國立交通大學

電信工程研究所

碩士論文

針對非即時訊務在 IEEE 802.16e 系統的一種排程式頻寬要求節能機制

A Schedule Based Power-Saving Bandwidth Request Mechanism for non-Real-Time Traffic in IEEE 802.16e System

研　究　生：林筠翰

指導教授：李程輝　教授

中華民國一佰年六月

針對非即時訊務在 IEEE 802.16e 系統的一種排程式頻寬要求節能機制

A Schedule Based Power-Saving Bandwidth Request Mechanism for non-Real-Time Traffic in IEEE 802.16e System

研 究 生：林筠翰　　Student：Yun-Han Lin

指導教授：李程輝　　Advisor：Tsern-Huei Lee

國 立 交 通 大 學

電信工程研究所

碩士論文

A Thesis

Submitted to Institute of Communications Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

In partial Fulfillment of the Requirements

for the Degree of

Master

in

Communications Engineering

June 2011

Hsinchu, Taiwan, Republic of China

中華民國一佰年六月

針對非即時訊務在 IEEE 802.16e 系統的一種排程式頻寬要求節能機制

學生: 林筠翰　　　　　　　　　　　　　　　　　　指導教授：李程輝

國立交通大學電信工程研究所碩士班

摘　　要

綠能議題在近幾年一直是一個熱門的主題，在 IEEE 802.16e 的架構之下，亦有不少針對節能而設計的功能。IEEEE 802.16e 本身就有提供一般訊務使用的節能功能（睡眠模式)，然而每種訊務在成立之前必經的頻寬要求程序目前卻沒有一套用以節能的機制。以往針對頻寬要求程序的研究多偏向於如何減低其碰撞機率，而我們並不朝這個方向去做，相對的，在符合向下相容的條件之下，我們利用已定義的節能信號來擬似一個回應機制，用以迴避因為頻寬資源不足而產生的能量浪費，同時加速發現碰撞的反應速度。同時我們也修正現存頻寬要求數學模型以預估節能機制的表現。在我們的模擬中，以額外的控制信號作為代價的情形之下，能夠節省將近 60%的能量浪費，同時每個頻寬要求平均上不超過 1.5 次的額外控制信號。

關鍵字：節能，頻寬要求，資源保留

A Schedule Based Power-Saving Bandwidth Request Mechanism for non-Real-Time Traffic in IEEE 802.16e System

Student: Yun-Han Lin Advisor: Prof. Tsern-Huei Lee

Institute of Communications Engineering
National Chiao Tung University

ABSTRACT

The issue of power saving is becoming more and more popular in recent years. In IEEE 802.16e system, there are many mechanism designed for the purpose of power saving. For regular traffic, IEEE 802.16e provides many functions of connection sleeping procedure. However, the unavoidable connection setup procedure, the bandwidth request mechanism, seems come with no specific power saving mechanism. Most existing studies focus on the issue of how to reduce the collision rate of bandwidth request mechanism. While maintaining backward compatibility, we try to make use of predefined functions in standard to produce a quasi-acknowledgement which avoids the waste of unnecessary power consumption resulting from insufficient bandwidth and hasten the collision detection procedure. At the same time, we also extend an existing analytical model to estimate the performance of the proposed algorithm. According to our simulation result, our proposed scheme reduces the waste of unnecessary power consumption by about 60% at the cost of less than 1.5 times control signaling per bandwidth request.

Keywords: power saving, bandwidth request mechanism, resource reservation

誌　謝

感謝交通大學電信工程研究所 NTL 實驗室的各位，郁文學長、孟諭學長、啟賢學長、梓洋學長、迺倫學姐、謙和、建男、佳心、家俊、亞蕾、承潔、順閔、孟哲、信宏、琮揚、運良、國書、煜傑、晴嬅，感謝你們陪我度過我的研究所，以及在這之間所提供給我的任何意見跟想法。

特別感謝我的指導教授 李 程輝 博士，在我的學業、研究方面的指導讓我在研究所兩年中獲益匪淺。與梓洋學長、琮揚同學與信宏同學在研究方面的相互討論更是讓我的研究能夠順利進行的一大助力。

最後感謝我的家人對我的付出與支持我才能走到今天。

謹將此論文獻給所有幫助過我的人

2011/06

Table of Contents

List of Figures

Symbols

Q_n ：queue of bandwidth request which active in n frames latter

$R[n]$ ：total traffic load of Q_n

$E[n]$ ：expectation of arrival traffic load in n frames latter

$E_r[n]$ ：expectation of real-time traffic load in n frames later

M_{T16} ：maximum value of T16 timer also the scheduling window

W ：a frame set of all frames which has available bandwidth in scheduling window

p ：collision rate of bandwidth request

q ：bandwidth allocation failure rate

c ：contention slots in a frame

b ：maximum backoff stage

M_k ：the maximum value of backoff count in backoof stage k in frame basis

λ ：mean of arrival rate of Poisson arrival process

λ' ：the mean arrival rate including retransmission

Chapter 1
Introduction

1.1 The issue of power saving

The issue of power saving is becoming more and more popular in recent years. Because of the evolution of network technologies and the rapid growth of wireless business, mobile devices are setting up its importance. However, power efficiency is a problem that cannot be ignored, since most mobile devices take batteries as the source of energy, and as we all know, batteries can provide only limited energy. Making good use of every iota of energy becomes an important issue, and there are many related studies focus on it. As we survey these studies, we find a potential problem.

1.2 System model of IEEE 802.16e

The knowledge of this section is from [1].

1.2.1 The power saving class

In IEEE 802.16e standard version 2009, there are three type of power saving class (PSC) which are type I, II and III. The power saving class is assigned to a connection, not a SS. That is a SS may have many connections and each connection may have its own power saving class. The SS could turn off its transceiver only when all connections belong to the SS are in sleep mode, and when one of these connections enters the active mode the SS shall turn on its transceiver immediately.

The type I PSC is designed for non-real-time and best effort traffic. While in type I PSC, the connection alternates between sleeping window and listening window, and the sleeping window is doubled time by time until receives traffic or reaches the maximum value which is a system parameter. While receive any traffic indicator at listen interval, the connection will reset the sleeping window to initial value and enter the active mode to transmit traffic.

Like the type I PSC, the type II PSC also alternates between sleeping window and active window. The difference is that the sleeping window is always the same and the connection sends traffic while active window. This type of PSC is designed for the real-time traffic which has a fixed inter arrival time.

The type III PSC which is different from type I and type II has only a sleep phase. When the sleeping phase finished, the connection enters a normal operation mode, active mode. The type III PSC is used for the purpose of connection management operation, and with this characteristic, it plays an important role in the proposed algorithm.

1.2.2 The bandwidth request mechanism

The knowledge of this section is from [2].

In IEEE 802.16e system, when a SS needs to ask for uplink bandwidth, it sends a message to the BS containing the immediate requirements of the requesting connection. Requests refer to the mechanism that SS use to indicate to the BS that they need uplink bandwidth allocation. To deserve to be mentioned, the request is made in a connection basis and the grant of request is made in a SS basis. That is the SS should decides whether a connection to transmit or not. Besides, SS may not send request at its pleasure. The system is traditionally consist of a BS and many SSs under a structure of peer to multi peers and centralized. BS will polls SS unicast, multicast or broadcast in some system, or SS may send contention token in a predefined period in the other system. Furthermore, every bandwidth request contains a timer, named T16, which setup while the bandwidth request transmitted and counts down every frame. When the timer counts down to zero, the bandwidth request is expired and shall be discarded from system. The setup value of T16 is a system parameter which is controlled by BS, and may vary from frames to frames. The frame structure shows as Figure 1. The downlink channel descriptor (DCD), uplink channel descriptor (UCD), downlink map (DL-MAP) and uplink map (UL-MAP) are the preamble, and stand for the channel condition parameter, system parameter and the detail of bandwidth allocation of both downlink and uplink. The bound between downlink sub-frame and uplink sub-frame are determined dynamically by the BS and are broadcast to the SSs through UL-MAP and DL-MAP messages at the beginning of each frame. The uplink sub-frame contains transmission opportunities scheduled for the purpose of sending bandwidth request (BW-REQ in figure 1) messages in which bandwidth request messages can be transmitted.

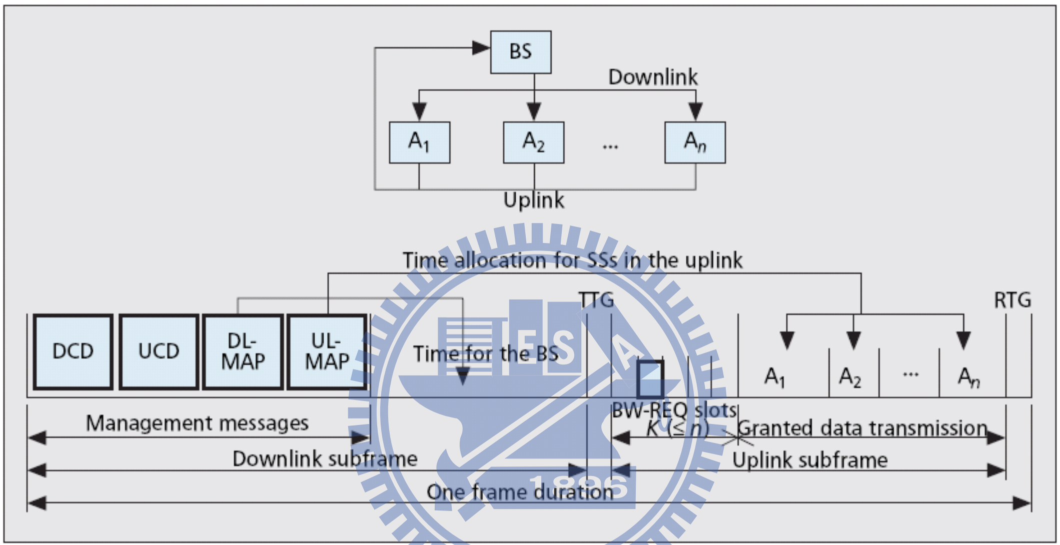


Fig. 1. Frame structure of IEEE 802.16e TDMA/TDD network in PMP mode.

