than or equal to $\ell$ waiting for transmission. Let us define packet in transmission and packet dropped as two absorbing states. Then, the Markov chain which illustrates the state transition process can be drawn as in Fig. 3. To analyze this absorbing Markov chain so as to derive $H_m(\ell)$, two different cases should be considered.

**Case I:** If the considered medical packet has the highest priority (i.e., $\ell = L$), the associated transition probabilities of the Markov chain are
\[
\alpha_m = \frac{N \mu + m \eta}{N \mu + \eta m}, \quad (29)
\]
\[
\beta_m = \frac{\eta}{N \mu + \eta(m + 1)} \quad (30)
\]
\[
1 - \alpha_m - \beta_m = 0. \quad (31)
\]
The transition probabilities in (29) – (31) imply that the states of the Markov chain can be converted in a single direction only (i.e., from $m + 1$ to $m$), as $1 - \alpha_m - \beta_m = 0$. Thus, we have
\[
H_0(L) = \alpha_0, \quad H_m(L) = \alpha_m H_{m-1}(L), \quad \forall m \geq 1. \quad (32)
\]
Substituting the expression of $\alpha_m$ in (29) into (32), we can get
\[
H_m(L) = \alpha_0 \alpha_1 \cdots \alpha_m = \frac{N \mu + \eta \cdots N \mu + m \eta}{N \mu + \eta m + 2 \eta \cdots N \mu + (m + 1) \eta} \quad (33)
\]
\[
= \frac{N \mu + (m + 1) \eta}{N \mu + (m + 1) \eta}.
\]

**Case II:** If the considered medical packet belongs to priority level $\ell, \forall \ell \in \{0, 1, \ldots, L - 1\}$, the associated transition probabilities of the Markov chain are
\[
\alpha_m = \frac{N \mu + m \eta}{N \mu + (m + 1) \eta + \sum_{h=\ell+1}^{L} \Lambda_h}, \quad (34)
\]
\[
\beta_m = \frac{\eta}{N \mu + (m + 1) \eta + \sum_{h=\ell+1}^{L} \Lambda_h}, \quad (35)
\]
\[
1 - \alpha_m - \beta_m = \frac{\sum_{h=\ell+1}^{L} \Lambda_h}{N \mu + (m + 1) \eta + \sum_{h=\ell+1}^{L} \Lambda_h}. \quad (36)
\]
The transition probabilities in (34) – (36) imply that the states of the Markov chain can be converted from $m$ to $m+1$ or vice versa (because newly arrived packets may have either higher or lower priorities than the tagged one). Hence, in this case, we have

\begin{align}
H_0(\ell) &= \alpha_0 + (1 - \alpha_0 - \beta_0)H_1(\ell), \\
H_m(\ell) &= \alpha_m H_{m-1}(\ell) + (1 - \alpha_m - \beta_m)H_{m+1}(\ell), \forall m \geq 1.
\end{align}

(37) \hspace{1cm} (38)

Substituting (34), (35) into (37) and (38) yields

\begin{align}
H_0(\ell) = \xi_1 H_1(\ell) &= \varphi_0 (1 - \xi_0 H_0(\ell)), \\
H_m(\ell) - \xi_{m+1} H_{m+1}(\ell) &= \varphi_m (H_{m-1}(\ell) - \xi_m H_m(\ell)),
\end{align}

(39) \hspace{1cm} (40)

where $\xi_m = \frac{\sum_{h=0}^{L} \lambda_h}{N \mu + m \eta}$ and $\varphi_m = \frac{N \mu + m \eta}{\sum_{h=0}^{L} \lambda_h}$ are introduced to simplify expressions. Based on (39) and (40), we can derive that

\begin{align}
H_m(\ell) &= \frac{H_{m-1}(\ell) - \xi_m H_m(\ell)}{\xi_m} + \frac{H_0(\ell)}{\xi_0} (1 - \xi_0) \\
&= \frac{H_{m-1}(\ell)}{\xi_m} + (H_0(\ell) - 1/\xi_0)(1/\xi_m + 1)
\end{align}

(41) \hspace{1cm} (42)

