End-to-End Delay Analysis for Multi-Hop Wireless Links Based on Martingale Theory

Hangyu Yan, Xuefen Chi, Wanting Yang, and Zhu Han, *Fellow, IEEE*

***Abstract*—In this paper, we derive a new method based on the martingale theory to analyse the end-to-end (E2E) delay of multi-hop wireless links (MWL). First, the arrival-martingales and service-martingales are constructed in the signal-to-noise ratio (SNR) domain where fading channel characteristics are easily incorporated into the node's service. Then, a martingale parameter called the joint service descriptor is proposed to reflect the joint service capability of MWL for a specific arrival flow. Based on the above, we decouple the joint service and total delay threshold of MWL, which transforms the complex E2E delay analysis into the simple delay analysis of all single nodes. Finally, a closed-form expression for the E2E delay violation probability bounds is derived. Simulation results verify the accuracy and generality of our derived method in E2E delay analysis.**

***Index Terms*—Multi-hop wireless links, martingale theory, E2E delay, SNR.**

I. Introduction

Multi-hop wireless links (MWL) are widely required in various scenarios, such as vehicular networks, self-organizing networks, and the Internet of things [1]. Accurate end-to-end (E2E) delay analysis can maximize the utilization of limited resources and ensure stringent quality-of-service (QoS) requirements for different applications. However, the fading characteristics of wireless channels result in the randomness of services at each wireless node in the MWL. Further, the mutual coupling of services between wireless nodes increases the difficulty of E2E analysis. It remains a challenge to model the joint services of MWL and to perform accurate E2E delay analysis [2].

Manuscript received 12 September 2023; revised 2 February 2024; accepted 4 April 2024. Date of publication 16 April 2024; date of current version 19 September 2024. This work was supported in part by the Natural Science Foundation of Jilin Province under Grant 20230101063JC, in part by NSF under Grant CNS-2107216, Grant CNS-2128368, Grant CMMI-2222810, and Grant ECCS-2302469, and in part by the U.S. Department of Transportation, Toyota Amazon and Japan Science and Technology Agency (JST) Adopting Sustainable Partnerships for Innovative Research Ecosystem (ASPIRE) under Grant JPMJAP2326. The review of this article was coordinated by Prof. Gabriella Olmo. *(Corresponding author: Xuefen Chi.)*

Hangyu Yan and Xuefen Chi are with the Department of Communications Engineering, Jilin University, Changchun 130012, China (e-mail: yanhy21@mails.jlu.edu.cn; chixf@jlu.edu.cn).

Wanting Yang is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore (e-mail: wanting_yang@sutd.edu.sg).

Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea (e-mail: hanzhu22@gmail.com).

Digital Object Identifier 10.1109/TVT.2024.3389704

To depict the stochastic service of wireless nodes, the authors in [3] mapped the arrival and service processes to the signal-to-noise ratio (SNR) domain by exponential operations, which avoids the difficulty of obtaining the service distribution in the bit domain due to the logarithmic operation. They further proposed a $(\min,\times)$ stochastic network calculus (SNC) to analyse the E2E delay of homogeneous MWL. Subsequently, many researchers extended their work to address more complex delay analysis problems in wireless scenarios. The authors in [4] investigated the impact of traffic dispersion, network densification and a hybrid scheme of both on the E2E delay of MWL in millimeter wave communication scenarios. They utilized $(\min,\times)$ SNC to derive the corresponding E2E delay violation probability bounds (DVPB). In [5], the authors improved the concatenation theorem, reducing the optimization dimensions required for calculating the E2E-DVPB using $(\min,\times)$ SNC. Additionally, they introduced a method based on the Meijer G-function to compute the moment generating function for various wireless fading channel service processes. The authors in [6] and [7] considered the impact of interference among accessing users on delay. Building upon this consideration, they studied the power allocation problem for uplink and downlink non-orthogonal multiple access under the delay constraint, respectively. In [8], the authors analyzed the E2E-DVPB of multiple wireless routing links, taking into account the interference among wireless routers and the reliability of wireless channel services. However, the existing SNC methods have flaws. Firstly, the use of the Boolean inequality in the derivation makes the obtained DVPB loose [9]. Furthermore, the presence of multiple nested Boolean inequalities in the analysis of E2E delay further amplifies the looseness of the obtained DVPB.

In recent years, the martingale theory has shown its superiority for delay analysis in the bit domain [10]. Compared to the Boolean inequality, the Doob's inequality in martingale can provide sharper stochastic bounds due to its tightness nature. The authors in [11] demonstrated the accuracy of the martingale-based method for delay analysis in single-node system. The derived DVPB are several orders of magnitude higher than those obtained using classical SNC methods. Furthermore, the authors in [12] employed martingales to derive the E2E-DVPB for multi-hop unmanned aerial vehicle (UAV) links. The authors in [13] derived the closed-form expressions for the delay and backlog DVPB of the ultra-reliable low-latency communications and enhanced mobile broadband multiplexing queuing system for two-hop communication links. However, these works require that both the arrival process and the service process at each node

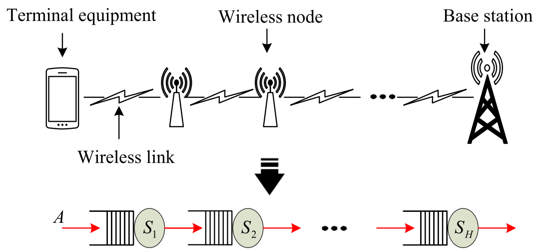

Fig. 1. Uplink multi-hop wireless communication scenario.

in the MWL are Markov processes [14], [15]. When considering the impact of the time-varying characteristics of the wireless channel on the E2E delay analysis, none of them are applicable.

In light of above, we propose a martingale-based method for analyzing the E2E delay of MWL. Our main contributions are summarized as follows.

