Abstract—In this paper, we model the location-dependent throughput and delay in wireless mesh networks. We analyze the packet-arrival and the packet-departure rates for the forwarding queues at relaying nodes and then derive the per-user throughputs and packet delays experienced by nodes at different hop-count distances from the gateway. Based on this analytical model, we propose two network design strategies to provide fair resource sharing and minimize the end-to-end delay in wireless mesh networks. We then conduct simulations to validate the correctness of our analytical model and evaluate the performance of the proposed network design strategies. The results show that our analytical framework can accurately model the location-dependent throughput and delay in wireless mesh networks and that our proposed strategies effectively and efficiently provide fair throughputs for nodes and reduce the end-to-end packet delay. This paper not only provides a framework for studying the performance of wireless mesh networks but also gives insights into the network design strategy for wireless mesh networks.

Index Terms—Delay, throughput, wireless mesh networks.

I. INTRODUCTION

WIRELESS mesh networks have received much attention in recent years due to its low up-front cost, easy network maintenance, robustness, and reliable service coverage [1]–[4]. In such networks, each mesh node plays both roles of a host and a router. Packets are forwarded in a multihop fashion to and from the gateway (connected to the Internet). It has been shown that the throughput and delay performances in wireless mesh networks are location-dependent [5]–[7]. Specifically, the per-node throughput may decrease and the end-to-end delay may dramatically increase with an increasing hop-count distance from the gateway (node). This leads to a fairness problem since the mesh nodes that are farther away from the gateway may be starved by the nodes that are closer to the gateway. Moreover, the long end-to-end packet delay due to multihop relaying may become unacceptable for delay-sensitive applications such as voice over IP (VoIP). Many mechanisms [8]–[10] have been proposed to alleviate these problems.

The tradeoff between the throughput and the delay in multihop wireless networks has been well addressed in the literature [11]–[15]. Early work in system performance [11]–[15] mainly focuses on determining the optimal capacity of the network. The factors under consideration include the number of nodes, the number of hops, the impact of mobility, and different routing strategies. In [11], the maximum throughput in a fixed wireless network is obtained. The network considered is a static ad hoc network, in which nodes are randomly distributed and the destination for each node is independently chosen. The authors then show that the per-node throughput in the network is bounded by \( \Theta(W/\sqrt{n \log n}) \), where \( W \) is the common transmission rate over the wireless channel for each node and \( n \) is the total number of nodes under consideration. In [12], the per-node throughput is shown to be dramatically increased by exploiting node mobility as a type of multiuser diversity. In [13], an analytical model is developed to obtain the optimal throughput-delay tradeoff by varying the number of hops, the transmission range, and the degree of node mobility in ad hoc networks. That model can also capture the characteristics of the model proposed by Gupta and Kumar [11] and that proposed by Grossglauser and Tse [12].

While many research results [11]–[15] have been reported for the throughput and delay performances in multihop wireless networks, most of the efforts are focused on the asymptotic case. Very few research works are devoted to modeling the statistical location-dependent throughput and delay performances in wireless mesh networks. However, such study is important because most of the problems in networks are concerned with the exact throughput and delay performances rather than the asymptotic case. Moreover, based on such study, the optimal design strategy can be determined. In this paper, we derive the per-node throughput and the end-to-end delay as functions of the hop-count distance from the gateway. We consider a static network consisting of \( N \) mesh nodes and one gateway that serves as the sink for the mesh nodes. In our system model, a node whose hop-count distance from the gateway node equals \( x \) is referred to as an \( x \)-hop node. We assume that the hop-count distribution of nodes is given (i.e., the ratio of the \( x \)-hop nodes in the network is given). To make our model more general, we do not assume any specific medium-access-control (MAC) mechanism but instead suppose a general MAC scheme with one parameter, i.e., the successful channel-access probability \( p(x) \) for the \( x \)-hop node. Each node is associated with two queues at the network layer, one for its local data and the other for relayed traffic. Each time a transmission opportunity is available, the mesh node forwards one packet from one of the two queues according to a probabilistic forwarding strategy. That is, the \( x \)-hop node will forward a relayed packet to the next hop with a probability of \( q(x) \) or a local packet with a probability of \( 1 - q(x) \); when either queue is empty, it will
forward a packet from the other queue with a probability of one. We analyze the intensity of relayed traffic at the $x$-hop node and derive the throughput and the delay for the $x$-hop node with the $M/M/1/K$ queuing model [16].

Based on this derivation, we further study how the channel-access probability $p(x)$ and the forwarding probability $q(x)$ influence the network performance. We then propose two network design strategies for fair resource sharing and delay minimization. We also conduct simulations to validate the analytical model and to evaluate the performance of our proposed network design strategies. The results show that our analysis can accurately model the location-dependent throughput and delay in the wireless mesh network. In addition, our strategy can provide a fair resource sharing for nodes under various spatial locations and network sizes. Moreover, our delay-minimization strategy can dramatically reduce the end-to-end delay without degrading the throughput performance.

The rest of this paper is organized as follows. In Section II, the system model is described. In Section III, the location-dependent throughput and delay are derived. In Section IV, two design strategies for fairness provisioning and delay minimization are introduced. In Section V, the correctness of our analytical model and the effectiveness of the proposed strategies are evaluated via simulations. Finally, this paper is concluded in Section VI.