To determine $H_0(\ell)$, let us take $m \to \infty$ on both sides of (42). Since it is intuitive that the transmission probability of a packet with an infinite number of other packets before it tends to zero, we have

\begin{align}
0 &= H_0(\ell) + \left( H_0(\ell) - \frac{1}{\xi_0} \right) \sum_{i=1}^{\infty} \prod_{j=1}^{i} \xi_j.
\end{align}

(43)

By some further mathematical manipulations, equation (43) can be rewritten as

\begin{align}
H_0(\ell) = \sum_{i=1}^{\infty} \prod_{j=1}^{i} \xi_j.
\end{align}

(44)

Substituting (44) back into (42) eventually results in

\begin{align}
H_m(\ell) &= \sum_{i=m+1}^{\infty} \sum_{j=m+1}^{i} \prod_{j=1}^{i} \xi_j, \forall \ell \in \mathcal{L} \setminus \{L\}.
\end{align}

(45)

Finally, after obtaining $H_m(\ell), \forall \ell \in \mathcal{L}$, as illustrated in (33) and (45), the packet transmission probability $H(\ell), \forall \ell \in \mathcal{L}$, can be expressed according to (28), and thus the packet loss probability $Q(\ell), \forall \ell \in \mathcal{L}$, can be directly calculated by

\begin{align}
Q(\ell) = 1 - H(\ell), \forall \ell \in \mathcal{L}.
\end{align}

(46)

B. Design of TMDC

After obtaining the explicit relationship between the QoS performance of the delay-constrained beyond-WBAN transmission scheduling system and the priority levels of medical packets, i.e., $Q(\ell), \forall \ell \in \mathcal{L}$, we are now ready to design the mechanism TMDC.

Assuming that all system requirements (i.e., (9) – (13)) are satisfied under the designed mechanism TMDC (i.e., $[\zeta, \pi]$), some important characteristics can be observed as follows.

**Proposition 1**: If TMDC (i.e., $[\zeta, \pi]$) satisfies (9) – (13), then we must have

\begin{align}
\pi(\ell) \leq \pi(\ell'), \forall \ell \leq \ell' \text{ and } \ell, \ell' \in \mathcal{L}.
\end{align}

(47)

**Proof**: We can prove this proposition by the way of contradiction. Let us assume that

\begin{align}
\pi(\ell) > \pi(\ell'), \forall \ell \leq \ell' \text{ and } \ell, \ell' \in \mathcal{L}.
\end{align}

(48)

Since TMDC guarantees truthfulness (i.e., satisfies constraint (10)), the expected utility of transmitting a medical packet in priority level $\ell$ is maximized when its priority level $\ell$ is reported truthfully, i.e.,

\begin{align}
E[U(\ell|\ell)] \geq E[U(\ell'|\ell')], \forall \ell \leq \ell' \text{ and } \ell, \ell' \in \mathcal{L}.
\end{align}

(49)

Besides, since TMDC also ensures that more critical medical packets (in a higher priority level) are granted with a higher transmission priority as required by (12), and given that $D$ is a random variable, it is intuitive that

\begin{align}
Q(\ell) \geq Q(\ell'), \forall \ell \leq \ell' \text{ and } \ell, \ell' \in \mathcal{L}.
\end{align}

(50)

This inequality can also be verified numerically as shown in Section V.

With (48) and (50), we can derive that

\begin{align}
v_\ell Q(\ell) + \pi(\ell) > v_\ell Q(\ell') + \pi(\ell'), \forall \ell \leq \ell' \text{ and } \ell, \ell' \in \mathcal{L},
\end{align}

(51)

which is equivalent to

\begin{align}
E[U(\ell|\ell)] < E[U(\ell'|\ell')], \forall \ell \leq \ell' \text{ and } \ell, \ell' \in \mathcal{L}.
\end{align}

(52)

Clearly, the inequality (52) contradicts the truthfulness condition (49), and thus the assumption (48) does not hold.

**Proposition 1** indicates a fact that a medical packet in a higher priority level so as to obtain a better service should be priced higher than the one in a lower priority level.

**Proposition 1** indicates a fact that a medical packet in a higher priority level so as to obtain a better service should be priced higher than the one in a lower priority level.