From [2].

The BS controls both the number of transmission opportunities for bandwidth request and data packet transmission through the UL-MAP message. With the knowledge above, we introduce the signaling procedure of a bandwidth request as below:

Scenario I (active by BS):

1. BS polls (unicast, multicast or broadcast) SSs.
2. The connection which is in the polled SS and has traffic to send randomly chooses a predefined bandwidth request contention slot and transmits its bandwidth request.
3. While receiving the bandwidth request, the BS will add the bandwidth request into its scheduling queue which is maintained by some policy such as FIFO. Furthermore, BS

will assigns bandwidth to higher priority connections in the queue until out of bandwidth or the queue is empty at the beginning of every frame.

4. The SS which contains the requesting connection keeps its transceiver active to wait the bandwidth grant of the request until the bandwidth request expired.
5. If the SS does not receive the bandwidth grant for the requesting connection when the bandwidth request expired, the connection which does not reach the maximum times of retransmission will enter the truncated binary exponential backoff procedure to retransmit the request, and the connection which reach

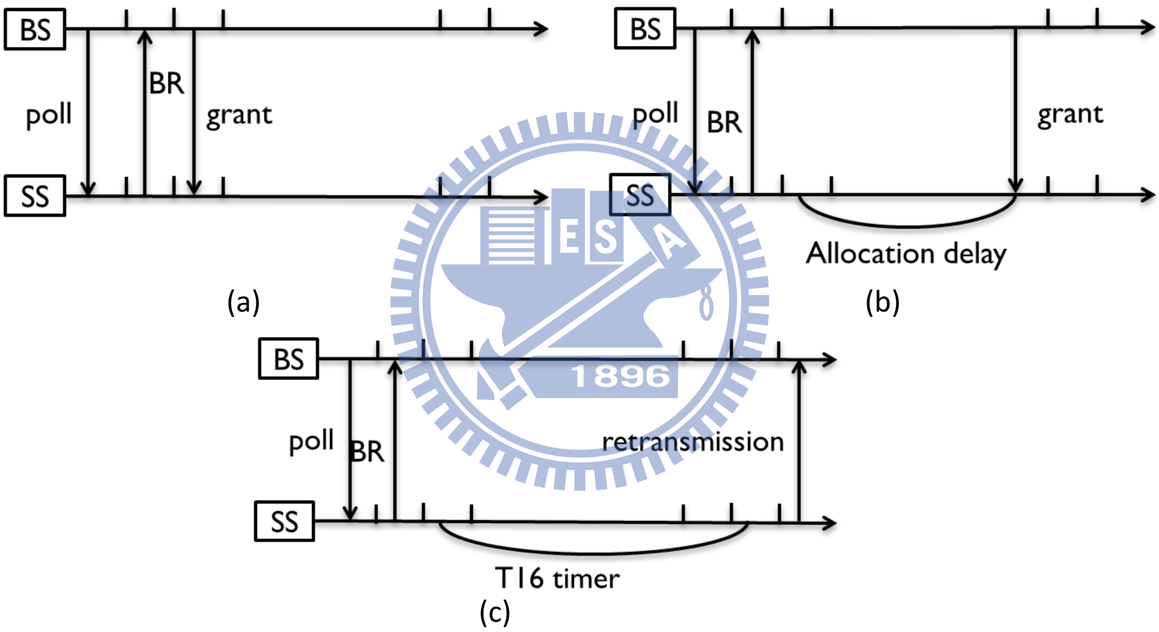


Fig. 2. The signaling procedure of bandwidth request (active by BS)

the maximum times of retransmission will be discarded and sends an error signal to upper layer.

Scenario II (active by SS):

1. The connection which has traffic to transmit sends a contention token which is much smaller than bandwidth request in a randomly chosen contention slot which is also much smaller than the bandwidth request contention slot.
2. If the BS receives the contention token, it will unicast polls the SS which contains the

requesting connection, and the follow is same to scenario I unicast case. If the token lost which may due to collision or bad channel condition, the requesting will enter backoff procedure to retransmit the token.

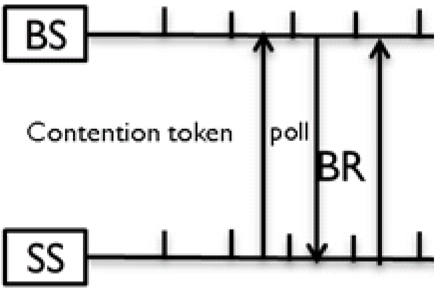


Fig. 3. The signaling procedure of bandwidth request (active by SS)

Intuitively, we can find that the total number of contention slot will affect the success rate of bandwidth request and the total amount of bandwidth resource also affect the probability of successful bandwidth allocation. In addition, both collision and bandwidth allocation failure consume unnecessary power with which our algorithm deals.

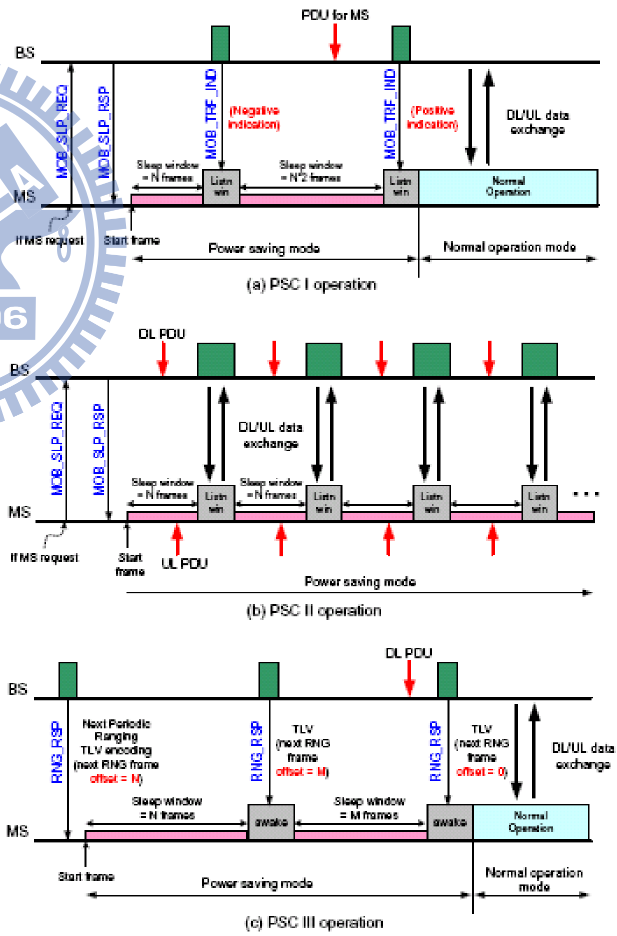


Fig. 4.operation of power saving class in 802.16 standard. From [4].

1.3 Related work

There are many studies discus about the power saving mechanism in IEEE 802.16e system. For example, [3] is discussing about how the traffic load affects the power saving mechanism, and mentions us to take traffic load into account. With the help of [3], we consider the load balance problem in the proposed algorithm. [4] is also a study discuss about power saving mechanism defined in standard. Although the algorithm in [4] has little help to the proposed algorithm, the detailed description about signaling process of the power saving sleep signal (figure 4.) really gives us an aid. As we mention before, there are many studies discus about how to reduce the collision rate of bandwidth request mechanism. For instance, [5] inspires us to the idea of using acknowledgment. In [5], the authors make use of a special bit, called "traffic indicator", to indicate subscriber station (SS) the condition of current traffic load. And by doing so, SS can skip the congestion period and transmits bandwidth request when traffic load decreases, resulting that the lower collision rate and also higher power efficiency. [6] is the fundamental reference of our study. The author introduces the power saving mechanism of IEEE 802.16e and the flexibility which we can control. Base on the reference, we find the possible signaling for quasi-acknowledgment, the type III sleep signal which is three-byte-long. [7] is a study which discus about resource reservation. In [7], the author tries to make a prediction of future traffic load and reserve the amount of predicted traffic load (Figure 5). With the algorithm in [7], the base station (BS) may optimize the throughput and reduce the delay of packet. By the inspiration of [7], we add an optional resource reservation function into the proposed algorithm. The function provides a tuneable parameter "expectation", and the more correct expectation we make, the better trade off of overhead signaling we have.

