Performance Analysis of Multi-hop Wireless Backhaul Networks Supporting Multiple-class Traffic

Ping-Chen Lin and Ray-Guang Cheng
Department of Electronic Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan, R.O.C.
Email: crg@mail.ntust.edu.tw

Li-Hung Liao
Networks & Multimedia Institute
Institute for Information Industry
Taipei, Taiwan, R.O.C.

Abstract—A multi-hop wireless backhaul network (WBN) is a high-speed wireless network that employs the wireless links provided by relay nodes to offer high-speed data access services in outdoor environment. In WBN, the chain topology is the commonly adopted network topology and IEEE 802.11 DCF is widely used medium access control protocol. However, use of current 802.11 DCF media access protocol in a chain-based WBN can result in very low throughput and severe unfairness problems. To resolve these problems, a Ripple protocol has been proposed to enhance the throughput of DCF and a time-based scheduling method was developed to deal with the unfairness problem. In this paper, we proposed a queuing model to derive the mean delay of the WBN adopting Ripple protocol and time-based scheduling method. The mean delay of the WBN accommodating single-class traffic and multiple-class traffic with strict-priority service discipline were derived. Simulation results indicate the accuracy of the analysis and the effectiveness of the proposed model.

Keywords - multi-hop wireless backhaul network, multiple-class traffic, times-based scheduling

I. INTRODUCTION

A wireless backhaul network (WBN) is a high-speed multi-hop wireless network which aims to offer low-cost public wireless access services in outdoor environment. In WBN, nomadic users may forward traffic to and from a gateway via wireless links provided by non-mobile relay nodes. Figure 1 illustrates the network architecture of an outdoor public WBN deployed along main streets of Taipei City, named ‘WIFLY Wireless Broadband Network Service.’ In this chain-based WBN, several access points (APs) act as relay nodes, Ni. The relay node that connects to the wired Internet is referred as a gateway (GW). Each relay node is equipped with two air interface cards. The relay node uses IEEE 802.11b/g interface to serve nomadic users within its coverage area and utilizes IEEE 802.11a interface to relay traffic for the other relay nodes.

Unfortunately, use of current 802.11 DCF media access protocol for such systems can result in very low throughput and severe unfairness problems. Li et al. [1] demonstrated that 802.11 DCF performs very poor in such a multi-hop environment due to severe packet collision. They showed that DCF with RTS/CTS can achieve only 1/7 of the maximum attainable physical layer throughput in a chain-based WBN. Gambiroza et al. found that DCF results in severe unfairness and even starvation for flows that are an increasing number of hops away from a wired Internet gateway [2]. They illustrated a so-called ‘Parking Lot’ problem in which the end-to-end throughput from a relay node to the gateway will depend on its relative location to the gateway.

Cheng et al. developed a collision-free MAC protocol, named Ripple [3], to enhance the throughput of DCF in a chain-based WBN. It was shown that Ripple protocol may achieve the ideal capacity (i.e., 1/3 of the maximum physical layer throughput) in an error-free wireless channel. However, it was demonstrated in [4] that ‘Parking Lot’ problem still results in serve unfairness even if Ripple protocol is adopted. Hence, Liao et al. [4] proposed a time-based scheduling (i.e., buffer management) method on top of Ripple in order to resolve the ‘Parking Lot’ problem. The time-based buffer management method ensures a first-come-first-served service policy in each relay node and thus, the gateway node may receive packets in the order as they enter the WBN. They further suggested an approximate queueing model to analyze the goodput and the mean delay of the time-based buffer management method [5]. However, the mean delay analysis is only available for single-class traffic and with errors. In this paper, we will extend their work and provide accurate mean delay analysis for single-class traffic and multiple-class traffic adopting strict-priority service discipline (i.e., bandwidth management). Simulations were conducted to verify the accuracy of the analysis and the results indicate the accuracy of the analysis. The proposed model can be used along with the call admission control policy to ensure the quality of services of users in such a chain-based WBN.