II. SYSTEM MODEL

A. Network Model

The network that we consider consists of $N$ mesh nodes and one gateway. Each node is equipped with one radio and has a common transmission range $r$. The capacity of the wireless channel is $W$ bits per second, i.e., for each node, the transmission rate is $W$ bits per second. This rate is constant and independent of the number of nodes in the network. Packets are forwarded in a multihop fashion to or from the gateway. For ease of explanation and without loss of generality, we consider unidirectional traffic, i.e., the traffic that only goes from the mesh nodes to the gateway, which is also referred to as upstream traffic. We consider the saturated condition, i.e., each node is always backlogged and always has local data to transmit to the gateway. With fairness consideration for relayed traffic [5]–[7], each node is associated with two queues, as shown in Fig. 1, i.e., $Q_r$ for the relayed packets and $Q_s$ for the locally generated packets. The forwarding rule at each $x$-hop node is described as follows.

1) If $Q_r$ is empty, it pops one packet from $Q_s$ (which is assumed backlogged) to send.

2) If $Q_r$ is not empty, it sends a packet from $Q_r$ with a probability of $q(x)$ or a packet from $Q_s$ with a probability of $1 - q(x)$.

In this paper, $q(x)$ is referred to as the forwarding probability for the $x$-hop node. We regard $q(x)$ as a network design parameter and discuss its influence on the network performance in Section IV.

B. Routing Strategy and Hop-Count Distribution

The routing scheme used in our model is shortest path routing. Since wireless mesh networks are generally considered robust and highly connected [1], [2], we assume that each mesh node can always find at least one path to the gateway. We further assume that on the average, each $x$-hop node needs to relay for the same number of $(x + 1)$-hop nodes and can ask the same number of $(x - 1)$-hop nodes to provide relaying service. When an $x$-hop node has more than one $(x - 1)$-hop neighbor, it can randomly select one of them as its relaying node on a per-packet or per-flow basis.

Denote the ratio for the number of $x$-hop nodes in the network by $h(x)$. The value of $h(x)$ is affected by many factors, including the routing algorithm, the node spatial distribution, and the transmission range, and can be obtained via simulations or by analysis as in [17] and [18]. Since the network area is bounded and the gateway may restrict the maximum hop-count distance within which the mesh nodes can be served, $h(x)$ approaches zero when $x$ is very large. For ease of explanation and without loss of generality, we let $H$ denote the maximum possible hop-count distance from the gateway in the network. This means that the $H$-hop nodes are the nodes farthest away from the gateway. Accordingly, there will be no relayed traffic to the gateway coming into the $H$-hop nodes.

C. Media Access Control

In this paper, we do not restrict our discussion to any specific MAC mechanism. Instead, our analytical model is developed based on a MAC scheme with one parameter $p(x)$ that specifies the successful channel-access probability for the $x$-hop node. To obtain $p(x)$ in wireless networks, many factors should be considered, including interference, the type of physical layer, the number of available channels in the network, the number of available radios of each node, and the MAC protocol [19], [20]. For example, in contention-free wireless networks, $p(x)$ may be controlled by the scheduling algorithm, whereas in contention-based systems, $p(x)$ is governed by the node density and the MAC protocol. In a time-division multiple-access (TDMA) system, $p(x)$ corresponds to the probability of the $x$-hop node being granted a transmission opportunity within one time slot. In a contention-based MAC scheme, such as IEEE 802.11 DCF, $p(x)$ is the probability of the $x$-hop node successfully occupying the media for one attempt. It is generally believed that a node closer to the gateway node should have a higher value of $p(x)$ due to its heavier amount of relayed traffic. In this paper, we regard $p(x)$ as a network design parameter and derive the performance result based on its best value in Section IV.
Fig. 2. Interval between two arrivals of the transmission opportunities for the x-hop node.

III. THROUGHPUT AND DELAY ANALYSES

In this section, we study the behavior of \( Q_r \) and \( Q_s \) and analyze the throughput and delay performances of the x-hop node. We consider a TDMA-based system in which each packet is transmitted within one time slot with length \( t_c \). This TDMA-based system is also characterized by a parameter \( p(x) \), which specifies the successful channel-access probability for the x-hop node. Note that the following analysis is still applicable to non-TDMA systems, in which \( t_c \) is replaced with the average amount of airtime occupied by a node’s successful channel access.

A. Service Rates of Packets for the Queues

Given the ratio of x-hop nodes in the network \( h(x) \), we can obtain the expected number of x-hop nodes in the network, which is denoted by \( N(x) \), as

\[
N(x) = N \cdot h(x). \tag{1}
\]

Since each x-hop node is assumed to relay for the same number of \((x+1)\)-hop nodes on the average, the expected number of \((x+1)\)-hop nodes for which each x-hop node has to relay, which is denoted by \( N_r(x) \), is given by

\[
N_r(x) = \begin{cases} 
\frac{N(x+1)}{N(x)} = \frac{h(x+1)}{h(x)}, & x = 1, 2, \ldots, H - 1 \\
0, & x = H. 
\end{cases} \tag{2}
\]

Let \( T_i(x) \) denote the interarrival time of transmission opportunities for the x-hop node (see Fig. 2). Since, within one time slot, the probability of the x-hop node being granted a transmission opportunity is equal to \( p(x) \), the probability that an x-hop node is not activated for \( m \) time slots is given by

\[
P[T_i(x) > mt_c] = (1 - p(x))^m.
\]

By replacing \( mt_c \) with \( t \), we can rewrite the aforementioned equation as

\[
P[T_i(x) > t] = (1 - p(x))^{\frac{t}{t_c}}. \tag{5}
\]

Since binomial probabilities can be approximated by Poisson probabilities with appropriate parameters [16], we approximate the arrival process of transmission opportunities as a Poisson process. In other words, we approximate the interarrival time of transmissions \( T_i(x) \) as an exponential random variable with mean \( 1/\mu(x) \). Thus, we have

\[
P[T_i(x) > t] = (1 - p(x))^{\frac{t}{t_c}} \approx e^{-\mu(x)t} \Rightarrow \mu(x) \approx \frac{1}{t_c} \ln \left( \frac{1}{1 - p(x)} \right). \tag{3}
\]