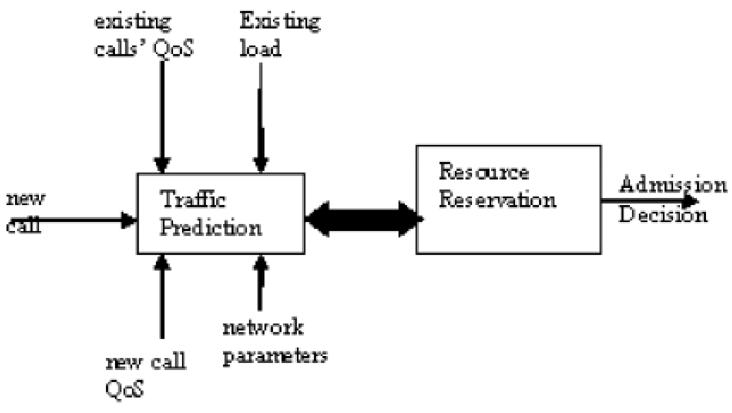


Fig. 5. Architecture of the proactive resource reservation system. From [7].

The idea of ours is every similar to [8], but not the same. In [8], the authors try to extend the fixed sleep length which is defined by standard to meet the optimal sleep length. By a making up state “extend sleep state”, the authors of [8] make use of statistic and find the optimal probability entering and leaving (Figure 6.) the “extend sleep state.” To say the difference, the proposed algorithm is scheduling based and extends the sleep period by exact number of frames which is calculated by program, and the proposed algorithm maintains backward compatibility. [9] provides us a mathematical system model of 802.11 bandwidth request mechanism. Although the model may not perfectly fit our system model (IEEE 802.16e), it gave us a hint of general idea of system framework. With the help of [9], we derive analytical model part of the proposed algorithm.

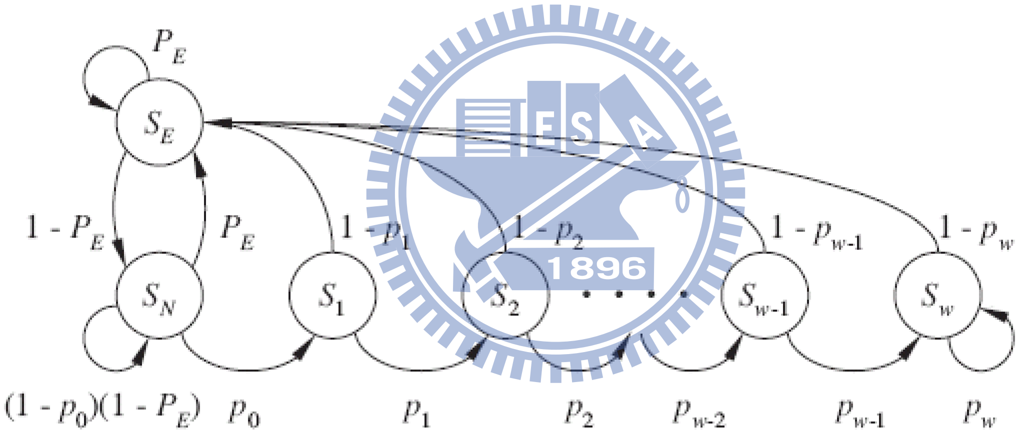


Fig. 6. The state transition diagram between S (sleep state) and S_E (extend sleep state).

From [8].

.

1.4 Problem Definition

As we know, there are many kinds of source of energy waste in wireless system. In [10], the authors mention us the main source should be collision, overhearing, control packet overhead, idle listen in 802.11 system. However, in IEEE 802.16e system, the problem of overhearing has been solved because of uplink map and downlink map and the problem of control packet overhead has also been eliminated as possible. From the previous section, we find that while bandwidth request fail, the system will consume additional power which is unnecessary. And the sources of bandwidth request failure are collision and bandwidth insufficient which stand for the two main sources of wireless system, the collision and idle listen. It is straightforward that bandwidth request collision stands for the collision type of power waste, but we may ask that why the lack of bandwidth stands for idle listen? In the bandwidth request mechanism in IEEE 802.16e system, while there is lack of bandwidth and BS still has bandwidth request in the scheduling queue, the BS will do nothing to the bandwidth request in queue and only discard it while expired. Meanwhile, the SS still actives its transceiver and is waiting a bandwidth grant which comes bulks of frames latter or even never come. In these cases, the SS actually does the idle listen and wastes power.

Besides, the two sources of power waste occur with different network condition. By the observation from simulation, we find that the idle listen will not occur while light load condition. Simply, we can image that while there is still bandwidth available, the bandwidth request shall not stay in scheduling queue, and the SS shall not do idle listen. Actually, the main reason of power waste in most condition, even in heavy load condition, is collision, and that is why there are so many studies discus about reducing collision rate. Then, we may ask when the condition of idle listen occurs? The answer which we obtain from simulation is when the traffic load nearly fulfills the total bandwidth or exceeds the channel capacity the idle listen may come into existence. Why we say that “may” come into existence? Because from [9], the author tells us that while the traffic increasing, the collision rate of bandwidth request is also increasing too. That is the increasing of bandwidth request may not bring the condition of insufficient bandwidth because of lack of successful bandwidth request. However, under the condition that the traffic load carried by a single bandwidth request large enough, the situation of insufficient bandwidth still have chance to take place. Therefor we shall take idle listen into account in the proposed algorithm.

To sum up in a word, the problem we deal with is to skip the consumption of unnecessary power which is caused by collision in light traffic load condition and by collision and insufficient bandwidth in heavy traffic load condition.

1.5 General idea of the proposed algorithm

With the assistance of many pioneers, we develop our algorithm while maintaining backward compatibility. To sum up in a word, the proposed algorithm is try to schedule successful bandwidth request for available resource, in most time intending for bandwidth, and turn off the transceiver of SS while congestion. To satisfy the purpose, we make use of TYPE III sleep signal for scheduling, and null response represents the condition of collision. With a structure like this, the first problem is how we schedule incoming bandwidth request to the proper frame, because of the inappropriate schedule may lead to bandwidth waste (there is available bandwidth and also bandwidth request, but the requesting connection is in sleep mode) and high delay. Also the scheduling window of schedule is an issue, because scheduling will lose it precision while time elapses, and the hardware may not allow us to extend the window as we wish. Besides, as we say before, the proposed algorithm maintains backward compatibility. The system model of IEEE 802.16e is PMP. That is there are a single BS and many SS in a cell. For the reason of economy, we design the proposed algorithm which making changes concentrate on BS under the condition that all SS which are not upgrade still work with no performance loss.

1.6 The composition of the rest

The rest of this thesis is organized as follow. Chapter 2 is the proposed algorithm. Chapter 3 is the analytical model of the proposed algorithm, and we derive the model from the model in [9]. Chapter 4 is the simulation part, and we discuss the use of some parameter in special condition. The final chapter, chapter 5, is conclusion and discuss.

Chapter 2
The Proposed Algorithm

2.1 Idea development

The problem we mention first chapter can be solved in many different direction. The straightforward idea is to reduce the collision rate, and in fact, there are many studies focus on this issue which is nearly complete. However, we think otherwise. The primal goal of previous studies which focus on reducing the collision rate may not focus on power saving. For them, the power saving is only a side effect. On the contrary, for us, the power saving is the only thing we care about, and we may trade something for the effect of saving power. Hence, the direction of the proposed algorithm is not follows these pioneers and we go our way.

The main idea of the proposed algorithm is that try to give right responses to unwanted situation which means collision and idle listen. As we say in chapter 1.4, the idea of making use of acknowledgement comes into our mind. And we not only try to get the information about condition of traffic load, but also try to info SS by having an acknowledgement of the state of the bandwidth request.

2.1.1 The acknowledgement

While maintaining backward compatibility, we try to find suitable signal in standard which contain three types of PSC. Obviously, the type I and II PSC are design for the connection which has been setup, and the purpose of them is to reduce the long term consumption of power. The bandwidth request mechanism, however, is not a long term procedure. In fact, the total life time of a bandwidth request is approximately equal to T16 length multiplied by the maximum backoff stage, and is about hundreds of microsecond. In such case the type III PSC is more suitable for management, because the controllable duration of the sleep interval. With the characteristic of sleep signal of type III PSC, we find the possibility of constructing a scheduling procedure, because the effect of it is to make the connection enter sleep mode for a while and go back to active mode after the period. Hence, we design the proposed algorithm with three acknowledgements which are bandwidth grant, sleep signal of type III PSC (called scheduling signal afterwards) and null response. They stand for successful bandwidth request with successful bandwidth allocation, successful bandwidth and waiting for bandwidth grant and collision in turn.

2.1.2 The scheduling window

After deciding the acknowledgement, the next question is that how we scheduling a bandwidth request and how long is the scheduling window? To decide the scheduling window, we have to know that the longer scheduling window the more improvement of power saving efficiency. Because we can let the requesting connection stays in sleep mode longer, and provide more flexibility in scheduling. However, we cannot extend the scheduling at will, because the limitation of hardware device and also, as we mentioned before, the life time of bandwidth request. Hence, while maintaining backward compatibility, we decide the scheduling window to be the maximum value of T16 timer which may differ from frame to frame. In such setting, we can guarantee that all receiving bandwidth requests are in the scope of the scheduling before they expired.

