Quality-of-service performance bounds in wireless multi-hop relaying networks

K. Zheng\(^1\)  L. Lei\(^2\)  Y. Wang\(^1\)  Y. Lin\(^1\)  W. Wang\(^1\)

\(^1\)Wireless Signal Processing and Network Lab, Key laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts & Telecommunications, P.O. Box 93, No. 10, Xi Tu Cheng Road, Beijing 100088, People’s Republic of China

\(^2\)Department of Computer Science and Technology, Tsinghua University, Beijing 100084, People’s Republic of China

E-mail: zkan@bupt.edu.cn

Abstract: The theoretical analysis on quality-of-service (QoS) performances is required to provide the guides for the developments of the next-generation wireless networks. As a good analysis tool, the probabilistic network calculus with moment generating functions (MGFs) recently can be used for delay and backlog performance measures in wireless networks. Different from the existed studies which mostly focused on the single-hop networks with single-user under a two state Markov channel model, this study develops an analytical framework for wireless multi-hop relaying networks under the finite-state Markov channel by using probabilistic network calculus with MGFs. By using the concatenation character of network calculus, the authors regard a two-hop wireless relaying channel as a single server equivalently, which consisting of two dynamic servers in series. When the single-user model is straightforwardly extended and applied in multi-user scenarios, the state space of service process is increased exponentially with the number of users, which is only applicable in case of very small user number. Then, in order to avoid the limitation of user number, the authors propose to reflect the multi-user effects by using the equivalent data rate of the modified service process, whose transition and stationary probabilities are kept unchanged with those in single-user scenarios. Next, delay and backlog bounds of multi-hop wireless relaying networks are derived with the proposed analytical framework. Simulation results show that analytical bounds match simulation results, whose accuracy depends on the required violation probability. The effectiveness of the relaying techniques in improving the performances is also demonstrated.

1 Introduction

Multi-hop cellular architecture with relaying is a cost-effective solution to support a wide range of services and applications in next-generation wireless networks. It can reduce the transmission distance and increase the amount of users under more favourable channel conditions, allowing for better channel quality and higher throughput. Standardisation efforts of integrating multi-hop relaying technologies to future cellular networks have commenced recently not only in the evolutorial mobile Worldwide Inter-operability for Microwave Access (WiMAX) networks but also in the third-generation (3G) long-term evolution advancements (LTE-A) networks [1, 2].

Although the wireless network technology has moved on, the analysis tools have not well matched new technological developments. It is expected to have the quality-of-service (QoS) analysis results in wireless multi-hop relaying networks [3, 4]. As a theory of deterministic queuing systems, network calculus has been widely used for performance measures in wired networks [5, 6]. It can be viewed as a system theory based on the min-plus algebra, which facilitates the efficient derivation of worst-case backlog and delay bounds. Originally developed for deterministic queuing systems, this theory has recently evolved as well to consider the statistical service guarantee. In recent years, many attempts have been made to develop a probabilistic network calculus. A probabilistic calculus
was derived in [7] and later used in [8, 9]. Meanwhile, the theory of effective bandwidths has been investigated, which provides a rich variety of probabilistic traffic models based on moment generating functions (MGF) [10, 11]. So, a probabilistic network calculus with MGFs was proposed in [7] and further elaborated in [12]. These probabilistic bounds are more reasonable for the applications in which the deployment of worst-case bounds causes inefficient usage of the resources such as the traffic over the wireless link.

However, there are few attempts on studying the wireless networks within the context of the network calculus. The wireless link model in the context of network calculus was first presented with an on–off impairment model in [13]. In [14], a modular dynamic approach was proposed to effectively model the behaviour of a two-state Markov wireless channel. Also, a methodology based on probabilistic network calculus with MGFs was used for performance analysis of fading channels and only analysis on the cases of a two-state Markov wireless channel was given in [15].

