Traffic Control for Time-bounded Services over IEEE 802.11 Wireless LAN

Po-Ning Chen, and Wen-Shuo Hsu

Abstract

In this paper, we study the problem of conveying two-way conversational traffics over a guaranteed quality-of-service (QoS) network access to an IEEE 802.11 wireless local area network (LAN) with point coordination function (PCF) capability. The (σ, ρ, d) parameter-guaranteed model is assumed at the guaranteed QoS network side, where σ, ρ, and d are the burst parameter, the rate parameter and the delay bound, respectively. Based on the QoS parameters, the admission control criteria for the standard-suggested round-robin polling scheme (RR-Polling) and the proposed earliest-deadline-first polling scheme with token generators (EDF/tg-Polling) are derived under the premise that the QoS of the down-link traffic (from the access point to the wireless station) is strictly guaranteed. As a result, the schedulable region of the EDF/tg-Polling scheme is almost twice of the RR-Polling schedulable region. For those admitted (i.e., QoS-guaranteed) down-link traffics, we also investigate the average delay of the respective up-link counterpart in a two-way conversation. Note that the up-link connections within the domain of IEEE 802.11 provide no QoS parameter-exchange mechanism, and hence, cannot be hard-guaranteed. Simulation results show that by employing a constant token generating rate to the EDF/tg-Polling scheme, a significant reduction in the average up-link delay than the RR-Polling scheme can be obtained. Pro-and-cons for different system settings, such as the maximum length of contention-free period (CFP) and the size of the data frames, are also addressed.

Keywords

Guaranteed QoS, wireless LAN, admission control, scheduling, point coordination function.

I. INTRODUCTION

The scope of the IEEE 802.11 standard [1] is the physical layer (PHY) and medium access control (MAC) layer implementation in a limited geographical area. The MAC scheme is designed to support three different PHY implementations, viz., baseband infrared transmission, radio transmission employing frequency hopping spread spectrum (FHSS), and radio transmission employing direct sequence spread spectrum (DSSS).

The work is supported by the National Science Council under Grant NSC 88-2213-E-009-089-.

P.-N. Chen is with the Department of Communications Engineering, National Chiao Tung University, Taiwan, R. O. C. E-mail: poning@cc.nctu.edu.tw.

W.-S. Hsu received his Master degree under the advisory of P.-N. Chen, and is currently serving his compulsory two-year military service.

The fundamental access method of the IEEE 802.11 MAC is a distributed coordination function (DCF) known as carrier sense multiple access with collision avoidance (CSMA/CA). To adapt to the growing demands for real-time services, the IEEE 802.11 MAC also provides an optional point coordination function (PCF) through the DCF services, as depicted in Figure 1. Enforced by the PCF, the standard is expected to support time-bounded transfer of communication data.

The PCF, as specified in IEEE 802.11, can only be operated in a network with at least one fixed access point (AP), or specifically, an infrastructure network (cf. Figure 4). In functionality, a PCF-enabled AP shall poll the PCF-enabled wireless stations in order by ascending association identifier (AID) values [1, Sec. 9.3.4.1]. Since the AID of a wireless station is a fixed number assigned by the AP during association, it turns out that a round-robin polling scheme is rendered, although the AP is allowed to issue polls only to a subset of the stations in the polling list.

We, however, observe that the round-robin-type polling scheme may not be efficient in bandwidth consumption, especially when the targeted time-bounded services are two-way real-time conversations. In order to meet the QoS demands of telephony-type services, a more efficient scheduling scheme may be necessary.

In this paper, we consider the scenario of conveying two-way conversational traffics over guaranteed QoS network access to a PCF-enabled IEEE 802.11 wireless LAN, as depicted in Figure 2. The guaranteed QoS network, which can be either wired or wireless, is a network with QoS-parameter-exchange mechanism. The QoS parameters exchanged prior to each communication include a burst parameter σ, a rate parameter ρ, and a delay request d. These parameters will be provided by the guaranteed QoS network to the access point (AP), where admission decision to this communication link is made. Two AP traffic schedulers will be studied: the standard-suggested round-robin polling scheme (RR-Polling), and our earliest-deadline-first polling scheme with token generators (EDF/tg-Polling). The admission decision is therefore made such that the adopted scheduler can strictly meet the delay demand d of the down-link traffic (from the AP to the wireless station) based on the given burst parameter σ and rate parameter ρ. The traffic patterns are assumed to fit the telephony-type two-way conversational models, i.e., at any time instance, exactly one of the two parties (caller and callee) generates voice data. This excludes the unlikely situations (for a normal conversation) that both parties talk or both keep silence at the same time.

Since only the down-link part of a conversation is delay-guaranteed, the system performances are therefore determined by the amount of conversations admitted and the average up-link delay experienced. We will investigate these two performance indices for the RR-Polling and the EDF/tg-Polling schedulers, respectively. Performance impact by employing different data frame size and contention-free period will also be studied.

The rest of the paper is organized as follows. In Section II, we give a brief overview of the PCF in IEEE 802.11 MAC specification. Section III presents the EDF/tg-Polling scheduler and its admission control criterion. Also covered in this section is the admission control criterion of the RR-Polling scheduler. Simulation results for their schedulable regions and average up-link delays are provided in Section IV. The system configuration and the source traffic model, which our simulations are based on, are also described in this section. Conclusion and future work appear in Section V.

II. Overview of point coordination function

The MAC sublayer of IEEE 802.11 consists of a mandatory contention-based distributed coordination function (DCF) and an optional contention-free point coordination function (PCF) (cf. Figure 1). Medium usage is alternated between contention period (CP) and contention-free period (CFP) as shown in Figure 3.

Three types of MAC frames are specified in the standard: *Control frames*, *Management frames* and *Data frames*. Control frames are basically used for handshaking and positive acknowledgements. It can also be used to forcefully end a CFP. Management frames are designed for association and disassociation with the AP, timing and synchronization, authentication and deauthentication. Data frames are used to transmit the MAC service data units (MSDUs) generated by the upper layers, and can be combined with polling and acknowledgements during the CFP. In an infrastructure network, a wireless station has to associate with an AP prior to the transmission of its data frames. After association, the wireless station can communicate with the other wireless stations under either DCF or PCF control.

The PCF performed by the AP in an infrastructure network provides contention-free services. During the CFP, the medium is fully controlled by the AP. Stations that are capable of responding to a contention free poll (*CF-Poll* control frame) are known as *CF-pollable* stations. The standard also allows a wireless station to associate as *receive-only* during the CFP. The AP shall maintain a polling list of the *CF-Pollable* stations for use in selecting stations that are eligible to receive *CF-Poll* control frames during the CFP.