2.1.3 The scheduling principle

The first procedure of scheduling is obviously fulfilling the current frame with available bandwidth request which is in the scheduling queue with the order predefined. By doing so, we can maximum the throughput as we can. From the observation of bandwidth request mechanism, we find that the best case of a bandwidth request is to active transceiver two frames which are the frame sending bandwidth request and the frame receiving grant in the view of power saving. With the knowledge of this, we should produce the “two frames” case as possible as we can. That is to saving more bandwidth as possible as we can in the coming frame. To say extremely, not to schedule any bandwidth request in the next frame will come to the most number of “two frame” case. However, the bandwidth request must be scheduled in one particular frame or it may do idle listen. If we schedule them to the frame next to next frame, then they will reduce the number of “two frames” case in the frame next to next frame. To make a good balance of every frame, a water filling scheduling mechanism may be the best way to do. That is we find the frame which has lowest scheduled traffic load in the scheduling window, and schedule the highest priority bandwidth request in the scheduling queue to the frame. As we repeat this procedure, finally we will have the condition that all frames in the scheduling window have nearly the same scheduling traffic load which is the best case of “two frame” case in long term statistic.

2.1.4 The optional bandwidth reservation

As we reading [7], an idea which comes to our mind is that why don’t we reserve some bandwidth in the coming frame to achieve more “two frame” case? That is if we know the traffic load in some future frame, and we reserve the amount of bandwidth in the frame. By doing so, the expecting coming bandwidth request will be the “two frame” case in the bandwidth-reserved frame. However, expecting arrival traffic is not an easy task, and the wrong reservation will lead to bandwidth waste which is the most intolerable mistake in scheduling procedure. With these reasons, we leave the expectation part as an external function and make the function of bandwidth reservation an optional function in the proposed algorithm. If we have a precise expecting function, then we may use this function to improve better performance, otherwise we just leave the expectation as zero traffic load to disable the expecting function.

2.2 The parameter of the proposed algorithm

Assume that there is a scheduling queue in BS which is standard defined, and we divide the queue into the maximum value of T16, M_{T16}, parts Q_n where $0 \leq n \leq M_{T16} - 1$. Each part of Q_n contains the scheduling queue of the connection which will be active in n frame latter and sort by the order of receiving slot. And in the BS, there are three vectors $R[n]$ which means the reserved bandwidth in the frame which is n frame latter, $E[n]$ which means the expectation traffic load in the frame which is n frame latter and $E_r[n]$ which means the expectation traffic load form the real-time traffic which is high priority and periodic. Assume $R[n]$ equal to the summation of the traffic load of Q_n. And we have $E[n]$ from external function as we mentioned before. The $E_r[n]$ is an expectation based on current information, and it may change while we have more information over time. For example, if we have a UGS connection which is x bytes every p frame coming f frame latter, then we may add x bytes into $E_r[f]$, $E_r[f+p]$...$E_r[f+p*k]$ where k is a positive integer and $f + p * k \leq M_{T16} - 1$, and if we get the information that the UGS connection will be terminated after the last transmission in t frame latter, then we shall subtract x bytes from $E_r[t]$, $E_r[t+p]$...$E_r[t+p*k]$ where k is a positive integer and $t + p * k \leq M_{T16} - 1$.

2.3 The procedure of the proposed algorithm

The proposed algorithm execute at the beginning of every frame by BS.

Phase I, make the most of current frame:

---

1. Adjust $E_r[n]$ based on the information we have last frame
2. Add the bandwidth request received last frame into Q_0 with sorting policy.
3. Set R, standing for reaming bandwidth, equal to the uplink bandwidth of current frame.
4. R=R-E_r [0].
5. **While** (Q_0 is not empty and R> 0)

6. Dequeue the highest priority connection from Q_0.

7. Assign bandwidth to the connection in current frame.

8. R=R- traffic of the connection

9. End while.

The first procedure of phase I is to reserve the bandwidth which is requested by real-time traffic which always has higher priority than non-real-time. The second procedure is to add the request we receive last frame into scheduling queue. The third and fourth procedures are to calculate the total bandwidth for non-real-time traffic. And finally the fifth to eighth procedures are to fill the bandwidth of current frame with the active connections which are in Q_0.

Phase II, schedule active connection with water filling mechanism:

1. Adjust $E[n]$ based on external function.

2. Set R equal to the bandwidth of uplink of current frame.

3. Set a frame set W which contains the frames in the scheduling window= $\{f_i | 1 \leq i \leq M_{T16} - 1 and\ R[i] + E[i] + E_r[i] < R, i \in N\}$.

4. While (Q_0 is not empty and $W \neq \emptyset$)

5. Find the scheduling frame f_s=arg $\max_{f_i \in W}\{R - R[i] - E[i] - E_r[i]\}$

6. Find the highest priority connection h whose T16 timer $\geq$ i from Q_0.

7. If ($h \neq$ null)

8. Dequeue h from Q_0.

9. Schedule bandwidth to h in the frame which is i frame latter by type III PSC sleep signal.

10. Add h to Q_i.

11. Set $R[i]$=$R[i]$+traffic of h.

12. **If** $(R[i] + E[i] + E_r[i] \geq R)$

13. Remove f_S from *W*.

14. **End if.**

15. **Else**

16. Remove f_S from *W*.

17. **End if.**

18. End while.

The first procedure of phase II is to adjust the expectation of future traffic based on current information by some external functions. This optional procedure can be omitted by setting the expectation to zero as we mentioned before. The second and third procedures are to construct a schedulable frame set *W* which consists of the frame which still has available bandwidth in scheduling window. The fourth to eighteenth procedures are the water filling mechanism. The fifth procedure is to find the frame f_s which has the most bandwidth in the schedulable frame set *W*. and the sixth procedure is to find the highest priority connection which can be scheduled to the frame f_s on the constraint of its T16 timer in the active connection queue. The seventh to eleventh procedures are the scheduling procedures work with type III PSC sleep signaling. The twelfth to fourteenth procedures remove the frame which has no available bandwidth after scheduling from *W*. and finally the fiftieth to seventeenth procedures remove the frame which still has available bandwidth but no active connection can be scheduled to the frame on the constraint of the T16 timer of the connection.

Phase III, discard the active connection which cannot be scheduled:

---

1. **If** (Q_0 is not empty)

2. Dequeu every connection in Q_0.

3. Sleep these connections to the frame at which their bandwidth request expired by type III PSC sleep signal.

4. End if.

The procedures of phase III are deal with the connections which have no available bandwidth before their T16 timer expiration. The handling procedures are simply sleeping them until the expiration of their T16 timer and discarding these connections from scheduling queue.

Phase IV, parameter shift:

1. For $\forall \mathrm{n}$, where $0 \leq \mathrm{j} \leq M_{T16} - 2$

2. *R[j]=R[j+1]*.

3. $R[M_{T16} - 1] = \sum_{the\ connections\ in\ Q_{M_{T16}}} traffic\ load$

4. *E[j]=E[j+1]*.

5. E_r *[j]*= E_r *[j+1]*.

6. End for.

This phase is simply shifting the vector *R[j+1]* to *R[j]*, and set $R[M_{T16} - 1]$ the traffic load of scheduled connections in the frame M_{T16} latter for the use of next frame. The same operation processes to *E[n]* and E_r *[n]*.

2.4 The state chart of connection

With the algorithm above, we now try to illustrate the state chart of connections and compare it to standard as below.