To the best of our knowledge, there is no lecture to analyse the QoS performance bounds of multi-hop relaying networks by using probabilistic network calculus with MGFs so far. Different from the existed studies which mostly focused on the single-hop networks with single-user under a two-state Markov channel model, an analytical framework is developed for multi-hop relaying networks under the finite-state Markov channel (FSMC) in this paper. Since only one relay station (RS) is assumed, the wireless channel in the first hop transmission can be regarded as a single-user server. More than one mobile users (MSs) are served by an RS in the second hop channel, which can be modelled as a multi-user server. Our main contributions in this paper are as follows:

1. By applying the concatenation character of network calculus, a multi-hop wireless relaying channel is equivalent to a single dynamic server, consisting of more than one individual dynamic server in series.

2. On the assumption of maximising signal-to-noise-ratio (Max-SNR) scheduling algorithm for multi-user transmission in the second hop transmission, a simplified method is proposed to avoid the exponentially enlarged state space, where the multi-user effects are only reflected by the equivalent data rate of the modified service process.

3. With the given arrival curve, we derive the analytical delay and backlog bounds of multi-hop wireless relaying networks and validate them by simulations.

The rest of this paper is organised as follows. Section 2 gives the brief description of a system model. The performances of delay and backlog in multi-hop relaying networks are analysed in Section 3. In Section 4, numerical and simulation results are presented and discussed. Finally, Section 5 gives the conclusion.

## 2 System model

### 2.1 MGF stochastic network calculus basics

As a conventional system theory, network calculus lies in the min–plus algebra. For good description, we briefly present the definitions in MGF stochastic network calculus. For a detailed discussion, please see [12]. The operations of infimum (‘inf’) and supremum (‘sup’) are similar to those of minimum (‘min’) and maximum (‘max’). The only difference is that the formers do not have to be in the set.

**Definition 1 (min–plus convolution and de-convolution):** The min-plus convolution and de-convolution of real-valued, bivariate function \( f(s, t) \) and \( g(s, t) \) are defined for \( t \geq s \geq 0 \) as

\[
(f \circ g)(s, t) = \inf_{t \geq \tau \geq s} \{ f(s, \tau) + g(\tau, t) \}
\]

\[
(f \circ g)(s, t) = \sup_{t \geq \tau \geq s} \{ f(\tau, \tau) - g(\tau, t) \}
\]

The discrete time model with \( t \in \mathbb{N}_0 = \{0, 1, \ldots \} \) is assumed in this paper. Let us define real-valued cumulative functions \( A(0, t) \) and \( D(0, t) \) as arrival and departure processes, respectively, which represent the amount of data observed in the interval \( (0, t) \). Denote that \( A(s, t) = A(0, t) - A(0, s) \) and \( D(s, t) = D(0, t) - D(0, s) \). In this paper, we assume that all processes are non-negative, increasing in \( t \), and decreasing in \( s \), and have zero value in case of \( s = t \), that is \( \forall t, A(t, t) = 0 \) and \( D(t, t) = 0 \).

**Definition 2 (dynamic server):** Assume \( A(0, t) \) and \( D(0, t) \) are the arrival and departure process of a lossless server, respectively. Let \( S(s, t) \) for \( t \geq s \geq 0 \) be a random process that is non-negative and increasing in \( t \). The server is called a dynamic server \( S(s, t) \) if for any fixed sample path it holds for all \( t \geq 0 \) that \( D(0, t) \geq A(0, t) \).

**Definition 3 (MGF):** The arrivals and services are assumed to be statistically independent, stationary random processes throughout this paper. Then, the distribution of \( A(s, s + t) \) equals that of \( A(0, t) \) for all \( s > 0 \) because of stationarity. The MGF of stationary arrival and service processes of the wireless transmission are defined as

\[
M_A(\theta, t) = E[e^{A(0, t)\theta}]
\]

\[
M_S(\theta, t) = E[e^{S(0, t)\theta}]
\]

where \( E[x] \) denotes the expectation of the random process \( x \). For notation, we define \( M_A(\theta, t) = M_S(-\theta, t) \).
2.2 Two-hop wireless network

For sake of analysis, we consider a simple two-hop downlink wireless communication network as shown in Fig. 1, where a relay station (RS) cooperates with a base station (BS) to transmit information to mobile stations (MSs).

