

무선랜에서의 실시간 및 비실시간 트래픽 스케줄링

이주희* · 이재영*

Scheduling of Real-time and Nonreal-time Traffics in IEEE 802.11 Wireless LAN

Ju-Hee Lee* · Chae Y. Lee*

■ Abstract ■

Media Access Control (MAC) Protocol in IEEE 802.11 Wireless LAN standard supports two types of services, synchronous and asynchronous. Synchronous real-time traffic is served by Point Coordination Function (PCF) that implements polling access method. Asynchronous nonreal-time traffic is provided by Distributed Coordination Function (DCF) based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol. Since real-time traffic is sensitive to delay, and nonreal-time traffic to error and throughput, proper traffic scheduling algorithm needs to be designed. But it is known that the standard IEEE 802.11 scheme is insufficient to serve real-time traffic.

In this paper, real-time traffic scheduling and admission control algorithm is proposed. To satisfy the deadline violation probability of the real time traffic the downlink traffic is scheduled before the uplink by Earliest Due Date (EDD) rule. Admission of real-time connection is controlled to satisfy the minimum throughput of nonreal-time traffic which is estimated by exponential smoothing.

Simulation is performed to have proper system capacity that satisfies the Quality of Service (QoS) requirement. Tradeoff between real-time and nonreal-time stations is demonstrated. The admission control and the EDD with downlink-first scheduling are illustrated to be effective for the real-time traffic in the wireless LAN.

Keyword : Wireless LAN, Call Admission, EDD Scheduling, Throughput

1. Introduction

In recent years, wireless LANs are becoming more widely recognized as a general-purpose connectivity alternative for a broad range of business community. Wireless LANs offer the productivity, convenience, and cost advantages over traditional wired networks because of mobility, installation speed and simplicity, installation flexibility, scalability and etc. On the other hand, in accordance with the evolution of Internet, transmitting real-time traffic like voice or video traffic through networks is growing. However, it is known that the standard IEEE 802.11 scheme is yet insufficient to serve real-time traffic [1].

The combined performance of data transmission with the Distributed Coordination Function (DCF) and voice with the Point Coordination Function (PCF) is evaluated in [3]. They improved the performance of queued data traffics by dropping idle voice stations. Performance of video and data transmission with the standard protocol is studied in [4]. The throughput and average MAC Protocol Data Unit (MPDU) delay is investigated for various contention free period (CFP) and CFP repetition interval. A Black burst contention period is suggested by [5] in which each station jams the channel with a number of black slots proportional to its waiting time. However, the scheme is difficult to implement and not compatible to the standard. In [6], a real-time connection establishment procedure is proposed within the framework of the PCF. A time window concept is proposed in which each station is polled and allowed to transmit a data frame in each superframe.

Considering two types of traffics : the real-

time traffic is sensitive to delay, and the non-real-time traffic is to error and throughput. For example, Voice over Internet Protocol (VoIP) is particularly sensitive to both packet loss and delay : 1% packet loss and 150ms delay are typically considered the maximum acceptable values [2]. Hence, a proper traffic scheduling algorithm needs to be examined that is compatible to the IEEE 802.11 standard.

In this paper, we propose a scheduling for the real-time traffic to satisfy the Quality of Service (QoS) requirements of both real-time and non-real-time traffics. The paper is organized as follows. Section 2 describes the standard media access control of the IEEE 802.11 wireless LAN. In Section 3, scheduling of real-time traffic and call admission strategy is developed which is compatible to the standard. Simulation environment and the results are discussed in Section 4. Section 5 concludes the paper.

2. Medium Access Control (MAC) in IEEE 802.11 Wireless LAN

The IEEE 802.11 wireless LAN system has a basic service set (BSS) which is defined as a group of stations that are under the direct control of a single coordination function i.e., DCF and/or PCF. Each station transmits directly to any other stations in the same BSS. On the other hand, to transmit data to another station belonging to a different BSS, the data must pass through an access point which is an inter-working unit implementing both the IEEE 802.11 and the distribution system MAC protocols. The distribution system can be thought of as a backbone network that is responsible for MAC-level

transport.

The IEEE 802.11 MAC protocol provides two service types : asynchronous and synchronous. These types can be provided on top of a variety of physical layers and for different data rates. The Distributed Coordination Function (DCF) provides the asynchronous type of service and the Point Coordination Function (PCF) provides synchronous type of service.

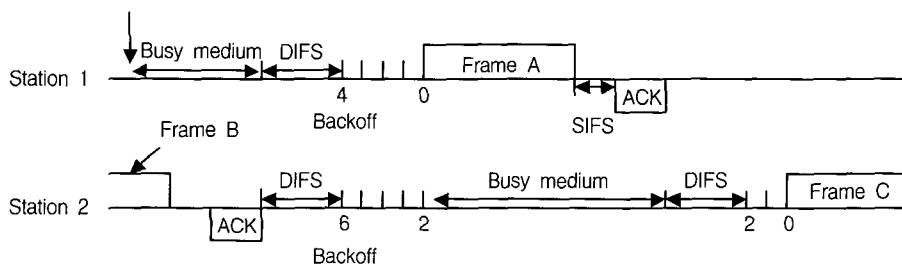
The PCF and the DCF can coexist in a manner that the two access methods alternate in a cycle called the Contention Free Period (CFP) repetition interval. The PCF provides synchronous type of service during the CFP and the DCF provides asynchronous type of service during the Contention Period (CP). A CFP and the following CP together are referred to as a super-frame.