No polling-list maintenance techniques are specified in the standard. The AP is only required to issue (at least one) polls to (a subset of) the stations in the polling list in order by ascending association identifier (AID) values during each CFP. Since the station AID is a fixed number assigned by the AP during association, a straightforwardly resultant polling-list maintenance technique from this requirement is the round-robin polling scheme, as graphically illustrated in Figure 4.

Two important system parameters regarding the PCF are the *CFPMaxDuration* and *CFPRepetitionInterval* (cf. Figure 3). The former indicates the maximum duration of a CFP; an AP, however, can issue a *CF-END* control frame to end a CFP prior to the expiration of its *CFPMaxDuration.* The latter that is formally named as *Contention-Free Repetition Rate* (CFPRate) in the standard represents the appearance period of the CFP.

At the scheduled beginning of each CFP repetition interval, the AP will gain the channel access by transmitting a *Beacon* management frame, after sensing the channel idle. The *Beacon* management frame contains the information of the *CFPMaxDuration* as well as the *CFPRepetitionInterval.* All other wireless stations will then update their network allocation vectors (NAV) to the *CFPMaxDuration.* An NAV is a time indicator during which a station cannot initiate its transmission, except a poll is received from its associated AP. To enhance the efficiency of the CFP usage, a data frame may "piggyback" the acknowledgement of the previously received data frame using particular data frame subtypes for this transmission. The piggybacked frames that are possibly transmitted from the AP are summarized below.

(a) *Data* frame: used to send data to a station when there is no previous data frame to acknowledge;

(b) *Data+CF-ACK* frame: used to send data to a station simultaneously with an acknowledgement to the previously received data frame from a CF-Pollable station;

(c) *Data+CF-Poll* frame: used to send data to a station that is the next station to be permitted to transmit, and there is no previous data frame to acknowledge;

(d) *Data+CF-ACK+CF-Poll* frame: used to send data to a station that is the next station to be permitted to transmit, and simultaneously acknowledge the previously received data frame from a CF-Pollable station;

(e) *CF-Poll* frame: used when there is no data for a station that is the next station to be permitted to transmit, and there is no previous frame to acknowledge;

(f) *CF-Poll+CF-ACK* frame: used when there is no data for a station that is the next station to be permitted to transmit, and when the AP needs to acknowledge the previously received data frame from a CF-Pollable station;

(g) *CF-ACK* frame: used when the AP has neither data to transmit nor stations to poll, but the AP needs to acknowledge the previously received data frame from a CF-Pollable station;

An example of the usage of the above piggybacked frames during the CFP is graphically demonstrated in Figure 5.

III. Earliest-deadline-first polling scheduler with token generators

The design of a service discipline with bounded delay guarantee for all packets has been the subject of many investigations [2][3][6][8][9][10][14]. Evaluation and improvement in performances of the IEEE 802.11 wireless LAN is also of great practical interest [4][11][13]. In this section, we will focus on the admission control criteria for two variations of the work-conserving service disciplines: round-robin and earliest-deadline-first schemes. They are modified to adapt to the characteristics of the IEEE 802.11 wireless LAN. We begin with the introduction of the source models.

A. Guaranteed QoS source models

Let n be the number of connections between the AP and the remote network terminals resided in the guaranteed QoS network (cf. Figure 2). The traffic arrivals from the remote terminal to the AP through connection j during the time interval $[t, t+\tau)$ is denoted by

$A_j[t, t+\tau]$. A frequently applied constraint on function $A_j[\cdot,\cdot]$ is:

$$A_j[t, t+\tau] \le A_j^*(\tau), \tag{1}$$

where $A_j^*(\tau)$ is a non-negative traffic constraint function, which is zero for $\tau < 0$. An upper constraint traffic satisfying (1) can be obtained by employing either a *traffic policer* that rejects traffic if it does not comply to $A_j^*(\tau)$ or a *rate controller* that buffers packets temporarily to ensure that the traffic conforms to the constraint. Usually, the traffic constraint functions are derived from deterministic traffic models, which characterize the worst-case traffic by a small set of parameters. One well-known example of these kinds of constraint functions is the (σ, ρ)-traffic model that describes the worst-case traffic by a burst parameter σ and a rate parameter ρ [5], i.e., $A_j^*(\tau) = \sigma + \rho\tau$. It has been shown that a traffic conformed to (σ, ρ)-constraint can be established by a leaky-bucket policing mechanism [12].

B. Specific characteristics of IEEE 802.11 wireless LAN environments

Before the introduction of the EDF/tg-Polling scheme, three specific and essential physical characteristics of the IEEE 802.11 wireless LAN environment are first noted.

1. Shared-medium characteristic: In the IEEE 802.11 wireless environment, all the wireless stations in the same Basic Service Set (BSS) share the transmission medium. Furthermore, the shared medium can only carry either up-link or down-link traffic at a time for one wireless station. In order to hard-guarantee the QoS demands of the down-link traffic, the worst-case overhead of the up-link traffic must be taken into consideration in the admission control criterion.

2. No signaling process for the up-link traffics: In the IEEE 802.11 MAC specification, there is no signaling process for the exchange of the traffic characterization, or specifically, the QoS parameters. Although it could be enhanced by the upper-layer or other protocols, we here assume that the AP is only armed with the IEEE 802.11. Hence, the AP has no knowledge about the up-link traffic characteristics. Under the assumption, we will focus on "to provide a small average delay for the guarantee-free up-link traffic subject to a delay-bound guarantee for the down-link connection according to its QoS (obtained from the guaranteed QoS network side)."

3. Contention-free and contention period alternation: Due to the necessity to support both the synchronous and asynchronous data transmissions, the IEEE 802.11 MAC specifies the time alternation of the CFP and the CP. As a consequence, the system service curve during the CFP is no longer a straight line as considered in the conventional service disciplines. A piece-wise linear service curve should be used instead, as shown in Figure 6.

C. Earliest-deadline-first polling scheduler with token generators (EDF/tg-Polling)

The EDF/tg-Polling scheduler consists of m queues respectively for the m associated PCF-enabled wireless stations, as depicted in Figure 7. The data unit in these queues is assumed to be a fixed multiple α of the data cells originated from the remote terminals resided in the guaranteed QoS network. In our simulations, we assume that the content of the data cell is of the same size as an ATM cell, i.e., 48-byte in length. An IEEE 802.11 data frame for PCF services therefore contains $(58 + 48\alpha)$ bytes, where the 58-byte corresponds to the length of the IEEE 802.11 frame header and CRC32 for DSSS PHY. Note that the $(58 + 48\alpha)$-byte frame-size constraint only applies to PCF services. The system can still use longer (variable-length) data frames, up to 2346 bytes as suggested in IEEE 802.11, for DCF services.