The discrete time model with $t \in \mathbb{N}_0 = \{0, 1, \ldots\}$ is assumed in this paper. Let us define real-valued cumulative functions $A_i(0, t)$ and $D_i(0, t)$, $i \in \{1, 2\}$ as arrival and departure processes of the $i$th hop transmission, respectively. The transmission of each hop in wireless relaying networks can be regarded as a dynamic server, which has probabilistic service curve $S_i(0, t)$, $i \in \{1, 2\}$. For all $t \geq 0$, it holds that

$$D_i(0, t) \geq \inf \{A_i(0, t) + S_i(t, t), \ i \in \{1, 2\}\}$$

(3)

The MGF of stationary arrival and service processes of each hop can be calculated by (2).

3 Backlog and delay bounds analysis

The two-hop wireless relaying network is equivalent to a single dynamic server $S(t, t)$, which consists of two dynamic servers $S_1(s, t)$ and $S_2(s, t)$ in series. Then, according to concatenation characteristic of network calculus in [12], the service process of this single equivalent server is expressed by

$$S(s, t) = (S_1 \otimes S_2)(s, t)$$

(4)

The corresponding MGF is upper bounded for $t > 0$ and all

$$\theta$$

according to

$$\mathcal{M}_S(\theta, t) \leq \sum_{r=0}^{t} \mathcal{M}_{S_1}(\theta, r) \mathcal{M}_{S_2}(\theta, t - r)$$

(5)

The total delay of the network includes the delay in the buffer and the time from when a packet leaves the buffer of the transmitter until it is decoded at the receiver. Usually, the second component of delay is number of frames, where accounts for the propagation delay and processing time. Since this quantity is fixed for each transmission, here we ignore this constant factor and only focus on the average delay in the buffer.

An upper backlog and delay bound on the assumption of first-in first-out (FIFO) scheduling in wireless relaying networks with violation probability $\epsilon \in (0, 1]$ are given by [12]

$$b = \inf_{\theta > 0} \left[ \frac{1}{\theta} \ln \sum_{j=0}^{\infty} \mathcal{M}_{A_1}(\theta, j) \mathcal{M}_{A_2}(\theta, s) - \ln \epsilon \right]$$

(6)

$$d = \inf_{\theta > 0} \left\{ \inf_{r > 0} \left[ \frac{1}{\theta} \ln \sum_{j=0}^{\infty} \mathcal{M}_{A_1}(\theta, j) \mathcal{M}_{A_2}(\theta, r) - \ln \epsilon \right] \leq 0 \right\}$$

(7)

These infinite sums can be calculated if bounds on the MGF of the arrival and service processes are known. Since the MGFs of different traffic models are widely studied in [7, 12 and reference therein], the arrival process is not our interest here. Therefore the main challenge is to obtain the probabilistic service curves that describe wireless fading channel and its effects on QoS performance bounds in wireless relaying networks. Respective models are not well investigated. We reuse the Markov models like [7, 12] to describe the random process $S(t, t)$ that denotes the service offered by single equivalent dynamic server of wireless relaying networks in the interval $(s, t]$.

3.1 Scenario 1: single RS/single MS

We first consider the simplest case where there is only single RS and single MS in wireless relaying networks. Then, the wireless channel in each hop can be regarded as a single-user server. The multi-rate transmission is achieved by adaptive modulation and coding (AMC) scheme where the transmitted data rate can be adjusted depending on the channel condition.

Assume that there are $L$ modulation and coding schemes (MCSs) available for AMC transmission in each hop. Correspondingly, an FSMC $S = \{s_1^{(0)}, s_2^{(0)}, \ldots, s_L^{(0)}\}$ with $L$ states is used to reflect the fading channel in the $i$th hop transmission. The data rate of the $i$th hop that can be processed in state $s_j$ is $b_j^{(0)}$. Then, the rate set,
corresponding to $L$-state of this channel, is denoted as $V = \{b_0^{(i)}, b_1^{(i)}, \ldots, b_{\bar{V}}^{(i)}\}$. Usually the rate of the worst channel state is set to be zero for reliable transmission, that is, $b_0^{(i)} = 0$. For the $i$th hop transmission, we denote the state transition probability from $s_l^{(i)}$ to $s_m^{(i)}$ and steady-state probability $s_l^{(i)}$ as $p_{l,m}^{(i)}$ and $\pi_l^{(i)}$, respectively. Furthermore, the transitions are assumed to be only happened between adjacent states, that is

$$p_{l,m}^{(i)} = 0, \quad \forall |l-m| > 1, \quad l, m \in \{1, 2, \ldots, L\}$$

