Time Slot Assignment Algorithms for Reducing Upstream Latency in IEEE 802.16j Networks

Shimpei Tanaka
Graduate School of Information Science and Technology
Osaka University, Japan
sinpei-t@ist.osaka-u.ac.jp

Go Hasegawa, Yoshiaki Taniguchi and Hirotaka Nakano
Cybermedia Center, Osaka University, Japan
{hasegawa, y-tanigu, nakano}@cmc.osaka-u.ac.jp

Abstract—In IEEE 802.16j wireless multihop networks, transmission latency from relay nodes to a gateway node is one of the important performance metrics. The transmission latency is mainly affected by a scheduling delay at every relay node, which is determined by algorithms for assigning time slots to wireless links between relay nodes. In this paper, we propose 2 kinds of time slot assignment algorithms for upstream wireless links in IEEE 802.16j multihop networks. One of the proposed algorithms assigns time slots considering the hop count from a gateway node, and the other takes the path from the relay node to the gateway node into account. We evaluated the performance of the proposed algorithms through simulation experiments and confirmed that our algorithms can decrease the upstream latency by up to 15% compared with the existing method, without increasing the average transmission latency of the entire network.

Keywords—IEEE 802.16j, wireless multihop network, upstream, time slot, latency.

I. INTRODUCTION

IEEE 802.16j wireless multihop networks [1] have received a significant amount of attention as a network technology providing a wide-area broadband wireless access environment at low cost. As depicted in Figure 1, in an IEEE 802.16j network, each relay node connects to other nodes with wireless links so that the overall topology becomes a tree structure, unlike the star structure in typical IEEE 802.11-based networks. In general, IEEE 802.16j consists of three kinds of nodes: gateway nodes that have wired connections to external networks, relay nodes that are interconnected with other nodes by wireless links, and user nodes that are connected to the nearest relay node [2]. Generally, the wireless channel used for communication between relay nodes and user nodes is different from that used among the relay nodes and the gateway nodes, and the wireless channel used for upstream communication is different from that used for downstream communication [3]. In this research, we ignore user nodes and focus on the communication between relay nodes.

One problem in wireless networks in general is that 2 nodes that exist in transmission range of each other cannot communicate simultaneously due to radio interference [4]. To solve this problem, IEEE 802.16j uses a time-division scheduling mechanism based on Orthogonal Frequency Division Multiplexing Access (OFDMA) at the MAC layer [5]. In the OFDMA-based mechanism, time is divided into constant intervals called frames, each of which consists of multiple time slots of constant time duration. The time slots are assigned to links in the network as communication opportunities, and communication on the links can take place only at the assigned time slots [6]. The gateway node performs centralized control of the time slot assignment for the links, while considering interference relationships in order to avoid radio interference.

In such wireless multihop networks using the time-division scheduling mechanism, we cannot ignore a scheduling delay at each relay node during packet transmission between relay nodes. The scheduling delay is defined as the period of time between the arrival of a packet at a relay node and the departure of the packet at the assigned time slot for the relay node. The end-to-end transmission latency between a relay node and a gateway node increases due to accumulation of scheduling delays at each relay node on the path between the relay node and the gateway node. The degree of scheduling delay is mainly dependent on the time slot assignment to the wireless links in the network.

We have already proposed a time slot assignment algorithm for reducing scheduling delay and evaluated its performance in [7]. The proposed algorithm tends to assign time slots to links in order of the density of interference relationships. Therefore, the links with small hop count from the gateway node obtain earlier time slots, and the links with large hop count from the gateway node obtain later time slots, when traffic demand is concentrated at the gateway node. As a result, the method in [7] can
decrease the scheduling delay and the transmission latency for downstream transmissions from the gateway node to the relay nodes, compared with random method. On the other hand, it decreases a little in the scheduling delay at upstream transmissions from the relay nodes to the gateway node.

Therefore, in this paper, we focus on the upstream transmission in IEEE 802.16j wireless multihop networks and propose 2 kinds of time slot assignment algorithms to give small transmission latency from the relay nodes to the gateway node. Our proposed algorithms aim to decrease the scheduling delay at each relay node on the path between the starting relay node and the gateway node. One of the proposed algorithms assigns time slots by considering the hop count from the gateway node, and the other takes the path from the starting relay node to the gateway node into account. Performance evaluation of the proposed algorithms was conducted through packet-level simulation experiments. The evaluation results showed that the proposed algorithms can improve the average transmission latency as compared with an existing algorithm described in [7] without increasing the average transmission latency of the entire network. The rest of this paper is organized as follows. In Section II, we describe the model of IEEE 802.16j wireless multihop networks. In Section III, we propose 2 kinds of time slot assignment algorithms. In Section IV, we present simulation evaluation results. Finally, in Section V, we conclude this paper and describe future work.