2.1 Distributed Coordination Function (DCF)

The DCF is the fundamental access method of the IEEE 802.11, known as carrier sense multiple access with collision avoidance, or CSMA/CA. CSMA/CD (Collision Detection) is not used because a station is unable to listen to the channel for collisions while transmitting. In a wireless environment, collision detection is impossible. The CSMA/CA protocol is designed to re-

duce the collision probability between multiple stations accessing one medium, at the point when collisions would most likely occur.

[Figure 1] illustrates DCF access mode. According to the DCF, a station must sense the medium before initiating the transmission of a frame. If the medium is sensed as being idle for a time interval greater than a Distributed Inter Frame Space (DIFS) then the station transmits the frame. Otherwise, the transmission is deferred and the backoff process is started. The station computes a random time interval, the backoff interval, uniformly distributed between zero and a maximum called Contention Window (CW). This backoff interval is used to initialize the backoff timer. After the medium becomes idle after a DIFS period, stations periodically decrements their backoff timer until the medium becomes busy again or the timer reaches zero. The decrement period is referred to as the slot-time, which corresponds to the maximum round-trip delay within the BSS. If the timer has not reached zero and the medium becomes busy, the station freezes its timer. When the timer is finally decremented to zero, the station transmits its frame. If two or more stations decrement to zero at the same time, a collision will occur. Because collision detection is not possible, a positive acknowledgement is used to notify the



[Figure 1] DCF access mechanism

sending station that the transmitted frame has been successfully received. The transmission of the acknowledgement is initiated at a time interval equal to the Short InterFrame Space (SIFS) after the end of the reception of the previous frame. Since the SIFS is, by definition, less than the DIFS the receiving station is given priority over other stations that are attempting to get transmission opportunities.

If the acknowledgement is not received the station assumes that the collision occurred, hence, schedules a retransmission and enters the backoff process again. However, to reduce the probability of collisions, after each unsuccessful transmission attempt, the Contention Window is increased until a predefined maximum value (CWmax) is reached. The CW values shall be sequentially ascending beginning with predefined minimum value (CWmin).

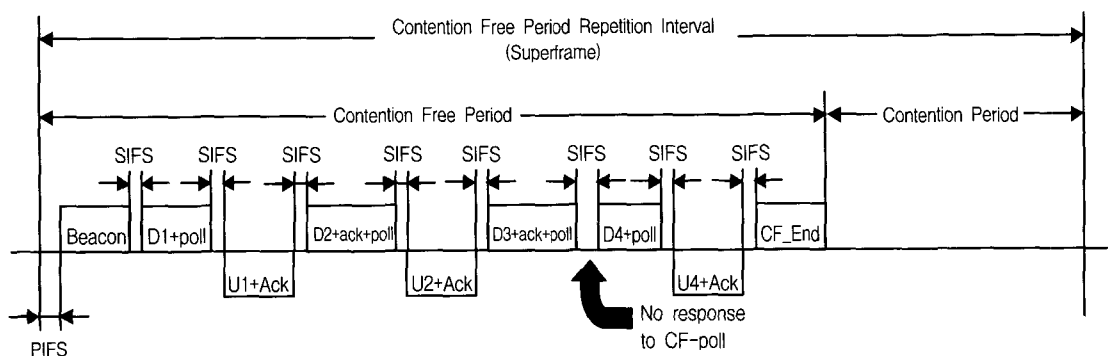
A refinement of the method may be used under various circumstances to further minimize collisions. Here the transmitting and receiving station exchange short control frames, request to send (RTS) and clear to send (CTS) frames, after determining that the medium is idle and after any deferrals or backoffs, prior to data transmission. The details of RTS/CTS exchanges

can be found in [7].

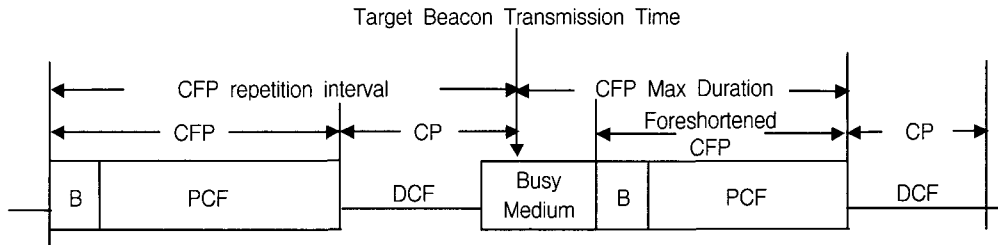
2.2 Point Coordination Function (PCF)

The PCF is an optional capability, which is connection-oriented, and provides contention-free (CF) frame transfer. The PCF relies on the point coordinator to perform polling, enabling polled stations to transmit without contending for the channel. The access point within each BSS usually performs the function of the point coordinator.

