Two-Tier Scheduling Algorithm for Uplink Transmissions in IEEE 802.16 Broadband Wireless Access Systems

Lin-Fong Chan and Hsi-Lu Chao
Department of Computer Science
National Chiao Tung University
Hsinchu, Taiwan
hlchao@cs.nctu.edu.tw

Zi-Tsan Chou
Networks and Multimedia Institute
Institute for Information Industry
Taipei 106, Taiwan
ztchou@nmi.iii.org.tw

Abstract

IEEE 802.16 Broadband Wireless Access (BWA) Systems support classes of traffic with differentiated Quality of Service (QoS). However, the detailed of how to schedule traffic are left unspecified. In this paper, based on this hierarchical QoS architecture, we propose a novel two-tier scheduling strategy, called two-tier scheduling algorithm (2TSA). The first tier is category-based and the second tier is weight-based. Both tiers are implemented at BSs. We evaluate the performance of 2TSA via simulation. The simulation results show that our proposed algorithm can achieve both QoS guarantee and fair bandwidth allocation.

Keywords–802.16, Scheduling, Quality of Service

1. Introduction

The IEEE 802.16 standard [1] is designed to satisfy various demands for higher capacity, higher data rate, and more advanced multimedia services. The advantages of IEEE 802.16 include rapid deployment, high speed data rate, high scalability, and so on.

An efficient scheduling algorithm is an essential technology to support real-time multimedia services with variable quality of service (QoS) demands. In IEEE 802.16 systems, there are four service classes defined: unsolicited grant service (UGS), real-time polling service (rtPS), non-real time polling service (nrtPS)m and best effort service (BE). Besides, IEEE 802.16 systems define several parameters for a connection to indicate its QoS demands, including minimum reserved rate, maximum sustained rate, maximum latency, and tolerated jitter. Based on traffic characteristics and supported applications, the QoS parameters of UGS connections are maximum sustained rate, maximum latency, and tolerated jitter. QoS parameters of rtPS, nrtPS, and BE, connections are (minimum reserved rate, maximum sustained rate, maximum delay), (minimum reserved rate, maximum sustained rate) and (maximum sustained rate), respectively. However, IEEE 802.16 does not have a designated scheduling algorithm to provide QoS guarantee. The majority of existing scheduling algorithms proposed for 802.16 BWA systems are strict priority-based and thus they incur the problem of starvation. In this paper, we will propose a novel scheduling algorithm to avoid starvation problem. Besides, our algorithm can provide QoS guarantee and fairness.

1.1. Related work

In [2], a QoS architecture for IEEE 802.16 BWA systems had been proposed. However, this architecture focused on admission control, and did not integrate with any scheduling protocol. In [3], the authors proposed a QoS architecture, too. Dissimilar to [3], they implemented a strict priority-based scheduling mechanism at a BS. Due to the characteristic of strict priority, [3] may have the problem of starvation. In [4], the authors proposed a strict-priority scheduling (SPS) protocol. The service priority is always UGS > rtPS > nrtPS > BE. The scheduling discipline adopted for rtPS, nrtPS, and BE is earliest deadline first (EDF), weighted fair queueing (WFQ), and first come first serve (FCFS), respectively. It is obvious that the drawback of SPS is the problem of starvation for low-priority service class. A new priority determination scheme was proposed in [5]. The priority of a connection determined by not only the service class, but also the signal-to-noise ratio (SNR). Afterward, according to all received bandwidth requests and re-calculated priorities, the BS will do bandwidth allocation. In [6], to avoid starvation, the authors suggested to set a threshold to specify the maximum bandwidth amount that each service class can use in a frame. Since [6] still adopts the concept of strict priority scheduling, BE connections may starve upon bandwidth shortage, too.

1.2. Problem statement

In this paper, we will study fair scheduling in IEEE 802.16 BWA systems. Our design goal is to guarantee each connection's minimum bandwidth requirement, and fairly allocate residual bandwidth to all connections.

The rest of this paper is organized as follows. Sec. II describes the proposed fair scheduling algorithm. Sec. III shows the simulation results. Finally, this paper is concluded in Sec. IV.

2. Two-tier scheduling algorithm (2TSA)

In this section, we describe our proposed scheduling algorithm, i.e., 2TSA, in detail. In our proposed algorithm, we only consider the TDD operating mode and uplink transmissions. Besides, we assume the uplink sub-frame occupies half a frame time. Due to the characteristic that the bandwidth for UGS connections is automatic allocated per frame, 2TSA does scheduling on rtPS, nrtPS, and BE services only.