Note that \( \mu(x) \) represents the arrival rate of transmission opportunities for the x-hop node and is equivalent to the service rate of packets for the x-hop node. For \( Q_r \) and \( Q_s \) at the x-hop node, the service rate of packets for either queue is equal to the product of \( \mu(x) \) and the probability that the queue is selected to send. Thus, the service rate of packets for \( Q_r \) at the x-hop node, which is denoted by \( \mu_r(x) \), is given by

\[
\mu_r(x) = \mu(x) \cdot q(x).
\]

According to the forwarding rules described in Section II-A, only when \( Q_r \) is nonempty will the transmission opportunity have a chance to come to \( Q_r \). Thus, the effective output rate of \( Q_r \) for the x-hop node (i.e., the effective departure rate of relayed packets that are forwarded to the next hop node), which is denoted by \( \sigma_r(x) \), is equal to the service rate multiplied by the probability of \( Q_r \) being nonempty, i.e.,

\[
\sigma_r(x) = \mu_r(x) \cdot [1 - P_0(x)], \tag{4}
\]

where \( P_0(x) \) denotes the probability of \( Q_r \) being empty at the x-hop node. We will derive \( P_0(x) \) in the next section.

For the service rate of packets for \( Q_s \) at the x-hop node, according to the forwarding rules described in Section II-A, when \( Q_r \) is empty, the transmission opportunity is always granted to \( Q_s \). Thus, the service rate of packets for \( Q_s \) at the x-hop node, which is denoted by \( \mu_s(x) \), is

\[
\mu_s(x) = \mu(x) - \sigma_r(x) = \mu(x) - \mu(x) \cdot q(x) \cdot [1 - P_0(x)].
\]

Recall that \( Q_s \) for each node is assumed to be backlogged, so the output distribution (i.e., the distribution of the time interval between two successive departures) of \( Q_s \) is identical to the service-time distribution of \( Q_s \). In other words, the effective output rate of \( Q_s \) at the x-hop node, which is denoted by \( \sigma_s(x) \), is equal to its service rate. Thus, we have

\[
\sigma_s(x) = \mu_s(x). \tag{5}
\]

From (4) and (5), we obtain the aggregate effective output rate for the x-hop node, which is denoted by \( \sigma(x) \), as

\[
\sigma(x) = \sigma_s(x) + \sigma_r(x) = \mu(x). \tag{6}
\]

Equation (6) states that the effective output rate of the x-hop node is equal to the arrival rate of the transmission opportunities for this node. This is definitely a necessary result if the network resource in the model is efficiently used at each node.

B. Arrival Rates of Packets for the Queues

To analyze the behavior of the queues, we need to know the packet-arrival rates for the queues. Denote the packet-arrival rate for \( Q_r \) at the x-hop node by \( \lambda_r(x) \). Note that since \( Q_s \) is assumed to be always backlogged, we only need to discuss the packet-arrival rates for \( Q_r \) at the x-hop node.

\( \lambda_r(x) \) can be obtained by applying the flow conservation law. Equation (7) states that the total departures of packets from the
(x+1)-hop nodes must be equal to the total arrivals of packets to the Q_r at the x-hop node, which is denoted by P_b(x). From the M/M/1/K formulas, we have

\[
P_b(x) = \begin{cases} 
\frac{1 - \rho(x) \rho(x)^K}{1 - \rho(x) \rho(x)^K + 1}, & \rho(x) \neq 1 \\
\frac{1}{K+1}, & \rho(x) = 1 
\end{cases}
\]

where \( \rho(x) \) is given by (10). Accordingly, the nonblocking probability for \( Q_r \) at the x-hop node is \( 1 - P_b(x) \). For a path, the end-to-end nonblocking probability is equal to the product of the nonblocking probabilities at all intermediate nodes; therefore, for the x-hop node, its throughput \( T(x) \) is equal to the product of the effective output rate of \( Q_s \) and the end-to-end nonblocking probability, i.e.,

\[
T(x) = \begin{cases} 
\sigma_s(1), & x = 1 \\
\sigma_s(x) \cdot \prod_{i=1}^{x-1} [1 - P_b(i)], & x = 2, \ldots, H 
\end{cases}
\]

Note that the blocking probabilities at two nodes may locally be correlated, but here, we ignore this correlation and assume that they are independent. In Section V, we will show that this assumption does not influence the accuracy of our analysis. From (12), we find that the throughput for a node varies with its location. In addition, since \( \sigma_s(x) \) and \( P_b(x) \) are functions of the successful channel-access probability \( p(x) \) and the forwarding probability \( q(x) \), the throughput is also a function of \( p(x) \) and \( q(x) \). In Section IV, we will study the influence of \( p(x) \) and \( q(x) \) on nodes’ throughputs.

Let \( T_{agg} \) denote the system throughput, i.e., the aggregate of the per-node throughputs, and \( T_{ave} \) denote the average per-node throughput. Since the number of x-hop nodes is \( N(x) \), we have

\[
T_{agg} = \sum_{x=1}^{x=H} [N(x) \cdot T(x)].
\]

Since all packets must be forwarded by one-hop nodes before they are received by the gateway, the aggregate throughput must be equal to the aggregate effective output rates of all one-hop nodes, i.e.,

\[
T_{agg} = N(1) \cdot \sigma(1) = N(1) \cdot \mu(1).
\]

Accordingly, the average per-node throughput is given by

\[
T_{ave} = \frac{T_{agg}}{N} = h(1) \cdot \mu(1).
\]

### D. Delay Analysis

To derive the end-to-end delay, we first investigate the expected number of packets queued in \( Q_r \) at the x-hop node. Let \( L_r(x) \) denote the steady-state queue size of \( Q_r \) for the x-hop node. According to the M/M/1/K formulas, we have

\[
L_r(x) = \begin{cases} 
\frac{\rho(x)}{1 - \rho(x)}, & \rho(x) \neq 1 \\
\frac{\rho(x)^{K(x+1)}}{2(K+1)}, & \rho(x) = 1 
\end{cases}
\]
where $\rho(x)$ is given by (10). Obviously, since $\rho(x)$ is a function of $p(x)$ and $q(x)$, the steady-state queue size of $Q_r$ for the $x$-hop node is also a function of $p(x)$ and $q(x)$.