The deadline of each data frame is marked as the sum of the arrival time[1] and

(1) either the delay request, if the data frame is from the guaranteed QoS network side; or

(2) the reciprocal[2] of the token generating rate, if no QoS is specified from the source of this data.

An earliest-deadline-first (EDF) sequencer will then pick up the head-of-line (HOL), among all queues, with the earliest deadline for transmission.

The description of the above scheduling discipline only concerns the transmission of the down-link traffics. To deal with the up-link traffics from wireless stations, we associate each queue with an independent polling-token generator. Whenever a token is dropped from the generator, the subtype of the first encountered *Data* frame in the respective queue

---

[1] The arrival time of a data frame that is formed by gathering α data cells is the arrival time of the first data cell.

[2] In case the token generating rate is zero, the deadline becomes "infinity," which implies that the data will be transmitted only at the absence of other HOLs with finite deadline.

will be modified to piggyback the poll command, which renders a *Data+CF-Poll* frame. Note that the length of the *Data* frame remains unchanged, since only the 4-bit subtype field in the header is modified. If, however, the queue is empty or contains only *Data+CF-Poll* frames at the time the token is dropped, a *CF-Poll* control frame will be added into the queue, and its deadline will be marked as the token generating time plus the reciprocal of the token generating rate. When the *CF-Poll* frame becomes the HOL with the earliest deadline, it will be transmitted, and an additional overhead will be introduced into the medium usage during transmission, since no data is carried in a control frame.

In the above description, we omitted the acknowledgement property since it is not relevant to the scheduling discipline introduced, and the inclusion of it may subtract the clarity of this presentation. Readers may presume that the *CF-ACK* subtype (for the acknowledgement of the previously received data frame) will be attached automatically, whenever necessary.

The objective of introducing the polling-token generator is to reserve the shared CFP bandwidth for the up-link traffics. As anticipated, each token will induce an up-link data frame, if there are traffics queued in the addressed recipient. The token generating rate can be adjusted according to the need of each up-link connection. Since no QoS is pre-specified for the up-link traffics, a reasonable specification on the token-generating rate is to fairly share the remaining CFP bandwidth among all up-link connections, i.e.,

$$r_i = \frac{1}{m_p}\left(C - \sum_{j=1}^{n} \rho_j\right),$$

if wireless station i is associated as CF-Pollable during the CFP, where C is the system service rate during the CFP, ρ_j is the rate parameter for connection j (of the guaranteed QoS network), n is the number of the guaranteed QoS connections, and m_p is the number of the associated CF-Pollable stations. The token-generating rate for a receive-only wireless station is apparently zero.

The main drawback of EDF schemes is perhaps the computational complexity in selecting the HOL with the earliest deadline, and yet in a wireless LAN environment, this complexity is reduced to an acceptable level, since the number of the associated wireless stations is usually small. Experiments show that an AP armed with 80386 CPU can easily handle such a computational complexity, if the number of the associated wireless stations are around twenty.

We close the subsection by pointing out four system assumptions made in our design and also the follow-up simulations.

(1) The AP only supports two-way call services during the PCF. In other words, the traffic from one of the admitted connections of the guaranteed QoS network will always destine to the same wireless station, and thereby, will be placed in the same queue. Meanwhile, no two source traffics (including those from the guaranteed QoS network, and those from the up-link connection of the wireless LAN) will be dispatched to the same queue.

(2) A wireless station requests to become CF-Pollable by either association or reassociation, only when it decides to take a call or initiate a call. After the call is ended, the station will dictate to be removed from the CF-Pollable list in AP (by reassociation) so that more PCF bandwidth can be saved for other stations.

(3) After the AP issues a *CF-Poll* control frame or a *Data+CF-Poll* frame, it will wait until the reception of the returned *Data* frame or *CF-ACK* control frame. Upon the reception of an up-link *Data* frame, the AP will immediately forward its content to the targeted destination.

(4) We did not consider the possible re-transmissions due to the unreliability of the wireless medium. Note that a hard-guaranteed down-link traffic is possible only when reliable transmissions are ensured after a bounded number of re-transmissions.

D. Admission control criteria for EDF/tg-Polling scheduler

The EDF/tg-Polling scheduler that we proposed is nonpreemptive, i.e., the scheduling decision will not be taken during the frame transmission on air. In our view, a nonpreemptive scheduling policy is more feasible for wireless medium. It has been shown in [7] that a nonpreemptive EDF scheduler is optimal for bounded delay services, since its schedulable region is the largest among all scheduling algorithms. The condition under which the traffics are nonpreemptive EDF-schedulable is also derived in the same paper based on (σ, ρ) traffic characterization, which is quoted (with necessary change in notations) below.

Theorem 1: Characterize the traffic from connection j by (σ_j, ρ_j, d_j), where d_j is the delay requirement, and the amount of traffic arrivals satisfies that $A_j[t, t+\tau] \leq \sigma_j + \rho_j \tau$ for $\tau \geq 0$. Assume without loss of generality that $d_i \leq d_j$ for $i < j$. Then $\{(\sigma_j, \rho_j, d_j)\}_{1 \leq j \leq n}$ is nonpreemptive EDF-schedulable, if, and only if,

$$t \geq \sum_{i=1}^{j} [\sigma_i + \rho_i(t - d_i)] + L_{data} \quad \text{for } d_j \leq t < d_{j+1} \text{ and } 1 \leq j < n; \tag{2}$$

and

$$t \geq \sum_{i=1}^{n} [\sigma_i + \rho_i(t - d_i)] \quad \text{for } t \geq d_n, \tag{3}$$

where L_{data} denotes the transmission time of a data frame, which is fixed in our design[3].

An implicit assumption for the validity of (3) is $\sum_{i=1}^{n} \rho_i \leq 1$, which can be justified by a re-formulation of (3) to

$$t \geq \left(\sum_{i=1}^{n} \rho_i \right) t + \sum_{i=1}^{n} [\sigma_i - \rho_i d_i], \tag{4}$$

and the observation that (4) should hold for all $t > d_n$. Accordingly, $\sum_{i=1}^{n} \rho_i < 1$ for every $1 \leq j < n$. By respectively drawing the curves of the left-hand-side and the right-hand-side of (2) and (3) (cf. Figure 8), we observe that the EDF schedulability condition in Theorem 1 holds for all t if, and only if, it is valid at $t = d_j$ for every $1 \leq j \leq n$. By taking these t values into (2) and (3), we conclude that a set of n connections, defined by $\{(\sigma_j, \rho_j, d_j)\}_{1 \leq j \leq n}$, is nonpreemptive EDF-schedulable, if, and only if,

$$d_j \geq \sum_{i=1}^{j} [\sigma_i + \rho_i(d_j - d_i)] + L_{data} \quad \text{for } 1 \leq j < n, \text{ and } d_n \geq \sum_{i=1}^{n} [\sigma_i + \rho_i(d_n - d_i)]. \tag{5}$$

Condition (5) can then serve as an admission control criterion for the nonpreemptive EDF scheduler under (σ, ρ, d) source models.