II. IEEE 802.16j Wireless Multihop Network

IEEE 802.16j uses an OFDMA-based mechanism for avoiding radio interference. When assigning time slots to links between relay nodes, the connections and the interference relationships between relay nodes are very important. In this section, we describe the network model and notation of the IEEE 802.16 wireless multihop network. We also explain the radio interference model and time slot assignment mechanisms based on TDMA.

A. Network Model

Figure 2(a) depicts a directed communication graph $G_c = (V_c, E_c)$ that indicates the communication relationship between relay nodes and a gateway node. $V_c = \{v_0, v_1, v_2, \ldots, v_n\}$ is a set of relay/gateway nodes deployed in a plane, and $v_0$ is the gateway node. We assume that there is only one gateway node in the network. $E_c$ is a set of the directed communication links $e_{i,j}$, which represents an edge directed from $v_i$ and $v_j$ when $\|v_i - v_j\| < t_i$. Here, $t_i$ is the communication range of $v_i$, and $\|v_i - v_j\|$ is the distance between $v_i$ and $v_j$. We define the hop count of a directed communication link $e_{i,j}$ as the larger hop count between 2 nodes $v_i$ and $v_j$.

In this paper, we assume that the gateway node connects to an external network, and each relay node communicates with the gateway node via other relay node(s) on the path between the relay node and the gateway node. The path between the relay node and the gateway node is determined by a routing algorithm, and the directed transmission graph $G_t = (V_t, E_t)$ is constructed as a tree-like graph whose root is the gateway node $v_0$, as shown in Figure 2(b). Here, $G_t$ is a subset of $G_c$. $V_t$ is a set of nodes satisfying $V_t = V_c$, and $E_t$ is a set of directed transmission links that is determined by the routing algorithm and satisfies $E_t \subseteq E_c$.

In what follows, the directed transmission links $e_{i,j} (\in E_t)$ are called links, the links on the path from a relay node to the gateway node are called upstream links, and the links on the path from the gateway node to a relay node are called downstream links. We assume that $G_t$ is given in advance by the routing algorithm, and each link $e_{i,j}$ has a link weight $w_{i,j}$ that represents the required time slots according to the traffic load. In this paper, we consider algorithms to assign time slots only to upstream links in the network.

B. Interference Model

In this paper, we use the radio interference model proposed in [8]. The model defines the interference relationship from $e_{i,j}$ to $e_{p,q}$ based on the distances among 4 vertices $v_i$, $v_j$, $v_p$, and $v_q$. Each relay node $v_i$ has the interference range $r_i$. The condition to determine the interference relationship is as follows: $e_{i,j}$ interferes with $e_{p,q}$ when and only when $\|v_i - v_j\| < r_i$. On the other hand, $e_{p,q}$ interferes with $e_{i,j}$ when and only when $\|v_p - v_q\| < r_p$. On the basis of these conditions, $e_{i,j}$ and $e_{p,q}$ are in the interference relationship; that is, they cannot communicate simultaneously when $\|v_i - v_j\| < r_i$ or $\|v_p - v_q\| < r_p$ is satisfied. Typically, $r_i > t_i$, and the ratio of interference range to communication range for node $v_i$, denoted as $\gamma_i = \frac{r_i}{t_i}$, is in the range of 2–4 in practice [8]. We define a function $I(e_{i,j}, e_{p,q})$ that indicates whether or not 2 links $e_{i,j}$ and $e_{p,q}$ are in the interference relationship. The function returns 1 if the 2 links are in the interference relationship, or returns 0 if the 2 links are not in the interference relationship.

We introduce the conflict graph $F_{G_t} = (F_V, F_E)$ obtained by applying the interference model to a directed transmission graph $G_t$, as depicted in Figure 3. $F_V = \{f_e | e_{i,j} \in E_t\}$ is a set of nodes which are all elements of a link set $E_t$ in $G_t$, and $F_E = \{f_{e_p} | f_{e_{i,j}} \in F_V\}$ is a set of links which exist when $I(e_{i,j}, e_{p,q}) = 1$, that is, $F_E$ represents the interference relationship among links in $G_t$. 