(8)

The entire received SNR region is partitioned into $L$ parts by region boundary of $\{\Gamma_l\}_{l=1}^{L+1}$, which are not only for the MCS selection but also taken as the SNR threshold value between the $\ell$th and $(\ell+1)$th states of the FSMC channel. Since the radio channels of two hops are assumed to be independent and statistically identical, the same set of threshold can be used for both hops and the index $i$ of the hop is omitted. These threshold values are in the increasing order with the packet is located in the range $[\Gamma_l, \Gamma_{l+1})$.

The adjacent-state transition probability over one frame $T_p$ can be calculated as [16]

$$p_{l,l+1}^{(i)} = \frac{\chi(\Gamma_{l+1})T_p^l}{\pi_l}, \quad l = 1, \ldots, L - 1$$

(9)

$$p_{l,l-1}^{(i)} = \frac{\chi(\Gamma_l)T_p^l}{\pi_l}, \quad l = 2, \ldots, L$$

(10)

Here, $\chi(\Gamma_l)$ denotes the level cross rate at an instantaneous SNR for one user in the $i$th hop transmission. Since usually the wireless channel is assumed to Rayleigh fading, $\chi(\Gamma_l)$ can be expressed by

$$\chi(\Gamma_l) = \sqrt{\frac{2\pi f_D^l}{\gamma(\gamma)^l}} \exp\left(-\Gamma_l/\gamma(\gamma)^l\right), \quad i \in \{1, 2\}$$

(11)

where $f_D^l$ denotes the mobility-induced Doppler spread in the channel of the $i$th hop, and $\gamma(\gamma)^l = \mathbb{E}[\gamma(\gamma)^l]$ is the average received SNR.

The stationary probability $\pi_l^{(i)}$ that the Markov chain is in state $s_l^{(i)}$ can be given by

$$\pi_l^{(i)} = \exp(-\Gamma_l/\gamma(\gamma)^l) - \exp(-\Gamma_{l+1}/\gamma(\gamma)^l).$$

(12)

Finally, $p_{l,j}^{(i)}$ can be derived from the normalising condition

$$\sum_{m=1}^{L} p_{l,m}^{(i)} = 1$$

as

$$p_{l,j}^{(i)} = \begin{cases} 1 - p_{l,j+1}^{(i)}, & (l = 2, \ldots, L - 1) \\ 1 - p_{l,j+1}^{(i)}, & (l = 1) \\ 1 - p_{l,j-1}^{(i)}, & (l = L) \end{cases}$$

(13)

For simplicity, only discrete-time rate matrix $\mathbf{H}(\theta) = \text{diag}(e^{\theta b_1^{(i)}}, e^{\theta b_2^{(i)}}, \ldots, e^{\theta b_{\bar{V}}^{(i)}}) \in \mathbb{C}^{L \times L}$ is considered in this paper. For all $t \geq 0$ and all $\theta$, the MGF of the random process $\mathbf{S}_t$ is

$$\mathbb{M}_\mathbf{S}(\theta, t) = \pi^{(i)}(\mathbf{H}(\theta)^{-1})^{t} \mathbf{H}(\theta)^{-1} \mathbf{1}$$

(14)

where $\pi^{(i)} = [\pi_1^{(i)}, \pi_2^{(i)}, \ldots, \pi_L^{(i)}] \in \mathbb{C}^{1 \times L}$ is stationary state distribution vector, $\mathbf{P}^{(i)} \in \mathbb{C}^{L \times L}$ represents the transition probability matrix with the element of $p_{l,m}^{(i)}$, $l, m \in \{1, 2, \ldots, L\}$ and $\mathbf{1}$ is a column vector of ones.

When $\mathbb{M}_\mathbf{S}(\theta, t)$ is available, the corresponding MGF of the equivalent single server for a two-hop relaying network can be calculated by (21). Then, given the known arrival process, the performances in terms of backlog and delay bounds are computed as in (6) and (7), respectively.