2.1. QoS parameter notations

(1) R^i_{min} : the minimum bandwidth reserved rate of connection *i* and its unit is Kbps.
(2) R^i_{max} : the maximum bandwidth sustained rate of connection *i* and its unit is Kbps.
(3) $R^i_{allocated}$: the bandwidth allocated rate of connection *i* and its unit is Kbps.
(4) w_i : the weight of connection *i*; its value is between [0, 1]. This parameter indicates a connection's satisfaction degree. Therefore, a connection being with a smaller weight has higher allocation precedence, compared with those connections in the same class.

2.2. Service category

Based on the allocated bandwidth, each connection is classified into either three categories:
(1) Unsatisfied: connection *i* is called unsatisfied means that the allocated bandwidth is less than its specified minimum requirement, i.e., $R^i_{allocated} < R^i_{min}$.
(2) Satisfied: connection *i* is satisfied while its allocated bandwidth is between [R^i_{min}, R^i_{max}].
(3) Over-satisfied: connection *i* is over-satisfied when its allocated bandwidth is larger than its specified maximum requirement, i.e., $R^i_{\max} < R^i_{allocated}$.

2.3. Weight calculation

Based on the current service category of connection *i*, its weight is calculated by (1).

$$w_i = \begin{cases} \dfrac{R^i_{allocated}}{R^i_{min}}, & if \quad R^i_{allocated} < R^i_{min} \\ \dfrac{R^i_{allocated} - R^i_{min}}{R^i_{max} - R^i_{min}}, & if \quad R^i_{min} \le R^i_{allocated} \le R^i_{max} \\ \dfrac{R^i_{allocated} - R^i_{max}}{R^i_{allocated}}, & if \quad R^i_{max} \le R^i_{allocated} \end{cases} . \quad (1)$$

Since our proposed algorithm first considers connections' categories while executing scheduling, weight calculation strongly depends on the category. For example, if the allocated bandwidth of a connection is less than its minimum demand, its weight indicates the shortage compared to this demand. However, weights of the other two categories indicate the corresponding satisfaction degree.

2.4. Initialization and bandwidth request

Initially, each service flow sends DSA/DSC messages, with QoS parameters (i.e., R^i_{min}, and R^i_{max}) encapsulated, to its BS to do handshake. Based on the received messages, the BS executes admission control to admit or deny the requested flows.

While an admitted connection being with frames to transmit, it sends a Bandwidth Request to the BS. According to all received Bandwidth Requests and adopted bandwidth allocation algorithm, the BS broadcasts an UL-MAP to all its served SSs to inform them the scheduled transmission time and allocated bandwidth. A SS cannot start to transmit frames until its designated time.

2.5. 2TSA operations

2TSA is implemented only at BSs. The objectives are to achieve both QoS guarantee and fairness. The first-tier and second-tier scheduling is category-based and weight-based, respectively.
(1) The first-tier allocation: based on the collected bandwidth requests and updated weights, the BS classifies all connections into three categories. Dissimilar to strict-priority allocation that a BS always allocates bandwidth on the service order of (UGS connections, rtPS connections, nrtPS connections, best effort connections), 2TSA first allocates bandwidth to the "unsatisfied" category. While still being with more available bandwidth, it then allocates bandwidth to connections belonging to "satisfied" category, and followed by "over-satisfied" category. Therefore, the first-tier bandwidth allocation is to ensure that each connection can be satisfied with their minimum requirement.
(2) The second-tier allocation: for a specific category, the received bandwidth is further distributed to connections based on the parameter of weight. The smaller weight of a connection, the higher bandwidth allocation priority it has. After finishing this two-tier bandwidth allocation, the BS generates the corresponding UL-MAP and broadcasts to all

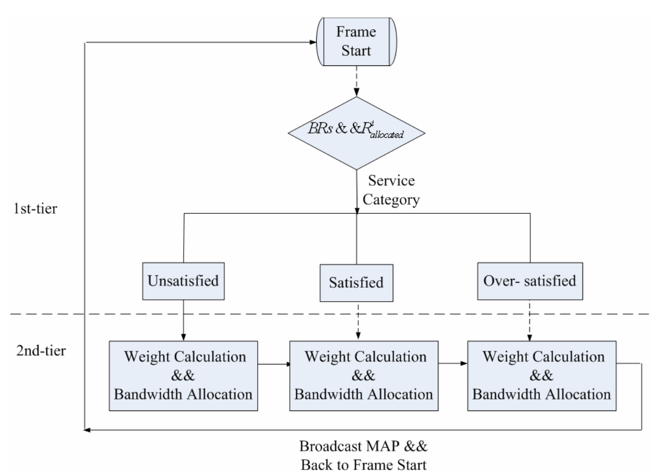

Figure1. Operation flowchart of 2TSA

served SSs. The flowchart of the proposed 2TSA is shown in Fig. 1.

3. Performance evaluation

In this section, we evaluate our proposed mechanism via simulation. We compare 2TSA with strict-priority scheduling (SPS) algorithm.