The end-to-end delay experienced by a packet is defined as the time between when the first bit of this packet is sent by its source and when the packet is entirely received by the destination (i.e., the gateway). We assume that the propagation delay is negligible. Thus, the end-to-end packet delay is equal to the sum of the waiting times spent in the intermediate nodes and the transmission times for traversing the $x$ hops. According to Little’s formula [16], we have

$$W_r(x) = \frac{1}{\mu_r(x)} + \frac{L_r(x)}{\lambda_r(x)[1 - P_b(x)]}, \quad x = 1, 2, \ldots, H - 1.$$  

(15)

Now, we can obtain $D(x)$ by summing the waiting times spent in the intermediate nodes and the transmission times for traversing the $x$ hops (see Fig. 3), i.e.,

$$D(x) = \left\{ \begin{array}{ll} t_c, & x = 1 \\ x \cdot t_c + \sum_{i=1}^{i=x-1} W_r(i), & x = 2, 3, \ldots, H. \end{array} \right.$$  

(16)

The average end-to-end packet delay, which is denoted by $D_{ave}$, is obtained by averaging the delays of packets which are successfully received by the gateway. Since, within one unit time, $T(x)$ represents the total number of packets generated by the $x$-hop node and successfully received by the gateway, the aggregate number of packets generated by all $x$-hop nodes and successfully received by the gateway is equal to $N(x)T(x)$. Thus, the aggregate delay of the successful packets generated by all $x$-hop nodes is $N(x)T(x)D(x)$. Consequently, we have

$$D_{ave} = \frac{\sum_{x=1}^{x=H} N(x)T(x)D(x)}{\sum_{x=1}^{x=H} N(x)T(x)} = \frac{\sum_{x=1}^{x=H} N(x)T(x)D(x)}{T_{agg}}.$$ 

IV. NETWORK DESIGN STRATEGY

In our system model, we assume that the successful channel-access probability $p(x)$ and the forwarding probability $q(x)$ for the $x$-hop node are given. From the analysis in the previous section, we find that the throughput and the delay are functions of these two parameters and, thus, are location-dependent. In this section, we study how these two parameters influence the throughput and the delay. In particular, we obtain the appropriate settings of these two parameters in order to achieve fairness provisioning and delay minimization.

A. Fairness Provisioning

A well-known problem in wireless mesh networks is the fairness problem [5]–[7], i.e., the nodes farther away from the gateway may experience a lower throughput than the nodes closer to the gateway. This problem arises when improper MAC schemes or bad forwarding strategies are employed. For example, nodes closer to the gateway should be allocated more radio resource due to their higher loads of relayed traffic. Therefore, in the TDMA system, if the scheduling algorithm does not grant more transmission opportunities to nodes which are closer to the gateway, buffer overflow may frequently happen at these nodes due to the unbalanced packet-arrival and -departure rates. On the other hand, if nodes are all selfish and each node refuses to relay other nodes’ packets, the nodes farther away from the gateway may suffer from starvation or an unacceptable end-to-end delay. In the following, we study how to set the successful channel-access probability $p(x)$ and the forwarding probability $q(x)$ when designing a wireless mesh network such that a fair throughput can be provided to all nodes no matter how far they are located from the gateway.

To address this fairness problem, we first investigate the condition under which the packets of the $H$-hop nodes are fairly treated by the $(H-1)$-hop nodes as the local packets of the $(H-1)$-hop nodes. In other words, we want to determine the condition under which the $H$-hop node can have the same throughput as the $(H-1)$-hop node. Since the average number of $H$-hop nodes for which an $(H-1)$-hop node has to relay is $N_r(H-1)$, to provide a fair throughput to the $H$-hop and $(H-1)$-hop nodes, the effective output rate of $Q_r$ at the $(H-1)$-hop node (i.e., $\sigma_r(H-1)$) must be $N_r(H-1)$ times the effective output rate of $Q_s$ at the $(H-1)$-hop node (i.e., $\sigma_s(H-1)$), i.e.,

$$\sigma_r(H-1) = N_r(H-1) \cdot \sigma_s(H-1).$$

From (2), we have $N_r(H) = 0$. Thus, the aforementioned equation can be rewritten as

$$\sigma_r(H-1) = [N_r(H-1) + N_r(H-1) \cdot N_r(H)] \cdot \sigma_s(H-1).$$

Now, consider the condition under which the packets of the $H$-hop and $(H-1)$-hop nodes are fairly treated by the $(H-2)$-hop nodes as the local packets of the $(H-2)$-hop nodes. Specifically, we want to determine the condition that for an $(H-2)$-hop node, the departure rate of packets coming from the $(H-1)$-hop nodes is equal to the departure rate of the $(H-2)$-hop node’s own packets. Since the average number of the $(H-1)$-hop nodes that the $(H-2)$-hop node has to relay for is $N_r(H-2)$, the total number of nodes that the $(H-2)$-hop node has to relay for, including the $H$-hop and...
(\(H - 1\))-hop nodes, is \(N_r(H - 2) + N_r(H - 2)N_r(H - 1)\), where \(N_r(H - 2)N_r(H - 1)\) represents the average number of \(H\)-hop nodes that the \((H - 2)\)-hop node has to relay for (via the \((H - 1)\)-hop nodes’ relaying). Consequently, the ratio of the effective output rate of \(Q_r\) to the effective output rate of \(Q_r\) at the \((H - 2)\)-hop node should satisfy the following:

\[
\sigma_r(H - 2) = [N_r(H - 2) + N_r(H - 2) \cdot N_r(H - 1)] \cdot \sigma_s(H - 2).
\]

