Queueing Analysis for Medical Data Transmissions with Delay-Dependent Packet Priorities in WBANs

Zhen Zhao, Changyan Yi, Jun Cai, Senior Member, IEEE, Huijin Cao
Department of Electrical and Computer Engineering, University of Manitoba Winnipeg, Canada R3T 5V6
Emails: zhaoz348@myumanitoba.ca, {changyan.yi, jun.cai}@umanitoba.ca, caoh3@myumanitoba.ca

Abstract—In this paper, we propose a queueing model for analyzing performances of prioritized data transmissions in Medical Wireless Body Area Networks (mWBANs). Unlike the existing works, we consider the medical packet priority jointly with the delay-dependent dynamic priority in transmission scheduling at sensor nodes. The interaction between priority queues and instantaneous delay limits make the analysis much more complicated. Based on the transient queueing analysis, we develop a buffer-overflowing queueing principle to approximate the over-deadline queueing principle in this prioritized transmission scheduling system. Numerical and simulation results show that the proposed approximation queueing model works as a good fit to this prioritized data transmission system in mWBANs, and the proposed delay-dependent falling priority principle makes medical packets' transmission more reasonable in terms of timeliness issues.

I. INTRODUCTION

Advances in wireless technology and low-cost miniaturized sensors have set the stage for the development of ubiquitous realtime healthcare monitoring systems [1]. Medical wireless body area networks (mWBANs) are key components in such systems and emerge as a promising solution to relieve the financial and social burdens resulting from the growth of aging population and rising healthcare costs [2]. An mWBAN usually consists of a central coordinator and several biosensors that are placed in, on or around the human body to collect vital biological signals. Sensors transmit the physical data to the coordinator via the wireless links, and then the coordinator processes the collected data locally and forwards patient data to remote servers for further diagnosis or storage.

Unlike the conventional wireless sensor networks (WSNs), mWBANs have more stringent quality of service (QoS) requirements in terms of energy efficiency and delay tolerance [3]. For this reason, lots of researches studied the novel MAC protocol designs with extreme emphases on energy-efficiency [4]. Nonetheless, another important issue, i.e. how to make full use of the limited resources to transmit the most emergent and valuable packets in mWBANs, has not been investigated sufficiently. In the traditional communication networks, we rarely consider the priority changes even if a packet suffers a long delay, i.e. fixed priority. On the contrary, packets that have been waiting for a long time period are considered to be of poor timeliness due to the medical features in healthcare applications. In our previous works, [5], [6], we studied the prioritized medical data transmission scheduling based on simple queueing models but ignored the detailed transient queueing behavior analysis. In [7], a dynamic priority principle, head-of-the-line with priority jumps, was analyzed as a continuous queueing systems, which does not fit practical mWBANs scenarios.

In this papers, we focus on the prioritized data transmission in mWBANs, by jointly considering the medical packet priority and delay-dependent dynamic priority. Moreover, we propose a over-deadline priority falling principle and construct a discrete-time priority queueing system to study such timeliness issues in mWBANs. Since dynamic changing of transmission priority breaks the service order from time to time, developing an appropriate queueing model and analyzing the performance theoretically become challenging. To address this issue, we transformed the formulated complex queueing model by analyzing the tail behavior of delay distribution. The major contributions of our work are summarized as follows:

- A delay-dependent dynamic priority scheduling based on over-deadline principles is proposed to model the timeliness issues of medical packets in mWBANs.
- This novel prioritized transmission is formulated as a two-class discrete-time preemptive priority queueing model with the over-deadline priority falling principle.
- A buffer-overflowing principle is designed to approximate the complex queueing model with over-deadline principle through tail behavior analysis. In addition, this approximation problem is formulated as an optimization problem. A comprehensive simulation is built up to verify its correctness in instantaneous and average senses.

The rest of this paper is organized as follows. Section II describes the proposed prioritized transmission system. In Section III, an over-deadline priority queueing model is constructed to analyze the transmission performances. Section IV presents the approximated queueing analysis in detail. The numerical results, as well as the discussion are presented in Section V. Finally, the paper is concluded in Section VI.