Figure 2. Directed communication graph $G_c$ and directed transmission graph $G_t$. 

[Diagram of communication and transmission graphs]

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-123-6
IEEE 802.16j controls transmission opportunities, called time slots, using the TDMA mechanism and assigns time slots to links in the network. Each link can communicate only in assigned time slots. IEEE 802.16j does not assign the same time slot to links that are in the interference relationship. Meanwhile, multiple links can communicate simultaneously in one time slot when the time slot is assigned to multiple links that are not in the interference relationship. This is called spatial reuse, which enhances the network throughput [9–13], and algorithms for assigning time slots are required to consider the interference relationship to increase the degree of spatial reuse. The algorithm in [7] increases the degree of spatial reuse; however, it is not applicable to upstream links in IEEE 802.16j because it would increase the upstream latency. Therefore, we propose algorithms in which priority is given to the time slot schedule over spatial reuse.

III. PROPOSED TIME SLOT ASSIGNMENT METHODS

On upstream transmissions in an IEEE 802.16j wireless multihop network, the time slot assignment to each link affects the scheduling delay and the end-to-end transmission latency. The transmission latency of the relay node that has large hop count from the gateway node is significantly affected by the time slot assignment.

In this section, we propose 2 kinds of methods for assigning time slots to reducing upstream transmission latency, based on the models described in Section II.

A. Hopcount-based Method

The hopcount-based method is based on the following idea. The order of the links for assigning time slots is determined based on the hop count of each link from the gateway node. The links with large hop count are assigned earlier time slots than those with small hop count. In detail, when a link $e_{i,j}$ is assigned a time slot(s), the hopcount-based method assigns $w_{i,j}$ time slot(s) that are not assigned to links in the interference relationship with $e_{i,j}$ and that are not assigned to links with larger hop count than $e_{i,j}$. Note that we utilize a greedy approach, meaning that we assign the earliest available time slots. Since the upstream transmissions in the network utilize the links on the path in the reverse order of hop count, we expect that the scheduling delay in upstream transmissions will be reduced by using this method.

Algorithm 1 represents the hopcount-based method in pseudo-code. $h_{i,j}$ is the hop count of $e_{i,j}$. $H_k$ is a set of the links with $k$ hop count from the gateway node, and $T_{e_{i,j}}$ is a set of the time slots assigned to $e_{i,j}$. $X_{e_{i,j}}$ is a set of the time slots assigned to the links in the interference relationship with $e_{i,j}$. $s_m$ is the $m$th time slot.

B. Path-based Method

The path-based method assigns time slots to links along with the paths from relay nodes to the gateway node. For determining the order of assigning time slots to links, the method first determines the order of paths to which the time slots are assigned. In detail, it orders the relay nodes in the network by visiting them in the depth-first order. When the method assigns time slot(s) to links, it chooses a relay node from the reverse order in which it visited relay nodes and assigns 1 time slot to each link on the path from the chosen relay node to the gateway node in descending order of hop count. We use a greedy approach in which the interference relationship is considered, as in the hopcount-based method. The method is applicable when the traffic demand on a path is determined only by the sender relay node, regardless of the path’s characteristics.

Algorithm 2 shows the pseudo-code of the path-based method. $T_{e_{i,j}}$ is a set of the time slots assigned to $e_{i,j}$. $X_{e_{i,j}}$ is a set of the time slots assigned to the links in the interference relationship with $e_{i,j}$. $s_m$ is the $m$th time slot, and $seq$ is an
algorithm 2 Path-based algorithm

INPUT: \( G_t = (V_t, E_t) \), \( F_{G_t} = (F V_t, F E_t) \), \( w_{i,j} \) of \( \forall e_{i,j} \)
OUTPUT: time slot assignment to \( \forall e_{i,j} \)