- Considering the random effects of wireless channel fading, we construct arrival-martingales and service-martingales in the SNR domain. Based on the constructed martingale processes, an innovative method for decoupling E2E delay is proposed. Firstly, it enables the decoupling of the joint service and total delay threshold of MWL, transforming the complex E2E delay analysis into the simple delay analysis of all single nodes. Secondly, it avoids the nesting operation of inequalities in the derivation of E2E-DVPB.
- We propose a martingale parameter called the joint service descriptor, which is collectively determined by the random characteristics of arrival flow and the service of all nodes in the MWL. It reflects the service capacity of the joint service for a specific arrival flow in martingale domain.
- By using the Doob's inequality, the DVPB for all single nodes in the MWL are derived. Then, based on the constraint of the total delay of MWL, we derive a closed-form expression for E2E-DVPB. The simulation results demonstrate that the accuracy of E2E-DVPB obtained using our derived method is several orders of magnitude higher than the existing methods in the literature.

II. System Model

We consider an uplink multi-hop wireless communication scenario, as shown in Fig. 1. The terminal equipment, the wireless nodes, and the base station (BS) are sequentially connected through wireless links. An arrival flow A, generated by the terminal equipment and following a specific distribution, is transmitted to the BS via multi-hop transmission. We assume that the wireless channels in all wireless links are block fading channels and have the same fading model, such as Rayleigh fading [16]. Each wireless node and BS operates independently without interference. We further assume the existence of a central controller that can acquire channel state information for all links in the MWL [6]. Power control techniques are then employed to ensure an equal average SNR is received by each wireless node and BS [17]. The communication scenario in Fig. 1 can be viewed as a discrete time tandem queuing system with H nodes. Let $\mathcal{H} = \{1, 2, \ldots, H\}$ be the set of all nodes in the queue system. The instantaneous SNR observed at the i-th time slot of node h, $h \in \mathcal{H}$, is denoted by $\gamma_h(i)$, which is assumed to be independent and identically distributed (i.i.d.) over all time slots, and it remains unchanged within one time slot [7]. Hence, the accumulated service process of node h during the time interval $(m, n]$ can be represented by a bivariate process as

$$S_h(m,n) = \sum_{i=m+1}^{n} T_b B \log_2(1+\gamma_h(i)), \tag{1}$$

where B is the bandwidth, and T_b is the duration of a time slot. Meanwhile, the accumulated arrival and departure processes for the tandem system can be expressed as $A(m,n) = \sum_{i=m+1}^{n} a(i)$ and $D_H(m,n) = \sum_{i=m+1}^{n} d_H(i)$, where $a(i)$ and $d_H(i)$ denote the arrival and departure at the i-th time slot. It is important to note that the departure process of the last node in the tandem queue is also considered as the departure process of the entire queue. As shown in (1), it is difficult to obtain the distribution of the service process in the bit domain due to the logarithmic operator. To facilitate delay analysis, we map the above stochastic process to the SNR domain [3]. The corresponding arrival, service and departure processes in the SNR domain can be represented as $\tilde{A}(m,n) = e^{A(m,n)}$, $\tilde{S}_h(m,n) = e^{S_h(m,n)}$, and $\tilde{D}_H(m,n) = e^{D_H(m,n)}$ respectively. With the help of the SNC theory, the joint service of the tandem queue in the time interval $(l_1, n]$ can be represented as [4]

$$\begin{aligned} &\tilde{S}_{net}(l_1,n) = \tilde{S}_1 \otimes \tilde{S}_2 \cdots \tilde{S}_H\,(l_1,n) \\ &= \inf_{0\le l_1\le l_2\cdots l_H\le n} \left\{ \tilde{S}_1\,(l_1,l_2)\cdot \tilde{S}_2\,(l_2,l_3)\cdots \tilde{S}_H\,(l_H,n) \right\}, \end{aligned} \tag{2}$$

where $\otimes$ is (min,$\times$) convolution, and $l_2, l_3, \ldots, l_H$ represents different time slots in the $(l_1, n]$ time interval that satisfy the corresponding conditions. Moreover, the arrival process, the joint service process and the departure process of the tandem system can be coupled using (min,$\times$) convolution, which can be expressed as

$$\begin{aligned} \tilde{D}_H(0,n) &\ge \tilde{A} \otimes \tilde{S}_{net}(0,n) \\ &= \inf_{0\le l_1\le n} \left\{ \tilde{A}\,(0,l_1)\cdot \tilde{S}_{net}\,(l_1,n) \right\}. \end{aligned} \tag{3}$$

The queuing delay of the tandem system can be given as [14]

$$\tilde{W}_{net}(n) = \inf\left\{ w \ge 0 : \tilde{A}(0,n) \le \tilde{D}_H(0,n+w) \right\}, \tag{4}$$

where $\tilde{W}_{net}(n)$ is the queuing delay at the n-th time slot, and w is the delay threshold. When the system is in a stationary condition, the E2E-DVPB can be described as

$$\begin{aligned} &P\left\{ \tilde{W}_{net}(n) \ge w \right\} \\ &= P\left\{ \tilde{A}(0,n) \ge \tilde{D}_H(0,n+w) \right\} \\ &= P\left\{ \tilde{A}(w,n) \ge \tilde{D}_H(0,n) \right\} \\ &\le P\left\{ \sup_{0\le l_1\le n-w} \left[\frac{\tilde{A}\,(l_1,n-w)}{\tilde{S}_{net}\,(l_1,n)} \right] \ge 1 \right\}. \end{aligned} \tag{5}$$

With an algebraic description of the E2E-DVPB in the SNR domain, we next investigate the calculation of the E2E-DVPB.

III. Martingale-Based E2E Delay Analysis in SNR Domain

A. Martingale Construction in the SNR Domain

The martingale is a powerful tool for the study of stochastic process. First, we give its definition.

Definition 3.1 (Martingale Process): If a discrete-time stochastic process $\{\mathcal{X}(n); n \geq 0\}$ is a martingale process, it needs to satisfy the following conditions [18]:

$$\begin{aligned} &\mathbb{E}\left[|\mathcal{X}(n)|\right] < \infty, \\ &\mathbb{E}\left[\mathcal{X}(n+1) \mid \mathcal{F}_n\right] = \mathcal{X}(n), \end{aligned} \tag{6}$$

where $\mathcal{F}_n$ is a filtration. The stochastic process $\{\mathcal{X}(n); n \geq 0\}$ needs to be $\mathcal{F}_n$ measurable. If $\{\mathcal{X}(n); n \geq 0\}$ is a super- or sub-martingale process, we can use $\leq$ or $\geq$ to replace the equal sign in (6).