[Figure 2] shows an example of PCF frame transfer. The CFP repetition interval (super-frame) is used to determine the frequency with which the PCF occurs. Within a repetition interval, a portion of time is allotted to contention-free traffic, and the remainder is provided for contention-based traffic. A nominal beginning of each CFP, the point coordinator senses the medium. When the medium is idle for one Point InterFrame Space (PIFS) period, the point coordinator transmits a beacon frame. Because the PIFS is shorter than DIFS and longer than SIFS, PCF is given priority over DCF frame transmission. The point coordinator starts contention free transmission after the beacon frame by send-



[Figure 2] Example of PCF frame transfer



[Figure 3] Foreshortened CFP

ing a CF-Poll, Data, or Data+CF-Poll frame. If a station receives a CF-Poll frame from the point coordinator, the station can respond to the point coordinator after a SIFS idle period with a CF-ACK or a Data+CF-ACK frame. If the point coordinator receives a Data+CF-ACK frame from a station, the point coordinator can send a Data+CF-ACK+CF-Poll frame to a different station, where the CF-ACK portion of the frame is used to acknowledge receipt of the previous data frame. The point coordinator can immediately terminate the CFP by transmitting a CF_End frame, which is common if the network is lightly loaded and the point coordinator has no traffic buffered.

The duration of the CFP repetition interval is a manageable parameter. It is up to the point coordinator to determine how long to operate the CFP during any given repetition interval. If traffic is very light, the point coordinator may shorten the CFP and provide the remainder of the repetition interval for the DCF. But the maximum size of the CFP is determined by the manageable parameter CFP_Max_Duration. The minimum value of CFP_Max_Duration is determined by the time required for the point coordinator to send one data frame to a station, while polling that station, and for the polled station to respond with one data frame.

The CFP may be shortened as in [Figure 3] if

DCF traffic from the previous repetition interval carries over into the current interval. If, for instance, a station with nonreal-time traffic starts transmission just before a superframe and lasts longer than the remaining contention period, the point coordinator has to defer the start of its real-time traffic transmission until the medium becomes free for a PIFS. The maximum amount of delay that can be incurred is the time it takes to transmit a maximum MPDU and ACK. [Figure 3] is the sketch of the delayed beacon and foreshortened CFP.

3. Scheduling of Real-time and Nonreal-time Traffics

The goal of traffic scheduling in the wireless LAN is to satisfy the QoS requirement of each type of traffic. In this study we consider end-to-end delay for the real-time voice traffic and the throughput of nonreal-time data traffic. For the real-time traffic the characteristic of widespread VoIP, G.723.1 [8] is considered.

The real-time traffic is assumed to have the known traffic characteristic (d_{max}, T_I, t_{MPDU}) , which represents that each source generates a frame (MAC Protocol Data Unit) of length t_{MPDU} every time interval T_I with end-to-end maximum allowable delay d_{max} . Then the QoS

requirement of the real-time traffic can be defined with the deadline violation probability as $p(d_{realtime} > d_{max}) \leq \alpha$, where $d_{realtime}$ is the end-to-end delay experienced by the real-time traffic and α is the threshold probability. For example, if $\alpha = 0.01$, it means that at least 99% of the real-time frames must be transmitted with the predefined end-to-end delay, d_{max} . For nonreal-time traffic, it is assumed that a throughput ρ_{min} has to be guaranteed for each active nonreal-time source.

Now, to support the QoS of both the real-time and nonreal-time traffics in the wireless LAN, the following three factors need to be considered.

- (1) CFP repetition interval and the duration of CFP
- (2) Real-time traffic scheduling in the CFP
- (3) Admission control of real-time traffic

Among the above three factors, the duration of the CFP and CP is important. Depending on the volume of the real-time and nonreal-time traffic, the CFP and CP are adjusted. Thus, the admission control of the real-time traffic is necessary to satisfy the QoS. In this paper the admission of the real-time traffic is controlled such that it satisfies the maximum limit of the duration of CFP and the throughput of the nonreal-time traffic. The admitted real-time traffic is then scheduled by EDD rule to minimize the deadline violation probability.

3.1 Determination of CFP repetition interval and CFP

The CFP repetition interval influences the delay of real-time traffic. Let T_{Rep} denote the

CFP repetition interval which represents the nominal length of a superframe. When T_{Rep} is short, the average delay of the real-time frame becomes small due to the frequent polls of each real-time station. On the other hand, when T_{Rep} is long, the average delay of real-time traffic becomes large. Hence, to minimize the delay experienced in the link between a station and the point coordinator, it is desirable to reduce T_{Rep} . Suzuki and Tasaka [4] investigated the relationship between CFP repetition interval and delay performance of real-time traffic numerically. They demonstrated that if the CFP repetition interval is set too long, the delay performance of the real-time video traffic deteriorates drastically.