1: \( n = (\text{the number of relay nodes}) \), \( w = 0 \)
2: while \( D_w \neq \emptyset \) do
3: \( \text{seq}[n] = v_m | m \in D_w \)
4: \( n = n - 1 \)
5: \( D_w \leftarrow D_w \setminus \{m\} \)
6: if \( D_m \neq \emptyset \) then
7: \( w = m \)
8: end if
9: if \( (D_w = \emptyset) \cap (w \neq 0) \) then
10: \( w = u_w \)
11: end if
12: end while
13: for \( a = 1 \sim ((\text{the number of relay nodes}) \) do
14: \( v_k = \text{seq}[a], l = w_{k,u_k} \)
15: repeat
16: for all \( e_{p,q} \in (E_t \cap (I(e_{u,u_k}, e_{p,q}) = 1)) \) do
17: \( X_{e_{p,q}} \leftarrow T_{e_{p,q}} \cup X_{e_{p,q}} \)
18: end for
19: for \( b = 1 \sim l \) do
20: \( m = \) (the earliest time slot not in \( X_{e_{p,q}} \))
21: \( T_{e_{p,q}} \leftarrow s_m \cup T_{e_{p,q}} \)
22: end for
23: \( k = u_k \)
24: until \( (v_k = v_0) \)
25: end for

array of the relay nodes in the reverse order in which the relay nodes were visited. \( D_w \) is a set of the node numbers of downstream nodes of \( v_i \) and \( u_i \) is the node number of the upstream node of \( v_i \).

We can easily implement these 2 methods and their computing overheads are following: overhead of hopcount-based method and the existing method discussed below is \( O((a + b \times w) \times n) \) and that of path-based method is \( O((a + \beta \times w \times n) \times n) \). \( a, b, w \) and \( n \) denote computing time of interference relationships, time of determining time slot to assign, average of link weight and the number of relay nodes, respectively. The path-based method is unsuitable for network that the topology or the traffic demand is frequently changed compared with the hopcount-based method and the existing method, because time slot assignment must be done when one of them is changed.

IV. Performance Evaluation

We show the evaluation results of our time slot assignment algorithms obtained by conducting packet-level simulation experiments.

A. Evaluation Environment

We randomly located 99 relay nodes uniformly in a \( 1 \times 1 \) square area and one gateway node at the center of the area. All relay nodes had a communication range of 0.2. As described in Section II-A, after obtaining a directed communication graph based on the node location and the communication range, we constructed a directed transmission graph as a tree-like graph rooted at the gateway node and optionally minimized the hop count from the gateway node to each node. Note that the detailed implementation of the algorithm of the directed transmission graph is outside the scope of this paper, and we used the method in [14]. We determined the interference relationship among links between relay nodes using the interference model explained in Section II-B. The traffic demand of the network was uniform, and we generated one packet from a randomly chosen relay node destined to the gateway node at regular intervals, which were equal to the time slot duration. This traffic demand setting means that the weight of each link was equal to the number of paths between relay nodes and the gateway node passing through the link. For packet-level simulation experiments, we implemented a wireless multihop network simulator that can simulate the packet-level behavior of IEEE 802.16j-based networks, including topology generation from the locations of relay/gateway nodes, TDMA-based time slot assignment, and store-and-forward packet transmission based on the FIFO principle. In each experiment, we ran the simulation until 5,000 packets were generated and arrived at the gateway node. For one parameter set we conducted 3,000 simulations by changing the relay nodes’ locations. When the interference ratio is 2.5, though they change according to the topology, the average of hop count is around 2.6–3.0 and the max of that is around 9–10.

We observed the frame size and the transmission latency as performance metrics. The frame size represents the total number of time slots needed for assigning time slots to all links and is desired smaller because the large one decreases network throughput. The transmission latency is defined as the time duration from when a packet is generated at the relay node to when the packet arrives at the gateway node. Note that in the packet-level simulation, some packets were queued at some relay nodes when congestion occurred, which may have increased the end-to-end transmission latency.

We evaluated the proposed methods and the existing method in [7] as the interference ratio \( y \) was changed, taking values 1.5, 2.5 and 3.5, for all relay nodes.

B. Existing Method for Comparison

Here we describe the existing method in [7] used for comparison purposes. The method first determines the order of assigning time slots to links by using a conflict graph. The assignment order is roughly the same as the order of the degree of nodes in the conflict graph. The links are assigned time slots along this order in a greedy manner, as in the proposed methods. Since the conflict graph is likely to be dense around the gateway node and have space around the nodes far from the gateway node, the method in [7] can decrease the transmission latency for downstream
transmission. We evaluated the upstream transmission of the method to compare it with the proposed methods.

Note that we also consider the modified algorithm from the above method to possibly decrease the upstream transmission latency. The method utilizes the reverse order of time slot assignment given by the above existing method. Although we do not show the results due to space limitation, we have confirmed that the modified method cannot outperform the proposed method in this paper.