Definition 3.2 (Mellin transform): For a non-negative stochastic variable $\mathcal{Z}$, its Mellin transform can be defined as

$$\mathcal{M}_{\mathcal{Z}}(\vartheta) = \mathbb{E}\left[\mathcal{Z}^{\vartheta-1}\right], \tag{7}$$

where ϑ is a free parameter.

With the above knowledge, we construct arrival-martingales and service-martingales in the SNR domain.

Lemma 3.1 (Arrival-martingales): For an i.i.d. arrival flow in the SNR domain, its accumulated arrival process can be expressed as $\tilde{A}(m,n) = \prod_{i=m+1}^{n} \tilde{a}(i), n \geq m \geq 0$, where $\tilde{a}(i)$ denotes the arrival at the i-th time slot. The Mellin transform of $\tilde{a}(i)$ is $\mathcal{M}_{\tilde{a}(i)}(1+\vartheta) = \mathbb{E}[(\tilde{a}(i))^{\vartheta}]$. Furthermore, as $\tilde{a}(i)$ is i.i.d., $\mathcal{M}_{\tilde{a}}(1+\vartheta) = \mathcal{M}_{\tilde{a}(i)}(1+\vartheta)$ is achieved for all $\tilde{a}(i)$. Fix $\vartheta > 0$ and define $\Phi_{\tilde{a}}(1+\vartheta) = \mathcal{M}_{\tilde{a}}(1+\vartheta)^{1/\vartheta}$. Then, the stochastic process

$$\mathrm{M}_{\tilde{a}}\left(m, n; \tilde{a}(i), \vartheta\right) = \left(\frac{\tilde{A}(m,n)}{\Phi_{\tilde{a}}\left(1+\vartheta\right)^{n-m}}\right)^{\vartheta} \tag{8}$$

is a martingale process, and it is also a super-martingale. ∎

Proof 3.1: See Appendix A.

Lemma 3.2 (Service-martingales): For the i.i.d. dynamic service provided by a wireless node h in the SNR domain, its accumulated service process can be represented as $\tilde{S}_h(m,n) = \prod_{i=m+1}^{n} \tilde{s}_h(i), n \geq m \geq 0$, where $\tilde{s}_h(i)$ represents the service at the i-th time slot of wireless node h. The Mellin transform of $\tilde{s}_h(i)$ is $\mathcal{M}_{\tilde{s}_h(i)}(1-\vartheta) = \mathbb{E}[(\tilde{s}_h(i))^{-\vartheta}]$. Due to $\tilde{s}_h(i)$ being i.i.d., we have $\mathcal{M}_{\tilde{s}_h}(1-\vartheta) = \mathcal{M}_{\tilde{s}_h(i)}(1-\vartheta)$ for all $\tilde{s}_h(i)$. Fix $\vartheta > 0$ and define $\Phi_{\tilde{s}_h}(1-\vartheta) = \mathcal{M}_{\tilde{s}_h}(1-\vartheta)^{-1/\vartheta}$. Then, the stochastic process

$$\mathrm{M}_{\tilde{s}_h}\left(m, n; \tilde{s}_h(i), \vartheta\right) = \left(\frac{\Phi_{\tilde{s}_h}(1-\vartheta)^{n-m}}{\tilde{S}_h(m,n)}\right)^{\vartheta} \tag{9}$$

is a martingale process, and it is also a super-martingale. ∎

Proof 3.2: See Appendix A.

B. E2E Delay Analysis in the SNR Domain

Theorem 3.1: Consider a MWL with H tandem wireless service nodes. Its arrival flow $\tilde{A}$ and the stochastic service $\tilde{S}_h, h = 1, 2, \ldots, H$, provided by each wireless node are independently and identically distributed random processes. Let $\mathrm{M}_{\tilde{a}}(l_1, n; \tilde{a}(i), \vartheta)$ and $\mathrm{M}_{\tilde{s}_h}(l_1, n; \tilde{s}_h(i), \vartheta)$ be the corresponding arrival-martingales and service-martingales. For any delay threshold $0 \leq w \leq n$, we have the E2E-DVPB as follows

$$\begin{aligned} &P\left\{\tilde{W}_{net}(n) \geq w\right\} \\ &\leq \sum_{h=1}^{H} \frac{\left(\vartheta^{\varepsilon} w \ln\left(\Phi_{\tilde{a}}(1+\vartheta^{\varepsilon})\right)\right)^{h-1}}{(h-1)!} e^{-\vartheta^{\varepsilon} w \ln\left(\Phi_{\tilde{a}}(1+\vartheta^{\varepsilon})\right)}, \end{aligned} \tag{10}$$

where $\tilde{W}_{net}(n)$ is the E2E delay of H-hop wireless links and ϑ^{ε} is the joint service descriptor. The martingale parameter ϑ^{ε} links the stochastic characteristics of arrival flow with the service process at each hop in the MWL, reflecting the joint service capability of H-hop wireless links for specific arrival flow in martingale domain. In a MWL, the mean arrival at each hop is greater than or equal to the mean arrival at the subsequent hop. Therefore, we choose $\Phi_{\tilde{a}}(1+\vartheta)$, which implicitly includes the first-hop mean arrival, to reflect the stochastic characteristics of arrival flow. Under the joint constraints of the martingale definition and queue stability conditions [19], we utilize ϑ^{ε} to establish a connection between the arrival process and the service process of MWL. It is essential that ϑ^{ε} meets the requirements of $\vartheta^{\varepsilon} := \sup\{\vartheta > 0 : \Phi_{\tilde{a}}(1+\vartheta) \leq \Phi_{\tilde{s}_h}(1-\vartheta), 1 \leq h \leq H\}$, where $\Phi_{\tilde{s}_h}(1-\vartheta)$ reflects the service capability of each node in the MWL and $\Phi_{\tilde{s}_1}(1-\vartheta) = \Phi_{\tilde{s}_2}(1-\vartheta) \cdots = \Phi_{\tilde{s}_H}(1-\vartheta)$. $\Phi_{\tilde{a}}(1+\vartheta)$ and $\Phi_{\tilde{s}_h}(1-\vartheta)$ are defined as shown in Lemma 3.1 and Lemma 3.2.