### 3.2 Scenario 2: single RS/multiple MSs

Usually there are more than one MSs served by one RS. Then, the wireless channels of the second hop transmission, that is $\text{RS} \rightarrow \text{MS}$ channel, is equivalent to a multi-user server, whereas that of the first hop transmission, that is $\text{BS} \rightarrow \text{RS}$ channel, is kept as a single-user server. So, the analytical method for a single-user server in the previous part can be applied directly in the analysis of the first hop transmission under this scenario. However, the performance of the second hop transmission in the multi-user channel has to be studied by constructing a new service process with appropriate Markov chain and corresponding rates.

Assume that $N$ MSs experience statistically identical radio channels in the second hop transmission. Each single-user channel can be modelled as an FSMC with $L$ states as before. Then, the state space in the multi-user channel of the second hop becomes $\mathbf{S} = \{s^{(2)}_0, s^{(2)}_1, \ldots, s^{(2)}_L\}$ with the size of $L^N$.

The $\ell$th, $1 \leq k \leq L$, state in this multi-user channel consists of $N$ sub-states, that is $s^{(2)}_k = (s^{(2)}_{k_1}, s^{(2)}_{k_2}, \ldots, s^{(2)}_{k_n})$ with $s^{(2)}_{k_n} \in \mathbf{S}$ and corresponding rate $\theta^{(2)}_{k_n} \in \mathbb{V}$, where $s_{k_n}, 1 \leq s_{k_n} \leq L$, is the state index of user $n, 1 \leq n \leq N$, in case that the multi-user channel is in state $s^{(2)}_k \in \mathbf{S}$.

The state transition probability from $s^{(2)}_{k-n}$ to $s^{(2)}_{k}$ and steady-state probability $\tilde{\pi}_k^{(2)}$ in the multi-user channel of the second hop can be, respectively, computed by

$$\tilde{p}_{l,j}^{(2)} = \sum_{k=1}^{N} \{\tilde{p}_{l,j}^{(2)}\}_{s_{k_n} = s_{k_n^*}}$$

(15)

$$\tilde{\pi}_k^{(2)} = \sum_{n=1}^{N} \tilde{\pi}_{k_n}^{(2)}$$

(16)

The stationary state distribution vector and the transition probability matrix in multi-user channel of the second hop
can be denoted as $\tilde{\pi}^{(2)} = [\tilde{\pi}_1^{(2)}, \tilde{\pi}_2^{(2)}, \ldots, \tilde{\pi}_N^{(2)}] \in \mathbb{C}^{1 \times L_N}$ and $\tilde{P}^{(2)} \in \mathbb{C}^{L_N \times L_N}$ with the element of $\tilde{P}_{k,j}^{(2)}$, $k, j \in \{1, 2, \ldots, L_N\}$.

The performance of the multi-user network depends on the applied scheduling strategies. The channel-aware scheduling algorithms aim at improving the performance by incorporating channel state information. Here only the greedy algorithm, also referred to as the Max-SNR algorithm, is assumed for the sake of analysis. This algorithm always selects the user with the best SNR for transmission, or equivalently, the best transmission data rate is guaranteed at every scheduling instant. For the Max-SNR algorithm, the data rate for user $n$, which can be processed in state $k$ of the second hop, is given by

$$\tilde{b}_{n,k}^{(2)} = \begin{cases} \frac{1}{\|M\|} b_{k,n}^{(2)}, & \text{if } n \in \mathcal{M} \\ 0, & \text{otherwise} \end{cases}$$ (17)

where $\mathcal{M} = \{m | b_{k,m}^{(2)} = \max(b_{k,1}^{(2)}, b_{k,2}^{(2)}, \ldots, b_{k,N}^{(2)})\}$ and $\|M\|$ represents the size of set $\mathcal{M}$. Here, it is assumed that the data rate is evenly divided among the selected users as suggested in [15].