C. Frame Size

We first evaluated the frame size. The results are shown in Figures 4(a), 4(b), and 4(c) for $\gamma = 1.5$, 2.5 and 3.5, respectively. The x-axis of the graphs shows the topology ID (1 ~ 3,000), which corresponds to the simulation experiments with 3,000 patterns of node locations. The results are sorted in ascending order of the frame size of the existing method.

From Figure 4, we can see that the frame size of the existing method was the smallest, and that of the hopcount-based method was the largest for all interference ratios. In the hopcount-based method, the wireless resource efficiency was lower than that in the other methods due to spatial reuse only among links with the same hop count. On the other hand, the frame size of the path-based method was smaller than that of the hopcount-based method because of spatial reuse among all links. However, since the path-based method assigns time slots to links along the path from the relay node to the gateway node in the descending order of hop count, the ratio of spatial reuse was less than with the existing method. As a result, the frame size of the path-based method was larger than that of the existing method.

On the one hand, by comparing Figures 4(a)–4(c), we can find as the interference ratio became larger, the frame size of all methods increased, and as the interference ratio became larger, the difference in the frame sizes among the three methods decreased. This is because the large interference ratio decreased the wireless resource efficiency due to less spatial reuse.

From these results, we conclude that our proposed methods are less effective than the existing method in terms of the frame size.

D. Transmission Latency

Figures 5(a), 5(b), and 5(c) are the results of the end-to-end transmission latency for $\gamma = 1.5$, 2.5 and 3.5, respectively. The x-axis of the graph is the hop count of the relay nodes from the gateway node at which packets are generated.

From Figure 5, we can see that the transmission latency of the path-based method was the smallest for all interference ratios. Since the path-based method sequentially assigns time slots to the links on a path from a relay node to the gateway node, the links on the path are assigned close time slots. Therefore, the transmission latency of path-based method decreased. In addition, when the interference ratio became large, the path-based method showed small latency, especially for large hop count. In particular, in Figure 5(c), when the interference ratio was 3.5, for the packets generated at the relay nodes with seven or larger hop count, the path-based method reduced the transmission latency by up to 15% as compared with the existing method. On the other hand, from Figure 5(c), we can also see that the transmission latency of the hopcount-based method was smaller than that of the existing method when the interference ratio was large. This is because the scheduling delay at each relay node decreased due to the time slot assignment of the hopcount-based method. However, when the interference ratio was small, the difference in transmission latency between the hopcount-based method and the existing method became small due to the large difference in frame size of the 2 methods, as depicted Figure 4(a).

We can also observe from Figure 5 that as the interference ratio became larger, the transmission latency of all methods increased. The reason for this is the increase of the frame size shown in Figure 4. The scheduling delay at each relay node increased because the communication opportunities per unit time for each link decreased when the frame size was large. Furthermore, the transmission latency of the packets with large hop count increased notably compared with small hop count since the number of relay nodes traversed was large.

From these results, we conclude that the hopcount-based method is more effective than the existing method when the interference ratio is large, and that the path-based method is superior to the existing method for all interference ratios in terms of the upstream end-to-end transmission latency. While both of them improve the upstream end-to-end transmission latency, they do not degrade average transmission latency of the entire network. Considering the increase in the frame length of the proposed algorithms in Figure 4, the proposed algorithms can make a good trade-off between the network throughput and latency.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed 2 kinds of time slot assignment algorithms for upstream wireless links in IEEE 802.16j multihop networks to reduce the upstream end-to-end transmission latency. One of the proposed algorithms is a method based on the hop count from the gateway node. The other takes the path from relay nodes to the gateway node into account. Through simulation experiments, we confirmed that the proposed methods can reduce the upstream transmission latency by up to 15% as compared with the existing method without degrading average transmission latency of the entire network, though the methods increase the frame size.

In future work, we need to evaluate the proposed methods in other cases where parameters other than the interference ratio change, and to consider the implementation complexity and the overhead. We plan to improve the method not only to reduce the transmission latency but also to enhance the network throughput and to reduce the cost of assigning time slots, to apply the methods to other radio interference.
model using Signal to Interference and Noise Ratio (SINR) that is more realistic than the radio interference model in this paper, and to develop the methods to a dynamic scheme adapting the change of topology, the change of traffic demand, condition of wireless channel and so on.

REFERENCES