Proof 3.3: See Appendix B. ∎

Equation (10) is a closed-form expression for the E2E-DVPB. Its complexity is primarily determined by the number of nodes H, resulting in a computational complexity of $O(H)$.

IV. Simulation and Verification

In this section, we use matlab to simulate and verify the derived E2E-DVPB analysis method in Section III. In the simulation, the Rayleigh fading model is assumed for all fading channel models. Furthermore, we emphasize that our method can be applied to analysis of Rician fading channels or Nakagami-m fading channels [6]. The probability density function of SNR for Rayleigh fading at any wireless node h, $h \in \mathcal{H}$, in the tandem system can be represented as

$$f(\gamma_h) = \frac{1}{\bar{\gamma}_h} e^{-\frac{\gamma_h}{\bar{\gamma}_h}}, \gamma_h \geq 0, \tag{11}$$

where γ_h is the instantaneous SNR of wireless node h, and $\bar{\gamma}_h$ is average SNR of wireless node h. Therefore, the Mellin transform of the service of a Rayleigh fading channel in one time slot can be obtained as follows

$$\mathcal{M}_{\tilde{s}_h}\left(1-\vartheta\right) = e^{\frac{1}{\bar{\gamma}_h}} \bar{\gamma}_h^{-\vartheta T_b B/\ln 2} \Gamma\left(1 - \vartheta T_b B/\ln 2, \frac{1}{\bar{\gamma}_h}\right), \tag{12}$$

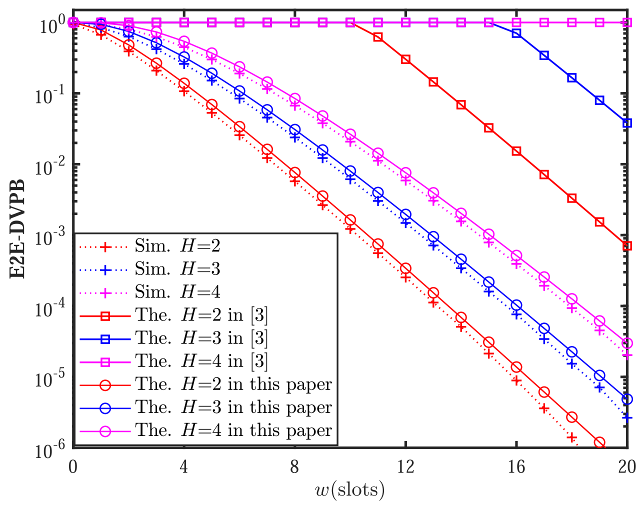

Fig. 2. Theoretical values and simulation results of E2E-DVPB under various hops H and delay thresholds w.

in which $\Gamma(1-\vartheta T_b B/\ln 2, \bar{\gamma}_h^{-1}) = \int_{\bar{\gamma}_h^{-1}}^{\infty} t^{-\vartheta T_b B/\ln 2} e^t dt$ is the upper incomplete Gamma function.

Proof 4.1: Based on (1) and (7), we can obtain

$$\begin{aligned} \mathcal{M}_{\tilde{s}_h}(1-\vartheta) &= \int_0^{+\infty} (1+\gamma_h)^{-\vartheta T_b B/\ln 2} \frac{1}{\bar{\gamma}_h} e^{-\frac{\gamma_h}{\bar{\gamma}_h}} d\gamma_h \\ &= e^{\frac{1}{\bar{\gamma}_h}} \bar{\gamma}_h^{-\vartheta T_b B/\ln 2} \int_{\frac{1}{\bar{\gamma}_h}}^{+\infty} t^{-\vartheta T_b B/\ln 2} e^t dt \\ &= e^{\frac{1}{\bar{\gamma}_h}} \bar{\gamma}_h^{-\vartheta T_b B/\ln 2} \Gamma\left(1-\vartheta T_b B/\ln 2, \frac{1}{\bar{\gamma}_h}\right), \end{aligned} \tag{13}$$

where variable substitution is applied for $t = (1+\gamma_h)/\bar{\gamma}_h$. ■

The arrival flow generated by the terminal equipment is modelled as a Bernoulli process [19]. The parameters we set satisfy the stability constraints of the tandem queue system, and some of these parameters are as follows: Bandwidth is 1 MHz [20], packet size is 500 bits [21], the number of time slots is 10^8, one time slot is 1 ms [7].

Firstly, we verify the superiority of the derived method for E2E-DVPB analysis. We compare the theoretical values obtained in (10) and [3] with the simulation results for different delay thresholds in 2-hop, 3-hop and 4-hop wireless links, as shown in Fig. 2. The SNC method derived in [3] is widely used in MWL delay analysis. The rate for the Bernoulli arrival model is $R = 3$ packets/slot with an arrival probability of $p = 0.8$. Each wireless node h service with $\bar{\gamma}_h = 5$ dB. As can be seen from Fig. 2, the theoretical values obtained from our derived (10) fit the simulation results very well and remain accurate even as the number of hops increases. In contrast, there is a large deviation between the theoretical values obtained by [3] and the simulation results, and the deviation increases with the number of hops. Comparing the two methods at different hop counts, the theoretical values obtained by (10) are several orders of magnitude more accurate than those obtained by [3]. This result demonstrates that our derived (10) is more suitable for the E2E-DVPB analysis of delay sensitive traffic.