Based on Theorem 1, we next derive an admission control criterion for EDF/tg-Polling scheduler operating over the IEEE 802.11 wireless LAN.

We first note that the right-hand-sides of (2) and (3) need to be modified by considering the latency due to the responded up-link data frame for each poll command (or specifically,

---

[3] The parameters—σ_i, ρ_i, and L_{data}—have been normalized by the system service rate C in the theorem. For example, given $C = 11$ Mbps, burtiness of 150 cells and rate of 83.3 cells/sec (cf. CLASS-A traffic pattern in Table III) are respectively normalized to $150 \times 48 \times 8/11\text{M} = 0.052$ C-sec and $150 \times 48 \times 8/11\text{M} = 0.029$ C in our analysis. The unit of delay requirement, 200 msec, remains unchanged. A data frame size of $58 + 48 \times 3 = 202$ bytes yields $L_{data} = 202 \times 8/11\text{M} = 0.1469$ msec.

polling token). The effect of *frame size*, as well as the overhead due to *CF-ACK* and *CF-Poll* control frames, should also be incorporated. This is done by replacing the right-hand-sides of (2) and (3) by:

$$\sum_{i=1}^{j}\left\{\min\left\{\sigma_i+\rho_i(t-d_i), r_i d_i\right\}\times 2+\left[\sigma_i+\rho_i(t-d_i)-r_i d_i\right]^+\times\left(1+\frac{L_{ACK}}{L_{data}}\right)\right\}$$
$$+\sum_{i=1}^{j} r_i(t-d_i)\left(1+\frac{L_{CF-Poll}}{L_{data}}\right)+2L_{data}-L_{ACK} \tag{6}$$

and

$$\sum_{i=1}^{n}\left\{\min\left\{\sigma_i+\rho_i(t-d_i), r_i d_i\right\}\times 2+\left[\sigma_i+\rho_i(t-d_i)-r_i d_i\right]^+\times\left(1+\frac{L_{ACK}}{L_{data}}\right)\right\}$$
$$+\sum_{i=1}^{n} r_i(t-d_i)\left(1+\frac{L_{CF-Poll}}{L_{data}}\right)-L_{ACK} \tag{7}$$

where $[x]^+ \stackrel{\triangle}{=} \max\{x,0\}$, $r_i=(1-\sum_{i=1}^{n}\rho_i)/m_p$, and L_{ACK} and $L_{CF-Poll}$ represent the transmission time of a *CF-ACK* control frame and a *CF-Poll* control frame, respectively.

Equations (6) and (7) can be justified as follows. In the worst case, all tokens generated during $[0,t-d_i)$ are transmitted as *CF-Poll* control frames, which result in the overhead corresponding to the second summation in (6). The tokens that generate during $[t-d_i,t)$ will contribute *Data+CF-Poll* frames in the number of $\min\{A_i[0,t-d_i], r_i d_i\}$ and *Data* frames in the number of $[A_i[0,t-d_i]-r_i d_i]^+$, which, together with $A_i[0,t-d_i]\le\sigma_i+\rho_i(t-d_i)$, cause the overhead corresponding to the first summation in (6). The third term $2L_{data}$ in (6) is exactly the nonpreemptive offset for the possible combination of a down-link *Data+CF-Poll* frame followed by an up-link *Data+CF-ACK* frame. The deduction of the last term L_{ACK} in (6) reflects the fact that the time for the last returned *CF-ACK* control frame from the wireless station can be ignored without violating the down-link hard-guaranteed delay bound. Although the last returned frame could also be a *Data+CF-ACK* frame, we take a pessimistic view by deducting only L_{ACK} to secure the hard-guaranteed down-link QoS. Similar arguments justify (7).

The expressions in (6) and (7), which involves the minimum and $[\cdot]^+$ operations may not be reduced to a simple admission control criterion formula as (5). We therefore take (slightly) larger quantities (i.e., (8) and (9)) to respectively replace (6) and (7) in our derivation of admission control criterion.

$$\sum_{i=1}^{j}[\sigma_i+\rho_i(t-d_i)-r_id_i]\times\left(1+\frac{L_{ACK}}{L_{data}}\right)+\sum_{i=1}^{j}2r_id_i$$
$$+\sum_{i=1}^{j}r_i(t-d_i)\left(1+\frac{L_{CF-Poll}}{L_{data}}\right)+2L_{data}-L_{ACK} \tag{8}$$

and

$$\sum_{i=1}^{n}[\sigma_i+\rho_i(t-d_i)-r_id_i]\times\left(1+\frac{L_{ACK}}{L_{data}}\right)+\sum_{i=1}^{n}2r_id_i$$
$$+\sum_{i=1}^{n}r_i(t-d_i)\left(1+\frac{L_{CF-Poll}}{L_{data}}\right)-L_{ACK} \tag{9}$$

As a result, a *sufficient* admission control criterion is yielded.

We now turn to the modification of the left-hand-sides of (2) and (3). As mentioned in Section III-B, the service curve of the IEEE 802.11 wireless LAN is no longer a straight line. Instead, one should introduce a piecewise linear function with slope being equal to one during the CFP, and zero during the CP (cf. Figure 6). According to the standard, the AP may end a CFP by issuing a *CF-END* control frame if all the stations in the polling list indicate no traffics for transmission. Hence, the minimum length of a CFP is $m_p(L_{CF-Poll}+L_{ACK})$, which immediately implies that the maximum length of a CP is $\mathit{CPLength}_{max}=\mathit{CFPRepetitionInterval}-m_p(L_{CF-Poll}+L_{ACK})$. Since $m_p(L_{CF-Poll}+L_{ACK})$ is small, it is reasonable to assume that $m_p(L_{CF-Poll}+L_{ACK})\leq\mathit{CFPMaxDuration}$.

To maximize the number of schedulable users, the length of the CFP should be extended to its largest possible value, i.e., *CFPMaxDuration*. However, the standard dictates that if the remaining time is not enough for the completion of a down-link *Data+CF-Poll* frame followed by an up-link *Data+CF-ACK* frame, the transaction must be deferred until the next CFP. Accordingly, we can only ensure that the maximum schedulable length of a CFP, denoted by $\mathit{CFPLength}_{greedy}$, is at least of the length $\mathit{CFPMaxDuration}-2L_{data}$. Note that the worst-case scenario should always be considered for a hard-guaranteed service discipline.

Again, we have to reduce the condition of the piece-wise linear service curve being no less than the function corresponding to (8) and (9) into a schedulability condition without argument t, as we did in (5). The two curves are plotted in Figures 9 and 10, respectively. We observe that it suffices to check the relation of the two curves only at $t = d_j$ (cf. Figure 9) and at the end of the CP immediately following $t = d_j$ (cf. Figure 10) for $1 \leq j \leq n$, if $\mathit{CFPLength}_{greedy}/\mathit{CFPRepetitionInterval} \geq \sum_{i=1}^{n}(\rho_i + r_i)$. As a consequence, a set of n connections, $\{(\sigma_j, \rho_j, d_j)\}_{1 \leq j \leq n}$, is nonpreemptive EDF/tg-Polling-schedulable over the IEEE 802.11 wireless LAN, if the following five conditions are satisfied.

$$(E1) \qquad \frac{\mathit{CFPLength}_{greedy}}{\mathit{CFPRepetitionIinverval}} \geq \sum_{i=1}^{n}(\rho_i + r_i)$$

$$(E2) \qquad s_j \geq \sum_{i=1}^{j}[\sigma_i + \rho_i(d_j - d_i) - r_i d_i] \times \left(1 + \frac{L_{ACK}}{L_{data}}\right) + \sum_{i=1}^{j} 2 r_i d_i + \sum_{i=1}^{j} r_i(d_j - d_i)\left(1 + \frac{L_{CF-Poll}}{L_{data}}\right) + 2L_{data} - L_{ACK} \quad \text{for } 1 \leq j < n;$$

$$(E3) \qquad s_n \geq \sum_{i=1}^{n}[\sigma_i + \rho_i(d_n - d_i) - r_i d_i] \times \left(1 + \frac{L_{ACK}}{L_{data}}\right) + \sum_{i=1}^{n} 2 r_i d_i + \sum_{i=1}^{n} r_i(d_n - d_i)\left(1 + \frac{L_{CF-Poll}}{L_{data}}\right) - L_{ACK} \quad \text{for } 1 \leq j < n;$$

$$(E4) \qquad \tilde{s}_j \geq \sum_{i=1}^{j}[\sigma_i + \rho_i(\tilde{d}_j - d_i) - r_i d_i] \times \left(1 + \frac{L_{ACK}}{L_{data}}\right) + \sum_{i=1}^{j} 2 r_i d_i + \sum_{i=1}^{j} r_i(\tilde{d}_j - d_i)\left(1 + \frac{L_{CF-Poll}}{L_{data}}\right) + 2L_{data} - L_{ACK} \quad \text{for } 1 \leq j < n;$$

$$(E5) \qquad \tilde{s}_n \geq \sum_{i=1}^{j}[\sigma_i + \rho_i(\tilde{d}_n - d_i) - r_i d_i] \times \left(1 + \frac{L_{ACK}}{L_{data}}\right) + \sum_{i=1}^{n} 2 r_i d_i + \sum_{i=1}^{n} r_i(\tilde{d}_n - d_i)\left(1 + \frac{L_{CF-Poll}}{L_{data}}\right) - L_{ACK},$$

where

$$s_j \triangleq \left\lfloor \frac{d_j - \mathit{CPLength}_{max}}{\mathit{CFPRepetitionInterval}} \right\rfloor \times \mathit{CFPLength}_{greedy} + a_j$$

is the effective service period[4] up to $t = d_j$ (cf. Figure 9),

$$a_j = \begin{cases} d_j - b_j, & \text{for } d_j - b_j < \mathit{CFPLength}_{greedy}; \\ \mathit{CFPLength}_{greedy}, & \text{otherwise,} \end{cases}$$

$$b_j \triangleq \left\lfloor \frac{d_j - \mathit{CPLength}_{max}}{\mathit{CFPRepetitionInterval}} \right\rfloor \times \mathit{CFPRepetitionInterval} + \mathit{CPLength}_{max},$$

---

[4] It is obvious from Figure 9 that d_j must be greater than $\mathit{CPLength}_{max}$, otherwise the delay bound could be violated.

$$\tilde{d}_j \triangleq \left\lceil \frac{d_j - CPLength_{max}}{CFPRepetitionInterval} \right\rceil \times CFPRepetitionInterval + CPLength_{max}$$

is the time at the end of the CP immediately following $t = d_j$, and

$$\tilde{s}_j \triangleq \left\lceil \frac{d_j - CPLength_{max}}{CFPRepetitionInterval} \right\rceil \times CFPLength_{greedy}$$

represents, again, the effective service period up to $t = \tilde{d}_j$ (cf. Figure 10).

E. Admission control criteria for RR-Polling scheduler

In studying the admission control criterion of the RR-Polling scheduler, we, again, assume that the AP only supports two-way call services for a fair comparison with the EDF/tg-Polling scheduler. Also, the situation of conducting two-way calls between wireless stations is excluded. Hence, the number of guaranteed QoS connections equals the number of associated wireless stations, and all the associated wireless stations are CF-Pollable.

In short, the RR-Polling scheduler will serve each associated CF-Pollable station in sequence. At each scheduled serving time of a wireless station, the AP will issue

(1) either a *CF-Poll* control frame, if there are no traffics queued in the AP for the targeted CF- Pollable wireless station;

(2) or a *Data+CF-Poll* frame, otherwise.

For consistency, the same notations in the EDF/tg-Polling schedulability conditions are also used here. The effective service curve and the maximum arrival curve of connection i are plotted in Figure 11. We then observe that a down-link QoS demand is hard-guaranteed if, and only if, for connection i, the effective service curve is no less than the maximum arrival at $t = d_i$ and also at the end of the CP immediately following $t = d_i$, provided that

$$\frac{1}{2n} \frac{CFPLength_{greedy}}{CFPRepetitionInterval} \geq \rho_i,$$

where $1/(2n)$ corresponds the service portion that can at least be allocated to this down-link traffic. We conclude that a set of n connections is nonpreemptive RR-Polling schedulable over the IEEE 802.11 wireless LAN, if for every $1 \leq i \leq n$,

$$(R1) \qquad \frac{1}{2n} \frac{CFPLength_{greedy}}{CFPRepetitionInterval} \geq \rho_i;$$

$$(R2) \qquad s_i \geq 2L_{data} n \sigma_i - L_{data};$$

$$(R3) \qquad \tilde{s}_i \geq 2L_{data} n[\sigma_i + \rho_i(\tilde{d}_i - d_i)] - L_{data},$$

where s_i, $\tilde{s}_i$, $\tilde{d}_i$ and L_{data} are defined the same as in the EDF/tg-Polling schedulability conditions except that $\mathit{CFPLength}_{greedy}$ for RR-Polling scheme now becomes

$$\lfloor \mathit{CFPMaxDuration}/(2L_{data}) \rfloor \cdot 2L_{data}.$$

The deduction of the last term L_{data} reflects the fact that for the last *Data+CF-Poll* frame, only the down-link transmission time affects the validity of the delay demand. Hence, the overestimate portion in the first term (i.e., $2L_{data}n\sigma_i$ in (R2) and $2L_{data}n[\sigma_i + \rho_i(\tilde{d}_i - d_i)]$ in (R3)) for the last transmission should be removed.

IV. Simulation results

A. System configurations

Our simulations are based on the following configurations.

(1) DSSS PHY implementation is employed.

(2) No multiple, overlapping, PCFs are operating on the same PHY channel.

(3) The wireless medium is assumed reliable during the PCF, i.e., the frame error rate for contention free services is zero.

(4) The propagation delay is neglected, which is a fairly realistic assumption if the transmission distances are within the order of 100 feet among wireless stations.

(5) The nominal data rate is assumed to be 11 Mbps, which is the proposed data transmission rate in the coming revised standard;

(6) The frame size of the *Beacon* management frames is assumed to be zero.

B. Schedulable regions

We now simulate the schedulable regions of the RR-Polling and the EDF/tg-Polling schedulers[5] for a set of two traffic characterizations. For convenience, these two traffic characterizations are respectively referred to as CLASS-1 and CLASS-2, as shown in Table I. Furthermore, four different CFP settings on *CFPMaxDuration* and *CFPRepetionInterval* are tested as listed in Table II.

---

[5] The schedulable region of the EDF/tg-Polling scheme obtained in this section is actually smaller than what can really achieve, since a sufficient schedulability condition is employed instead of the optimal one (cf. Section III).

TABLE I

Typical source characterizations.

	Burstiness σ (cells)	Ave. rate ρ (cells/s)	Delay bound d (ms)
CLASS-1	120	55.6	500
CLASS-2	150	83.3	250

TABLE II

The CFP settings used in simulations.

	CFPMaxDuration	*CFPRepetitionInverval*
Default	390 ms	410 ms
Modified-1	20 ms	25 ms
Modified-2	5 ms	10 ms
Modified-3	5 ms	6 ms

In our experiments, we fix the number of CLASS-1 connections, and determine the maximum number of CLASS-2 connections that are schedulable after the settlement of these CLASS-1 connections.

We observe that under the Default CFP setting, both RR-Polling and EDF/tg-Polling schemes can admit at most four CLASS-1 users to the system, and schedule no CLASS-2 users. Thus, the typical Default CFP setting as suggested in the standard is by no means suitable for these two classes of time-bounded services.

We then turn to the experiments on the schedulable regions for the Modified-1 CFP setting, which are summarized in Figure 12. We first note, from these results, that a significant improvement in schedulable region for the Modified-1 CFP setting over the Default CFP setting is obtained, which matches the intuition that a smaller repetition cycle for the CPF is more suitable for real-time services. Secondly, the EDF/tg-Polling schedulable region is approximately twice (in its area) of the RR-Polling schedulable region. The third observation is that a smaller frame size does not necessarily yield a larger schedulable region. This is because smaller frame size may not be sufficient to compensate the 58-byte overhead of the frame header. Figure 13 indicates that a three-cell frame (or 202 bytes) is a better choice in the schedulable regions. Finally, we found that the number of schedulable users actually depends on the order in which these users enter the system. For example, in the figure of "frame size = 3 data cells," only six CLASS-2 users are allowed to enter the system after the admission of one CLASS-1 user, but the system can accept up to seven CLASS-2 users following the admission of two CLASS-1 users. This effect can be roughly explained by the fact that the former case reserves more bandwidth for the up-link traffic due to a larger token generating rate, and hence, the delay demand of the 7th CLASS-2 user can no longer be guaranteed.

In the last experiment, we investigate whether a further improvement on schedulable region can be obtained by further reduction of the CFP repetition cycle. The simulation results are plotted in Figures 13 and 14, respectively. We conclude that as long as the ratio between the *CFPMaxDuration* and *CFPRepetitionInterval* remains nearly the same, a slight improvement can be reached by taking smaller *CFPRepetitionInterval*. For example, the ratios of Modified-1 and Modified-3 CFP settings are respectively 4/5 and 5/6, which are close to each other. As a result, the Modified-3 CFP setting yields a slightly larger schedulable region. On the contrary, if this ratio is decreased, say 1/2 for Modified-2 CFP setting, degradation in schedulable region may arise even if a smaller *CFPRepetitionInterval* is employed.

C. Average up-link delays

In the simulations of average up-link delays, we assume that the wireless station is in a conversation with the guaranteed QoS remote terminal, where at each time instance, exactly one of the two parties is in a talking state. We also assume that a perfect voice-activity detector is implemented for both call participants, and thereby, voice data is generated only when the call party is speaking. Four kinds of conversational-type source patterns are adopted, which are listed in Table III from the view of guaranteed QoS remote terminals. The *ON* and *OFF* state periods are assumed to be exponentially distributed, where the traffic arrival is at the peak rate during the *ON* state. In order to ensure the conformity to (σ, ρ) upper constraint, leaky-bucket policing mechanisms are applied at the

TABLE III

The conversational-type source patterns.

Group	Burstiness σ (cells)	Ave. rate ρ (cells/s)	Delay bound d (ms)	Peak rate (Kbps)	Ave. *ON* period (s)	Ave. *OFF* period (s)
CLASS-A	150	83.3	200	64	0.4	0.6
CLASS-B	240	111.1	300	64	1	1.35
CLASS-C	75	41.7	400	32	0.35	0.65
CLASS-D	120	55.6	500	32	1	1.35

guaranteed QoS network side.

By uniformly selecting from these four classes of source patterns, we simulate the average up-link delay for different frame sizes under the Modified-1 CFP setting. The results are summarized in Figures 16 and 17. The order in which the source patterns enter the system are marked above each node. For example, in the figure for frame size = 1 data cell, a CLASS-C source is first admitted; then a CLASS-D source enters the system after the previous CLASS-C source is settled, followed by a request from another CLASS-D source, and so forth. The x-axis is the accumulated data rate for all admitted source patterns up to this node. Specifically, at the first node, the CLASS-C source will contribute at most 32 Kbps data rate. At the second node, the CLASS-C and CLASS-D sources together sum to 64 Kbps data offered load. At the third node, there are totally 32 Kbps + 32 Kbps + 32 Kbps = 96 Kbps data offered load, etc. The simulation on the average up-link delay for each curve continues until the next selected source pattern cannot be schedulable.

Based on these graphs, we have the following remarks.

(a) The EDF/tg-Polling scheduler provides less average up-link delay than the RR-Polling, especially for small frame size.

(b) The larger the frame size, the longer the average up-link delay.

(c) The average up-link delay decreases in spite of the increment of the traffic load at some range of the curves. In exploration, we first note that a deep decrement often happens at the time a CLASS-A or CLASS-B user enters the system. By tracing back the simulations, we found that when the data rate of the new user is higher than the current average system data rate, the average up-link delay decreases. Hence, the introduction of a higher data-rate user will improve the average up-link delay. We also observe that the up-link delay will grow, when more *CF-END* control frames are issued by the AP. Specifically, if an up-link data frame is generated immediately after the generation of a *CF-END* control frame, it has to be deferred for transmission until the next CFP. This is the main reason that the data frames under a light traffic load tend to suffer a larger average up-link delay than those data frames under a heavy traffic load.

V. Conclusion and future work

In this paper, we proposed an EDF/tg-Polling scheduler for conveying two-way conversational traffics over the PCF-enabled IEEE 802.11 wireless LAN. Simulation results show that our scheme can provide around twenty two-way calls with guaranteed down-link latency and acceptable up-link delay, and is more robust for variant QoS than the RR-Polling scheme. A side observation during this research is that a better choice of the frame size is perhaps 200-byte in length due to the frame header overhead. Also, smaller CFP repetition period is more suitable for time bounded services. Further improvement on the system performance may be obtained by a more deliberate design of token generating rates.

The design premise that the AP only supports two-way call services can in fact be relaxed, if a multiple two-way call implementation is taken for multi-way call services (cf. Figure 18 as an example). Simulation results show that the performance remains the same for system with (such a) multi-way call service extension. Combination of the multiple two-way calls into a true multi-way conference call could be an interesting extension, if multicasting technique is available.

References

[1] IEEE P802.11, *Draft Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Standards Department, D6.1, May 1997.

[2] C. M. Aras, J. F. Kurose, D. S. Reeves and H. Schulzrinne, "Real-time communication in packet- switched networks," *Proceedings of the IEEE*, Vol. 82, pp. 122–139, January 1994.

[3] C.-S. Chang, K.-C. Chen, M.-Y. You and J.-F. Chang, "Guaranteed quality-of-service wireless access to ATM networks," *IEEE J. Select. Areas Commun.*, Vol. 15, No. 1, pp. 106–118, January 1997.

[4] B. P. Crow, I. Widjaja, J. G. Kim, and P. Sakai, “Investigation of the IEEE 802.11 Medium Access Control (MAC) Sublayer Functions,” in *Proc. IEEE INFOCOM’97*, pp. 126–133, 1997.

[5] R. L. Cruz, “A calculus for network delay, Part I: Network elements in isolation,” *IEEE Trans. Inform. Theory*, Vol. 37, No. 1, pp. 114–131, January 1991.

[6] D. Ferrari and D. C. Verma, “A scheme for real-time channel establishment in wide-area networks,” IEEE J. Select. Areas Commun., Vol. 8, No. 2, pp. 368–379, September 1993.

[7] L. Georgiadis, R. Guerin, and A. Parekh, “Optimal multiplexing on a single link: delay and buffer requirements,” *IEEE Trans. Inform. Theory*, Vol. 43, No. 5, pp. 1518–1535, September 1997.

[8] T.-L. Ling and N. Shroff, “Scheduling real-time traffic in ATM networks,” in *Proc. IEEE INFOCOM’96*, pp. 198–205, 1996.

[9] C. L. Liu and J. W. Layland, “Scheduling algorithms for multiprogramming in hard real time environment,” *J. Assoc. Comput. Math.*, Vol. 20, No. 1, pp. 46–61, January 1973.

[10] A. K. Parekh and R. G. Gallager, “A generalized processor sharing approach to flow control in integrated services networks—the single-node case,” *IEEE/ACM Trans. Networking*, Vol. 1, pp. 344–357, June 1993.

[11] G. Sfikas, R. Tafazolli and B. G. Evans, “ATM cell transmission over the IEEE 802.11 wireless MAC protocol,” in *PIMRC’96*, pp. 173–177, 1996.

[12] J. S. Turner, “New directions in communications,” *IEEE Communication Magazine*, Vol. 25, No. 8, pp. 8–15, October 1986.

[13] M. A. Visser and M. El Zarki, “Voice and data transmission over an 802.11 wireless network,” in *PIMRC’95*, pp. 648–652, 1995.

[14] Z. Zhang, I. Habib, T. Saadawi and N. Mir, “Supporting multimedia traffic over microcellular networks,” in *GLOBECOM ’97*, pp. 1328–1332, 1997.

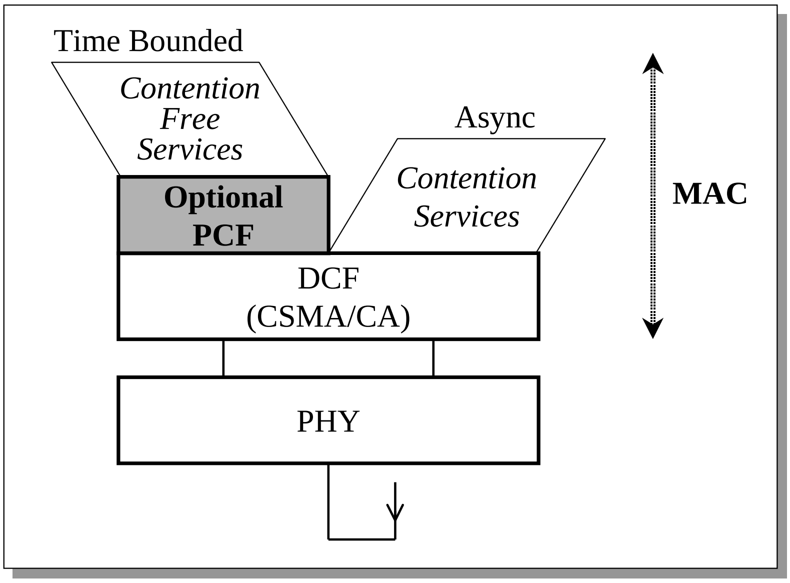


Fig. 1. The MAC architecture of IEEE 802.11.