Next, given user $n$, $1 \leq n \leq N$, the corresponding rate matrix in the second hop transmission can be written as $\tilde{H}_n^{(2)}(\theta) = \text{diag}(e^{i\theta_{1,n}}, e^{i\theta_{2,n}}, \ldots, e^{i\theta_{N,n}}) \in \mathbb{C}^{1 \times L_N}$. Therefore, the MGF of user $n$’s random process in the second hop transmission can be calculated by

$$\tilde{M}_{\tilde{B}_n}(\theta) = \tilde{\pi}^{(2)}(\tilde{H}_n^{(2)}(-\theta)\tilde{P}^{(2)})^{-1}\tilde{H}_n^{(2)}(-\theta)\mathbf{1}$$ (18)

However, the exponentially enlarged state space leads to the high computational complexity in calculating the MGF of the random processes in case of a large user number. Since the channel processes of all the users are assumed to be independent with each other and identically distributed, the statistical characters of each user are expected to be same. As shown in (14), the MGF of random process $S_i$ in single-user scenarios is determined by the stationary state distribution vector, the transition probability matrix and the data rate matrix. So, as one of possible solutions to avoid the high computational complexity in multi-user scenarios, the multi-user effects can be reflected only by modifying the data rate matrix with the $L$-state FSMC used in single-user scenarios. A simplified method is proposed to extend the single-user model to the multi-user scenarios with the acceptable computation complexity.

In this method, the stationary state distribution vector and the transition probability matrix of each user in the multi-user scenarios are assumed to be kept same as those in the single-user scenarios. Meanwhile, the Max-SNR scheduling affects the data rate of each user, which depends not only on its own channel state but also on channel states of other users in multi-user scenarios. Then, the equivalent data rate of user $n$, corresponding to the channel state $s_i^{(2)} \in S$, can be computed by

$$\tilde{b}_{n,k}^{(2)} = \sum_{k=1}^{\|M\|} \frac{\tilde{b}_{k,n}^{(2)}}{\|M\|}, \quad 1 \leq l \leq L, \quad 1 \leq n \leq N$$ (19)

Now, the modified rate matrix $\tilde{H}_n^{(2)}(\theta) = \text{diag}(e^{i\theta_{1,n}}, e^{i\theta_{2,n}}, \ldots, e^{i\theta_{N,n}}) \in \mathbb{C}^{1 \times L}$ is used together with $\tilde{\pi}^{(2)}$ and $\tilde{P}^{(2)}$ of the $L$-state FSMC to calculate the MGF of user $n$’ random process in the second hop transmission, that is

$$\tilde{M}_{\tilde{B}_n}(\theta, t) = \tilde{\pi}^{(2)}(\tilde{H}_n^{(2)}(-\theta)\tilde{P}^{(2)})^{-1}\tilde{H}_n^{(2)}(-\theta)\mathbf{1}$$ (20)

Next, the MGFs of random process in the first hop can be generated by (14) whereas that of the second hop is calculated as in (20). Then, the MGF of a two-hop server can be generated by the concatenation character of network calculus, that is

$$\tilde{M}_{\tilde{B}_{\text{tot}}}(\theta, t) \leq \sum_{t=0}^{\infty} \tilde{M}_{\tilde{B}_n}(\theta, t) = \tilde{M}_{\tilde{B}_n}(\theta, t - \tau)$$ (21)

Finally, the delay and backlog performances of wireless multi-hop relaying networks in the multi-user scenarios are derived by (6) and (7), respectively.

## 4 Numerical and simulation results

In the previous section, we derive the delay and backlog bounds of a multi-hop relaying network. In particular, the simplified method to avoid the exponentially enlarged state space in multi-user scenarios is proposed. In this section, we provide the simulation and analytical results, demonstrating the effectiveness of our derived bounds under different scenarios.

### 4.1 Network configuration

#### 4.1.1 RS location:

As shown in Fig. 2, we assume that the RS is located between the BS and the MSs, and all the MSs have the same distance to the RS. In the two-hop relaying communication, we denote the BS $\rightarrow$ RS distance as $d$ and the RS $\rightarrow$ MS distance as $1 - d$, respectively, which is normalised by the distance of BS $\rightarrow$ MS link in case that the RS is on the straight line connecting them. Furthermore, the same long-term average transmit power