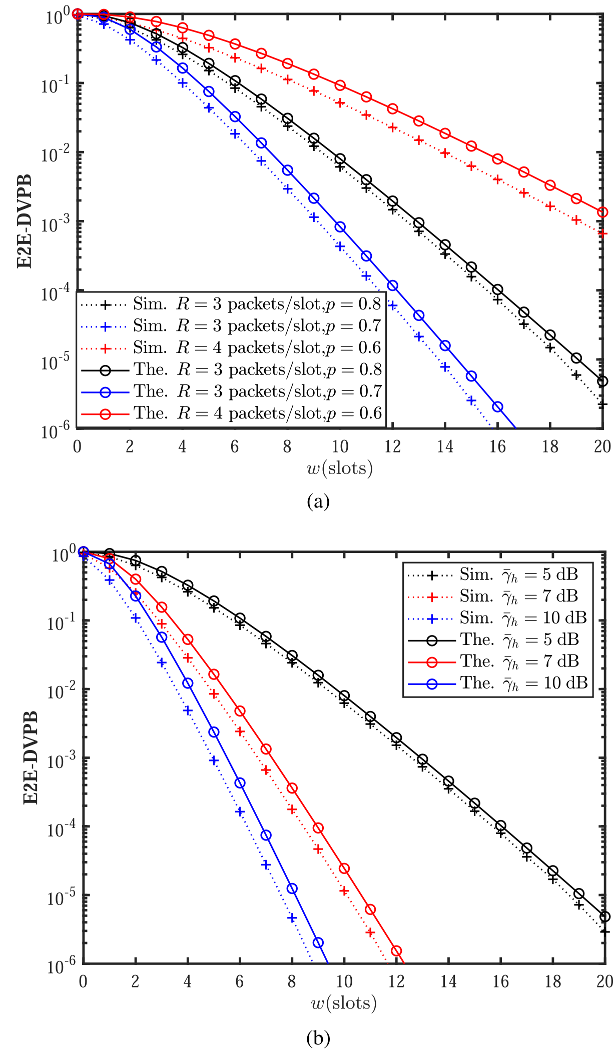

Fig. 3. Comparison of theoretical values and simulation results for 3-hop wireless links with different arrival and service parameters. (a) Fixing service parameters, varying arrival parameters. (b) Fixing arrival parameters, varying service parameters.

Further, we conduct two experiments as shown in Figs. 3(a) and (b). A 3-hop wireless links is considered in the experiments. In Fig. 3(a), we set each wireless node h with $\bar{\gamma}_h = 5$ dB and vary the arrival parameters. In Fig. 3(b), we fix the arrival parameters of $R = 3$ packets/slot and arrival probability of $p = 0.8$, and vary the average SNR of each wireless node. It can be observed that under different parameter settings, the E2E-DVPB derived from (10) align well with the simulation results, demonstrating the generality of our approach. Additionally, from the figures, we can also obtain the influence of model parameters on the E2E-DVPB derived from (10). In Fig. 3(a), it can be observed that the alignment between theoretical values and simulation results varies under different parameters. This is due to the varying levels of dispersion of arrival flows caused by different arrival parameters [22]. For example, when comparing the black and blue curves, it can be observed that, as the arrival probability p decreases, the fit between the simulation results and the theoretical values slightly decreases. This is attributed to the increased dispersion of arrival flow, leading to a reduction in the accuracy of theoretical modeling of the arrival process. Ultimately, this reduction in accuracy contributes to the slightly looser E2E-DVPB estimated in (10). By comparing the red and black curves, we can also find that the dispersion has an important impact on the guarantee of E2E delay. When the average arrivals are the same, a larger service rates are required for arrivals with high dispersion to guarantee the same QoS requirements. Similarly, in Fig. 3(b), we observe that with the increase in average SNR $\bar{\gamma}_h$, the fitting between theoretical values and simulation results becomes slightly looser. This is attributed to the increased dispersion of the service process at each node as the average SNR increases.

V. Conclusion

In this paper, we introduce a martingale-based delay analysis method for assessing the E2E latency of MWL. In the SNR domain, corresponding martingale processes are constructed for the arrival and service processes, and joint service descriptors are introduced to link them. Subsequently, the E2E delay is decomposed and approximated using the delay of each node in the MWL. Based on this, we derive a closed-form expression of the E2E-DVPB for MWL. Through simulation results, we validate that the E2E-DVPB obtained using our proposed method are more accurate than those in existing literature. In addition, our proposed method contributes to a better optimisation of resource allocation, which improves the QoS of the wireless network. In the future, we will consider applying the proposed analytical method to UAV communication scenarios.

Appendix A
Proof of Lemma 3.1 and Lemma 3.2

In order to prove that $\mathrm{M}_{\tilde{a}}(m,n;\tilde{a}(i),\vartheta)$ is a martingale process, it needs to satisfy the definition of martingale. The expectation for arrival flow is limited, therefore, the remaining condition needs to be satisfied. We can derive

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{M}_{\tilde{a}}\left(m,n+1;\tilde{a}(i),\vartheta\right) \mid \tilde{a}(1),\tilde{a}(2),\ldots,\tilde{a}(n)\right]\\
&\overset{a}{=}\mathbb{E}\left[\left(\frac{\tilde{A}(m,n)}{\Phi_{\tilde{a}}\left(1+\vartheta\right)^{n-m}}\right)^{\vartheta}\cdot\left(\frac{\tilde{a}(n+1)}{\Phi_{\tilde{a}}\left(1+\vartheta\right)}\right)^{\vartheta}\right]\\
&\overset{b}{=}\left(\frac{\tilde{A}(m,n)}{\Phi_{\tilde{a}}\left(1+\vartheta\right)^{n-m}}\right)^{\vartheta}\cdot\frac{\mathbb{E}\left[\left(\tilde{a}(n+1)\right)^{\vartheta}\right]}{\mathcal{M}_{\tilde{a}}\left(1+\vartheta\right)}\\
&\overset{c}{=}\mathrm{M}_{\tilde{a}}\left(m,n;\tilde{a}(i),\vartheta\right),
\end{aligned}\tag{14}$$

where we use $\mathbb{E}[(\tilde{a}(n+1))^{\vartheta}]/\mathcal{M}_{\tilde{a}}(1+\vartheta)=1$ and the property of i.i.d. can get step c from step b.

The proofs of the service-martingales and arrival-martingales are similar. Based on the assumption that service is i.i.d., for $\vartheta>0$, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{M}_{\tilde{s}_h}\left(m,n+1;\tilde{s}_h(i),\vartheta\right) \mid \tilde{s}_h(1),\tilde{s}_h(2),\ldots,\tilde{s}_h(n)\right]\\
&=\mathbb{E}\left[\left(\frac{\Phi_{\tilde{s}_h}(1-\vartheta)^{n-m}}{\tilde{S}_h(m,n)}\right)^{\vartheta}\cdot\left(\frac{\Phi_{\tilde{s}_h}\left(1-\vartheta\right)}{\tilde{s}_h(n+1)}\right)^{\vartheta}\right]\\
&=\left(\frac{\Phi_{\tilde{s}_h}(1-\vartheta)^{n-m}}{\tilde{S}_h(m,n)}\right)^{\vartheta}\cdot\mathbb{E}\left[\left(\tilde{s}_h(n+1)\right)^{-\vartheta}\right]\mathcal{M}_{\tilde{s}_h}\left(1-\vartheta\right)^{-1}\\
&=\mathrm{M}_{\tilde{s}_h}\left(m,n;\tilde{s}_h(n+1),\vartheta\right),
\end{aligned}\tag{15}$$

where $\mathcal{M}_{\tilde{s}_h}(1-\vartheta)^{-1}\cdot\mathbb{E}[(\tilde{s}_h(n+1))^{-\vartheta}]=1$.