II. SYSTEM MODEL

A. Basic Framework

We concentrate on the continuous healthcare monitoring system where each person is equipped with a medical WBAN, and the general architecture is based on the standard 802.15.6 [3]. In particular, we mainly focus on up-link transmission in the intra-WBANs (communications from sensors to the coordinator), where N various medical sensors are deployed on a human body, and a powerful personal device, such as a smart phone, works as a coordinator. Usually, these
nodes form a star-topology network, i.e. low-power sensors and the coordinator can exchange data directly in a one-hop communication.

We consider a discrete-time system and divide time up into finite intervals with equal lengths. The communications between sensors and their coordinator are based on the TDMA. And the coordinator determines time slot allocations and each sensor transmits the sensed information to their coordinator with a constant power during its transmission slots.

B. Wireless Channel Model

We assume that the wireless link is characterized by a slow flat fading channel. The channel state can be characterized by the received signal-to-noise (SNR) ratio, and it follows an independent and identically distributed (i.i.d.) Rician-square (the envelop follows a Rician distribution) distribution among time slots but remains unchanged within each time slot [8]. Under the Rician model, the probability density function of the received SNR \( \gamma \) is given by

\[
p_\gamma(\gamma) = \frac{\kappa + 1}{\bar{\gamma}} \exp[-\frac{\gamma(\kappa + 1)}{\bar{\gamma}} - \kappa] I_0(2\sqrt{\frac{\kappa \gamma(\kappa + 1)}{\bar{\gamma}}}), \tag{1}
\]

where \( I_0(\cdot) \) is the modified Bessel function of the first kind and 0th order, \( \kappa \) and \( \bar{\gamma} \) denote Rician K-factor and the average SNR, respectively.

C. Prioritized Transmission

In such continuous monitoring system, the sensed data traffic in an arbitrary sensor is relatively high. Therefore, sensors always have lots of data to transmit, especially when the transmission channel suffers from deep fading. As to these energy and buffer limited sensors, transmitting the most valuable and emergent data would be the first consideration. In this paper, we define two transmission priorities to solve this problem:

1) Medical Packet priority: Similar to [5], [6], we divide the medical packets into 2 categories: emergency or high priority (HP) packets and non-emergency or low priority (LP) packets. Moreover, we consider the preemptive-resume service process [9] to ensure the most emergent packets’ transmissions. This serve process also reflects the absolute priority rules in mWBANs. Specifically, no LP packets can start receiving services before HP packets in the system; if an LP packet is suffering from a long waiting time is less valuable than new arrival packets, in contrast to the earliest deadline transmission policy [10]. Thus, the priority discipline should be flexible enough to adjust the relative priorities of medical packets according to their delay constraints and actual experienced delay in the system [7]. More specifically, the priority of each buffered packet will decrease when its waiting time in the queue exceeds its delay limit, which is defined as the over-deadline queueing principle. In practice, these delay limits can be predefined according to the QoS requirements in IEEE standard 802.15.6.

2) Service Process: According to the IEEE Standard 802.15.6, different priority classes and different types of sensors have distinct QoS requirements. Specifically, \( R_{\text{Qos}}^H(i) \), \( R_{\text{Qos}}^L(i) \) and \( T_{\text{Qos}}^H(i) \), \( T_{\text{Qos}}^L(i) \) denote the transmission rates and delay limits for HP and LP packets, respectively. Moreover, the maximum achievable rate with constant transmission power can be calculated based on the Shannon Theorem:

\[
r_i = \log_2(1 + \gamma). \tag{2}
\]

We consider a geometric distribution to model this service process with successful probability \( b_{i,H} \) and \( b_{i,L} \) for HP and LP classes, where

\[
b_{i,k} = Pr\{r_i > R_{\text{Qos}}^k(i)\} = \int_{\gamma_{\text{Qos}}^k(i)}^{\infty} p_r(\gamma) d\gamma, \quad k = H, L. \tag{3}
\]

Here, \( \gamma_{\text{Qos}}^k(i), \quad k = H, L \) represent the minimum SNR for HP and LP packet transmission respectively, which can be written as \( \gamma_{\text{Qos}}^k(i) = 2 \frac{R_{\text{Qos}}^k(i)}{\bar{\gamma}} - 1 \).