![Figure 2 Network setting for analysis and simulation](image)
The limitation is assumed at both the BS and the RS. For a fixed pathloss exponent $\alpha = 4$, the average received power is dependent on the distance between the transmitter and receiver [17]. The performances of the networks without RS, that is, the direct transmission (DT) from the BS to the MS, are also given for comparison. Then, the average SNR ratio between the first hop, second hop and DT is given by $\bar{\gamma}(1) = d^{-1}$, $\bar{\gamma}(2) = d^{-4}$. In our analysis and simulations, the received SNRs of the BS → RS link and RS → MS link are set as $\bar{\gamma}(1) = 20$ dB and $\bar{\gamma}(2) = 20$ dB in case of $d = 0.5$, and the corresponding received average SNR of the network with DT is $\bar{\gamma} = 8$ dB.

4.1.2 FSMC model: The five state FSMC model is adopted to reflect the wireless fading characters, in which each state provides different data rate by using different MCS. The data rate is represented by the number of packets that can be processed in time period of $T_p = 2$ ms. On the assumption of target block error rate (BLER) $= 10^{-2}$ with the system parameters in [18], the SNR threshold and data rate for each state are shown in Table 1. Only one channel is used for the first hop transmission whereas the channels of all the users in the second hop transmission are assumed to be independent and identically distributed. If not specified, the Doppler spread of the wireless channel is assumed to be 10 and 30 Hz for the first and second hop transmission, respectively, that is, $f_D^{(1)} = 10$ Hz and $f_D^{(2)} = 30$ Hz.

4.1.3 Arrival process: The arrivals and services are assumed to be statistically independent, stationary random processes in this paper. Therefore only one of typical arrival processes in this paper. This source generate G packets of workload at times $\{\alpha T_i + i T_s, \ i = 0, 1, 2, \ldots\}$ with time period of $T_s$ frames, where the initial start time $\alpha$ is uniformly distributed over the interval of $[0, 1)$. Then, the MGF of this arrival process can be given by [10]

$$M_{\lambda}(\theta, t) = e^{\theta G T_s} \left[ 1 + \left( \frac{t}{T_s} - \left\lfloor \frac{t}{T_s} \right\rfloor \right) (e^{\theta G} - 1) \right] \quad (22)$$

| Table 1 Rate and SNR thresholds in FSMC |
|-----------------|-----------------|-----------------|
| State | Rate | MCS | $[\Gamma_i, \Gamma_{i+1}]$ |
| 1 | 0 | QPSK, $R_c = 1/2$ | $[0, 1.7414]$ |
| 2 | 1 | 16QAM, $R_c = 1/2$ | $[1.7414, 5.7187]$ |
| 3 | 2 | 16QAM, $R_c = 3/4$ | $[13.8867, 31.908]$ |
| 4 | 3 | 64QAM, $R_c = 2/3$ | $[31.908, \infty]$ |

$R_c$ denotes the coding rate of convolutional Turbo code.

In our simulations, the parameters of the source per user are set as $T_0 = 8$ and different $G$, resulting in the packet arrive rate of $\lambda_p = G/T_0$. Usually, the simulation results are collected over $10^5$ source periods.

4.2 Performance comparison and analysis

4.2.1 Scenario 1 (single RS/single MS): The delay performances of the two-hop wireless networks for violation probabilities $\epsilon \in \{10^{-7}, 10^{-6}, \ldots, 10^{-3}\}$ with the packet arrive rate $\lambda_p = 0.5$ are given through the simulation and the proposed analytical method in Fig. 3, where the number of users in the second hop transmission is assumed to be one. The analytical bounds of the delay match with those by simulations, which proves the feasibility of applying the concatenation character in analysing the performance of multi-hop relaying networks. However, since these bounds are derived by using Chernoff’s inequality, the accuracy is dependent on the violation probability. The smaller the violation probability is, the tighter the analytical bounds become. In order to clearly show this trend, the relative error between the analytical delay bound, that is, $d_{ana}$, and the delay by simulation, that is, $d_{sim}$ is calculated as $(d_{ana}/d_{sim} - 1)$ and also shown in the figures. The error curve becomes floor after the violation probability is smaller than the given value, for example, less than $10^{-5}$. Therefore the derived bounds are more applicable in case of the small violation probability. The similar conclusion can be drawn for the backlog performances.