The rest of the paper is organized as follows. The system model adopted by this paper and the detail of the analysis were described in Section II. The numerical results were given in Section III. Conclusions and future work are finally drawn in Section IV.
II. SYSTEM MODEL

In this paper, a chain-based WBN adopting Ripple protocol under an error-free wireless channel is considered. Without losing generality, this work illustrates the cases of unidirectional uplink transmission, where packets are only sent from a leaf node (certain relay node) to the root node (gateway). Bidirectional transmission is possible either with time-division duplex (TDD) using one channel or frequency-division duplex (FDD) using two individual channels [3].

A system model of a chain-based WBN accommodating m classes of traffic is illustrated in Fig. 2. There are n relay nodes and each relay node has m individual queues to accommodate m classes of traffic. The packet arrival of class j traffic at the relay node i is assumed to follow a Poisson process with rate $\lambda_{j,i}$, $i = 1, \ldots, n$, and $j = 1, \ldots, m$.

In each relay node, a two-level QoS scheduler hierarchy is adopted to schedule the transmission of multiple-class packets. At the buffer management level, a time-based scheduling method is used to enforce a first-come-first-served service policy for each individual queue. At the bandwidth management level, each relay node may choose a strict-priority or a weighted-fair-queueing (WFQ) service discipline to allocate bandwidth among m classes of traffic.

WBN acts as interconnected queues with exponential arrival rates $\lambda_{j,i}$ ($i = 1, \ldots, n$; $j = 1, \ldots, m$) and constant service rate $\mu$ due to Ripple protocol [5]. Jackson theorem cannot be applied to analyze such a network because the service time of the relay nodes is not exponentially distributed and the transmission order of packets are altered by two-level QoS scheduler at each queue.

In this paper, a TDM-based queueing model as shown in Fig. 3 is utilized to model the queueing behavior of the chain-based WBN. We focus on the first relay node $N_1$ and model the remaining relay nodes as a single queue. As mentioned, relay node $N_1$ adopts Ripple protocol to access the shared wireless medium. In Ripple, each relay node circulates among TX, RX, and Listen states in sequence. Each relay node spends a fixed time interval $T$ to stay in each of the TX, Listen, and RX states. $T$ is the time required to transmit a DATA frame, which is given by

$$T = T_{RTS} + T_{CTS} + T_{DATA} + T_{ACK} + 4SIFS,$$

where $T_{RTS}$, $T_{CTS}$, $T_{DATA}$, and $T_{ACK}$ are the time required to transmit and frames, respectively. SIFS is the inter-frame space duration defined in IEEE 802.11.

In other words, packets in the relay node $N_1$ are time-division multiplexed in a scheme whereby the time axis is divided into three-slot frames with only one slot dedicated to the packet transmission. As shown in Fig. 3, each frame is $3T$ time units long and the packet transmissions can start only at the beginning of the TX slot. Each slot is $T$ time units long and can carry a single DATA frame. The queueing delay of this TDM system can be modeled as an M/D/1 queue with vacations [6]. When there are no packets in the queue at the beginning of a frame, the server takes a vacation for $3T$ time units. The queueing delay of the TDM system is given by [6]

$$W_{TDM} = \frac{3T}{2(1 - 3T\lambda)},$$

where $\lambda$ is the arrival rate of the TDM system. Note that the server time takes to transmit a packet in the TDM system is only $T$ time units.

In the following two sub-sections, the mean delay that each packet spent in the WBN will be derived for single-class and multiple-class traffic, respectively. The case of single-class traffic ($m = 1$) will be first investigated and the results are then extended to accommodate multiple-class traffic.