**Proposition 2**: Let TMDC (i.e., $[\zeta, \pi]$) be a mechanism that meets the truthfulness condition (10). If the expected utility of a medical packet in the priority level $\ell_0$ is non-negative, then the expected utility of any medical packet in a higher priority level $\ell, \forall \ell > \ell_0$, is also non-negative.

**Proof**: For any medical packet in the priority level $\ell_0$, we have

\begin{align}
E[U(\ell_0|\ell_0)] = v_{\ell_0} (1 - Q(\ell_0)) - \pi(\ell_0) \geq 0.
\end{align}

(53)
By the definition in (2), we have \( v_\ell > v_{\ell_0}, \forall \ell > \ell_0 \), and hence
\[
v_\ell (1 - Q(\ell_0)) - \pi(\ell_0) > v_{\ell_0} (1 - Q(\ell_0)) - \pi(\ell_0).
\] (54)

Moreover, the satisfaction of truthfulness constraint (10) implies that
\[
v_\ell (1 - Q(\ell_0)) - \pi(\ell_0) = \mathbb{E}[U(\ell_0|\ell)] \leq \mathbb{E}[U(\ell|\ell)].
\] (55)

Combining inequalities (53) – (55) yields
\[
\mathbb{E}[U(\ell|\ell)] > \mathbb{E}[U(\ell_0|\ell_0)] \geq 0,
\] (56)
which completes this proof.

Proposition 2 reveals an important design idea of TMDC, i.e., individual rationality will be guaranteed automatically if the expected utility of any medical packet with the lowest priority (in priority level 0) can be constructed as a non-negative function.

Based on the above characteristics of the mechanism, i.e., Propositions 1 and 2, and along with the packet loss probabilities of the queueing scheduling system, i.e., \( Q(\ell), \forall \ell \in \mathcal{L} \), derived in Section IV-A, the pricing rule \( \pi \) of TMDC can be formulated as follows.

**Pricing rule:** For any medical packet in priority level \( \ell, \forall \ell \in \mathcal{L} \), the service charge for its beyond-WBAN transmission is
\[
\pi(\ell) = v_\ell (1 - Q(\ell)) - \sum_{h=0}^{\ell} (1 - Q(h))\delta, \quad \forall \ell \in \mathcal{L}.
\] (57)

Note that the pricing rule \( \pi = [\pi(0), \ldots, \pi(L)] \) in (57) is constructed to satisfy Propositions 1 and 2, and it only relies on the performance of the queueing scheduling system, i.e., \( Q(\ell), \forall \ell \in \mathcal{L} \).

In summary, the implementation time-line of TMDC is described as follows.

1) Given the statistics of IoT-based healthcare networks, the BS formulates and solves a mechanism design problem for delay-constrained beyond-WBAN transmission scheduling.

2) The BS decides to employ a non-preemptive priority queueing discipline \( \zeta \) and an efficient pricing rule \( \pi \) formulated in (57), for maximizing the expected revenue \( \mathcal{R} \), while guaranteeing truthfulness, individual rationality and medical-grade QoS.

3) The BS broadcasts the determined TMDC, i.e., \( [\zeta, \pi] \), to all associated gateways.

4) When a gateway collects a medical packet from its biosensors, it will immediately send a beyond-WBAN transmission request to the BS by reporting the priority level of this packet.

5) The BS manages the long-term beyond-WBAN transmission scheduling following TMDC.

6) All medical packets are stored in gateways’ buffers before they are scheduled for transmission. If a packet has waited longer than its required delay limit, it will be dropped and leave the scheduling system.

To verify the feasibility and the efficiency of the proposed TMDC, i.e., \( [\zeta, \pi] \), in the following, we prove that it can indeed meet all system requirements imposed in Section III-B, i.e., medical-grade QoS, individual rationality, truthfulness and revenue maximization (optimality).

**Theorem 1 (Medical-grade QoS):** By implementing TMDC, if all medical packets are reported truthfully with their actual priority levels, then packets with higher priorities will always be transmitted prior to the ones with lower priorities, i.e., \( O_\zeta(\ell) \geq O_\zeta(\ell'), \forall \ell \geq \ell', \forall \ell, \ell' \in \mathcal{L} \).