Since the communications among different nodes are independent, for expression simplicity, we just need to focus on a
single sensor node. We use \(a_H, b_H\) and \(a_L, b_L\) to denote the arrival and departure probability for HP and LP classes at a certain node, respectively.

3) Queuing Principle: In this discrete-time queuing system, we assume that the sensor can only transmit one packet at a time slot in a first come first served order. Since the chronological order of sensed data is also useful for diagnosis of patients’ medical conditions, last-come first-served is not an appropriate solution. Besides, the HP priority packets can preempt the services of LP priority packets, and the packets dropped from HP queue will join the tails of LP queue. We assume no buffer size limitation for each class queue and there are infinite size of the sensed packets.

It is obvious that our over-deadline principle makes this queuing model much more complicated due to the priority changing from time to time. Nevertheless, we observe that the head of line (HOL) packet in the queue always has the longest waiting time, and hence, the over-deadline principle can be mathematically modeled by the HOL priority falling (HOL-PF) principle, as shown in Fig. 1, where the delay limits for \(k\)-th class are \(t_{bnd}(k), k = H, L\). Consequently, we just need to check whether the HOL packet is over-deadline, where the extra overhead is considered to be small.

B. Performance Metrics

1) Dropping Probability: In this model, the buffer size is unlimited, and thus a transmitted packet is dropped only if it becomes worthless after waiting too long in the buffer, i.e. when the over-deadline behavior occurs in the LP queue. So the dropping probability at a node \(i\) can be written as

\[
p_d(i) = \frac{\text{over-deadline packets in LP}}{\text{all sensed arrival packets}}.
\]

2) Delay: The delay of a sensed packet at node \(i\) in this paper is defined as waiting time in the queue, which is the duration of the time between a packet’s arrival and when it starts to receive service. Furthermore, we need to consider a service vacation period due to the MAC scheduling, which is considered to be a fixed value in TDMA manner. As a Result, the mean delay can be expressed as

\[
W_q(i) = W_T_{\text{queue}} + \frac{(N-1)}{N} T_{\text{frame}},
\]

where, for simplicity, the frame payload for each sensor is considered to be equal. Additionally, the queueing delay bounds are calculated as

\[
t_{bnd}(i) = T_{Qos}(i) - \frac{(N-1)}{N} T_{\text{frame}}[T_{Qos}(i) / T_{\text{frame}}].
\]

Based on these definitions, we will explore the their closed-form expressions and analyze the system performances in following sections.

IV. BUFFER-OVERFLOWING QUEUEING ANALYSIS

Despite that the over-deadline principle is transformed into HOL-PF queuing principle, the queuing analysis is still too difficult to analyze directly because of its HOL dropping process, constant delay limits and interactions between two priority queues. For this reason, we study this model in further steps through the analysis of tail behavior of the delay distributions. Firstly, we develop a buffer-overflowing queuing principle to approximate the over-deadline principle by finding the optimal buffer length; subsequently, we analyze this new priority queue in specific details.

A. Buffer-overflowing Approximation

Inspired from the fact that queueing length can indicate the waiting time in some sense, we develop a buffer-overflowing principle and aim at looking for an optimal buffer size \(L_k^*\) through tail analysis. This buffer-overflowing queuing model has the same queueing elements as over-deadline queuing model, except the priority falling principle: the HOL packet will leave the HP queue and join the LP queue when the buffer is full. So \(L_k^*\) can minimize the tail probability of the delay distribution for different classes beyond their delay limits.