A. Single-class Traffic

For single-class traffic ($m=1$), WBN is modeled as a TDM system with single queue, as shown in Fig. 4. The aggregate arrival rate of the class 1 traffic, $\lambda_1$, is given by

$$\lambda_1 = \sum_{j=1}^{m} \lambda_{j,i}.$$

The queueing delay of the WBN, $W_{a1}$, consists of two parts. The first part is the queueing delay introduced at each individual queue, which is denoted as $W_{TDM,i}$ and can be obtained from the TDM system by substituting $\lambda$ by $\lambda_i$ in (3). It gives

$$W_{TDM,i} = \frac{3T}{2(1 - 3T\lambda_i)}.$$

The second part is transmission delay due to multi-hop relay (i.e., from the entry relay node to the first relay node $N_1$), which is denoted as $W_{R1}$. $W_{R1}$ can be obtained by finding an
entry node $N_i$ and calculating the time it required to relay a packet to the first relay node $N_i$. Note that the transmission time used to relay a packet from $N_i$ to $N_j$ is $(i-1)T$. Hence, $W_{R,i}$ is given by

$$W_{R,i} = \sum_{j=1}^{n} \lambda_{j,i} \times (i-1)T.$$  \hfill (5)

From (4) and (5), we can have

$$W_q = W_{TDM,i} + W_{R,i} = \frac{3T}{2(1-3\lambda_i)} + \sum_{j=1}^{n} i\lambda_{j,i}T. \hfill (6)$$

The mean delay of a packet spent in the chain-based WBN in second, which denotes as $W_i$, is equal to the mean queueing delay plus the service time, which gives

$$W_i = W_q + T = \frac{3T}{2(1-3\lambda_i)} + \sum_{j=1}^{n} i\lambda_{j,i}T. \hfill (7)$$

B. Multiple-class Traffic

The queuing model for multiple-class traffic is illustrated in Fig. 5. There are $m$ individual queues, which is responsible for holding class $j$ traffic. The aggregate arrival rate of the class $j$ traffic, $\Lambda_j$, is given by

$$\Lambda_j = \sum_{i=1}^{n} \lambda_{j,i}. \hfill (8)$$

Figure 5. The TDM queuing model for multiple-class traffic.

In this system, a strict-priority service discipline is adopted to allocate bandwidth for each traffic class. Assume the class 1 traffic has the highest priority and class $m$ has the lowest priority. This system can be modeled by an M/D/1 queue with vacations combined with non-preemptive priority queue [7]. The mean queueing delay of class $j$ traffic in this system is given by [7]

$$W_{TDM,j} = \frac{3T}{2(1-3\lambda_j)(1-3\sum_{i=1}^{n} \lambda_{i,j})}. \hfill (9)$$

Similarly, the mean queueing delay of the $j$ th traffic class, $W_{R,j}$, consists of two parts. The first part is the mean queueing delay introduced at each individual queue, which is denoted as $W_{TDM,j}$ and can be obtained from the TDM system by substituting in (8)

The second part is transmission delay due to multi-hop relay (i.e., from the entry relay node to the first relay node $N_j$), which is denoted as $W_{R,j}$. $W_{R,j}$ can be obtained by finding an entry node $N_j$ and calculating the time it required to relay a packet to the first relay node $N_j$. $W_{R,j}$ is given by

$$W_{R,j} = \sum_{i=1}^{n} \frac{\lambda_{j,i}}{\lambda_j} \times (i-1)T. \hfill (10)$$

From (9) and (10), we can have

$$W_{q,j} = W_{TDM,j} + W_{R,j} = \frac{3T}{2(1-3\sum_{i=1}^{n} \lambda_{i,j})(1-3\sum_{j=1}^{m} \lambda_j)} + \sum_{i=1}^{n} \frac{i\lambda_{j,i}T}{\lambda_j}. \hfill (11)$$

The time that a class $j$ packet spent in the chain-based WBN in second, which denotes as $W_j$, is equal to the mean queueing delay plus the service time, which gives

$$W_j = W_{q,j} + T = \frac{3T}{2(1-3\sum_{i=1}^{n} \lambda_{i,j})(1-3\sum_{j=1}^{m} \lambda_j)} + \sum_{i=1}^{n} \frac{i\lambda_{j,i}T}{\lambda_j}. \hfill (12)$$

III. NUMERICAL RESULTS

In this section, the accuracy of the analysis will be verified via computer simulation. The simulation results are obtained based on a C-based platform. In this section, the number of the relay nodes is set to be 7 ($n=7$). Both single-class traffic and multi-class traffic are investigated. In the following figures, each sample was obtained by averaging 1000 outcomes and each outcome was collected within 30 seconds.