3.1. Simulation environment and performance metrics

The simulation program is based on a self-configured environment with g++ compiler in Linux. There are five UGS, seven rtPS, seven nrtPS, and seven best effort connections in the subnet served by a BS. The QoS parameters specified by each type of connections are listed in Table 1. A frame length is set to 10ms and the simulation time is 1000 frames, i.e., 10 seconds. We assume all connections are always backlogged and their sending rate is bounded by the maximum sustained rate in the first five seconds. In the next five seconds, some connections will try to send more data in order to have better throughput. That is, we assume some connections are with greedy behavior. The parameter settings are summarized and tabled in Table 1. There are two scenarios in our simulation: one is with a total bandwidth of 8Mbps, and the other is with 12Mbps.

The performance metrics measured in the simulation include the average throughput, share degree, and fairness degree.

(1) Average throughput: this parameter is defined as the average allocated bandwidth per time unit for a connection of a specific class. That is, for class i, and connection j,

$$average\_throughput_i = \frac{\sum_{j \in i} bandwidth\_usage_j}{n_i},$$

where $bandwidth\_usage_j$ is the allocated bandwidth of connection j, and n_i is the number of connections in class i.

(2) Fairness degree (FD): this parameter indicates how fair the residual bandwidth is shared by all connections for each approach, and is defined as $FD = \frac{\left[\sum_{i=1}^{n} SD(i)\right]^2}{n \sum_{i=1}^{n} [SD(i)]^2}$,

where n is the number of connections. The FD value is within [0,1]. The larger FD value, the fairer bandwidth allocation.

3.2. Simulation results

(1) Scenario 1

The available UL bandwidth is 8Mbps. This scenario is to show how 2TSA performs while the summation of all connections' maximum sustained rate is larger than 8Mbps. We assume all rtPS connections will generate lot of data from the sixth second. We found that no matter how much traffic a connection generates, 2TSA can guarantee each connection to have its minimum bandwidth demand (i.e., 0.5Mbps, 0.3Mbps, and 0 for rtPS, nrtPS, and BE, respectively), and fairly distribute the residual bandwidth to all connections (i.e., each with 0.1Mbps), as shown in Fig. 2. In contrast, since SPS always first allocates bandwidth to rtPS connections, all connections of nrtPS and BE service classes starve after the fifth seconds. Therefore, 2TSA has a much higher fairness degree value, compared with SPS, shown in Fig. 3.

Table 1. Parameter settings

Type	R^i_{min} $(Kbps)$	R^i_{max} $(Kbps)$	Max latency (ms)	Number of flows
UGS	60	60	20	5
rtPS	500	700	50	7
nrtPS	300	500	100	7
BE	0	200	200	7

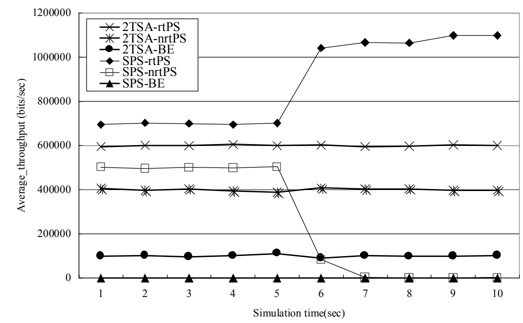

Figure 2. Average throughput of scenario 1

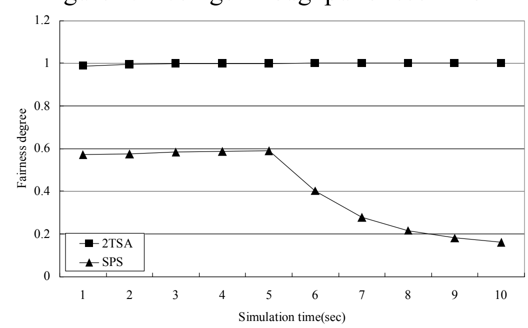

Figure 3. Fairness Degree of scenario 1

(2) Scenario 2

The experiment is to evaluate how fair 2TSA and SPS can achieve for the residual bandwidth allocation. The available UL bandwidth is set to 12Mbps. We further run two cases. The first case is that all rtPS connections are greedy by sending lots of data to use out the bandwidth after the fifth second. The second case is all connections are greedy after the fifth second. Figs. 4 and 5 are the average throughput of both cases. In Fig. 4, since the available bandwidth can afford all connections to send data at their maximum sustained rates, 2TSA first allocates each connection with its maximum sustained rate, and then rtPS connections with the residual bandwidth. For the second case, 2TSA still achieve fair share of residual bandwidth, as shown in Fig. 5. However, the problem of starvation occurs again in both cases for SPS. The corresponding fairness degrees of these two cases are in Fig. 6, and 2TSA outperforms SPS in fairness while incurring greedy connections.