According to the CSMA/CA protocol, the point coordinator might be unable to control the channel at the nominal beginning of the superframe. Thus, to serve all of the real-time traffic in queue during the CFP the experienced maximum delay becomes $T_{Rep} + t_{MaxMPDU}$, where $t_{MaxMPDU}$ represents the maximum size of the nonreal-time traffic MPDU as discussed in [Figure 3]. Let d_{target} denote the target delay between a station and the point coordinator. Then T_{Rep} is restricted by $T_{Rep} + t_{MaxMPDU} \leq d_{target}$. T_{Rep} also influences the delay jitter of real-time frames. It is desirable to set T_{Rep} to be equal to the frame generation interval T_I to minimize the delay jitter.

The duration of the CFP also influences the delay of real-time frames. Let T_{CFP} denote the average duration of CFP in each superframe. Then as shown in [Figure 2], T_{CFP} can be estimated with the PIFS time, T_{PIFS} the duration

of beacon frame (T_B), the number of real-time stations in the system (N_r), the SIFS time (T_{SIFS}) and the duration of CF_End frame (T_{CF_END}). The TCFP is increased as the number of real-time stations increases. By assuming every pair of up and down link traffic in the contention free period, the T_{CFP} is given by

$$T_{CFP} = T_{PIFS} + T_B + N_r \times \frac{T_{Rep}}{T_I} \times (T_{SIFS} + t_{MPDU}) \times 2 + T_{SIFS} + T_{CF_End} \quad (1)$$

However, the maximum size of the CFP is limited to the T_{MaxCFP} which is the CFP_Max_Duration. Note that the maximum difference between the actual and the nominal start time of a superframe is $t_{MaxMPDU}$. Thus, in order to serve real-time traffic in each superframe T_{CFP} is restricted by $T_{CFP} + t_{MaxMPDU} \leq T_{MaxCFP}$.

The above discussion can be summarized with the following restriction for the duration of T_{Rep} and T_{CFP} .

1. The TRep has to satisfy the delay constraint :

$$T_{Rep} \leq d_{target} - t_{MaxMPDU}$$

- A. If $T_I \leq d_{target} - t_{MaxMPDU}$, then it is recommended to keep T_{Rep} as $T_I \leq T_{Rep} \leq d_{target} - t_{MaxMPDU}$. To minimize the delay and delay jitter experienced in the link we propose to set $T_{Rep} = T_I$.
- B. If $T_I \geq d_{target} - t_{MaxMPDU}$, set $T_{Rep} \leq d_{target} - t_{MaxMPDU}$. Since the traffic interval is relatively long, it is desirable to set the superframe interval relatively short to reduce the delay.

2. The T_{CFP} has to satisfy the capacity constraint : $T_{CFP} + t_{MaxMPDU} \leq T_{MaxCFP}$

3.2 Real-time traffic scheduling

To provide the QoS guarantee in wireless LAN, several traffic scheduling schemes have been proposed including deficit round robin [9] and distributed deficit round robin [10]. However, it is hard to satisfy QoS requirement with simple round-robin scheme or fair queueing scheduling algorithm. This is because the real-time traffic generally requests end-to-end delay bound. By assuming that real-time connections are established with stations in different basic service sets or different backbone networks the end-to-end delay for downlink traffic becomes critical depending on the traffic scheduling. As an example, even if any two frames arrived at the point coordinator at the same time, their remaining service time to satisfy the end-to-end delay bound may be different due to the location of source station and the route of the frame. Hence, the First-In First-Out (FIFO) scheduling policy or round robin scheduling may not satisfy the QoS requirement.

The Earliest Due Date (EDD), also denoted as Earliest Deadline First (EDF), is a mechanism to provide absolute delay differentiation. It is well known that EDD policy is optimal in terms of minimizing the maximum latency of packets if the deadline can be associated with the packet [11]. Choi and Shin [12] used improved EDF scheduling algorithm for real-time traffic scheduling. The frame with earlier deadline gets the higher priority. However, the scheduling is based on the delay between the wireless radio link which is a fraction of end-to-end delay.

In this paper, the downlink traffic is scheduled first by the EDD rule that applies the frame generation time at the other end. In the implementation the frame generation time may well be approximated with the hop count by utilizing the Time-to-live field in the IP (Internet Protocol) header. The uplink traffic is scheduled after the downlink traffic. A cyclical scheduling algorithm as in [3] is employed. ACK is not used for real-time service.

The uplink and downlink scheduling of the real-time traffic is as follows.

1. Polling table maintenance

The point coordinator maintains polling list. An arriving connection request, if accepted, is placed at the end of the polling list and a closing connection is deleted from the list.

2. Scheduling

For uplink traffic, all the real-time stations are polled only once during each superframe. The point coordinator polls real-time stations in order of polling list.

For downlink traffic, the point coordinator calculates service due of frames in queue. Let the frame i be generated at t_i , then the service due of frame i becomes $t_i + d_{\max}$. Thus the point coordinator serves the frames in the order of service due.

3. Fundamental rules

No ACK is used.