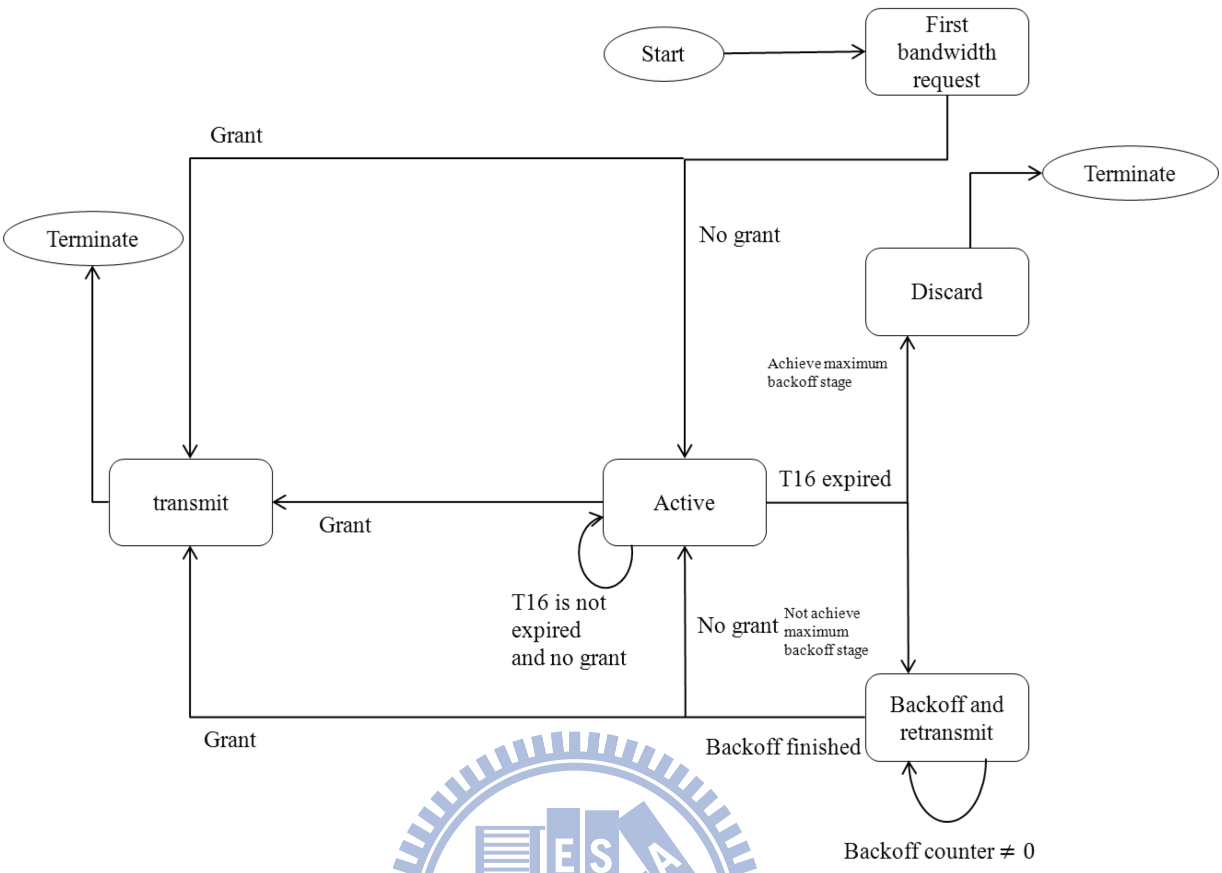


Fig. 7. The connection state chart of IEEE 802.16e bandwidth request mechanism

Comparing the difference between Fig. 7 and Fig. 8, the first difference is the information connection has. In Fig. 7, the connection does not know the condition of bandwidth request which may have been collide. And it is clearly in Fig. 8. With the additional information, the connection which is collided will enter sleep mode until T16 expired, and result in the line in right hand side in Fig. 8. The second difference is the scheduling procedure which makes the connection enter sleep mode while no available bandwidth. The heavy line box in Fig.8 means that the connection is in sleep mode and is the contribution of the proposed algorithm.

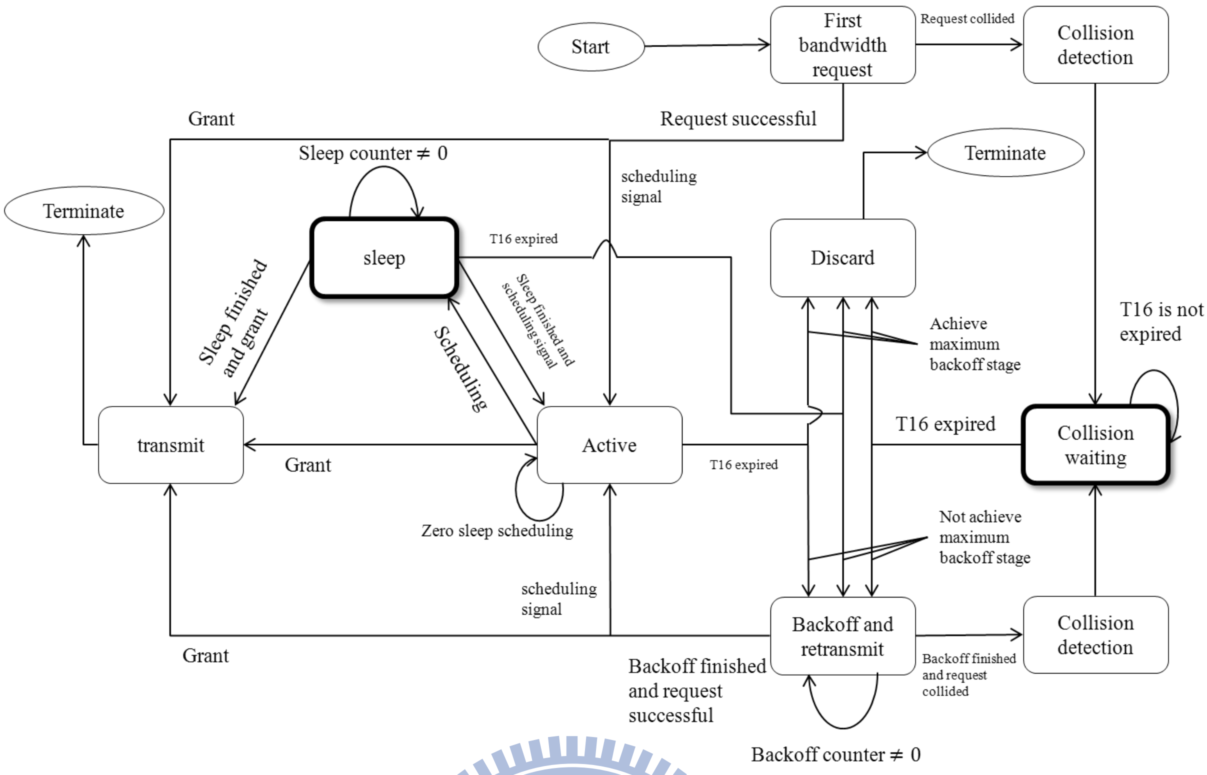


Fig. 8. The connection state chart of the proposed bandwidth request mechanism, the heavy box means the connection in sleep mode.

Chapter 3
Analytical model

3.1 Analytical model framework

There are many studies focus on the performance of bandwidth request mechanism, and we take [9] which is one of the most cited references as the framework of our model. In [9], the author derives the mathematical model of bandwidth request mechanism of 802.11 system which is similar to IEEE 802.16e system. However, in [9], or even in most studies focus on performance analysis of bandwidth request mechanism, the bandwidth allocation failure does not be taken into account usually. The bandwidth allocation failure which means that a bandwidth request transmitted successfully but no available bandwidth is clearly a factor which infects the collision rate of bandwidth request, because the raising of bandwidth allocation failure rate results in the more retransmission, and the more retransmission leads to more collision. Because of the reason above, we slightly extend the model with bandwidth allocation failure rate q which is paired with p, the collision rate.

Assume c is the contention slot of a frame which is a system parameter, and b is the maximum backoff stage which still is a system parameter. *k(t)* is the backoff stage of the connection at time t where $0 \leq k(t) \leq b$, and $m_k(t)$ is the number of backoff counter of the connection at time t in frame basis where $0 \leq m_k(t) \leq M_k$ where $M_k = \left\lfloor \frac{c+2^k-1}{c} \right\rfloor$ which means the maximum backoff frame in backoff stage k. I which means the initial transmitting slot is a random variable which is uniformly distributed between 1 and c. B_k which means the number of backoff slot in backoff stage k is also a random variable which is uniformly distributed between 0 and 2^k-1. T_k which means the total backoff period from the first contention slot at time t equal to *I* adds to B_k. Finally, the $r_{m,k}$ which means probability of backoff to the frame which is m frame latter in backoff stage k is equal to $P(m * c < T_k \leq (m+1) * c)$. As shown in Fig. 9, we have a bi-dimensional Markov chain *{k(t),m(t)}*.

3.2 The bi-dimension Markov chain

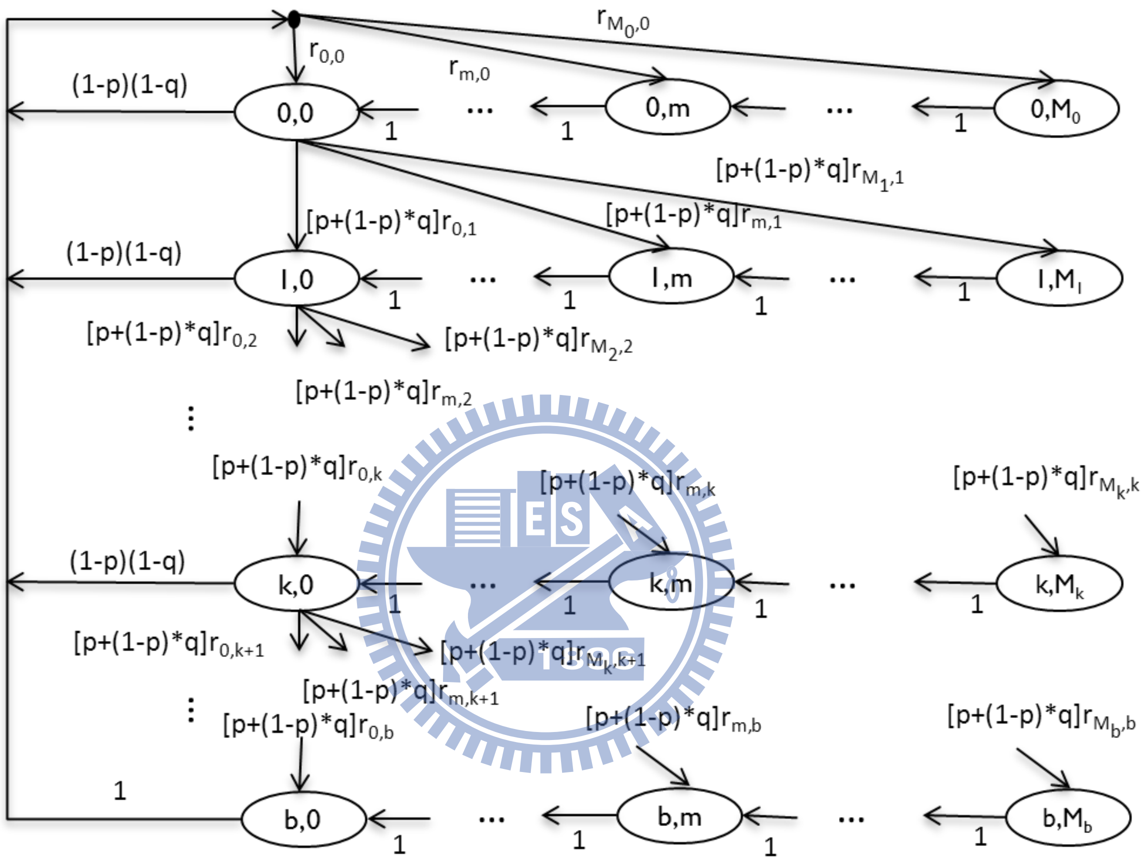


Fig. 9. Markov chain of bandwidth request mechanism.