4. Conclusion

In this paper, we proposed a two-tier scheduling algorithm (2TSA) for uplink transmission in IEEE 802.16 BWA systems of TDD mode. Based on the specified QoS parameter and allocated bandwidth, we classify each connection into either "unsatisfied", "satisfied", or "over-satisfied" category. The first-tier allocation algorithm is category-based, and the unsatisfied category has the highest service priority, followed by the satisfied category, and finally the over-satisfied category. The second-tier allocation scheme, contrarily, is weight-based. For connections within the same category, the one with the smallest weight value has the highest service priority. Compared with strict-priority bandwidth allocation algorithm, the simulation results show that 2TSA can guarantee connections' QoS demands, avoid starvation of lower-priority service class, and achieve better fairness degree.

In the future, we will investigate bandwidth allocation considering other QoS metrics, such as delay or delay jitter, for 802.16 BWA systems.

Acknowledgement

This research was supported by the New Generation Broadband Wireless Communication Technologies and Applications Project of Institute for Information Industry and sponsored by MOEA, R.O.C.

Reference

[1] IEEE 802.16 Standard-Local and Metropolitan Area Networks-Part 16. IEEE 802.16-2004.

[2] Hamed S. Alavi, Mona Mojdeh, Nasser Yazdani, "A quality of service architecture for IEEE 802.16 standards", *2005 Asia-Pacific Conference on Communications, Perth, Western Australia*, Page(s):249 – 253, 03-05 Oct. 2005.

[3] Dong-Hoon Cho, Jung-Hoon Song, Min-Su Kim, and Ki-Jun Han, "Performance analysis of the IEEE 802.16 wireless metropolitan area network", *Distributed Frameworks for Multimedia Applications, 2005, First International Conference,* Page(s):130 – 136, Feb. 2005.

[4] Kitti Wongthavarawat, Aura Ganz, "IEEE 802.16 based last mile broadband wireless military networks with quality of service support", IEEE *Military Communications Conference, 2003,* Page(s):779 - 784 Vol.2, 13-16 Oct. 2003.

[5] Qingwen Liu, Xin Wang and Georgios B. Giannakis, "Cross-layer scheduler design with QoS support for wireless access networks", *Quality of Service in Heterogeneous Wired/Wireless Networks, 2005. Second International Conference*, Aug. 2005.

[6] Jianfeng Chen, Wenhua Jiao, Hongxi Wang, "A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode", *Communications, 2005 IEEE International Conferenc,* Page(s):3422- 3426, May 2005.

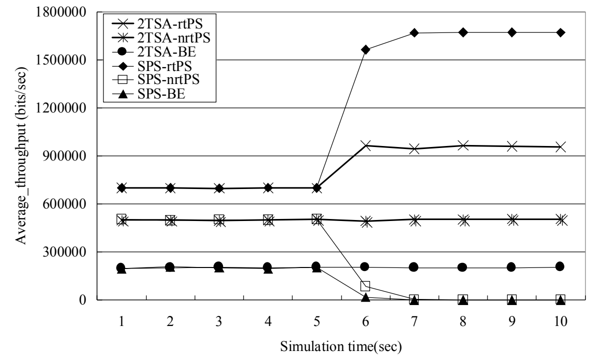

Figure 4. Average throughput of scenario 2

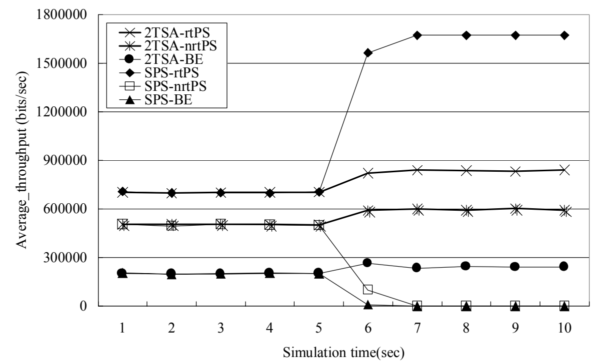

Figure 5. Average throughput of scenario 2

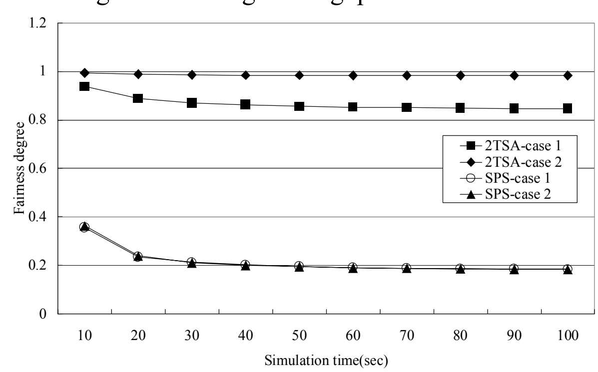

Figure 6. Long-term fairness degree of scenario 2