Again, the aforementioned equation can be rewritten as

\[
\sigma_r(H - 2) = [N_r(H - 2) + N_r(H - 2) \cdot N_r(H - 1) + N_r(H - 2) \cdot N_r(H - 1) \cdot N_r(H)] \cdot \sigma_s(H - 2).
\]

Similarly, we can obtain the condition under which a fair throughput can be provided to the \((x + 1)\)-hop and the \(x\)-hop nodes:

\[
\sigma_r(x) = \sigma_s(x) \cdot \sum_{i=x}^{i=x} \prod_{j=x}^{j=x} N_r(i), \quad x = 1, 2, \ldots, H - 1.
\]

To simplify the equations, we define

\[
R(x) = \sum_{i=x}^{i=x} \prod_{j=x}^{j=x} N_r(i), \quad x = 1, 2, \ldots, H - 1.
\]

Thus, we can rewrite (17) as

\[
\sigma_r(x) = R(x) \cdot \sigma_s(x), \quad x = 1, 2, \ldots, H - 1.
\]

Note that by definition, we know that \(R(x)\) is independent of \(p(x)\) and \(q(x)\). By substituting (4) and (5) into (18), we obtain

\[
\mu(x) \cdot q(x) \cdot [1 - P_0(x)] = R(x) \cdot \{\mu(x) - \mu(x) \cdot q(x) \cdot [1 - P_0(x)]\}.
\]

Based on (9), when \(\rho(x) < 1\) and \(K\) (i.e., the buffer size of \(Q_r\)) is large enough, we have the following approximation:

\[
P_0(x) \approx 1 - \rho(x).
\]

By substituting (10) and (20) into (19), we obtain

\[
N_r(x) \cdot \mu(x + 1) = R(x) \cdot [\mu(x) - N_r(x) \cdot \mu(x + 1)].
\]

By rearranging the aforementioned equation, we obtain

\[
\frac{\mu(x)}{\mu(x + 1)} = \frac{N_r(x) \cdot \left[1 + \frac{1}{R(x)}\right]}{N_r(x) \cdot \left[1 + \frac{1}{R(x)}\right]}
\]

By substituting (3) into the aforementioned equation, we obtain

\[
\ln\left(1 - p(x)\right) = N_r(x) \cdot \left[1 + \frac{1}{R(x)}\right].
\]

When \(p(x)\) is a small number, we have

\[
\ln(1 - p(x)) \approx -p(x).
\]

Thus, we can rewrite (21) as

\[
\frac{p(x)}{p(x + 1)} \approx N_r(x) \cdot \left[1 + \frac{1}{R(x)}\right], \quad x = 1, 2, \ldots, H - 1.
\]

Equation (22) shows that if the ratio of the successful channel-access probabilities is carefully controlled, it is possible to provide a fair throughput to the nodes despite their distances from the gateway.

Recall that (20) is derived based on the assumption that the traffic intensity for each node is less than one, i.e.,

\[
\rho(x) < 1.
\]

This assumption is desirable because when the traffic intensity for each node is larger than or equal to one, many packets may be blocked, leading to a waste of radio resource and a poor system throughput. In the following, we derive the condition under which this assumption is valid. Based on (10), we rewrite (23) as

\[
q(x) \cdot \mu(x + 1) \approx N_r(x) \cdot \frac{p(x + 1)}{p(x)}.
\]

By substituting (22) into the aforementioned equation, we obtain

\[
q(x) > 1 - \frac{1}{1 + R(x)}.
\]

Inequality (24) specifies the lower bound of the forwarding probability \(q(x)\) for the \(x\)-hop node, with which we can ensure that (23) is valid, and thus, we can utilize (22) to provide a fair throughput to all of the nodes irrespective of how far they are from the gateway.

The aforementioned discussion is focused on the per-node throughput. For the network throughput, we review (13), which states that the aggregate network throughput is equal to the number of one-hop nodes times the arrival rate of the transmission opportunities for the one-hop node. This product can be interpreted as the total resource allocated to the one-hop nodes on uplink (note that in this paper, we only consider the upstream traffic). By combining (1) and (3) into (13), we obtain

\[
T_{agg} = N \cdot h(1) \cdot \frac{1}{t_e} \ln\left(\frac{1}{1 - p(1)}\right) \approx \frac{N \cdot h(1) \cdot p(1)}{t_e}.
\]

Obviously, as the successful channel-access probability for the one-hop node, i.e., \(p(1)\), increases, the aggregate network throughput increases. However, this increase in throughput may be at the expense of fairness. Specifically, due to the limited amount of total radio resource, if the successful channel-access probability for the one-hop node is too high such that (22) is violated, it is impossible to provide fairness. Thus, an appropriate approach to maximizing the system throughput is to maximize \(p(1)\) with the constraints given by (22).
B. Delay Minimization

Next, we study the delay-minimization problem based on the result of our analytical model. Similar to the system-throughput maximization problem mentioned in the previous section, we consider the delay-minimization problem with constraints on fairness. Specifically, the relation between \( p(x) \) and \( p(x + 1) \) must satisfy (22) to provide fairness, and the problem is to determine the best value of \( q(x) \) which must still be lower bounded by (24) such that the end-to-end delay for the \( x \)-hop node is minimized.