Appendix B
Proof of Theorem 3.1

The necessary knowledge for the proof is provided initially. Inspired by (2), firstly, we consider that each service node in a tandem queuing system has an impact on the delay. Based on this, the delay threshold w of a system constituted by H tandem service nodes can be expressed as the sum of the delay thresholds of each node. We can obtain that

$$w=\sum_{h=1}^{H}w_h,\tag{16}$$

where w_h represents the delay threshold allocated to node h. According to (2), each delay threshold w_h needs to satisfy the following conditions $0\le w_h\le l_{h+1}-l_h$, where $n=l_{H+1}$ is specified. The natural logarithm function f is a continuous nondecreasing and injective function in its domain of definition. For stochastic sequences $\{X(n);n\ge 0\}$, we have the following conclusion

$$f\left(\sup_{0\le t\le n}X(t)\right)=\sup_{0\le t\le n}f(X(t)),\tag{17}$$

where $X(t)>0$, $0\le t\le n$.

Proof B.1: Based on the properties of the natural logarithm function f, we can first derive

$$\begin{aligned}
&X(s)\le\sup(X(t)),\quad 0\le t\le n, 0\le s\le n\\
&\Rightarrow f(X(s))\le f(\sup(X(t))),\quad 0\le t\le n, 0\le s\le n\\
&\Rightarrow \sup_{0\le t\le n}\left(f(X(t))\right)\le f\left(\sup_{0\le t\le n}\left(X(t)\right)\right).
\end{aligned}\tag{18}$$

By using (18), we can further derive

$$\begin{aligned}
&f(X(s))\le\sup(f(X(t))),\quad 0\le t\le n, 0\le s\le n\\
&\Rightarrow X(s)\le f^{-1}(\sup(f(X(t))),\quad 0\le t\le n, 0\le s\le n\\
&\Rightarrow \sup_{0\le t\le n}\left(X(t)\right)\le f^{-1}\left(\sup_{0\le t\le n}\left(f(X(t))\right)\right.\\
&\Rightarrow f\left(\sup_{0\le t\le n}\left(X(t)\right)\right)\le\sup_{0\le t\le n}\left(f(X(t)\right).
\end{aligned}\tag{19}$$

Combining (18) and (19), we can obtain (17). ■

The E2E-DVPB that we derived is shown in (22) shown at the bottom of the next page. In the following, we explain the derivation process in detail. According to (5), step a and step b

are obtained. We construct step c by using the arrival-martingales in Lemma 3.1, the service-martingales in Lemma 3.2 and (2). Since MWL provides homogeneous services, we can get

$$\Phi_{\tilde{s}_1}(1-\vartheta) = \Phi_{\tilde{s}_2}(1-\vartheta) \cdots = \Phi_{\tilde{s}_h}(1-\vartheta), \tag{20}$$

where $\Phi_{\tilde{s}_h}(1-\vartheta) = \mathcal{M}_{\tilde{s}_h}(1-\vartheta)^{-1/\vartheta}$, $1 \le h \le H$, and $\mathcal{M}_{\tilde{s}_h}(1-\vartheta)$ is the Mellin transform of the instantaneous service of node h. In order to describe the joint service capability of the H-hop wireless links for the arrival flow in step c, we propose the joint service descriptor, which needs to satisfy $\vartheta^\varepsilon := \sup\{\vartheta > 0 : \Phi_{\tilde{a}}(1+\vartheta) \le \Phi_{\tilde{s}_h}(1-\vartheta), 1 \le h \le H\}$. For ease to understand, we use $\Phi_{\tilde{s}}(1-\vartheta^\varepsilon)$ to replace $\Phi_{\tilde{s}_h}(1-\vartheta^\varepsilon)$ in subsequent content. Based on (16), we can obtain step d. Based on (17), we can get step e from step d. To derive step f, we use the inequality as follows [23]

$$\sup_{0<t\le n}\{Y(t)+Z(t)\} \le \sup_{0<t\le n} Y(t) + \sup_{0<t\le n} Z(t). \tag{21}$$

Through the aforementioned derivation process, the complex E2E delay analysis is decoupled into the delay analysis of each node in the H-hop wireless links. Combining step f and (17), the DVPB for any one node in the tandem queuing system can be obtained as shown in (23) shown at the bottom of this page. The

$$\left(\frac{\tilde{A}\left(l_h, l_{h+1}-w_h\right)}{\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{l_{h+1}-l_h-w_h}} \cdot \frac{\Phi_{\tilde{s}}(1-\vartheta^\varepsilon)^{l_{h+1}-l_h}}{\tilde{S}_h\left(l_h, l_{h+1}\right)}\right)^{\vartheta^\varepsilon}$$

in (23) is a super-martingale. Based on Doob's inequality, e.g.,

$$P\left\{\sup_{0\le t\le n}|\mathcal{X}(t)| > x\right\} \le \frac{\mathbb{E}\left[|\mathcal{X}(0)|\right]}{x},$$

we can obtain the final result of (23). It can be concluded from (23) that the DVPB of each node in the tandem system obeys an exponential distribution of $\lambda = 1$. Thus, we can obtain that the sum of all exponentially distributed random variables with the same parameter λ on the left-hand side of the inequality sign in step f obeys the Erlang distribution [19]. Based on this conclusion, we can derive step g from step f.