We denote \(C_k^1\) as the event that the maximum waiting time of served packets in the queue is larger than \(t_{bnd}^k\), and denote \(C_k^2\) as the event that the minimum waiting time of dropped packets is less than \(t_{bnd}^k\). Furthermore, we observe the optimal buffer length can make both of event \(C_k^1\) and \(C_k^2\) less likely to happen, i.e. \(P[|C_k^1|/L_k^*] \rightarrow 0\) and \(P[|C_k^2|/L_k^*] \rightarrow 0\). In order to find the \(L_k^*\), we can formulate an optimization problem:

\[
L_k^* = \arg \min_{L \in \mathbb{Z}} (P_k^1 + P_k^2)
\]

s.t. \(L \in [1, L_{bnd}]\),

where \(P_k^1, P_k^2\) represent the probabilities that the event \(C_k^1\) and \(C_k^2\) happen, respectively. Moreover, it is obvious that the optimal buffer size is less than \(t_{bnd}^k\) due to the random arrival and departure process, thus we set \(L_k^* = t_{bnd}^k\).

The distribution of maximum waiting time can be indicated by the arrival packets that are most likely to suffer a long delay, defined as Late Packets. The queuing behavior of such late packets can be separated into three periods: 1) becoming the HOL packet with the buffer full; 2) staying at HOL without packet arrivals and departures; 3) leaving the queue because of either a prior packet departure or a new packet arrival. Accordingly, the \(P_1^H\) can be calculated as

\[
P_1^H = 1 - \sum_{j=L-1}^{L} \sum_{t_{j} = L-1}^{t_{bnd}} P_a(t_1)P_b(t - t_1)P_c(1),
\]

where

\[
P_a(t) = \frac{(t-1)}{L-2}[(a_H)^L - (1 - a_H)^{t-L+1}],
\]

\[
P_b(t) = (1-a_H)^t(1-b_H)^t,
\]

\[
P_c(1) = b_H + (1-b_H)a_H.
\]
According to previous discussion, \( P_2^H \) can be calculated as
\[
P_2^H = \sum_{i=L}^{L_0} \left( \frac{i - 1}{L - 1} \right) (a_H)^{L - 1} (1 - a_H)^{i - L}.
\]
(12)

Though the closed-form expression of the optimal result cannot be obtained, we can find the optimal solution through exclusive searching owing to the finite values of \( L \). After that, because of the preemptive services, the dropping probability of HP queue can be given directly as
\[
p_d^H = \pi L_H,
\]
(13)

\[
\pi_0^H = [1 + \frac{a_H \theta_L^H}{(1 - a_H)b_H} + \sum_{i=1}^{L_H} \frac{\theta_i^H}{1 - b_H}]^{-1},
\]
(14)

\[
\pi_i^H = \frac{\pi_0^H \theta_i^H}{1 - b_H}, i = 1, 2, \ldots, L_H,
\]
(15)

where \( \theta_H = \frac{a_H (1 - b_H)}{(1 - a_H) b_H} \), and \( \pi_i^H \) is the stationary distribution of the HP queue. The waiting time distribution of dropped packets is
\[
WT_{dH}^{ij} = \sum_{i=L_H-1}^{j-1} (a_H)^{i+L_H}(1 - a_H)^{i-L_H}(1 - b_H)^{j-i}.
\]
(16)

Similar analysis can be performed on LP queue to find \( L^*_L \). And we need to consider the service effect from HP packets when analyze the waiting time of LP packets. Here, we just give the expression of \( P_L^L \) and \( P_L^H \) directly.
\[
P_L^L = 1 - \sum_{i=L-1}^{L_0} \sum_{j=L-1}^{i} \left( \frac{j - 1}{L - 2} \right) \left( a_L^{(j,L-1)} (1 - a_L^{(j-1,L)} - b_L^{(j-1,L)}) \right),
\]
\[
x \sum_{i=L}^{L_0} \sum_{j=L-1}^{i} \left( \frac{j - 1}{L - 2} \right) \left( a_L^{(j,L-1)} (1 - a_L^{(j-1,L)} - b_L^{(j-1,L)}) \right),
\]
\[
WT_{dL}^{ij} = \sum_{i=L_H-1}^{j-1} (a_L)^{i+L_H}(1 - a_L)^{i-L_H}(1 - b_L)^{j-i}.
\]
(18)

where \( a_L^{(j,L-1)} = a_L + a_L \theta_L^H \) represents the compound arrival probability in LP queue. Then similarly, the optimal buffer length of LP queue, \( L^*_L \), can be acquired through traversing all the possible values.