Recall that (16) specifies the end-to-end delay experienced by a packet generated by the one-hop node is minimized. By (23) and a large \( K \), the steady-state queue size of \( Q_r \) at the \( x \)-hop node, i.e., \( L_r(x) \), which is given by (14), has the following approximation:

\[
L_r(x) = \frac{\rho(x)}{1 - \rho(x)} - \frac{\rho(x) [K \rho(x)^K + 1]}{1 - \rho(x)^{K+1}} \approx \frac{\rho(x)^2}{1 - \rho(x)}.
\]

(25)

Similarly, the blocking probability for \( Q_r \) at the \( x \)-hop node given by (11) approximates to zero, i.e.,

\[
P_b(x) = \frac{[1 - \rho(x)] \rho(x)^K}{1 - \rho(x)^{K+1}} \approx 0.
\]

(26)

Substituting (25) and (26) into (15), we obtain

\[
W_r(x) = \frac{1}{\mu_r(x)} + \frac{L_r(x)}{\lambda_r(x) [1 - P_b(x)]} \\
\approx \frac{1}{\mu_r(x)} + \frac{\rho(x)^2}{1 - \rho(x)} \cdot \frac{1}{\lambda_r(x)} \\
= \frac{1}{\mu(x) \cdot q(x) - N_r(x) \cdot \mu(x + 1)}, x = 1, 2, \ldots, H - 1.
\]

Since \( N_r(x) \) is constant and the values of \( \mu(x) \) and \( \mu(x + 1) \) have been determined in the previous section, a larger \( q(x) \) leads to a smaller \( W_r(x) \). This means that we can let \( q(x) = 1 \) to minimize the end-to-end packet delay. If all nodes’ forwarding probabilities are set to one, each time a node obtains a transmission opportunity, it will check its \( Q_r \) first, and only when no relayed packets are queued in the \( Q_r \) will it send its own packet. Although this approach theoretically provides the minimum end-to-end packet delay, each relaying node may risk being starved by other nodes in real-world networks due to such excessive selflessness. For instance, each one-hop node may always be busy in relaying other nodes’ packets since its \( Q_r \) has a low probability of being empty. To eliminate this risk, we suggest using a large \( q(x) \), but a minimum throughput is still guaranteed for each node. Observe the throughput at the one-hop node given by (12).

\[
T(1) = \sigma_s(1) = \mu(1) - \mu(1) \cdot q(1) \cdot [1 - P_b(1)].
\]

Suppose that the minimum guaranteed throughput is \( T_{\text{min}} \). For a one-hop node, the following inequality must hold:

\[
T(1) = \mu(1) - \mu(1) \cdot q(1) \cdot [1 - P_b(1)] \geq T_{\text{min}}.
\]

Now, suppose that more relayed traffic than expected comes to the one-hop node, i.e.,

\[
P_b(1) \to 0.
\]

The following inequality must hold to guarantee a minimum throughput:

\[
\lim_{P_b(1) \to 0} (\mu(1) - \mu(1) \cdot q(1) \cdot [1 - P_b(1)]) \geq T_{\text{min}}.
\]

Thus, the upper bound of \( q(1) \) is given by

\[
q(1) \leq 1 - \frac{T_{\text{min}}}{\mu(1)}.
\]

(27)

Similarly, we can obtain the upper bound of the forwarding probability for the \( x \)-hop node. Thus, we obtain

\[
q(x) \leq 1 - \frac{T_{\text{min}}}{\mu(x)} \equiv q_u(x), \quad x = 1, 2, \ldots, H.
\]

(28)

Setting \( q(x) \) to the upper bound specified in (28) minimizes the end-to-end delay for each node while a minimum throughput is still guaranteed. However, this approach may be too conservative. Setting \( p(x) \) based on (22) leads to

\[
\mu(x) \leq \mu(x - 1), \quad x = 2, 3, \ldots, H
\]

which yields

\[
1 - \frac{T_{\text{min}}}{\mu(x)} \leq 1 - \frac{T_{\text{min}}}{\mu(x - 1)}, \quad x = 2, 3, \ldots, H,
\]

and thus

\[
q_u(x) \leq q_u(1), \quad x = 2, 3, \ldots, H.
\]

However, since the one-hop nodes have heavier traffic load than the other nodes, \( Q_r \) at the one-hop node is most likely nonempty, i.e., \( P_b(1) \) is most likely zero. Thus, we only need to consider the risk of starvation at one-hop nodes rather than at the other nodes in the network. In other words, for \( x = 2, 3, \ldots, H \), setting \( q(x) \) to its upper bound specified in (28) may be too conservative, degrading the reduction of delay. Thus, we propose a more progressive approach which sets the forwarding probability for each node to \( q_u(1) \), i.e., we let

\[
q(x) = 1 - \frac{T_{\text{min}}}{\mu(1)}, \quad x = 1, 2, \ldots, H.
\]

(29)

V. Simulation

In this section, we validate our analytical model and evaluate the effectiveness of our network design strategies proposed in Section IV. We consider two scenarios in the simulation. In the first scenario, we fix the total number of nodes in the network...
and compare the influence of different settings of the successful channel-access probability $p(x)$ and the forwarding probability $q(x)$ on the throughput and the delay. In the second scenario, we vary the number of nodes in the network to see the influence of the network size on the performance of our strategies.