4.2.2 Scenario 2 (single RS/multiple MSs): Fig. 4 gives the delay performances of the case that more than one user exist in the second hop transmission, that is, $N = 5$. The delay performances become worse with the increase number of users. Similar to the single MS scenario, the relative error is decreased and becomes floor with the smaller violation probability. For simplicity, the
violation probability of delay and backlog bound is set to $10^{-5}$ in the following analysis.

Then, we compare the analytical bound and simulation results of the delay and backlog performances with the different packet arrival rate in Fig. 5. For simplicity, the number of users is still set to three. With the increase of the packet arrival rate per user, the delay and backlog become larger because the data rate provided is limited by the wireless channel. The slope of the curves becomes larger with the increase of the packet arrival rate, which means that the performance degradation becomes more serious. If the total arrival rate (i.e. $N\lambda_p$) is increased to be larger than the maximum data rate that the wireless channel of the first hop can provide (i.e. 4), the delay and backlog are expected to become infinite. It is noted that the analytical results are in line with those simulation results, which also validates the rightness of our proposed analytical method.

Next, Fig. 6 gives the delay performances of the networks with the two-hop transmission and DT with the respect to

the number of users, where the packet arrival rate is fixed to be 0.5, that is, $\lambda_p = 0.5$. Compared with the direction transmission, the qualities of both BS $\rightarrow$ RS and RS $\rightarrow$ MS channels become much better with the cost of repetition in the time domain. So, an equivalent two-hop dynamic server can process more packets, reducing the waiting time in the buffer. Therefore the delay of the networks with two-hop transmission is less than that with DT, which is confirmed by our results. For example, the delay of the network with the two-hop transmission is only 70%, 40% and 22% of that with the DT in case of $N = 2$, 3 and 4, respectively.

On the assumption of the given parameters, we compare the delay performances with different Doppler spread in the second hop channel as shown in Fig. 7. When the channel is in poor state, the data rate that it can provide is smaller than that the network requires so that the delay happens in the buffer. On the other hand, the channel in good state with high data rate only causes quite small or even no delay in wireless networks. Therefore the delay performance is mostly determined by the duration of the
channel in poor states. With the increase of Doppler spread, the probability that the channel transfers from the poor channel to the good channel becomes larger. So, in Fig. 7, the delay in case of $f_D^{(2)} = 30$ Hz is smaller than that in case of $f_D^{(2)} = 20$ Hz.

5 Conclusion

In this paper, the analytical models by using probabilistic network calculus with MGFs are applied to study the delay and backlog performance bounds in wireless multi-hop relaying networks. The bounds of two-hop wireless relaying network can be calculated by the concatenation of those in each hop transmission. We first point out that the analytical method for single-user scenarios is hard to be used directly for multi-user scenarios because of exponentially enlarged state space. Then, on the assumption of Max-SNR scheduling algorithm, we propose to reflect multi-user effects only by the equivalent data rate without increasing the state size of the service process. Simulation results show that our proposed analytical framework is applicable in various scenarios. The analytical bounds match with those by simulations, which proves the feasibility of applying the concatenation character in analysing the performance of multi-hop relaying networks. Moreover, the accuracy is dependent on the violation probability. The smaller the violation probability is, the tighter analytical bounds become. The relative error becomes floor when the violation probability is smaller than the given value, for example $10^{-5}$. It is shown that the delay performance of two-hop transmission outperforms the DT when the qualities of both BS $\rightarrow$ RS and RS $\rightarrow$ MS channels are much improved by deploying the RS. For example, the delay of the network with the two-hop transmission is only 70%, 40% and 22% of that with the DT in case of $N = 2, 3$ and 4, respectively. In addition, we also demonstrate that the delay performance is mostly determined by the duration of the channel in poor states but not good states.

6 Acknowledgments

This work was supported in part by China NSFC under Grant 60702009 and National Key Technology R&D Program of China under Grant 2009ZX03003-008-01.

7 References

[1] IEEE 802.16m-07/002r10: ‘IEEE 802.16m System requirements’. January 2010


[18] IEEE 802.16m-08/004r5: ‘IEEE 802.16m evaluation methodology document (EMD)’. January 2009