And the transition probability formulas are shown below:

$$P\{i+1, j|i, 0\} = [p + (1-p) * q]r_{j,i+1}\ , where\ \ i \in [0, b-1], j \in [0, M_{i+1}]$$

$$P\{i, j-1|i, j\} = 1\ , where\ \ i \in [0, b], j \in [1, M_i]$$

$$P\{0, j|i, 0\} = (1-p)(1-q)r_{j,0}, where\ \ i \in [0, b-1], j \in [0, M_0]$$

$$P\{0, j|b, 0\} = r_{j,0}, where\ \ j \in [0, M_0]$$

With transition probability formulas, we can represent state probability of every state with *p* and *q*, and derive *p* as a function of *q* by the method in [9] which derive *p* actually. And then we find the actually *p* and *q* by following procedure:

1. Assume q=0, we have the actual value of p from [9].
2. With the mean of arrival, λ, in unit time, we can find the total transmitting user λ' including retransmitting from (1).

$$\lambda' = \lambda * \sum_{n=0}^{maximum\ backoff\ stage} ((1-p)*q+p)^n \quad (1)$$

3. With the total transmitting user, we can derive the overall successful bandwidth α request from (2).

$$\alpha = p * \lambda' \quad (2)$$

4. Assume the rate of mean of traffic load carried by a bandwidth request divided by total bandwidth in unit time is β. We have the valid traffic γ from (3).

$$\gamma = \alpha * \beta \quad (3)$$

5. If $\gamma \leq 1$, then the throughput of the system will be less than one. We have the final value of q which is equal to zero and p which is derived in step 1.

   On the contrary, if $\gamma > 1$, then we can find the first-stage q from (4).

$$q = \frac{\gamma - 1}{\gamma} \quad (4)$$

   And repeat the procedure with new q recursively until p and q are precise enough.

After having p and q, we can derive some objective parameter we want, such as throughput, delay, discarded rate etc.

Chapter 4
Simulation result

4.1 Environment

In this chapter, we present our simulation result based on a C simulator. The simulation duration is 5000 frame which we make sure the scheduling process achieves steady state. There are three instances which are standard, the proposed algorithm with traffic expectation equal to traffic arrival mean and the proposed algorithm with traffic expectation equal to zero. Traffic of connections is generated with a Poisson arrival process and go through a collision test before enter to scheduler. The connection which is collided will process a truncated binary exponential backoff mechanism. And we record power consumption per successful bandwidth request, request dropped rate, throughput, signaling overhead per bandwidth request, delay, and collision rate under the different traffic load.

We run the simulation in the following scenario. The collision with padding scenario is the most practical one. And we remove the collision waiting state from algorithm in the collision without padding scenario. That is the connection collide will process the truncated binary exponential backoff mechanism immediately rather than waiting until the T16 expired which the padding means. By doing so, we may considerably reduce the delay time, but weaken the resistance of burst traffic of connection. The collision free scenario is the collision free version of the previous two. As we mentioned before, there are many studies focus on reducing the collision rate of bandwidth request mechanism. This scenario assumes that someday we may reach the degree of collision free, and show the performance under the condition of collision free. Besides, the number of contention slots is also an issue, the more contention slots the lower collision and also the higher throughput. However, the proposed algorithm work under the constraint that throughput is nearly 100%, because the scheduling procedure will not be active while there is still bandwidth available. Hence, to simulate the proposed algorithm, we purposely produce an environment where the number of contention slots larger enough. Even so, we are not shirk from the problem, we simulate two more scenarios under the lower number of contention slots to show the performance of the proposed algorithm still good because of the effect of null response which is designed for collision.

The parameters of our simulation are listed below:

Simulation frame time: 5000 frames

SS number: 1500 (may not concurrent transmit)

Traffic load per connection: 1/15 UL bandwidth of frame.

T16 value: 10 frames

Initial back-off window: 2 slots

Maximum back-off stage: 6 times

Contention slot in a frame: 64 slots

And the title explanations are listed below:

Power consumption: the total active frames of all SS divided by bandwidth successfully transmitting traffic.

Throughput: total used bandwidth divided by total available bandwidth.

Overhead: the number of scheduling signal transmitted divided by bandwidth request successfully transmitting traffic.

Delay: the time interval between bandwidth request generated and traffic transmitted.

Collision rate: collision rate of all generated bandwidth request.