B. Queueing Stationary Analysis

In a long terms, HOL packet dropping is same as end-of-the-line packet dropping except that the HOL dropping packets already have a waiting time when joining the LP queue. Therefore, after obtaining the optimal buffer length \( L^*_H \) and \( L^*_L \), a new queueing model is reformulated as shown in Fig. 2, which is a relatively standard Geo/Geo/1 priority queue. Let us write the stationary distribution, \( x_i = [x_{i,0}, x_{i,1}, x_{i,2}, \ldots, x_{i,L_H+1}] \), \( x = [x_0, x_1, x_2, \ldots, x_{L_H+1}] \). Consider the following state space \( \Delta = \{(i,1,2), i_1 \in [0, L_H+1], i_2 \in [0, L_L+1]\} \). We define \( a_{0,0} = (1 - a_L)(1 - a_H), a_{1,0} = (1 - a_L)a_H, a_{0,1} = a_L(1 - a_H), a_{1,1} = a_La_H \). The associated transition matrix is of the form,
\[
P_t = \begin{bmatrix}
B_{0,0} & B_{0,1} & A_0 & A_0 & \cdots & A_0 \\
B_{1,0} & A_1 & A_0 & A_0 & \cdots & A_0 \\
A_2 & A_1 & A_0 & A_0 & \cdots & A_0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
A_2 & A_1 & A_0 & A_0 & \cdots & A_2 \\
A_2 & A_1 & A_0 & A_0 & \cdots & A_1 \\
\end{bmatrix},
\]

where the block matrices are
\[
B_{0,0} = \begin{bmatrix}
a_{0,0} & a_{0,1} & A_0 & A_0 & \cdots & A_0 \\
B_{2,0} & B_{1,0} & A_0 & A_0 & \cdots & A_0 \\
A_1 & B_{0,1} & A_0 & A_0 & \cdots & A_0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
A_1 & B_{0,1} & A_0 & A_0 & \cdots & A_0 \\
A_0 & B_{0,1} & A_0 & A_0 & \cdots & A_0 \\
\end{bmatrix},
\]

\[
B_{0,1} = \begin{bmatrix}
a_{0,0}b_L & a_{0,1}b_L & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
B_{2,0} & B_{1,0} & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
A_1 & B_{0,1} & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
A_1 & B_{0,1} & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
A_0 & B_{0,1} & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
\end{bmatrix},
\]

\[
B_{1,0} = \begin{bmatrix}
a_{0,0}b_L & a_{0,1}b_L & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
B_{2,0} & B_{1,0} & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
A_1 & B_{0,1} & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
A_1 & B_{0,1} & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
A_0 & B_{0,1} & a_{0,0}b_H & a_{0,1}b_H & \cdots & a_{0,0}b_H \\
\end{bmatrix},
\]

with \( B_{0,1} = a_{1,1}(1 - b_L), B_{0,1} = a_{1,0}b_L, B_{0,1} = a_{1,0}(1 - b_L) + a_{1,1}b_L, B_{0,0} = a_{0,1}(1 - b_L), B_{0,0} = a_{0,0}b_L, B_{0,0} = a_{0,0}b_L + a_{0,1}b_L \).