The CFP can be last up to T_{MaxCFP} . When the queue is cleared out for the downlink traffic and all stations in the polling list are polled or it has passed T_{MaxCFP} since the nominal start time of the previous superframe, the point coordinator closes contention free period and switches to contention period.

3.3 Admission control in the wireless LAN

In this section, the admission control for real-time traffic is considered. The admission of real-time traffic largely affects the QoS of the wireless LAN traffics. If excess real-time traffic is admitted, the transmission delay of each real-time traffic is increased. Also, the throughput of nonreal-time stations is diminished due to the reduced contention period. The objective of admission control is twofold. It is to keep the deadline violation probability of real-time traffic below the predefined threshold and to guarantee minimum throughput for individual nonreal-time station.

To satisfy above QoS requirement in the wireless LAN, it is necessary to estimate the deadline violation probability of real-time traffic and the throughput of the nonreal-time traffic. Depending on the result of the estimation, the admission will be accepted or denied for the incoming real-time traffic.

3.3.1 Estimation of the throughput of nonreal-time stations

Cali et al. [14] considered the following aggregated throughput of IEEE 802.11 MAC protocol for a system only with DCF.

$$\rho_{agg} = \frac{\bar{m}}{E(t_v)} \quad (2)$$

In the equation t_v is the time interval between two successful transmissions from a nonreal-time station, which is referred to as virtual transmission time, and \bar{m} represents average length of nonreal-time frame. Since $E(t_v)$ is the average virtual transmission time, we have

$$E(t_v) = E(T_W) + E(T_{TX})$$

where $E(T_W)$ is the average waiting time to have a successful transmission and $E(T_{TX})$ represents the average time taken to transmit a frame and to exchange ACK successfully. Then \bar{m} is written as

$$\bar{m} = E(T_{TX}) - T_{SIFS} - T_{ACK}$$

where T_{ACK} is the duration of ACK frame.

Now, assuming N_{nr} is the active nonreal-time stations in the system, the throughput ρ of each active nonreal-time station can be represented as

$$\rho = \frac{\rho_{agg}}{N_{nr}} \quad (3)$$

Thus to estimate the throughput, we need to estimate $E(T_W)$ and $E(T_{TX})$. In this paper we propose the following exponential smoothing method which is known to be an excellent estimation with small amount of historical data.

$$E(T_W^{i+1}) = \beta E(T_W^i) + (1 - \beta) T_W^i \quad (4)$$

$$E(T_{TX}^{i+1}) = \beta E(T_{TX}^i) + (1 - \beta) T_{TX}^i \quad (5)$$

$E(T_W^i)$ and $E(T_{TX}^i)$ are respectively the approximation of $E(T_W)$ and $E(T_{TX})$ at the end of the i_{th} transmission attempt. T_W^i is the time taken for a successful transmission at the i_{th} transmission attempt, and T_{TX}^i is the time taken to transmit a frame and to exchange ACK successfully at the i_{th} transmission. β is a smoothing factor. For each virtual transmission time, the point coordinator measures T_W^i and T_{TX}^i , and calculates $E(T_W^{i+1})$ and $E(T_{TX}^{i+1})$, which

finally leads to the throughput of a nonreal-time station in the system.

3.3.2 Estimation of the deadline violation probability of real-time traffic

As in Section 3.1, the QoS requirement of real-time traffic is to maintain the deadline violation probability below the threshold α . i.e., $p(d_{realtime} > d_{max}) \leq \alpha$.

Since all stations in the system are guaranteed to be polled once in a superframe, no deadline violations happen for uplink frames. The threshold probability, α only relates to the downlink traffic. To estimate the probability at every superframe the following exponential smoothing is employed.

$$p^{j+1}(d_{realtime} > d_{max}) = \gamma p^j(d_{realtime} > d_{max})_j + (1 - \gamma) q^j \quad (6)$$

In the equation q^j is the portion of downlink frames that experienced delay more than d_{max} in superframe j , $p^j(d_{realtime} > d_{max})_j$ represents the approximation at the end of the superframe j , and γ is a smoothing factor.

In every superframe, the point coordinator measures q_j and calculates the deadline violation probability $p^{j+1}(d_{realtime} > d_{max})$ with equation (6).

3.3.3 Estimation of the throughput of nonreal-time stations with a new real-time connection

Even if the QoS requirement of both real-time and nonreal-time traffic is satisfied, the throughput of nonreal-time stations may be diminished due to the reduced contention period by the admission of the new real-time connection. Thus to decide the admission of the new real-time connection into the system, it is necessary to

estimate the throughput assuming the real-time traffic is admitted.

Let T_{CP} denote the average duration of contention period in each superframe. Then we have $T_{CP} = T_{Rep} - T_{CFP}$. Now, let T'_{CP} and T'_{CFP} denote the expected value of T_{CP} and T_{CFP} by accepting a new real-time connection. From equation (3.1) T'_{CFP} is given by

$$T'_{CFP} = T_{PIFS} + T_B + (N_r + 1) \times \frac{T_{Rep}}{T_I} \times (T_{SIFS} + t_{MPDU}) \times 2 + T_{SIFS} + T_{CF_End} \quad (7)$$

Since T_{Rep} is fixed, we have

$$T'_{CP} = T_{Rep} - T'_{CFP} \quad (8)$$

Finally the throughput is updated to ρ' which is given by

$$\rho' = \rho \frac{T'_{CP}}{T_{CP}} \quad (9)$$

Based on the estimation of the throughputs ρ and ρ' and the deadline violation probability we

propose the following admission control algorithm for a new real-time traffic.

Proposed Admission Control Algorithm

If a real-time connection request is received,

Step 1. Check if the QoS requirements of on-going connections are satisfied

IF $p(d_{realtime} > d_{max}) < \alpha$ and $\rho > \rho_{min}$, go to Step 2.

ELSE, reject the request.

Step 2. Check if the QoS of existing connections are satisfied after admitting the new real-time connection request

IF $T'_{CFP} \leq T_{MaxCFP} - t_{MaxMPDU}$ and

$\rho' > \rho_{min}$, accept the request.

ELSE, reject the request.

4. Simulation of the Traffic Scheduling

The system parameters for the simulation environment are reported in <Table 1> as specified in the IEEE 802.11b standard [15]. To simplify the simulation, the radio link propagation delay is assumed zero with no transmission

<Table 1> System parameters for the IEEE 802.11b standard

| Attribute | Symbol | Value |
|---------------------------|-------------------------|------------|
| Channel rate | CR | 11 Mbps |
| Maximum contention window | CW_{min} | 31 |
| Minimum contention window | CW_{max} | 1023 |
| ACK frame size | $T_{ACK} \times CR$ | 14 octets |
| CF-End frame size | $T_{CF_End} \times CR$ | 20 octets |
| Slot time | T_{ST} | 20 μs |
| SIFS time | T_{SIFS} | 10 μs |
| PIFS time | T_{PIFS} | 30 μs |
| DOFS time | T_{DIFS} | 50 μs |

errors.

The parameters for the real-time traffic are summarized in <Table 2>. Main characteristics of the real-time traffic are taken from the G.723.1 protocol [8]. The frame length of real-time traffic is set to 300 octets considering the overheads of upper layer protocols. To reflect the end-to-end delay bound of real-time traffic, the remaining due, which represents the remaining time to the service deadline, between a station and the pointer coordinator is considered instead of end-to-end delay between two communicating stations. The remaining dues of real-time frame are generated from a uniform distribution over the interval (due_{min} , due_{max}). Each downlink frame is assumed to arrive at the access point differently with service due from 30ms to 40ms.

The target delay between a station and the pointer coordinator, d_{target} is set by 35ms as in

[10]. Since d_{target} and T_I are set to 35ms and 30ms respectively, T_{Rep} is set to 30ms, to satisfy the scheme proposed in Section 3.1. The allowed deadline violation probability α is set to 0.01 as in [2, 10]. The CFP_Max_Duration is set to 28ms considering maximum size of MPDU and other parameters including the slot time and the size of ACK frame [7].

<Table 3> shows the parameters for the non-real-time traffic. The maximum and minimum MPDU sizes are taken from the IEEE 802.11 wireless LAN standard [7]. The MPDU sizes of non-real-time traffic are obtained from a uniform distribution over the interval ($MPDU_{min}$, $MPDU_{max}$). Each MPDU by each nonreal-time station is generated by following the Poisson process with the arrival rate $\lambda = 1/30ms$ such that the expected number of traffic by each station is one MPDU per superframe.

<Table 2> Parameters for the real-time traffic

| Attribute | Symbol | Value |
|--------------------------------|---------------|------------|
| Real-time traffic frame length | $MPDU_{real}$ | 300 octets |
| Qos requirement | α | 0.01 |
| Target delay | d_{target} | 35 ms |
| Traffic generation interval | T_I | 30 ms |
| CFP repetition interval | T_{Rep} | 30 ms |
| CFP_Max_Duration | T_{MaxCFP} | 28 ms |
| Maximum remaining due | due_{max} | 40 ms |
| Minimum remaining due | due_{min} | 30 ms |

<Table 3> Parameters for the nonreal-time traffic

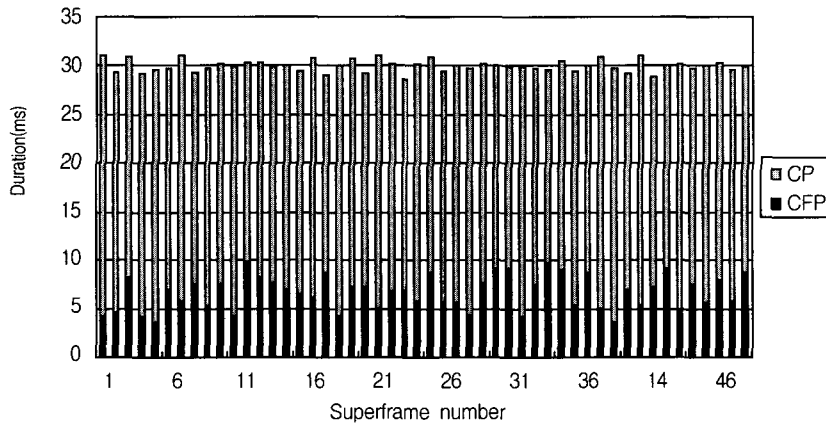
| Attribute | Symbol | Value |
|--------------------------|--------------|-------------|
| Maximum MPDU size | $MPDU_{max}$ | 2346 octets |
| Minimum MPDU size | $MPDU_{min}$ | 34 octets |
| Minimum throughput bound | ρ_{min} | 0.05 |

The smoothing factors β and γ for estimating the throughput and the deadline violation probability are set to 0.99. It means that 99% of the current estimate is from the estimate just one time period ahead. It is desirable to set the smoothing factor a large value when the system is not changing dynamically.

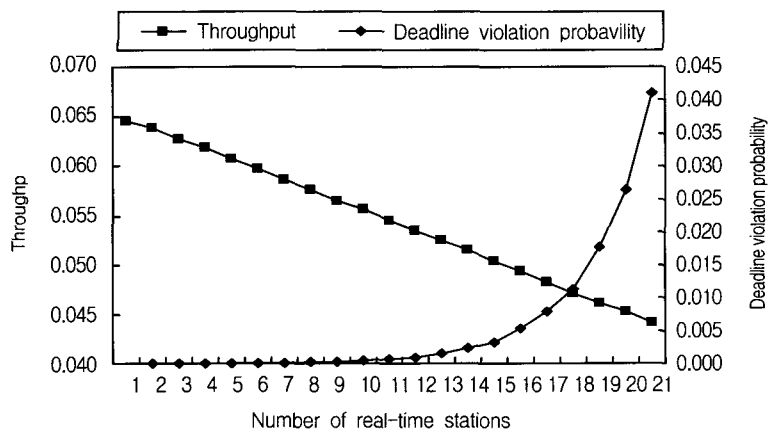
[Figure 4] shows the duration of CFP and CP in 50 superframes with 15 real-time stations and five nonreal-time stations. Each bar in the figure stands for one superframe. The dark and the light part represent the duration of CFP and CP respectively. Though the CFP repetition inter-

val, T_{Rep} is set to 30ms, the duration of a superframe cannot be fixed due to the variation of the CFP. The CFP fluctuates as the number of arriving downlink frames changes in each superframe.

The effect of different number of real-time stations is experimented with 60,000 superframes in [Figure 5]. No admission control is performed to the real-time stations. The number of non-real-time stations is fixed to ten. The throughput of a nonreal-time station decreases as the real-time traffic increases. The increase of the deadline violation probability of real-time frames



[Figure 4] Duration of CFP and CP



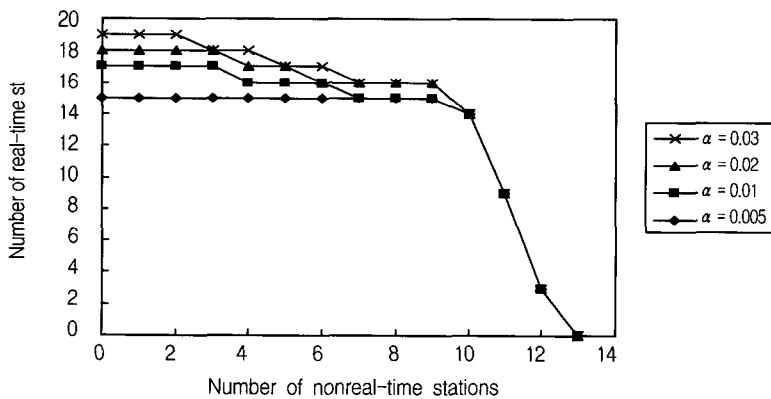
[Figure 5] Performance of real and nonreal time traffics (without admission control)

is dramatic as the number of real-time connections exceeds 15. It illustrates that a proper admission control is necessary to satisfy the traffic requirements and to prevent system collapse.

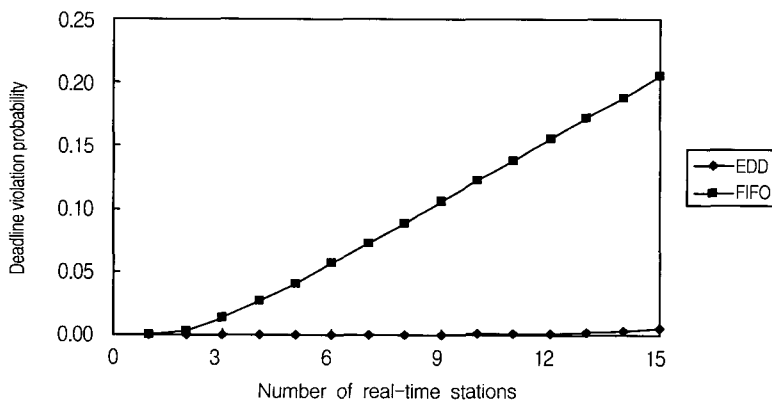
[Figure 6] shows a proper system capacity to guarantee the QoS. The maximum number of the acceptable real-time and nonreal-time stations is experimented for different threshold probability α . The proposed admission control is applied to satisfy the QoS of on-going connections. The arrival rate of real-time connections is given 5 requests/second with average 3 minutes duration of a connection. 120,000

superframes (corresponds to 60 minutes) are simulated. The figure well illustrates the trade-off of real-time and nonreal-time traffics. The maximum number of the acceptable real-time stations decreases as the number of nonreal-time stations increases. The number of the acceptable real-time stations is sharply reduced as the number of nonreal-time stations exceeds ten. It is mainly due to the required minimum throughput by the nonreal-time traffics.

[Figure 7] shows the performance of the proposed real-time scheduling with EDD policy. The percentage of frames that exceeds the service due is dramatically reduced compared to the



[Figure 6] System capacity with admission control



[Figure 7] Comparison of EDD and FIFO

FIFO. It is clear that the EDD with the downlink-first scheduling in a superframe is effective for the real-time traffic in the wireless LAN.

5. Conclusion

Real-time traffic scheduling is discussed that guarantees the delay and throughput in the wireless LAN. The CFP repetition interval is determined such that the interval is shorter than the target delay and the real-time frame generation interval. In the CFP downlink traffic is scheduled before uplink traffic. The downlink frames are scheduled by EDD while the uplink frames are serviced in the order of polling list.

Admission control algorithm is suggested such that it satisfies both the deadline violation probability for the real-time connections and the throughput for the nonreal-time stations. The two measures in each superframe are estimated with the exponential smoothing.

Simulation is performed by applying the proposed scheduling in the superframe. The increase of real-time stations linearly degrades the throughput of nonreal-time stations without admission control. However, the required minimum throughput is satisfied even with the increase of the nonreal-time stations, when the admission control algorithm is applied.

REFERENCE

- [1] A. Visser and E. Zarki, "Voice and data transmission over an 802.11 wireless network," *Wireless : Merging onto the Information Superhighway, Sixth IEEE International Symposium*, Vol.2(1995), pp.648-652.
- [2] J. Feigin and K. Pahlavan, "Measurement of Characteristics of Voice over IP in a Wireless LAN," *1999 IEEE International Workshop on Mobile Multimedia Communications*, pp.236-240, 1999.
- [3] B. P. Crow, I. Widjaja, L.G. Kim and P.T. Sakai, "IEEE 802.11 wireless local area networks," *IEEE Communications Magazine*, Vol.35(September 1997), pp.116-126.
- [4] T. Suzuki and S. Tasaka, "Performance evaluation of integrated video and data transmission with the IEEE 802.11 standard MAC protocol," *GLOBECOM '99*, Vol. 1b(1999), pp.580-586.
- [5] R. Prasad, "Performance comparison of voice over IEEE 802.11 schemes," *VTC 1999*, vol. 5(1999), pp.2636-2640.
- [6] C. Coutras, S. Gupta and N.B. Shroff, "Scheduling of real-time traffic in IEEE 802.11 wireless LANs," *Wireless Networks*, (June 2000), pp.457-466.
- [7] IEEE Standard 802.11-1997 Information Technology-Telecommunications And Information exchange Between Systems-Local And Metropolitan Area Networks-Specific Requirements-part 11 : Wireless LAN Medium Access Control (MAC) And Physical Layer (PHY) Specifications, 18 (November 1997).
- [8] D. Minoli and E. Minoli, "Delivering voice over IP networks," New York : Wiley, 1998.
- [9] M. Shreedhar and G. Varghese, "Efficient Fair Queuing Using Deficit Round Robin," *IEEE/ACM Trans. on Networking*, Vol.4 No.3(1996), pp.375-85.
- [10] R.S. Ranasinghe, L.L.H. Andrew and D. Everitt, "Impact of polling strategy on capacity of 802.11 based wireless multimedia

- LANs," *IEEE International Conference on Networks (ICON '99)*, (1999) pp.96-103.
- [11] C.L. Liu and J.W. Layland, "Scheduling Algorithms for Multiprogramming in a Hard-Real-Time Environment," *Journal of the Association for Computing Machinery*, Vol.20, No.1(January 1973), pp.46-61.
- [12] S. Choi and K.G. Shin, "A Unified Wireless LAN Architecture for Real-time and Non-real-time Communication Services," *IEEE/ACM Transactions on Networking*, Vol.8, Issue 1(Feb. 2000), pp.44-59.
- [13] S. Bodamer, "A new scheduling mechanism to provide relative differentiation for real-time IP traffic," *GLOBECOM '00*, Vol.1 (2000), pp.646-650.
- [14] F. Cali, M. Conti and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Transactions Networking*, Vol. 8, No.6(December 2000), pp.785-799.
- [15] IEEE Standard 802.11b-1999 Information Technology - Telecommunications And Information Exchange Between Systems-Local And Metropolitan Area Networks-Specific Requirements-Part 11 : Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 2000.