4.2 scenarios

Scenario I: collision with padding

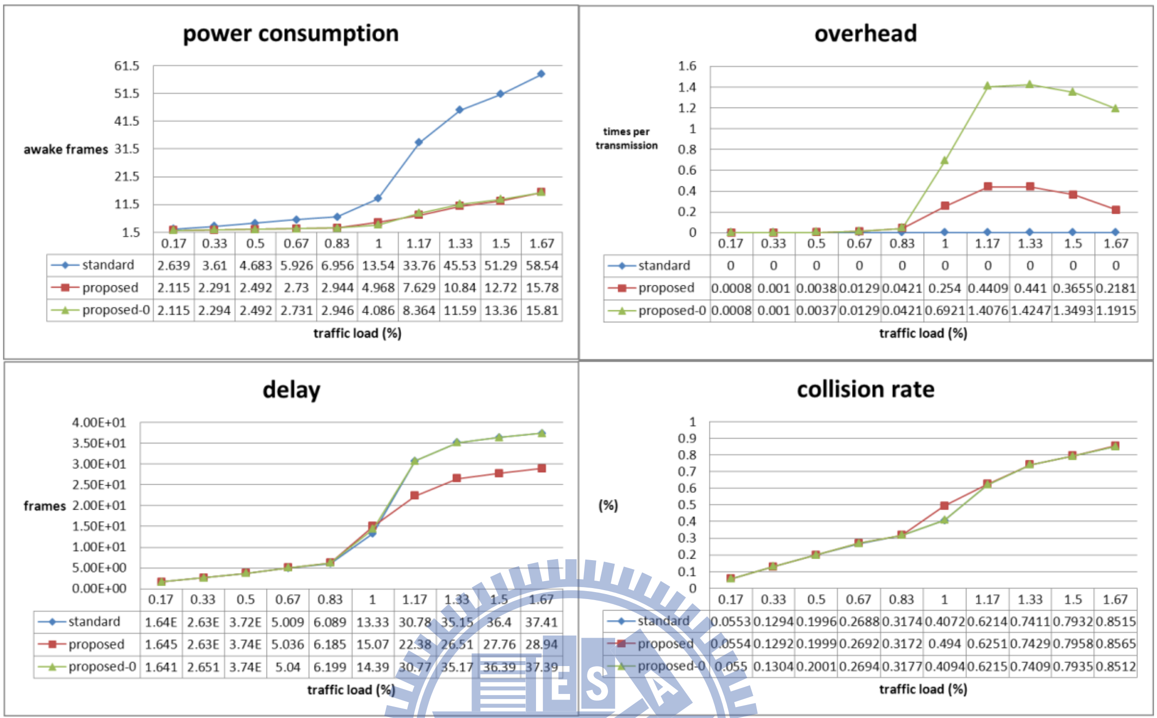


Fig. 10 simulation result under collision test active with padding

Scenario II: collision without padding

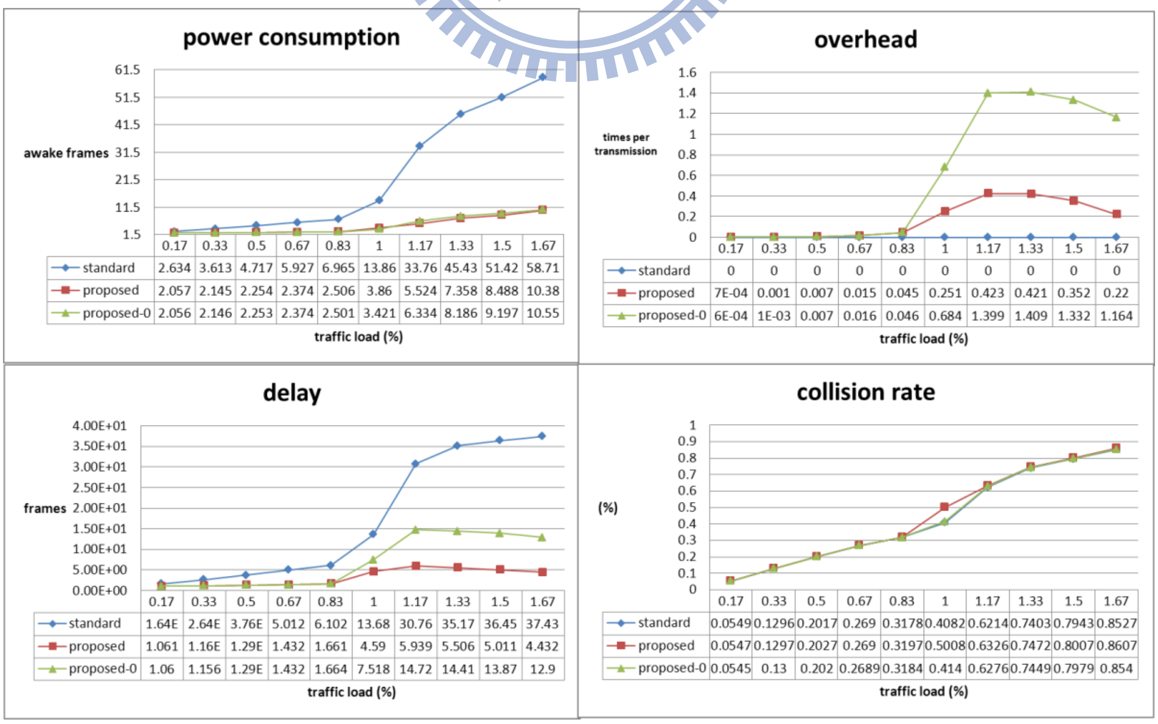


Fig. 11 simulation result under collision test active without padding

Scenario III: collision free

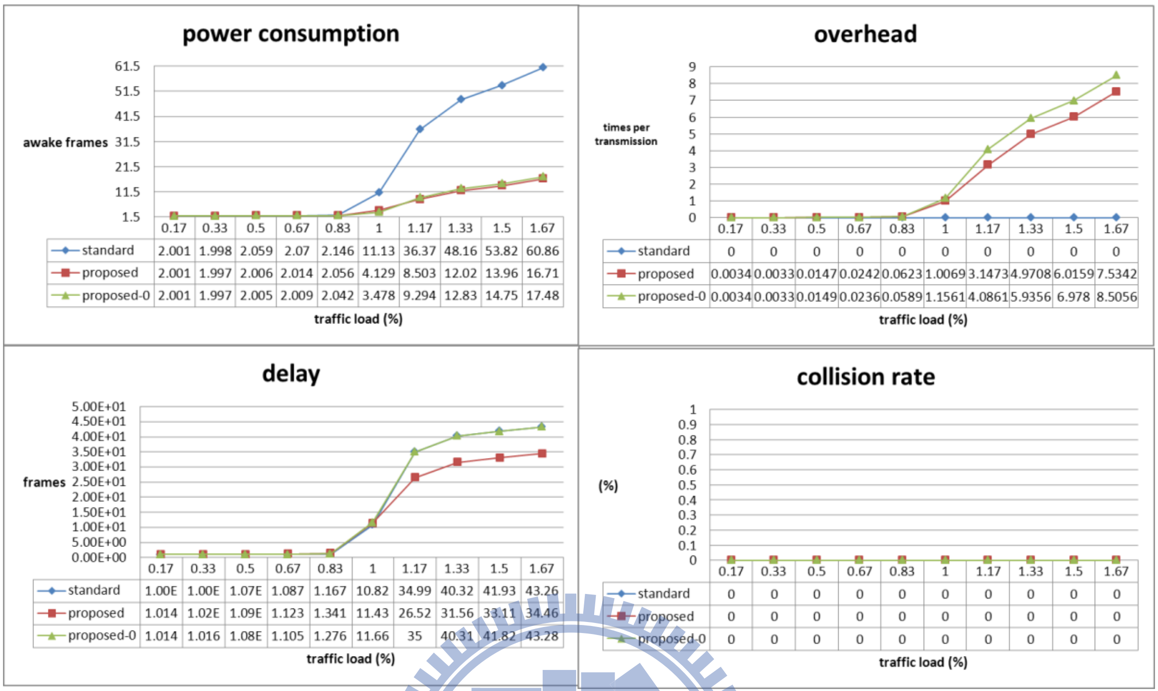
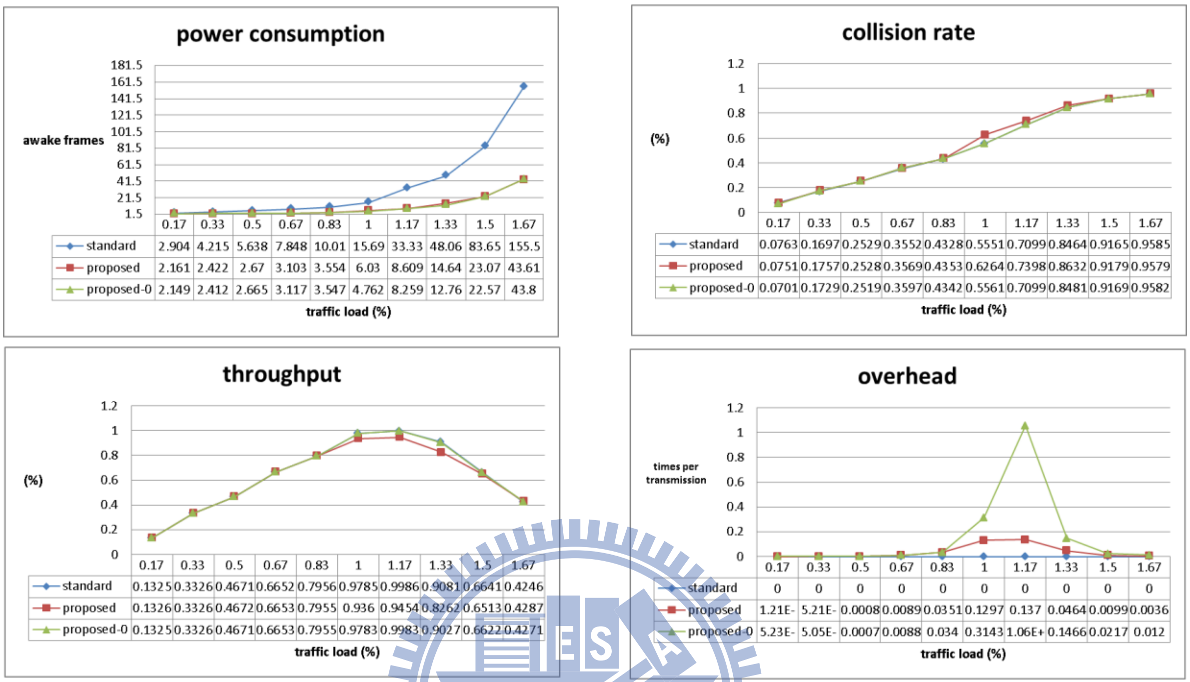


Fig. 12 simulation result under collision test de-active

In Fig.10, 11 and 12, the result of power consumption can be divided into two parts. The first part is while the traffic load is lower than 1, and in this part, the main difference between the standard and the proposed algorithm is the effect of collision which is no difference in Fig. 12. The second part is while the traffic load is not lower than 1, and in this part the main reason which causes the difference between standard and the proposed algorithm is the idle listen which dominated the power consumption in Fig 10, 11 and 12. Overhead shows us when the scheduling signal of the proposed algorithm works, and as we predicted, the scheduling procedure takes place while throughput is about 1 in Fig.10 and 11. And the overhead will be more and more with traffic load increasing in Fig.12, because the more successful bandwidth requests the more the more bandwidth allocation failure. The delay is affected by two factors, one is the expectation and the second is the padding. The first reason results in the difference of the proposed algorithm with expectation equal to mean arrival rate, and the second one results in the difference between the proposed algorithm and standard.