$$\begin{aligned}
&P\left\{\tilde{W}_{net}(n) \ge w\right\}\\
&\overset{a}{=} P\left\{\tilde{A}(w,n) \ge \tilde{D}_H(0,n)\right\}\\
&\overset{b}{\le} P\left\{\sup_{0\le l_1\le n-w}\left[\frac{\tilde{A}\left(l_1, n-w\right)}{\tilde{S}_{net}\left(l_1, n\right)}\right] \ge 1\right\}\\
&\overset{c}{\le} P\left\{\sup_{0\le l_1\le\cdots\le l_H\le n}\left[\left(\frac{\tilde{A}\left(l_1, n-w\right)}{\prod_{h=1}^{H}\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{l_{h+1}-l_h}\cdot\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{-w}}\cdot\frac{\prod_{h=1}^{H}\Phi_{\tilde{s}_h}(1-\vartheta^\varepsilon)^{l_{h+1}-l_h}}{\prod_{h=1}^{H}\tilde{S}_h(l_h, l_{h+1})}\right)^{\vartheta^\varepsilon}\right] \ge \Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{\vartheta^\varepsilon w}\right\}\\
&\overset{d}{\le} P\left\{\sup_{0\le l_1\le l_2-w_1\le\cdots\le l_{H+1}-w_H\le n}\left[\left(\prod_{h=1}^{H}\frac{\tilde{A}\left(l_h, l_{h+1}-w_h\right)}{\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{l_{h+1}-l_h-w_h}}\cdot\frac{\Phi_{\tilde{s}}(1-\vartheta^\varepsilon)^{l_{h+1}-l_h}}{\tilde{S}_h\left(l_h, l_{h+1}\right)}\right)^{\vartheta^\varepsilon}\right] \ge \Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{\vartheta^\varepsilon w}\right\}\\
&\overset{e}{\le} P\left\{\sup_{0\le l_1\le l_2-w_1\le\cdots\le l_{H+1}-w_H\le n}\left[\ln\left(\left(\prod_{h=1}^{H}\frac{\tilde{A}\left(l_h, l_{h+1}-w_h\right)}{\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{l_{h+1}-l_h-w_h}}\cdot\frac{\Phi_{\tilde{s}}(1-\vartheta^\varepsilon)^{l_{h+1}-l_h}}{\tilde{S}_h\left(l_h, l_{h+1}\right)}\right)^{\vartheta^\varepsilon}\right)\right] \ge \ln\left(\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{\vartheta^\varepsilon w}\right)\right\}\\
&\overset{f}{\le} P\left\{\sum_{h=1}^{H}\sup_{0\le l_h\le l_{h+1}-w_h\le n}\left[\ln\left(\left(\frac{\tilde{A}\left(l_h, l_{h+1}-w_h\right)}{\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{l_{h+1}-l_h-w_h}}\cdot\frac{\Phi_{\tilde{s}}(1-\vartheta^\varepsilon)^{l_{h+1}-l_h}}{\tilde{S}_h\left(l_h, l_{h+1}\right)}\right)^{\vartheta^\varepsilon}\right)\right] \ge \vartheta^\varepsilon w\ln\left(\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)\right)\right\}\\
&\overset{g}{\le} \sum_{h=1}^{H}\frac{\left(\vartheta^\varepsilon w\ln\left(\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)\right)\right)^{h-1}}{(h-1)!}e^{-\vartheta^\varepsilon w\ln(\Phi_{\tilde{a}}(1+\vartheta^\varepsilon))}.
\end{aligned} \tag{22}$$

$$\begin{aligned}
&P\left\{\sup_{0\le l_h\le l_{h+1}-w_h\le n}\left[\ln\left(\left(\frac{\tilde{A}\left(l_h, l_{h+1}-w_h\right)}{\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{l_{h+1}-l_h-w_h}}\cdot\frac{\Phi_{\tilde{s}}(1-\vartheta^\varepsilon)^{l_{h+1}-l_h}}{\tilde{S}_h\left(l_h, l_{h+1}\right)}\right)^{\vartheta^\varepsilon}\right)\right] \ge \vartheta^\varepsilon w_h\ln\left(\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)\right)\right\}\\
&= P\left\{\sup_{0\le l_h\le l_{h+1}-w_h\le n}\left[\left(\frac{\tilde{A}\left(l_h, l_{h+1}-w_h\right)}{\Phi_{\tilde{a}}(1+\vartheta^\varepsilon)^{l_{h+1}-l_h-w_h}}\cdot\frac{\Phi_{\tilde{s}}(1-\vartheta^\varepsilon)^{l_{h+1}-l_h}}{\tilde{S}_h\left(l_h, l_{h+1}\right)}\right)^{\vartheta^\varepsilon}\right] \ge e^{\vartheta^\varepsilon w_h\ln(\Phi_{\tilde{a}}(1+\vartheta^\varepsilon))}\right\}\\
&\le e^{-\vartheta^\varepsilon w_h\ln(\Phi_{\tilde{a}}(1+\vartheta^\varepsilon))}.
\end{aligned} \tag{23}$$

REFERENCES

[1] D. Zhou, M. Sheng, J. Li, and Z. Han, "Aerospace integrated networks innovation for empowering 6G: A survey and future challenges," *IEEE Commun. Surv. Tut.*, vol. 25, no. 2, pp. 975–1019, Secondquarter 2023.

[2] J. Lou, X. Yuan, P. Sigdel, X. Qin, S. Kompella, and N.-F. Tzeng, "Age of information optimization in multi-channel based multi-hop wireless networks," *IEEE Trans. Mobile Comput.*, vol. 22, no. 10, pp. 5719–5732, Oct. 2023.

[3] H. Al-Zubaidy, J. Liebeherr, and A. Burchard, "Network-layer performance analysis of multihop fading channels," *IEEE/ACM Trans. Netw.*, vol. 24, no. 1, pp. 204–217, Feb. 2016.

[4] G. Yang, M. Xiao, and H. V. Poor, "Low-latency millimeter-wave communications: Traffic dispersion or network densification?," *IEEE Trans. Commun.*, vol. 66, no. 8, pp. 3526–3539, Aug. 2018.

[5] P. Cui, S. Han, X. Xu, J. Zhang, P. Zhang, and S. Ren, "End-to-end delay performance analysis of industrial Internet of Things: A stochastic network calculus perspective," *IEEE Internet Things J.*, vol. 11, no. 3, pp. 5374–5387, Feb. 2024.