**Proof:** This can be directly observed from the adopted queueing discipline \( \zeta \) of TMDC.

**Theorem 2 (Individual rationality):** With the implementation of TMDC, the expected utility for transmitting any medical packet that has been reported truthfully with its actual priority will always be non-negative, i.e.
\[
\mathbb{E}[U(\ell|\ell)] = v_\ell (1 - Q(\ell)) - \pi(\ell) \geq 0, \quad \forall \ell \in \mathcal{L}.
\] (58)

**Proof:** Substituting the formulated pricing function (57) into \( \mathbb{E}[U(\ell|\ell)], \forall \ell \in \mathcal{L} \), we have
\[
\mathbb{E}[U(\ell|\ell)] = v_\ell (1 - Q(\ell)) - \pi(\ell)
\] (59)
\[= \sum_{h=0}^{\ell} (1 - Q(h))\delta \geq (1 - Q(0))\delta \geq 0,
\]
where above inequalities hold because \( Q(0) \geq Q(h), \forall h \in \mathcal{L} \), and \( Q(0) \) (which stands for a packet loss probability) must be less than or equal to 1.

**Theorem 3 (Truthfulness):** By implementing TMDC, all gateways in the beyond-WBAN will always behave truthfully by reporting the actual priority levels of all medical packets, because doing so can maximize the expected utility of each medical packet transmission, i.e.,
\[
\mathbb{E}[U(\ell|\ell)] \geq \mathbb{E}[U(\ell'|\ell')], \quad \forall \ell, \ell' \in \mathcal{L}.
\] (60)

**Proof:** With the pricing function (57), \( \mathbb{E}[U(\ell'|\ell')] \) can be expressed as
\[
\mathbb{E}[U(\ell'|\ell')]
\] (61)
\[= v_{\ell'} (1 - Q(\ell')) - \pi(\ell')
\]
\[= (v_\ell - v_{\ell'}) (1 - Q(\ell')) + \sum_{h=0}^{\ell'} (1 - Q(h))\delta, \quad \forall \ell, \ell' \in \mathcal{L}.
\]
Besides, as illustrated in (59), \( \mathbb{E}[U(\ell|\ell)] = \sum_{h=0}^{\ell} (1 - Q(h))\delta \).
Hence, we have
\[
\mathbb{E}[U(\ell|\ell)] - \mathbb{E}[U(\ell'|\ell')]
\] (62)
\[= \sum_{h=0}^{\ell} (1 - Q(h))\delta - \sum_{h=0}^{\ell'} (1 - Q(h))\delta - (v_\ell - v_{\ell'}) (1 - Q(\ell')).
\]
Next, we consider three different cases.

1) If \( \ell' = \ell \): It is obvious that \( \mathbb{E}[U(\ell|\ell)] - \mathbb{E}[U(\ell'|\ell')] = 0 \).

2) If \( \ell' < \ell \): According to (2), we have \( v_{\ell'} < v_\ell \), and thus (62) can be further rewritten as
\[
\mathbb{E}[U(\ell|\ell)] - \mathbb{E}[U(\ell'|\ell')]
\] (63)
\[= \sum_{h=\ell'+1}^{\ell} (1 - Q(h))\delta - (v_\ell - v_{\ell'}) (1 - Q(\ell'))
\[\geq \sum_{h=\ell'+1}^{\ell} (1 - Q(\ell'))\delta - \sum_{h=\ell'+1}^{\ell} (1 - Q(\ell'))\delta = 0,
\]
where the above inequality holds because $\zeta$ leads to $Q(\ell') \geq Q(h), \forall h \in \{\ell' + 1, \ldots, \ell\}$, and $v_\ell - v_{\ell'} = \sum_{h=\ell'+1}^{\ell} \delta$ by definition.