### A. Scenario 1

In this scenario, there are 126 nodes and one gateway in the network, as shown in Fig. 4(a). Each node has a common transmission range of 100 m and has one radio. The routing and forwarding strategies are the same, as described in Section II. We adopt a TDMA-based system, in which each time slot is granted to an $x$-hop node with probability $p(x)$. For simplicity, we do not consider spatial reuse in the simulation. Thus, only one node is allowed to send within one time slot. All nodes operate on the same channel. The channel capacity is 75 Mb/s. The packet size is set to 1500 B. The length of a time slot is set to the amount of airtime needed for transmitting one packet, i.e., $1500 	ext{ B} / 75 	ext{ Mb/s} = 0.16 \text{ ms}$. The buffer size of $Q_r$ for each node is fixed at 64 packets.

We consider the following four different combinations of $p(x)$ and $q(x)$ in the first scenario.

1) Each node has a common successful channel-access probability, i.e., $p(1) = p(2) = \cdots = p(6)$, and a common forwarding probability of 0.7.
2) The setting of $p(x)$ is the same as 1) but the common forwarding probability is 0.9.
3) $p(x)$ is set based on (22) for providing a fair throughput, and $q(x)$ is set to the lower bound specified in (24).
4) $p(x)$ is set to the same value as in 3) but $q(x)$ is set based on (29) for delay minimization.

We abbreviate these four settings by CF0.7-N126, CF0.9-N126, FT-N126, and FTDM-N126, respectively. The four settings are described in more detail in Tables I and II. Note that in our simulations, we do not allow idle time slots, i.e., each time slot is granted to at least one node; thus, the above four settings all satisfy the following equation:

$$
\sum_{x=1}^{x=H} N(x)p(x) = 1.
$$

Fig. 5 shows the average per-node throughput for the one-to-six-hop nodes. We can see that the analytical curves (labeled with prefix "a-") well match the simulation results (labeled with prefix "s-"). From the curve of CF0.7-N126, we find that when the nodes employ a moderate common forwarding probability (i.e., 0.7) and have a common successful channel-access probability, the throughput rapidly decreases as the distance from the gateway increases. However, when a larger forwarding probability (i.e., 0.9) is used, the decrease in throughput is mitigated. Moreover, the nodes farther away from the gateway may have a higher throughput than the nodes closer to the gateway. This is because when the forwarding probability is large, the packets of nodes farther away from the gateway may

---

### Table I

**SETTINGS OF THE SUCCESSFUL CHANNEL-ACCESS PROBABILITIES**

<table>
<thead>
<tr>
<th>Setting</th>
<th>$p(1)$</th>
<th>$p(2)$</th>
<th>$p(3)$</th>
<th>$p(4)$</th>
<th>$p(5)$</th>
<th>$p(6)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF0.7-N126</td>
<td>7.937</td>
<td>7.937</td>
<td>7.937</td>
<td>7.937</td>
<td>7.937</td>
<td>7.937</td>
</tr>
<tr>
<td>CF0.9-N126</td>
<td>7.937</td>
<td>7.937</td>
<td>7.937</td>
<td>7.937</td>
<td>7.937</td>
<td>7.937</td>
</tr>
<tr>
<td>FT-N126</td>
<td>38.46</td>
<td>18.32</td>
<td>10.99</td>
<td>6.868</td>
<td>4.029</td>
<td>1.852</td>
</tr>
<tr>
<td>FTDM-N126</td>
<td>38.46</td>
<td>18.32</td>
<td>10.99</td>
<td>6.868</td>
<td>4.029</td>
<td>1.852</td>
</tr>
<tr>
<td>CF0.7-N168</td>
<td>5.952</td>
<td>5.952</td>
<td>5.952</td>
<td>5.952</td>
<td>5.952</td>
<td>5.952</td>
</tr>
<tr>
<td>FTDM-N168</td>
<td>28.85</td>
<td>13.74</td>
<td>8.242</td>
<td>5.151</td>
<td>3.022</td>
<td>1.374</td>
</tr>
</tbody>
</table>

**Table II**

**SETTINGS OF THE FORWARDING PROBABILITIES**

<table>
<thead>
<tr>
<th>Setting</th>
<th>$q(1)$</th>
<th>$q(2)$</th>
<th>$q(3)$</th>
<th>$q(4)$</th>
<th>$q(5)$</th>
<th>$q(6)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF0.7-N126</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>CF0.9-N126</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>FT-N126</td>
<td>0.962</td>
<td>0.909</td>
<td>0.842</td>
<td>0.741</td>
<td>0.551</td>
<td>0</td>
</tr>
<tr>
<td>FTDM-N126</td>
<td>0.975</td>
<td>0.975</td>
<td>0.975</td>
<td>0.975</td>
<td>0.975</td>
<td>0</td>
</tr>
<tr>
<td>CF0.7-N168</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>FTDM-N168</td>
<td>0.975</td>
<td>0.975</td>
<td>0.975</td>
<td>0.975</td>
<td>0.975</td>
<td>0</td>
</tr>
</tbody>
</table>
have a higher probability of being successfully forwarded to the gateway. On the other hand, when the forwarding probability is large, the nodes closer to the gateway may have less chance of sending its own data packets. For the FT-N126 and the FTDM-N126 settings, a fair and favorable throughput is provided to the nodes irrespective of where they are located. It shows that our strategy for fairness is effective and efficient.

Fig. 6 shows the average number of packets buffered in $Q_r$ for the one- to five-hop nodes (note that the six-hop node has no relayed traffic; thus, its $Q_r$ is always empty). We find that with a common successful channel-access probability, $Q_r$ is almost full at any one- to five-hop nodes. This is because the traffic pattern is skewed in the network, i.e., the nodes closer to the gateway have higher relayed traffic load. As a result, using a common successful channel-access probability may overload a relaying node, leading to a perpetually full $Q_r$. On the other hand, our strategy for fairness results in a small queue size for each node. This is because our strategy adjusts the successful channel-access probability according to the expected load of the node such that the service and arrival rates of packets for each node can be balanced. Moreover, from the curve of FTDM-N126, we find that with our delay-minimization settings, the queue size can be further reduced. It proves that our strategy is effective.