The parameters used in the simulation are summarized in Table I, which can be obtained from 802.11a Standard [8]. From [8], it is found that SIFS is equal to 16 $\mu$s and the time required to transmit a given Frame is derived by

$$T_{\text{(Frame)}} = 20 + \frac{16 + \text{Length} + 6}{N\text{bps}} \times 4,$$  \hfill (13)

\text{(unit: $\mu$s)}.$$

<table>
<thead>
<tr>
<th>Frame Type</th>
<th>Length (bit)</th>
<th>Link Speed (Mbps)</th>
<th>Ndbps (bit/symbol)</th>
<th>T_{\text{(Frame)}} (\mu s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTS</td>
<td>160</td>
<td>24</td>
<td>96</td>
<td>28</td>
</tr>
<tr>
<td>CTS or ACK</td>
<td>112</td>
<td>24</td>
<td>96</td>
<td>28</td>
</tr>
<tr>
<td>DATA</td>
<td>12000</td>
<td>54</td>
<td>216</td>
<td>252</td>
</tr>
</tbody>
</table>

The time required to transmit a DATA frame, $T$, can be calculated from Table I and (1), which gives $T = 400 \mu$s. According to [8], it can be found that the maximum goodput of WBN, $g_{\text{max}}$, is equal to 10 Mbps. Hence, the aggregated offered load was varied from 1 Mbps to 9 Mbps, which results in a utilization factor ranging from 0.1 to 0.9.

In the following, we use a traffic pattern matrix $\Gamma$ to identify the relationship among $\lambda_{i,j}$, where $\Gamma$ is defined as

$$\Gamma = \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} & \gamma_{1,4} & \gamma_{1,5} \\ \gamma_{2,1} & \gamma_{2,2} & \gamma_{2,3} & \gamma_{2,4} & \gamma_{2,5} \\ \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} & \gamma_{3,4} & \gamma_{3,5} \\ \gamma_{4,1} & \gamma_{4,2} & \gamma_{4,3} & \gamma_{4,4} & \gamma_{4,5} \\ \gamma_{5,1} & \gamma_{5,2} & \gamma_{5,3} & \gamma_{5,4} & \gamma_{5,5} \end{bmatrix}. \hfill (14)$$

And the $\lambda_{i,j}$ is given by

**TABLE I.** PARAMETERS BETWEEN SIMULATION AND SIMULATION
The case of single-class traffic was first investigated. Four cases illustrating different combinations of the arrival rates in each relay node were studied. The four traffic pattern matrices are listed below:

- Case 1: \( \Gamma = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \);
- Case 2: \( \Gamma = [7 \ 6 \ 5 \ 4 \ 3 \ 2 \ 1] \);
- Case 3: \( \Gamma = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7] \);
- Case 4: \( \Gamma = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1] \).

Figure 6 shows the results considering the traffic pattern specified in case 3. The mean delay obtained by Liao [5] and by (7) in this paper. It can be found in Fig. 6 that Liao’s results have some errors. In contrast, the proposed approach, which is referred as Lin’s Model herein, is coincided with the simulation results. Figure 7 shows the mean delay for different combinations of the arrival rates in each relay node. The mean delay of the four cases are increased in an ascending order because \( W_{R,i} \) of case 1, case 2, case 3, and case 4 are equal to 0, 2\( T \), 4\( T \), and 6\( T \), respectively.