3) If $\ell' > \ell$: We have $v_{\ell'} > v_\ell$, and thus (62) becomes

$$E[U(\ell'|\ell)] - E[U(\ell'|\ell')]$$

$$= -\sum_{h=\ell+1}^{\ell'} (1 - Q(h)) + (v_\ell - v_{\ell'}) (1 - Q(\ell'))$$

$$\geq -\sum_{h=\ell+1}^{\ell'} (1 - Q(h)) + \sum_{h=\ell+1}^{\ell'} (1 - Q(\ell')) \delta = 0. \tag{64}$$

This is because $Q(\ell') \leq Q(h), \forall h \in \{\ell + 1, \ldots, \ell'\}$ and $v_\ell - v_{\ell'} = \sum_{h=\ell'+1}^{\ell'} \delta$.

Therefore, we always have $E[U(\ell'|\ell)] - E[U(\ell'|\ell')] \geq 0$.

To prove the optimality of TMDC on maximizing the BS’s expected revenue $R$, let us denote $P(\ell)$ as the probability that packets in the beyond-WBAN are in priority level $\ell, \forall \ell \in \mathcal{L}$, such that $P(\ell) = \Lambda_\ell / \Lambda, \forall \ell \in \mathcal{L}$, where $\Lambda = \sum_{\ell=0}^{L} \Lambda_\ell$. Then, by substituting the pricing function (57) of TMDC into the expression of $R$ in (8), and doing some mathematical manipulations, we have

$$R = \sum_{\ell=0}^{L} \Lambda_\ell \pi(\ell)$$

$$= \Lambda \sum_{\ell=0}^{L} P(\ell) \left( v_\ell (1 - Q(\ell)) - \sum_{h=0}^{\ell} (1 - Q(h)) \delta \right) \tag{65}$$

$$= \Lambda \sum_{\ell=0}^{L} P(\ell) (1 - Q(\ell)) \left( v_\ell - \sum_{h=\ell}^{L} P(h) \frac{P(h)}{P(\ell)} \delta \right).$$

Let $\psi(\ell) = \frac{\sum_{h=\ell}^{L} P(h)}{P(\ell)}$, so that (65) can be rewritten as

$$R = \Lambda \sum_{\ell=0}^{L} P(\ell) (1 - Q(\ell)) (v_\ell - \psi(\ell) \delta). \tag{66}$$

**Theorem 4 (Optimality):** The proposed TMDC can maximize the BS’s expected revenue $R$, if the distribution of $P(\ell), \forall \ell \in \mathcal{L}$, satisfies a mild condition that $\psi(\ell)$ monotonically decreases with the increase of priority level $\ell$.

**Proof:** Consider medical packets in any two different priority levels $\ell_i$ and $\ell_j$, such that $0 \leq \ell_i < \ell_j \leq L$. According to the scheduling discipline $\zeta$, we have $Q(\ell_i) > Q(\ell_j)$. Besides, if $\psi(\ell)$ monotonically decreases with the increase of priority level $\ell$, we must have $\psi(\ell_i) > \psi(\ell_j)$. Clearly, if we modify the mechanism by exchanging the transmission orders of medical packets in priority levels $\ell_i$ and $\ell_j$, their packet loss probabilities will also be exchanged. Denote $R^{TMDC}$ and $R^{M}$ as revenues of the BS under the proposed TMDC and the modified mechanism, respectively. We then can calculate the change of $R$ due to such modification as

$$\Delta = R^{TMDC} - R^{M}$$

$$= (1 - Q(\ell_i)) (v_{\ell_i} - \psi(\ell_i) \delta) + (1 - Q(\ell_j)) (v_{\ell_j} - \psi(\ell_j) \delta)$$

$$- (1 - Q(\ell_i)) (v_{\ell_j} - \psi(\ell_j) \delta) + (1 - Q(\ell_j)) (v_{\ell_i} - \psi(\ell_i) \delta)$$

$$= (Q(\ell_i) - Q(\ell_j)) \left( (v_{\ell_j} - v_{\ell_i}) + (\psi(\ell_i) - \psi(\ell_j)) \delta \right) > 0.$$  

Since $\Delta > 0$, we can conclude that any modification on TMDC is not preferred. In other words, the expected revenue of the BS is maximized when TMDC is applied. Note that most common distributions (e.g., uniform, geometric and Poisson distributions) satisfy this mild condition that their resulted $\psi(\ell)$ is decreasing when $\ell$ increases.

**V. SIMULATION RESULTS**

In this section, simulations are conducted to evaluate the performance of our proposed mechanism TMDC in managing delay-constrained transmissions in IoT-based healthcare networks. Table III lists the values of main simulation parameters. Similar settings have also been employed in [17], [24], [25]. Note that all results obtained are averaged over 100 runs.

Fig. 4 shows the packet loss probabilities of beyond-WBAN medical packet transmissions with different traffic arrival rates in the designed mechanism TMDC. From this figure, we can first observe that the curves produced by the analytical results well match those generated by simulations. This verifies the theoretical derivations presented in Section IV-A. Besides, this figure illustrates that the packet loss probability of medical packets in a certain priority level increases with the aggregate traffic arrival rate $\Lambda$. Intuitively, a larger $\Lambda$ implies a heavier traffic load caused by more beyond-WBAN medical packet transmission requests, and thus the scheduling system will become more congested, leading to a higher packet loss probability. In addition, by comparing the curves standing for packet loss probabilities of medical packets in different priority levels (i.e., $\ell = 2, 4, 6$, respectively), we can see that, given a fixed $\Lambda$, the higher priority level the medical packet has, the lower packet loss probability it obtains, which proves
transmission service charge increases with the packet priority for beyond-WBAN transmissions and the priority level of medical packets in TMDC. It is shown in this figure that the rate Λ also indicate that a larger aggregate traffic arrival for beyond-WBAN transmission. From this figure, we can clearly see that the mean delay of medical packets in beyond-WBAN implementing TMDC, intelligent gateways can strategically report the packet priority level so as to maximize the transmission utility of each medical packet. The trend of the curves in Fig. 7 shows that the transmission utility of a medical packet first increases with the reported priority level ℓ′. This is because reporting a larger ℓ′ can result in a lower packet loss probability as demonstrated in Fig. 4, so that a higher utility may be obtained. However, after a certain point (i.e., the reported packet priority level ℓ′ equals the actual one ℓ), the transmission service charge, which increases exponentially with the packet priority level as illustrated in Fig. 6, becomes dominant so that the utility decreases. Intuitively, the optimal decision (of gateways) is to report each packet with its actual priority which can maximize its transmission utility, and thus Theorem 3 holds. Moreover, since it is defined in (2) that the value of a more critical packet is higher than that of a normal one, given the same packet loss probability, the transmission utility of a higher-priority packet is obviously higher.

For comparison purpose, two existing scheduling mechanisms, i.e., DTM-L mechanism [25] and non-priority mechanism [45], are simulated as benchmarks. DTM-L mechanism was designed for managing multi-class delay-sensitive medical packet transmissions, where an extra delay control scheme was
required for differentiating beyond-WBAN services among different packet priorities. While, the non-priority mechanism aimed to manage all beyond-WBAN medical packet transmissions based on a first-come first-serve manner. Both of them ignore delay constraints for medical packet transmissions so that arbitrarily long delays may happen.

Fig. 8 depicts the beyond-WBAN transmission probabilities of medical packets with different priorities in the delay-constrained network scenario. Here, the transmission probability is defined as the probability that a medical packet is transmitted in the beyond-WBAN within its required delay limit. From this figure, we can observe that when the proposed TMDC or DTM-L mechanism is employed, medical packets in a higher priority level have a higher transmission probability. This is because both TMDC and DTM-L mechanisms are priority-aware, namely a better beyond-WBAN transmission service is always granted for medical packets with a higher priority. However, due to the extra delay control which introduces additional delays for medical packets in lower priority levels, the overall performance of DTM-L is worse than that of our proposed TMDC. Furthermore, the non-priority mechanism treats all transmission requests equally, so that the transmission probability remains unchanged for medical packets in any priority level. As a consequence, the QoS of emergent medical information deliveries is completely unprotected, which may cause serious issues in healthcare.