Fig. 7 shows the average end-to-end packet delay for nodes. Compared with the FT-N126 and FTDM-N126 settings, using a common successful channel-access probability leads to a higher end-to-end packet delay for nodes. In addition, for the nodes farthest from the gateway, our strategy for fairness can reduce the end-to-end packet delay by about 29% compared with using the CF(0.9) settings, and our strategy for delay minimization can further reduce the end-to-end packet delay by about 60%.

Figs. 8 and 9 show the average per-node throughput and the average end-to-end packet delay, respectively. They show that our strategy for fairness outperforms the approaches of using the common forwarding probabilities and has a 384% throughput improvement and a 61% delay improvement. More importantly, our delay-minimization mechanism can further
reduce the average end-to-end packet delay by about 29% without any negative impact on the throughput.

B. Scenario 2

In this scenario, we increase the number of nongateway nodes from 126 to 168 in the network, as shown in Fig. 4(b). We compare the following four different settings.

1) There are 126 nodes in the network. All nodes employ a common successful channel-access probability, i.e., $p(1) = p(2) = \ldots = p(6)$, and a common forwarding probability of 0.7.

2) There are 168 nongateway nodes in the network. All nodes employ a common successful channel-access probability and a common forwarding probability of 0.7.

3) There are 126 nongateway nodes in the network. $p(x)$ is set based on (22) for fairness, and $q(x)$ is set based on (29) for delay minimization.

4) There are 168 nongateway nodes in the network. $p(x)$ is set based on (22) for fairness, and $q(x)$ is set based on (29) for delay minimization.

Similarly, we use CF0.7-N126, CF0.7-N168, FTDM-N126, and FTDM-N168 to label these four settings, respectively. The detailed values of these four settings can be found in Tables I and II. Again, we do not allow idle time slots so that the above four settings all satisfy (30).

Fig. 10 shows the average per-node throughput for the one-to-six-hop nodes. It shows that the analytical curves well match the simulation results in all cases. It also shows that our strategy for fairness can effectively guarantee a fair throughput for all nodes no matter where they are and no matter how many nodes are placed in the network. Fig. 11 shows the average number of packets queued in $Q_r$ for intermediate nodes at different distances from the gateway. Obviously, our strategies produce a lower queue size both in the cases of 126 and 168 nodes compared with the CF(0.7) settings. The average end-to-end packet delay is shown in Fig. 12. Compared with the CF(0.7) settings, our strategies successfully reduce the end-to-end packet delay for the six-hop nodes by 92% and 93% in the cases of 126 and 168 nodes, respectively.

VI. CONCLUSION

In this paper, we model the location-dependent throughput and delay in wireless mesh networks. Our analysis is based on a general network model with two key parameters, namely, the successful channel-access probability and the forwarding probability for the $x$-hop node. We analyze the packet-arrival and the packet-departure distributions for the relaying nodes. We then derive the location-dependent throughput and the end-to-end packet delay for the $x$-hop node. Based on this derivation, we further study the influence of the successful channel-access probability and the forwarding probability on the network performance, including the throughput for the $x$-hop node, the system throughput, and the end-to-end packet delay for the $x$-hop node. We determine the settings of the successful channel-access probability and the forwarding probability for each node, with which a fair throughput can be provided to the nodes and the end-to-end packet delay can be
can dramatically reduce the delay without any negative impact on the network. In addition, our strategy for delay minimization matches the simulation results and, thus, successfully models the location-dependent throughput and delay in wireless mesh networks. Moreover, with our strategy for fairness provisioning, a fair throughput can be provided to all nodes no matter where they are located and no matter how many nodes are placed in the network. In addition, our strategy for delay minimization can dramatically reduce the delay without any negative impact on the throughput.

Fig. 13. Average per-node throughput (Scenario 2).

Fig. 14. Average end-to-end packet delay (Scenario 2).

minimized. The simulation results show that our analysis well matches the simulation results and, thus, successfully models the location-dependent throughput and delay in wireless mesh networks. Moreover, with our strategy for fairness provisioning, a fair throughput can be provided to all nodes no matter where they are located and no matter how many nodes are placed in the network. In addition, our strategy for delay minimization can dramatically reduce the delay without any negative impact on the throughput.

REFERENCES


Wanjun Liao (S’96–M’97–SM’06) received the B.S. and M.S. degrees from the National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1990 and 1992, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1997.

She has been with the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, as an Assistant Professor since 1997, where she has been a Full Professor since August 2005. Her research interests include wireless networks, multimedia networks, and broadband access networks.

Dr. Liao is currently an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON MULTIMEDIA. She has served as the Technical Program Committee Chair/Cochair of many international conferences, including the Tutorial Cochair of IEEE INFOCOM 2004, the Technical Program Vice Chair of the IEEE Globecom 2005 Symposium on Autonomous Networks, and the Technical Program Cochair of the IEEE Globecom 2007 General Symposium. She has been the recipient of many research awards. Papers she has coauthored with her students received the Best Student Paper Award at the First IEEE International Conference on Multimedia and Expo in 2000 and the Best Paper Award at the First IEEE International Conference on Communications, Circuits, and Systems in 2002. She was the recipient of the K. T. Li Young Researcher Award from the Association for Computing Machinery in 2003 and the Distinguished Research Award from the National Science Council, Taiwan, in 2006.

Authorized licensed use limited to: National Taiwan University. Downloaded on January 20, 2009 at 23:23 from IEEE Xplore. Restrictions apply.