Owing to limited space, we refer to [9] to solve the vector \( x \) by applying the Matrix-analytic approach based on the relationship \( x = xP_t, x_1 = 1 \). After that, the mean number of HP and LP packets in the queue are given as
\[
E[X_H^q] = \sum_{i=1}^{L_H+1} \sum_{j=0}^{L_H+1} (i - 1)x_{i,j},
\]
(19)

\[
E[X_L^q] = \sum_{j=1}^{L_L+1} \sum_{i=0}^{L_L+1} (j - 1)x_{i,j}.
\]
(20)

Then, the expected waiting time in the queue for HP packets can be obtained by using the Little’s Law as
\[
WT_{queue}^H = \frac{E[X_H^q]}{a_H(1 - \sum_{j=0}^{L_H+1} x_{i,j})}.
\]
(21)
However, as for LP packets, we need to consider the extra waiting time due to the arrivals dropped from HP queue. Thus the expected waiting time in the queue for LP packets is

\[ WT_{\text{queue}}^L = \frac{E[X_i^L]}{a_L \sum_{i=1}^{L_i^H+1} x_i, L_i^L + x_0, L_i^L+1} + a_H \frac{H_i}{H_L} \sum_{j=1}^{L_i^H} j \cdot WT_{dH}^j. \]  

(22)

Consequently, we can also write the dropping probability in a closed-form expression as

\[ p_d \approx p_d^L = \sum_{i=1}^{L_i^H+1} x_i, L_i^L + x_0, L_i^L+1. \]  

(23)

V. NUMERICAL RESULT

In this section, numerical results are presented to verify the correctness of the proposed buffer-overflowing analytical model and to illustrate the impact of delay-dependent dynamic priorities in mWBANs. The simulation at a sensor node is configured as shown in Table I, referring to [3]. The service rates and delay limitations can be calculated: \( b_H = b_L \approx 0.6, \) \( t_H^{bnd} = 6 \) slots and \( t_L^{bnd} = 12 \) slots.

We verify our proposed buffer-overflowing queueing principle from two viewpoints: stationary distribution and mean delay of the HP queue. Table II shows stationary distributions with over-deadline (OD) and buffer-overflowing (BO) principles when \( a_H = 0.25 \) and \( L_i^H = 3. \) We can see that BO principle similarly describes the properties of OD principle except for several tail probabilities that only contribute less than 0.001 in total. In Table III, the mean delay of HP queue is evaluated under fixed \( b_H \) and different \( a_H. \) It is obvious that the performance gap between OD and BO principle increases w.r.t \( a_H. \) However, the difference is relatively small in practice because \( \frac{a_H}{b_H} \ll 1. \) Hence, the performance of OD principle can be approximately analyzed by our proposed BO principle.

For comparison purpose, we simulate two normal transmission scheduling schemes, i.e. non-priority scheme with over-deadline principle (NoPOD) and fix-priority scheme with over-deadline principle (FixPOD). NoPOD scheme considers same delay limitations for all packets, and FixPOD scheme operates in a same way as our proposed scheme expect for ignoring the interactions between priority queues. Fig. 3 compares performances of three distinct schemes in terms of packet dropping probability and expected waiting time in the queue. It can be seen from the figure that, comparing with the FixPOD scheme, our proposed scheme can achieve a lower dropping probability with similar mean delay in HP queue through sacrificing a slightly higher mean delay in LP queue. Besides, though the NoPOD scheme has lower dropping probability than our proposed scheme, the mean delay in the queue for all packets is much higher than that for HP queue packets in our proposed scheme. Therefore, we can conclude that our proposed scheme outperforms FixPOD and NoPOD schemes in improving the performances of critical medical data transmissions.

VI. CONCLUSION

In this paper, we proposed and analyzed the prioritized data transmission with both medical packet and delay-dependent dynamic priorities in mWBANs. The transmission scheduling jointly takes into account timeliness and emergency features of medical packets. We constructed a discrete-time preemptive priority queue with the over-deadline priority falling principle, and developed an optimization problem to find an optimal buffer length to transform the over-deadline principles into buffer-overflowing principles. The analytical and simulation results showed that our proposed prioritized transmission scheduling could improve the performance of the most emergent and valuable packets’ transmission in mWBANs.

REFERENCES