[6] C. Xiao, J. Zeng, W. Ni, R. P. Liu, X. Su, and J. Wang, "Delay guarantee and effective capacity of downlink NOMA fading channels," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 508–523, Jun. 2019.

[7] J. Zeng, C. Xiao, T. Wu, W. Ni, R. P. Liu, and Y. J. Guo, "Uplink non-orthogonal multiple access with statistical delay requirement: Effective capacity, power allocation, and α fairness," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1298–1313, Feb. 2023.

[8] C. Jing, X. Zhu, and X. Liu, "Performance analysis model and deterministic routing decision algorithm for broadband real-time services in wireless multi-hop networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 12113–12123, Sep. 2023.

[9] M. Mei, M. Yao, Q. Yang, M. Qin, K. S. Kwak, and R. R. Rao, "Delay analysis of mobile edge computing using poisson cluster process modeling: A stochastic network calculus perspective," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2532–2546, Apr. 2022.

[10] T. Liu, H. Zhou, J. Li, F. Shu, and Z. Han, "Uplink and downlink decoupled 5G/B5G vehicular networks: A federated learning assisted client selection method," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2280–2292, Feb. 2023.

[11] F. Poloczek and F. Ciucu, "Scheduling analysis with martingales," *Perform. Eval.*, vol. 79, pp. 56–72, Sep. 2014.

[12] B. Picano and R. Fantacci, "A combined stochastic network calculus and matching theory approach for computational offloading in a heterogenous MEC environment," *IEEE Trans. Netw. Serv. Manage.*, vol. 21, no. 2, pp. 1958–1968, Apr. 2024.

[13] H. Peng, C.-C. Hsia, Z. Han, and L.-C. Wang, "A generalized delay and backlog analysis for multiplexing URLLC and eMBB: Reconfigurable intelligent surfaces or decode-and-forward?," *IEEE Trans. Wireless Commun.*, early access, Sep. 18, 2023, doi: 10.1109/TWC.2023.3313964.

[14] R. Fantacci and B. Picano, "End-to-end delay bound for wireless uVR services over 6G terahertz communications," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 17090–17099, Dec. 2021.

[15] Y. Hu, H. Li, Z. Chang, and Z. Han, "End-to-end backlog and delay bound analysis for multi-hop vehicular ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6808–6821, Oct. 2017.

[16] R. McEliece and W. Stark, "Channels with block interference," *IEEE Trans. Inf. Theory.*, vol. 30, no. 1, pp. 44–53, Jan. 1984.

[17] 3GPP "Technical specification group radio access network; NR; physical layer procedures for control," 3GPP, Sophia Antipolis, France, Tech. Specification 38. 213, 2022, version 17.4.0. [Online]. Available: https://www.3gpp.org/ftp/specs/archive/38_series/38.213

[18] W. Yang, X. Chi, L. Zhao, and R. Qi, "Predictive two-timescale resource allocation for VoD services in fast moving scenarios," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10002–10017, Oct. 2021.

[19] J. F. Shortle, J. M. Thompson, D. Gross, and C. M. Harris, *Fundamentals of Queueing Theory*. Hoboken, NJ, USA: Wiley, 2018, vol. 399.

[20] H. S. Dhillon, H. Huang, and H. Viswanathan, "Wide-area wireless communication challenges for the Internet of Things," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 168–174, Feb. 2017.

[21] S. Peng, T. Wang, and C. P. Low, "Energy neutral clustering for energy harvesting wireless sensors networks," *Ad Hoc Netw.*, vol. 28, pp. 1–16, Jan. 2015.

[22] D. D. Boos and C. Brownie, "Comparing variances and other measures of dispersion," *Stat. Sci.*, vol. 19, no. 4, pp. 571–578, Nov. 2004.

[23] Y. Jiang and Y. Liu, *Stochastic Network Calculus*. Berlin, Germany: Springer-Verlag, 2008.

Hangyu Yan received the B.S. degree from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2017, and the M.S. degree from Southwest University, Chongqing, China, in 2020. He is currently working toward the Ph.D. degree with the College of Communication Engineering, Jilin University, Changchun, China. His research interests include wireless network QoS analysis, queuing theory, and resource allocation.

Xuefen Chi received the B.Eng. degree in applied physics from the Beijing University of Posts and Telecommunications, Beijing, China, in 1984, and the M.S. and Ph.D. degrees from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, in 1990 and 2003, respectively. She was a Visiting Scholar with the Department of Computer Science, Loughborough University, Loughborough, U.K., in 2007, and the School of Electronics and Computer Science, University of Southampton, Southampton, U.K., in 2015. She is currently a Professor with the Department of Communications Engineering, Jilin University, Changchun. Her research interests include machine type communications, indoor visible light communications, random access algorithms, delay-QoS guarantees, and network modeling theory and its applications.

Wanting Yang received the B.S. and Ph.D. degrees from the Department of Communications Engineering, Jilin University, Changchun, China, in 2018 and 2023, respectively. She is currently a Research Fellow with the Singapore University of Technology and Design, Singapore. Her research interests include wireless semantic communication, generative artificial intelligence, learning, martingale, and predictive resource allocation. She was a Technical Program Committee Member for flagship conferences, such as WCNC, GLOBECOM, and VTC.

Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively. From 2000 to 2002, he was an R&D Engineer with JDSU, Germantown, MD. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, Boise, ID, USA. He is currently a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department as well as with the Computer Science Department, University of Houston, Houston, TX, USA. His main research focuses on the novel game-theory related concepts critical to enabling efficient and distributive use of wireless networks with limited resources. His other research interests include wireless resource allocation and management, wireless communications and networking, quantum computing, data science, smart grid, carbon neutralization, security and privacy. Dr. Han was the recipient of the NSF Career Award in 2010, Fred W. Ellersick Prize of the IEEE Communication Society in 2011, EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (best paper award in IEEE JSAC) in 2016, IEEE Vehicular Technology Society 2022 Best Land Transportation Paper Award, and several best paper awards in IEEE conferences. He was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018 and ACM Distinguished Speaker from 2022 to 2025, AAAS Fellow since 2019, and ACM Fellow since 2024. He is a 1% highly cited Researcher since 2017 according to Web of Science. He was also the recipient of the 2021 IEEE Kiyo Tomiyasu Award (an IEEE Field Award), for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks."