무선랜 시스템에서의 하향우선 실시간 음성트래픽 스케쥴링

정동운, 이채영

한국과학기술원 산업공학과

Downlink-first Scheduling of Real-time Voice Traffic in IEEE 802.11 Wireless LANs

Dong W. Jeong and Chae Y. Lee

Department of Industrial Engineering, KAIST, 373-1 Kusung Dong, Taejon, Korea
Tel : +82-42-869-2916
FAX : +82-42-869-3110
Email : {wander, cylee}@mail.kaist.ac.kr

Abstract

The IEEE 802.11 MAC (Media Access Control) Protocol supports two modes of operation, a random access mode for nonreal-time data applications processed by Distributed Coordinated Function (DCF), and a polling mode for real-time applications served by Point Coordinated Function (PCF). It is known that the standard IEEE 802.11 is insufficient to serve real-time traffic. To provide Quality of Service (QoS) of real-time traffic, we propose the Downlink-first scheduling with Earliest Due Date (EDD) in Contention Free Period (CFP) with suitable admission control. The capacity and deadline violation probability of the proposed system is analyzed and compared to the standard pair system of downlink and uplink. Analytical and simulation results show that the proposed scheme is remarkably efficient in view of the deadline violation probability.

***Keywords*:** WLAN, MAC, PCF, Downlink-first scheduling, Deadline violation probability

1. Introduction

Wireless local area networks (LAN) have been growing in popularity, and many products of wireless LAN have been commercially available. With these backgrounds, the IEEE 802.11 committee has developed a wireless LAN standard to satisfy the needs of wireless access. The scope of the standard is MAC (Media Access Control) and physical layers. The first standard allows data rates of up to 2Mbps in the 2.4GHz band. Then, the IEEE 802.11a and IEEE 802.11b committees have developed wireless standards for higher data rates of up to 54Mbps in 5GHz band and 11Mbps in the 2.4GHz band, respectively. Furthermore, the IEEE 802.11e committee is currently working to enhance the 802.11 MAC to expand support for application with QoS requirements (Srinivas Kandala, 2002).

Task group E of the IEEE 802.11 working group are currently working on an extension to the IEEE 802.11 standard called IEEE 802.11e. The goal of this extension is to enhance the access mechanisms that can provide service differentiation. All the details have not yet been finalized, but a new access mechanism called Enhanced DCF (EDCF), which is an extension of the basic DCF mechanism, and Hybrid Coordination Function (HCF) have been selected. Stations, which operate under the 802.11e is called QoS stations (QSTAs). A QoS station, which works as the centralized controller for all other stations within the same Basic Service Unit (BSS), is called the Hybrid Coordinator (HC). The HC will typically reside within an 802.11e access point (AP).

At present, the IEEE 802.11 standard MAC protocol supports two kinds of access methods: DCF and PCF. The DCF is designed for asynchronous data transmission by using CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) and must be implemented in all stations. On the order hand, the PCF is intended for transmission of real-time traffic as well as that of asynchronous data traffic. This access method is optional and is based on polling controlled by the AP. When both DCF and PCF are used, the IEEE 802.11 standard MAC is a hybrid protocol of random access and polling. In this case, a wireless channel is divided into superframe consisting of a CFP for the PCF and CP for the DCF.

The performance of the DCF (H.S. Chhaya and Gupta, 1997) and the combined performance of the DCF and PCF (B.P. Crow, 1997) were evaluated. With regard to the PCF, several traffic scheduling schemes to provide QoS were proposed, including Deficit Round Robin (M. Shreedhar, 1996) and Distributed Deficit Round Robin (R. Ranasinghe, 2001). However, it is hard to satisfy QoS requirement with simple round-robin scheme or fair queuing scheduling algorithm, because real-time traffic generally requests to keep end-to-end delay bound. It is reasonable to assume that real-time traffic connections are established with stations in different BSSs or DS (Distributed System) because the size of BSS is relatively small.

In this paper, we focus on the real-time voice traffic in PCF and propose Downlink-first scheme in which all downlink traffics are processed earlier than the uplink traffics in CFP. The capacity and the deadline violation probability are analyzed using order statistics and simple queuing model. Comparison of the performance of the proposed scheme and that of the standard is discussed. It is shown that the proposed scheme is effective in providing QoS of voice traffic.

2. Point Coordinator Function (PCF) in IEEE 802.11 Standard

The PCF mode provides contention-free frame transfer and the time period in which the LAN is operated in the PCF mode is known as the CFP. The AP performs the function of the point coordinator by gaining control of the medium at the beginning of the CFP after sensing the medium to be idle for PIFS period. During the CFP, CF_Pollable stations are polled by the AP. On receiving the poll the station transmits its data after a Short Interframe Space (SIFS) interval.

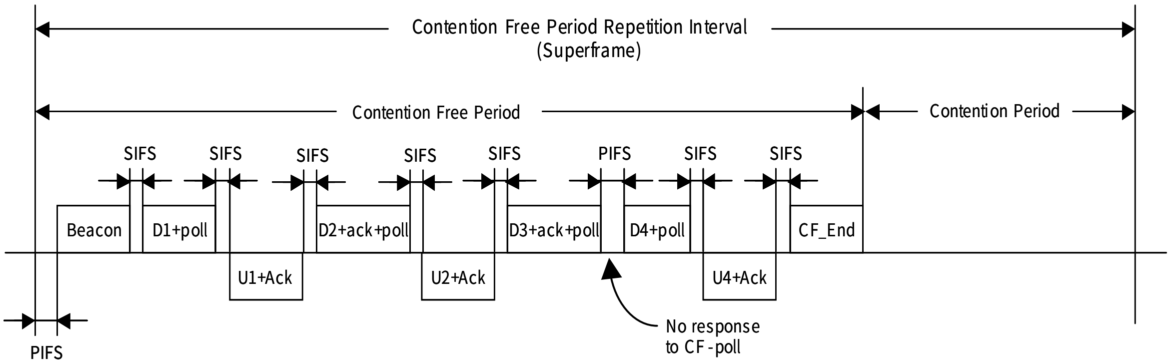

Figure 1. Example of PCF transfer in standard system.

The AP initiates the CFP by transmitting a Beacon frame. If the traffic during the CFP is light and/or the AP has completed polling all the stations on the polling list, it ends the CFP by transmitting a CF_End frame. At the nominal start of the CFP, the point coordinator (PC) senses the medium. If the medium remains idle for a PIFS interval, the PC transmits a beacon frame to initiate the CFP. The PC starts CF transmission at the SIFS interval by sending a CF_Pollable, Data, or Data+CF-Poll frame. If a CF-aware station receives a CF-Poll frame from the PC, the station can respond to the PC after a SIFS idle period, with a CF_ACK or a Data+CF_ACK frame. If the PC receives a Data+CF_ACK frame from a station, the PC 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 ability to combine polling and acknowledgement frames with data frames, transmitted between stations and the PC, was designed to improve efficiency. If the PC transmits a CF_Poll frame and the destination station does not have a data frame to transmit, the station sends Null Function frame back to the PC. **Figure 1** illustrates the transmission of frames between the PC and a station, and vice versa. If the PC fails to receive an ACK for a transmitted data frame, the PC waits a PIFS interval and continues transmitting to the next station in the polling list.

3. Downlink-first Scheduling with EDD and Admission Control

In IEEE 802.11 standard, the real-time traffic is served by PCF and the downlink and uplink transmission is performed as a pair for each connection. In other words, the down/up transmission of a connection is performed after the paired transmission of the previous connection as shown in **Figure 1**. This may cause the serious downlink delay problem of the real-time traffic in the wireless LAN. To improve the delay problem, the downlink-first transmission is proposed. After the beacon frame in the CFP, the downlink transmission of each connection is performed first. Then the uplink traffic is processed as shown in **Figure 2**. The EDD rule is applied to the downlink traffics to improve the delay problem. In the downlink service, the traffic directed to a station is acknowledged from the station by the ACK indication after a SIFS. When all of the downlink traffic that belongs to a polling list is served, then the uplink traffic is served. For uplink traffic, the AP polls a station using Poll or Poll+ACK. Then the polled station may send a data frame to its destination. The uplink data frame to the AP is then acknowledged by the next Poll+ACK frame transmitted after one SIFS interval.

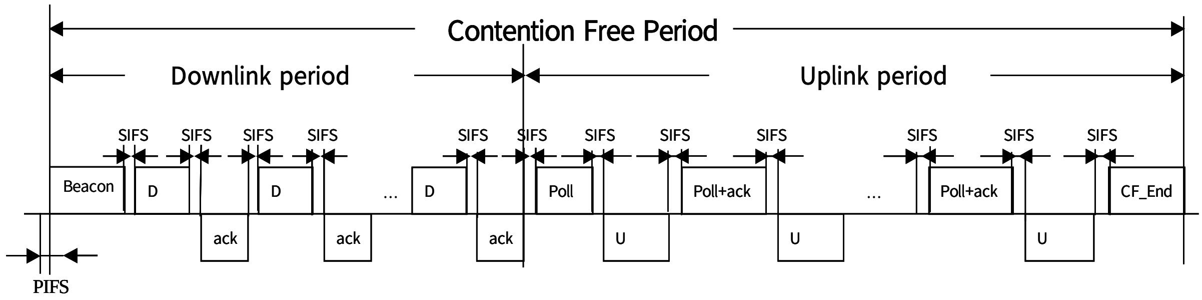

Figure 2. Example of PCF transfer in proposed system.

Now, admission control is necessary to balance the real-time and nonreal-time traffic in wireless LAN. If excess real-time traffic is admitted, the throughput of each nonreal-time station is diminished. Also the transmission delay of real-time traffic is expected. The objective of admission control is to maintain a suitable number of real-time downlink traffic such that the deadline violation probability satisfies a certain limit and to guarantee minimum throughput bound for nonreal-time stations.

In wireless LAN, the PC monitors the state of system continuously and checks the QoS requirement of on-going connections. Thus, the admission control algorithm applied to the coordination function will successfully enhance the QoS of real-time and nonreal-time traffics. For the admission control in the Wireless LAN, we consider the number of real-time stations that can be served in the CFP.

To obtain the maximum number of stations that can be served in a CFP, the following notations are employed with regard to the time intervals given in **Figure 2**.

T_B : transmission time of beacon

T_{CF_END} : CF_END frame

T_{SIFS} : SIFS interval

T_{MPDU} : downlink or uplink real-time traffic frame without piggybacking

T_{ACK} : ACK frame transmission time

T_{Poll} : transmission time of poll frame

T_{PA} : transmission time of poll frame piggybacking ACK

Then by letting $N_{\max}$ be the maximum number of real time traffics during the maximum duration of CFP, T_{CFP_Max} is given by

$$\begin{aligned} T_{CFP\_Max} &= T_B + T_{SIFS} + (2T_{MPDU} + T_{ACK} + T_{Poll} + 4T_{SIFS}) \\ &\quad + (2T_{MPDU} + T_{ACK} + T_{PA} + 4_{SIFS})(N_{\max} - 1) + T_{CF\_END} \end{aligned}$$

The third term in the equation is for the first scheduled downlink and uplink traffic and the forth term is for other traffics. Thus, the max capacity $N_{\max}$ of real-time traffic is obtained as

$$N_{\max} = \left\lfloor \frac{T_{CFP\_Max} - (T_B + T_{SIFS} + T_{CF\_END} - T_{ACK})}{2T_{MPDU} + T_{ACK} + T_{PA} + 4T_{SIFS}} \right\rfloor$$

Since we have the maximum real time capacity in a CFP, the following relationship holds among traffics generated N_G and traffics transferred N_T to the next CFP that exceed the capacity.

$$N_T^{t+1} = \begin{cases} (N_G^t + N_S^t) - N_{\max} & \text{if } N_G^t + N_S^t > N_{\max} \\ 0 & \text{otherwise} \end{cases}$$

N_G^t is the traffics generated during $t-1$ superframe and N_S^t is the traffics transferred from $t-1$ to t whose delay bound is not violated at the start of period *t*. Traffics whose delay bound is violated is discarded.

By giving priority to the transferred traffics that are within the delay bound the admission can be controlled with the traffic generated N_G^t. That is, traffics generated at superframe $t-1$ are accepted as far as they satisfy the following limit.

$$N_G^t \le N_{\max} - N_S^t$$

4. Analysis of System Capacity and Deadline Violation Probability

The proposed Downlink-first with EDD and the standard paired system is compared in terms of capacity and deadline violation.

4.1. Capacity Analysis

To compare two systems, it is assumed that all stations that are active and belong to the polling list have the traffics to transmit and receive. With regard to the standard system given in **Figure 1**, the following notations are additionally employed.

T_{DAP} : transmission time of downlink frame piggybacking ACK and Poll

T_{UA} : transmission time of uplink frame piggybacking ACK

T_{DP} : transmission time of downlink frame piggybacking Poll

Let T_{CFP}^1 and T_{CFP}^2 be respectively the duration of CFP at Downlink-first system and standard system. The difference of T_{CFP} in two systems is due to the transmission time of the uplink and downlink frames. Due to the piggybacked ACK frame and Poll frame, T_{CFP} in standard system is less than that in the Downlink-first system. By letting the number of real-time traffics be N, the two CFPs are given as follows.

$$\begin{aligned} T_{CFP}^1 &= T_B + T_{SIFS} + (2T_{MPDU} + T_{ACK} + T_{poll} + 4T_{SIFS}) \\ &\quad + (2T_{MPDU} + T_{ACK} + T_{PA} + 4T_{SIFS})(N-1) + T_{CF\_END} \end{aligned}$$

$$T_{CFP}^2 = T_B + T_{SIFS} + (T_{DP} + T_{UA} + 2T_{SIFS}) + (T_{DAP} + T_{UA} + 2T_{SIFS})(N-1) + T_{CF\_END}$$

By assuming $T_{DP} = T_{MPDU} + T_{poll}$, $T_{UA} = T_{MPDU} + T_{ACK}$, and $T_{DAP} = T_{MPDU} + T_{PA}$, we have $T_{CFP}^1 - T_{CFP}^2 = 2NT_{SIFS}$.

Note that the duration to process a connection that consists of uplink and downlink traffic is $2T_{MPDU} + T_{ACK} + T_{PA} + 2T_{SIFS}$ in the standard system. Thus, when the saved time $2NT_{SIFS}$ exceeds the duration, one more connection can be served in the standard system. Considering the maximum data rate of 11Mbps of the IEEE 802.11b standard and 300-octet voice frame, we have $T_{MPDU} = 218\mu\sec$. By applying $T_{SIFS} = 10\mu\sec$, one more connection can be processed when the number of real time connections $N \ge 25$. Now, from $T_{CFP\_Max} = 28ms$, the maximum number of real time connections in the CFP is computed as 55 in the standard and as 53 in the Downlink-first system respectively. Therefore, it can be concluded that even if the standard system has piggybacking efficiency of polling message, the capacity difference for the real time connection is negligible.

4.2. Analysis of the Deadline Violation Probability

For the analysis of the deadline violation probability the downlink traffics are generated following the Poisson process with the arrival rate λ. Each frame generated is assumed to have uniformly distributed remaining due over $[Due_{\min}, Due_{\max}]$ at the point coordinator. Note that a frame with its remaining due less than the length of a superframe may probably be discarded at the PC. We thus assume that the $Due_{\min}$ is equal to the length of a superframe.

Let X_i, $i = 1, 2, \ldots, N_G$ be the random variable of remaining due of the i-th generated frame to be scheduled at the PC. Note that X_i is i.i.d. uniform random variable over $[Due_{\min}, Due_{\max}]$. Since the traffics are scheduled by EDD, let $X_{(j)}$, $j = 1, 2, \ldots, N_G$ be the j th smallest remaining due of the $X_1, X_2, \ldots, X_G$. That is, $X_{(1)}, X_{(2)}, \ldots, X_{(N_G)}$ are the order statistics corresponding to $X_1, X_2, \ldots, X_{N_G}$. Then the density function of $X_{(j)}$ is given by

$$f_{(j)}(x_{(j)}) = \frac{N_G!}{(j-1)!(N_G - j)!}[F(x_{(j)})]^{(j-1)}[1 - F(x_{(j)})]^{(N_G - j)} f(x_{(j)})$$

Since we assume the $Due_{\min}$ is equal to a superframe length, the deadline of a real-time traffic may be violated when the traffic is transferred to the next CFP. Let k, $k = N_G - N_T + 1, \ldots, N_G$ be the index of transferred traffic, then $X_{(k)}$ is the random variable of the remaining due of the transferred traffic. Accordingly, the traffic transferred from previous superframe has the remaining due, $Y_{(l)}$ given by

$$Y_{(l)} = X_{(k)} - Due_{\min}, \quad l = k - N_G + N_T .$$

Thus, the deadline violation probability of the l-th traffic is represented as $P_{(l)}(T_{(l)} > Y_{(l)})$, where $T_{(l)}$ is the scheduled time of traffic l at the transferred superframe. When the deadline of a frame is violated, the traffic is discarded, and the following traffics are served. Thus, the scheduled time $T_{(l)}$ may be different from the initial schedule $t_{(l)}$ which is the scheduled time for the l-th traffic before the frame is discarded due to deadline violation. Therefore, $P_{(l)}(T_{(l)} > Y_{(l)})$ is represented as the conditional probability. As an example, consider two transferred traffics to be served, $T_{(l)}$ is given by

$$T_{(1)} = t_{(1)}$$

$$T_{(2)} = \begin{cases} t_{(1)} & : \text{if the first frame is discarded} \\ t_{(2)} & : \text{else} \end{cases}$$

The deadline violation probability $P_{(l)}(T_{(l)} > Y_{(l)})$ is given by

$$P_{(1)}(T_{(1)} > Y_{(1)}) = P(t_{(1)} > Y_{(1)})$$

$$\begin{aligned} P_{(2)}(T_{(2)} > Y_{(2)}) &= P(t_{(1)} > Y_{(2)} \mid t_{(1)} > Y_{(1)})P(t_{(1)} > Y_{(1)}) + P(t_{(2)} > Y_{(2)} \mid t_{(1)} < Y_{(1)})P(t_{(1)} < Y_{(1)}) \\ &= P(t_{(1)} > Y_{(2)},\ t_{(1)} > Y_{(1)}) + P(t_{(2)} > Y_{(2)},\ t_{(1)} < Y_{(1)}) \end{aligned}$$

By assuming each real-time traffic has the same frame length, the initial schedule $t_{(l)}$ by the EDD rule is determined as follows.

Initial schedule $t_{(l)}$ in the standard system

$$t_{(l)} = \begin{cases} T_B + T_{SIFS}, \text{ if } \ l = 1 \\ t_{(1)} + (T_{DP} + T_{UA} + 2T_{SIFS}), \ \text{ if } \ l = 2 \\ t_{(2)} + (T_{DAP} + T_{UA} + 2T_{SIFS})(l-2), \ \text{ if } \ l \geq 3 \end{cases}$$

Initial schedule $t_{(l)}$ in the Downlink-first system

$$t_{(l)} = \begin{cases} T_B + T_{SIFS}, \text{ if } \ l = 1 \\ t_{(1)} + (T_{MPDU} + T_{ACK} + 2T_{SIFS})(l-1), \text{ if } \ l \geq 2 \end{cases}$$

To obtain the deadline violation probability the following probability needs to be computed.

$$P_{(l)}(t_{(l)} > Y_{(l)}) \ = \ P_{(k)}(t_{(l)} > X_{(k)} - Due_{\min})$$

$$= \int_0^{t_{(l)}+Due_{\min}} f_{(k)}(x_{(k)})dx_{(k)} \ = \int_0^{t_{(l)}+Due_{\min}} \frac{N_G!}{(k-1)!(N_G-k)!} F(x_{(k)})^{(k-1)}(1-F(x_{(k)}))^{(N_G-k)} f(x_{(k)})dx_{(k)}$$

Now, from the deadline violation probability the expected number of discarded frame in a superframe can be obtained by $\sum_{k=N_G-N_T+1}^{N_G} P_{(k)}(T_{(k)} > Y_{(k)})$. **Figure 3** shows the expected number of discarded frames for each pair of (N_G , N_T) by the standard system and the proposed Downlink-first system. For fair comparison the EDD rule is also applied to the standard system. From the figure it is clear that more frames are discarded as the number of transferred traffic N_T increases. Better performance by the proposed Downlink-first is illustrated compared to the standard.

5. Simulation results of the Real-time Traffic Scheduling

The system parameters for simulation are reported in **Table** 1 as specified in the IEEE 802.11b standard. To simplify the simulation the propagation delay, transmission errors are not considered.

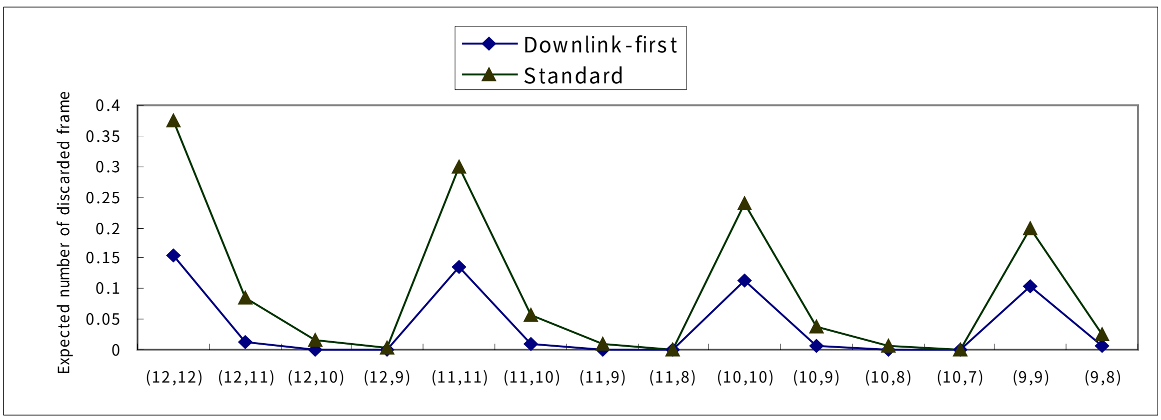

Figure 3. Expected number of discarded frames due to deadline violation

Table 1. Default attribute value from IEEE 802.11b standard

Attribute	Symbol	Value
Channel rate	CR	11Mbps
ACK frame size	$T_{ACK} \times CR$	14 octets
CF-End frame size	$T_{CF\_End} \times CR$	20 octets
Poll frame size	$T_{Poll} \times CR$	20 octets
Slot Time	T_{ST}	20 μs
SIFS Time	T_{SIFS}	10 μs
PIFS Time	T_{PIFS}	30 μs
DIFS Time	T_{DIFS}	50 μs

The superframe length is assumed to be 30ms with $Due_{\max} = 40ms$ and $Due_{\min} = 30ms$. Main characteristics of the real-time traffic are taken from G.723.1 protocol (D. Minoli et al, 1998). At each station real-time frames are generated by following the Poisson process with the arrival rate $\lambda = 0.6 \sim 1.0/30m\sec$.

Figure 4 shows the expected number of discarded frames in a superframe. The number of active real-time stations are given by N = 18, 19, 20 with the system capacity $N_{\max} = 15$. 120,000 superframes that corresponds to 60 minutes are simulated both for the standard and the proposed Downlink-first systems. The EDD rule is also applied to the two systems. The increase of the number of discarded frames is far degraded by the proposed method that assigns the downlink traffic in front of the uplink in a superframe. The figure also shows that the

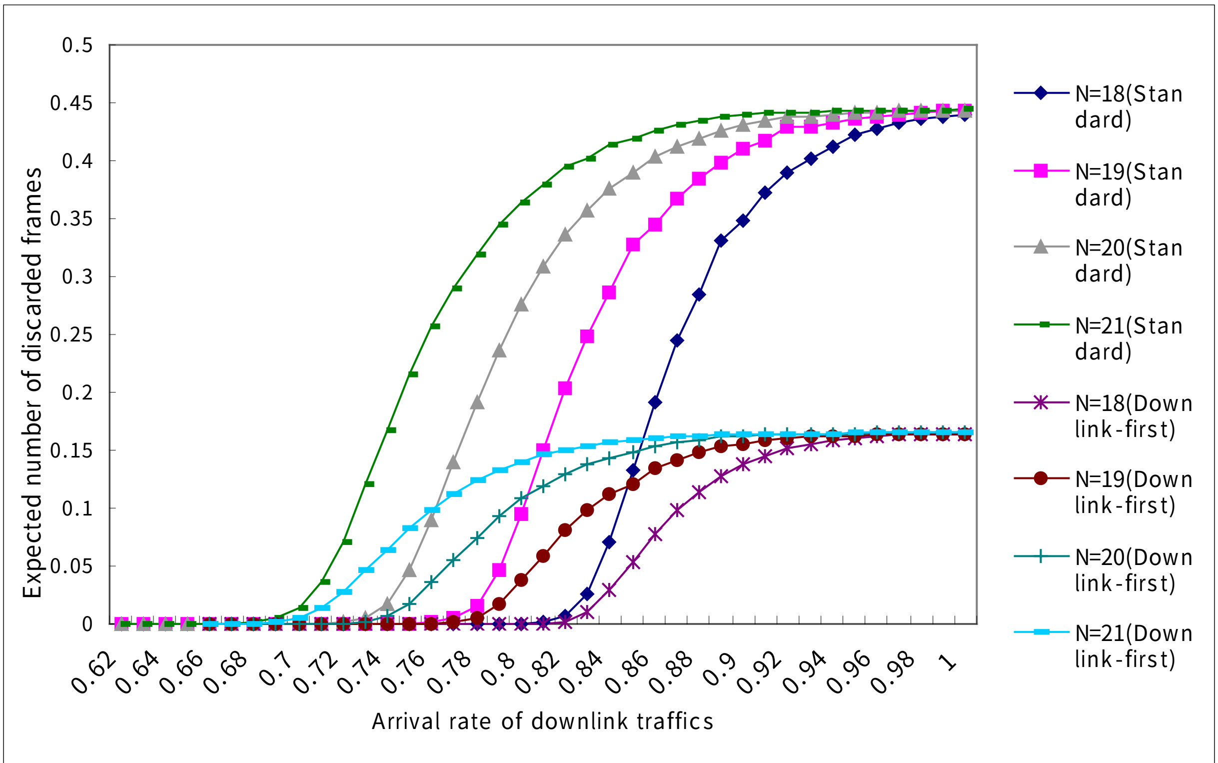

Figure 4. Performance of the Downlink-first vs. Standard

proposed admission strategy effectively controls new connections. The expected number of discarded frames by the total stations converges to a limit even if the traffic arrival rate is increased. Per station frame discarded rate is less than 1% with the proposed system.

The blocking probabilities in the two systems are compared in **Figure 5**. The figure shows the performance with $N = 20$ real-time stations when system capacity is fixed to $N_{\max} = 15$. The blocking probability is obtained by checking the number of traffics blocked by admission control among 120,000 superframes.

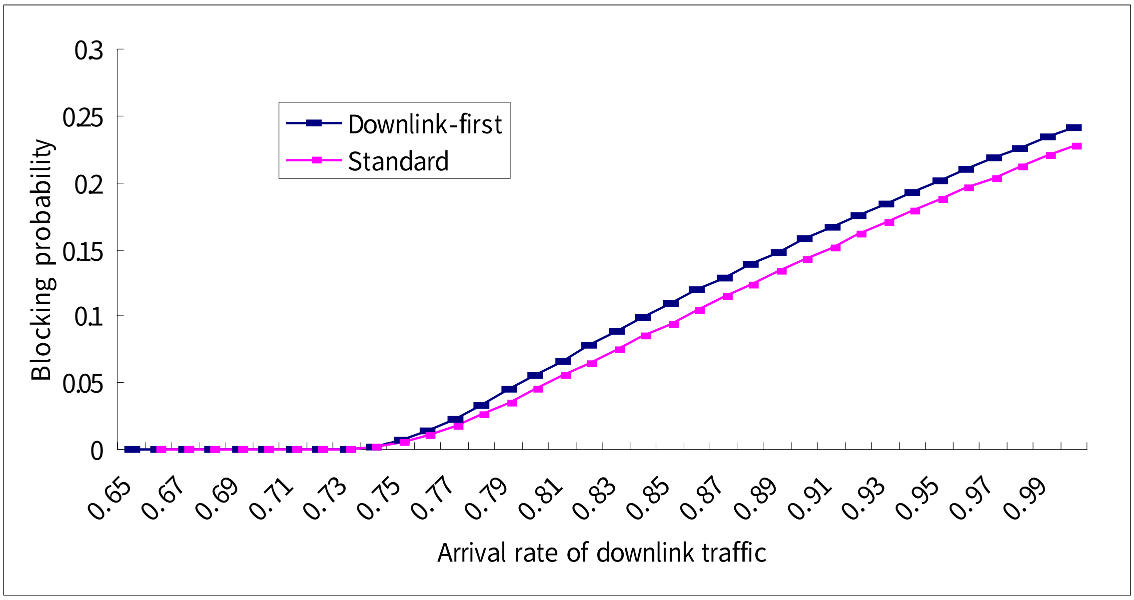

Figure 5. Blocking probabilities

From the figure it is clear that no traffics are blocked when the generated traffic λN is less than the system capacity $N_{\max}$. As the traffic exceeds the capacity part of it is blocked by the admission control. Higher blocking probability by the Downlink-first well explains the reduced number of discarded frames as shown is **Figure 4**.

6. Conclusion

A Downlink-first scheduling is proposed to reduce the delay of the real-time traffic in the WLAN. The uplink traffics are scheduled after the downlink in order of polling list. Admission control algorithm is also suggested such that it satisfies both the deadline violation probability for the real-time connections and the throughput for the nonreal-time stations. The acceptable number of downlink real-time traffic is controlled by the number of frames transferred from the previous superframe and the maximum number of frames that can be processed at a superframe.

The proposed Downlink-first with EDD is compared to the standard system by analyzing the system capacity and the deadline violation probability. The analysis proves that the proposed Downlink-first with EDD outperforms the standard. The number of discarded frames that violate the deadline is dramatically reduced compared to the standard system where the uplink and downlink transmission is paired for each connection. The same result is obtained with the simulation. Due to the admission control the frame discard rate converges to a threshold less than 1% even with the increased downlink traffics.

References

B. P. Crow, I. Widjaja, L. G. Kim and P. T. Sakai, (1997), IEEE 802.11 wireless local area networks, *IEEE Communications Magazine*, .**35**, 116-126.

D. Minoli and E. Minoli, (1998), Delivering voice over IP networks, *New York: Wiley*

F. Cali, M. Conti. and E. Gregori, (2000), Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit, *IEEE/ACM trans. Networking*, **8**(6), 785-799.

H. S. Chhaya and S. Gupta, (1997), Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol, *Wireless Networks*,.**3**, 217-234.

Ju-Hee Lee and Chae Y. Lee, (2002), Scheduling of Real-time and Nonreal-time Traffics in IEEE 802.11 Wireless LAN, *submitted to the Journal of the Korean OR/MS society.*

M. Shreedhar and G. Varghese, (1996), Efficient Fair Queuing Using Deficit Round Robin, *IEEE/ACM Trans. On Networking* , **4**(3), 376-85.

R. Ranasinghe, (2001), Distributed contention-free traffic scheduling in IEEE 802.11 multimedia networks, *Selected Papers, 10th IEEE Workshop on Local and Metropolitan Area Networks*, 18-28.

S. Bodamer, (2002), A new scheduling mechanism to provide relative differentiation for real-time IP traffic, *GlOBECOM'00*, **1,** 646-650.

Srinivas Kandala, Partho Mishra, FrankHowley and Rolf Devegt, Matthew Sherman, Bob Meier and Menzo Wentik, (2002), *Normative Text for Tge Consensus Proposal, IEEE 802.11-02/604r0*.