Scenario V and VI: collision with padding and the number of contention slot is 48, 32



power consumption (awake frames)

traffic load (%)	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67
standard	2.904	4.215	5.638	7.848	10.01	15.69	33.33	48.06	83.65	155.5
proposed	2.161	2.422	2.67	3.103	3.554	6.03	8.609	14.64	23.07	43.61
proposed-0	2.149	2.412	2.665	3.117	3.547	4.762	8.259	12.76	22.57	43.8

collision rate (%)

traffic load (%)	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67
standard	0.0763	0.1697	0.2529	0.3552	0.4328	0.5551	0.7099	0.8464	0.9165	0.9585
proposed	0.0751	0.1757	0.2528	0.3569	0.4353	0.6264	0.7398	0.8632	0.9179	0.9579
proposed-0	0.0701	0.1729	0.2519	0.3597	0.4342	0.5561	0.7099	0.8481	0.9169	0.9582

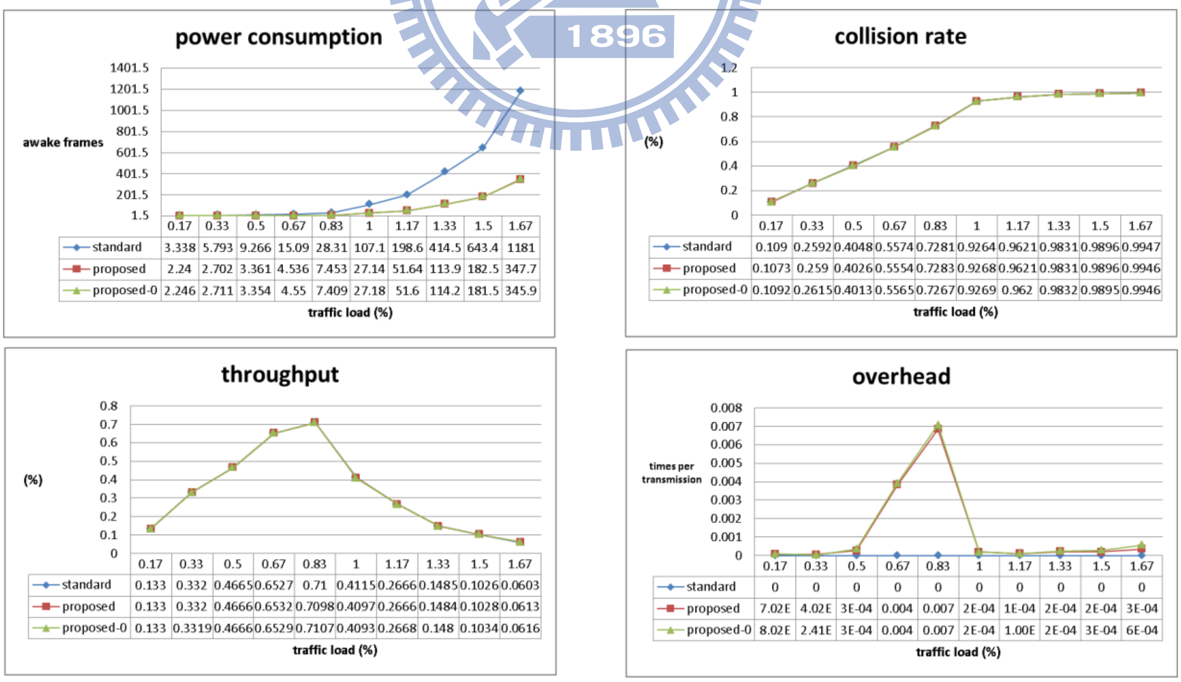
throughput (%)

traffic load (%)	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67
standard	0.1325	0.3326	0.4671	0.6652	0.7956	0.9785	0.9986	0.9081	0.6641	0.4246
proposed	0.1326	0.3326	0.4672	0.6653	0.7955	0.936	0.9454	0.8262	0.6513	0.4287
proposed-0	0.1325	0.3326	0.4671	0.6653	0.7955	0.9783	0.9983	0.9027	0.6622	0.4271

overhead (times per transmission)

traffic load (%)	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67
standard	0	0	0	0	0	0	0	0	0	0
proposed	1.21E-	5.21E-	0.0008	0.0089	0.0351	0.1297	0.137	0.0464	0.0099	0.0036
proposed-0	5.23E-	5.05E-	0.0007	0.0088	0.034	0.3143	1.06E+	0.1466	0.0217	0.012

Fig. 13 simulation result under 48 contention slots



power consumption (awake frames)

traffic load (%)	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67
standard	3.338	5.793	9.266	15.09	28.31	107.1	198.6	414.5	643.4	1181
proposed	2.24	2.702	3.361	4.536	7.453	27.14	51.64	113.9	182.5	347.7
proposed-0	2.246	2.711	3.354	4.55	7.409	27.18	51.6	114.2	181.5	345.9

collision rate (%)

traffic load (%)	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67
standard	0.109	0.2592	0.4048	0.5574	0.7281	0.9264	0.9621	0.9831	0.9896	0.9947
proposed	0.1073	0.259	0.4026	0.5554	0.7283	0.9268	0.9621	0.9831	0.9896	0.9946
proposed-0	0.1092	0.2615	0.4013	0.5565	0.7267	0.9269	0.962	0.9832	0.9895	0.9946

throughput (%)

traffic load (%)	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67
standard	0.133	0.332	0.4665	0.6527	0.71	0.4115	0.2666	0.1485	0.1026	0.0603
proposed	0.133	0.332	0.4666	0.6532	0.7098	0.4097	0.2666	0.1484	0.1028	0.0613
proposed-0	0.133	0.3319	0.4666	0.6529	0.7107	0.4093	0.2668	0.148	0.1034	0.0616

overhead (times per transmission)

traffic load (%)	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67
standard	0	0	0	0	0	0	0	0	0	0
proposed	7.02E	4.02E	3E-04	0.004	0.007	2E-04	1E-04	2E-04	2E-04	3E-04
proposed-0	8.02E	2.41E	3E-04	0.004	0.007	2E-04	1.00E	2E-04	3E-04	6E-04

Fig. 14 simulation result under 32 contention slots

Fig. 13 and 14 are the condition where the contention slot is insufficient. We can see the overhead is nearly zero which means that there is few scheduling signal. The main factor affecting the power consumption in Fig. 13 and 14 is the null response of collision, and the collision causes the waste of power while counting down the T16 timer which is avoided by null response in the proposed algorithm.

The final part of simulation is the comparison between the analytical model and simulation. The table below shows that the values extracted from simulator fit the results of procedure of find p and q. and we can find that the results are similar.

Traffic load	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67
q(simulation)	0	0	0.001	0.006	0.011	0.021	0.157	0.262	0.351	0.416
q(analytical)	0	0	0	0	0	0	0.145	0.248	0.333	0.401

Chapter 5
Conclusion

From the simulation last chapter, we can image the efficiency of the proposed algorithm and the overhead it paid. We develop the proposed algorithm for the purpose of saving power and create a tradeoff between power and overhead.

As the result of simulation, the proposed algorithm work well with little overhead in the most practical simulation, and work with higher overhead while the condition of collision free. However, a simple simulation provides nothing but a rough estimation of the proposed algorithm. The implement may not work as well as we expect.

There are many solution of the problem we deal with, such as more moderate polling mechanism. The proposed algorithm is only a straightforward achievement. There are still much to go, such as the issue of packing and the fair priority of bandwidth request. The first issue means that how we make a decision that which connection is better to scheduling while there are two connections that one fits the remain bandwidth but the other has higher priority. The second issue is discus about how do we decide which bandwidth request first for fairness. The rule of the proposed algorithm is to sort bandwidth requests with the order of the time they transmitted, but there are many factors which may affect the sorting procedure, and we have not considered yet. Also, the analytical model is still incomplete, and we may try to derive it completely. With more survey, we may fix the proposed algorithm properly one day.

References

[1] IEEE Standard 802.16-2009 IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems, May 2009.

[2] Q. Ni , Vinel A. , Y. Xiao, Turlikov A., and T. Jiang, "WIRELESS BROADBAND ACCESS: WIMAX AND BEYOND - Investigation of Bandwidth Request Mechanisms under Point-to-Multipoint Mode of WiMAX Networks," Communications Magazine, IEEE, Vol.45, pp. 132, May 2007.

[3] L. L Wang, and H. M. Xu, "A Load Balance Based Sleep Mode Management Strategy for VoIP Services in IEEE 802.16 Systems," Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on Wireless Communications, pp. 1, Oct. 2008.

[4] S. W. Kwon, and D. H. Cho, "Dynamic Power Saving Mechanism for Mobile Station in the IEEE 802.16e Systems," VTC Spring 2009. IEEE 69th Vehicular Technology Conference, pp. 1, April 2009.

[5] S. C. Shiu, and J. Y. Pan, "Energy-Efficient Congestion Control on Contention-Based BW-REQ in WiMAX System," APCC 2008. 14th Asia-Pacific Conference on Communications, pp. 1, Oct. 2008.

[6] R. Y. Kim, and S. Mohanty, "Advanced power management techniques in next-generation wireless networks," Communications Magazine, IEEE, Vol. 48, pp. 94, May 2010.

[7] A. Ukil, and J. Sen, "Proactive resource reservation in next-generation wireless networks," 2010 National Conference on Communications (NCC), pp. 1, Jan. 2010.

[8] J. H. Chen, and Y. L. Li, "Adaptive traffic indication algorithm for energy efficiency in IEEE 802.16e systems," 2010 5th International ICST Conference on Communications and Networking in China (CHINACOM), pp. 1, Aug. 2010.

[9] Bianchi G., "Performance analysis of the IEEE 802.11 distributed coordination function," IEEE Journal on Selected Areas in Communications, Vol.18, pp. 535, Mar 2000.

[10] W. Ye, J. Heidemann, and D. Estring, "Medium access control with coordinated adaptive sleeping for wireless sensor networks," IEEE/ACM Transactions on Networking, Vol.12, pp. 493, June 2004.

100

碩士論文

針對非即時訊務在IEEE802.16e系統的一種排程式頻寬要求節能機制

交通大學
電機學院
電信工程研究所

林筠翰