For multiple-class traffic, the case of \( m = 4 \) was studied as an example. The mean delay of a class \( j \) traffic, which is given in (12), was investigated. Three cases with the following traffic pattern matrices were considered:

- Case A: \( \Gamma = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 & 4 & 4 & 4 \end{bmatrix} \);
- Case B: \( \Gamma = \begin{bmatrix} 4 & 4 & 4 & 4 & 4 & 4 & 4 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \);
- Case C: \( \Gamma = \begin{bmatrix} 6 & 6 & 6 & 6 & 6 & 6 & 6 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 14 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \).

Figures 8, 9 and 10 show the mean delay for case A, case B, and case C, respectively. Cases A and B illustrates a uniformly distributed traffic pattern. In both cases, the traffic arrival rate for each class of traffic is identical in each relay node. The main difference between case A and case B is that different among of arrival rate are assigned to the four-class traffic. In case A, the highest priority traffic contributes the least arrival packets while the lowest priority traffic provides the most arrival packets. In case B, the distribution of the four-class traffic is reversed. Note that, although the aggregated arrival rate of cases A and B are the same, their mean delays are quite different. In cases A and B, the highest priority class traffic has the lowest mean delay and the lowest priority traffic has the highest delay, which matches our anticipation for the strict-priority service discipline. In all of the three cases, the simulation results are all coincided with the analytical results, which verify the accuracy of the analysis for multiple-class traffic.

In Fig. 8, the mean delays of the top three traffic classes are quite small. It is because that the lowest priority dominates 40% of the offered load and thus, the WBN has spare capacity to serve the top three traffic classes. In Fig. 9, the mean delays of the four traffic classes are quite different. It is because the top two priority traffic classes dominates 70% of the offered load and thus, it results in very high mean delay for the rest of the two traffic classes. The results can be explained based on the results obtained from (9) to (12). It can be found from (10) that \( W_{R,j} \) is identical for both cases A and B. The main difference between case A and case B comes from \( W_{TDM,j} \) given in (9). From (9), it
can be found that the mean delay of lower priority traffic is highly dependent on the arrival rate of higher priority traffic. That is the reason why the mean delay of the class 4 traffic in case B is much higher than that in case A.

The mean delay for case C is illustrated in Fig. 10. This case demonstrates a randomly distributed traffic pattern, which could be the most possible case in a real application environment. In this example, only the first relay node $N_1$ is injected by the class 4 traffic. All of the relay nodes are also injected with unequally distributed higher class traffic. It can be found from Fig. 10 that the service discipline of strict-priority seems to be violated under low utilization factor. However, it can be found from (12) that $W_{R,j}$ is a dominate component of $W_j$ if the utilization factor is low. In case C, we have zero $W_{R,4}$ and non-zero $W_{R,j}$ ($j = 1, 2, 3$). Class $j$ ($j = 1, 2, 3$) packets need $W_{R,j}$ to reach $N_1$ and thus, $N_1$ may have no higher class packets buffered for transmission. In this case, class 4 packets have a higher chance to be served and may experience a lower mean delay.

![Figure 8. The mean delay adopted strict-priority (Case A)](image)

![Figure 9. The mean delay adopted strict-priority (Case B)](image)

![Figure 10. The mean delay adopted strict-priority (Case C)](image)

### IV. CONCLUSIONS

In our work, we propose a queueing model to analyze the mean delay of a chain-based WBN accommodating single-class traffic and multiple-class traffic with strict-priority service discipline. The accuracy of analysis is verified via computer simulation. The proposed queueing model can also be applied to analyze the delay and goodput of the chain-based WBN accommodating multiple-class traffic with WFQ service discipline. However, due to the space limitation, the analysis of the WFQ will be left as our future work.

### ACKNOWLEDGMENT

This work was supported in part by the National Science Council, Taiwan under Contracts, NSC 95-2219-E-011-006 and NSC 96-2219-E-011-005

### REFERENCES