Table IV further compares the performance of three mechanisms (i.e., the proposed TMDC, DTM-L and non-priority mechanisms) in terms of the mean waiting delays experienced by successful beyond-WBAN medical packet transmissions. It is shown that the mean delay of successful beyond-WBAN transmissions decreases with the increase of the packet priority level for both TMDC and DTM-L mechanisms, while remaining unchanged for all priority levels under the non-priority mechanism. Moreover, from this table, we can also see that the proposed TMDC mechanism can not only guarantee priority-aware beyond-WBAN transmission services (compared to the non-priority mechanism), but also reduce the waiting delay of packets in each priority level (compared to the DTM-L). It is worth noting that all these observations can be explained following the same justifications for Fig. 8.

Fig. 9 shows expected revenues of the BS when different mechanisms are implemented. It can be observed that the expected revenue first increases and then decreases with the aggregate traffic arrival rate $\Lambda$. Even though a larger $\Lambda$ implies more beyond-WBAN transmission requests, the service charge for each individual packet transmission decreases with the increase of $\Lambda$ due to the heavier system congestion, as examined in Fig. 6. Thus, when $\Lambda$ increases over a certain value, reductions on transmission charges become dominant so that the revenue of the BS decreases. Moreover, this figure demonstrates that our proposed TMDC obviously outperforms both DTM-L and non-priority mechanisms. This is because i) TMDC results in higher packet transmission probabilities than DTM-L mechanism, as illustrated in Fig. 8; and ii) TMDC can guarantee more desirable QoS for high-priority packet transmissions than the non-priority mechanism.

VI. Conclusion

In this paper, the management of delay-constrained medical packet transmissions in IoT-based healthcare networks has been studied. To characterize the dynamic nature of wireless transmission scheduling integrating the medical-grade QoS requirements, a priority-aware queueing system is formulated and analyzed. Considering the intelligence of gateways in the IoT-based beyond-WBAN, we propose a truthful and efficient mechanism, i.e., TMDC, which can guarantee all gateways to honestly report the actual priority levels of their medical packets, while incentivizing the BS to participate in the beyond-WBAN scheduling. Both theoretical and simulation results show that the proposed mechanism can meet all design requirements and outperform the counterparts in terms of packet transmission probability and network revenue.

REFERENCES

Changyan Yi (S'16-M'18) received the B.Sc. degree from Guilin University of Electronic Technology, China, in 2012, the M.Sc. and Ph.D. degrees from University of Manitoba, MB, Canada, in 2014 and 2018, respectively. He is currently working as a Research Associate in Electrical and Computer Engineering, University of Manitoba, Canada. He was awarded Chinese Government Award for Outstanding Students Abroad in 2017, A. Keith Dixon Graduate Scholarship in Engineering for 2017-2018, Edward R. Toporeck Graduate Fellowship in Engineering for 2014-2017 (four times), University of Manitoba Graduate Fellowship (UMGF) for 2015-2018, and IEEE ComSoc Student Travel Grant for IEEE Globecom 2016. His research interests include algorithmic game theory, queuing theory and their applications in radio resource management, wireless transmission scheduling and network economics.

Jun Cai (M’04-SM’14) received the B.Sc. and the M.Sc. degrees from Xi’an Jiaotong University, Xi’an, China, in 1996 and 1999, respectively, and the Ph.D. degree from the University of Waterloo, ON, Canada, in 2004, all in electrical engineering. From June 2004 to April 2006, he was with McMaster University, Hamilton, ON, as a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellow. Since July 2006, he has been with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada, where he is currently a Professor. His current research interests include energy-efficient and green communications, dynamic spectrum management and cognitive radio, radio resource management in wireless communications networks, and performance analysis. Dr. Cai served as the TPC Co-Chair for IEEE VTC-Fall 2012 Wireless Applications and Services Track, IEEE Globecom 2010 Wireless Communications Symposium, and IWCMC 2008 General Symposium; the Publicity Co-Chair for IWCMC in 2010, 2011, 2013, and 2014; and the Registration Chair for QShine in 2005. He also served on the editorial board of the Journal of Computer Systems, Networks, and Communications and as a Guest Editor of the special issue of the Association for Computing Machinery Mobile Networks and Applications. He received the Best Paper Award from Chinacom in 2013, the Rh Award for outstanding contributions to research in applied sciences in 2012 from the University of Manitoba, and the Outstanding Service Award from IEEE Globecom in 2010.