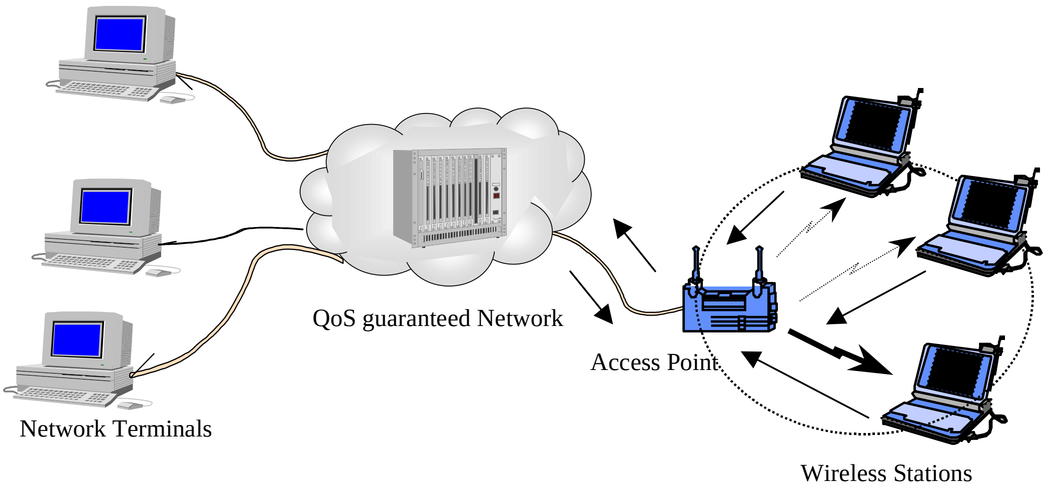


Fig. 2. Guaranteed QoS network access to an IEEE 802.11 wireless LAN.

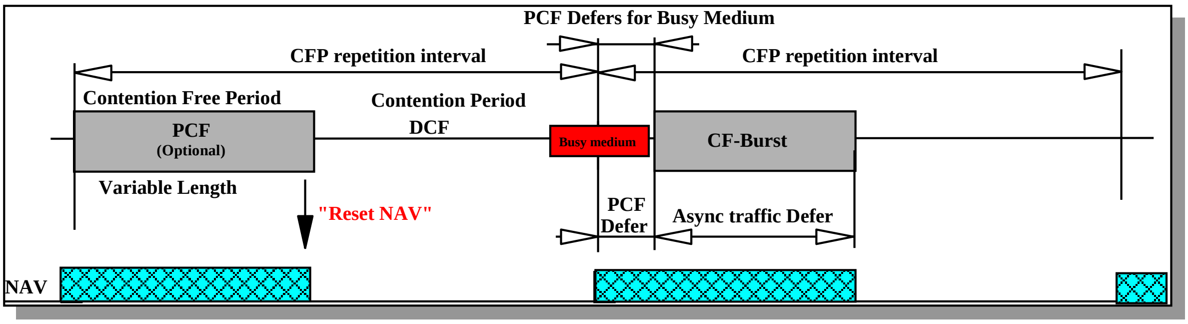


Fig. 3. Operation of alternating Contention Free and Contention Period.

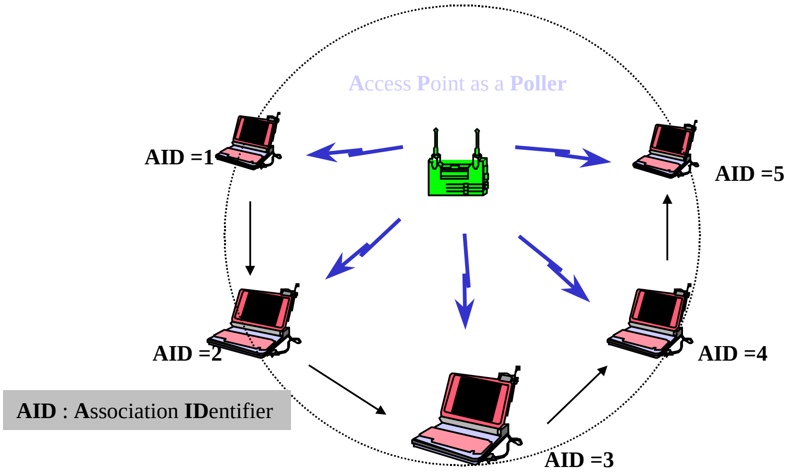


Fig. 4. Round-robin polling scheme (RR-Polling).

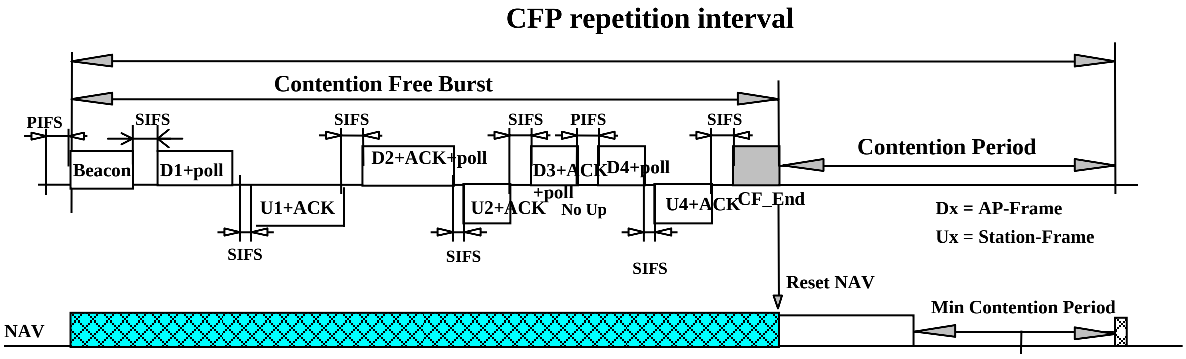


Fig. 5. Example of operation in the CFP. (SIFS and PIFS respectively represent the *Short InterFrame Space* and *PCF InterFrame Space*, which are assumed zero in our simulation.)

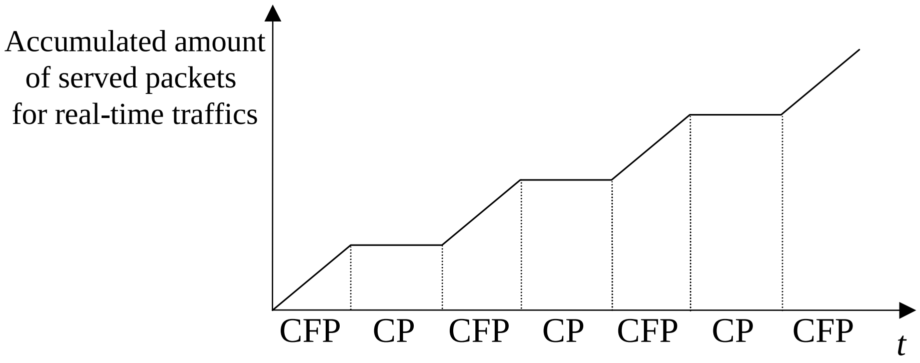


Fig. 6. The service curve of an IEEE 802.11 wireless LAN.

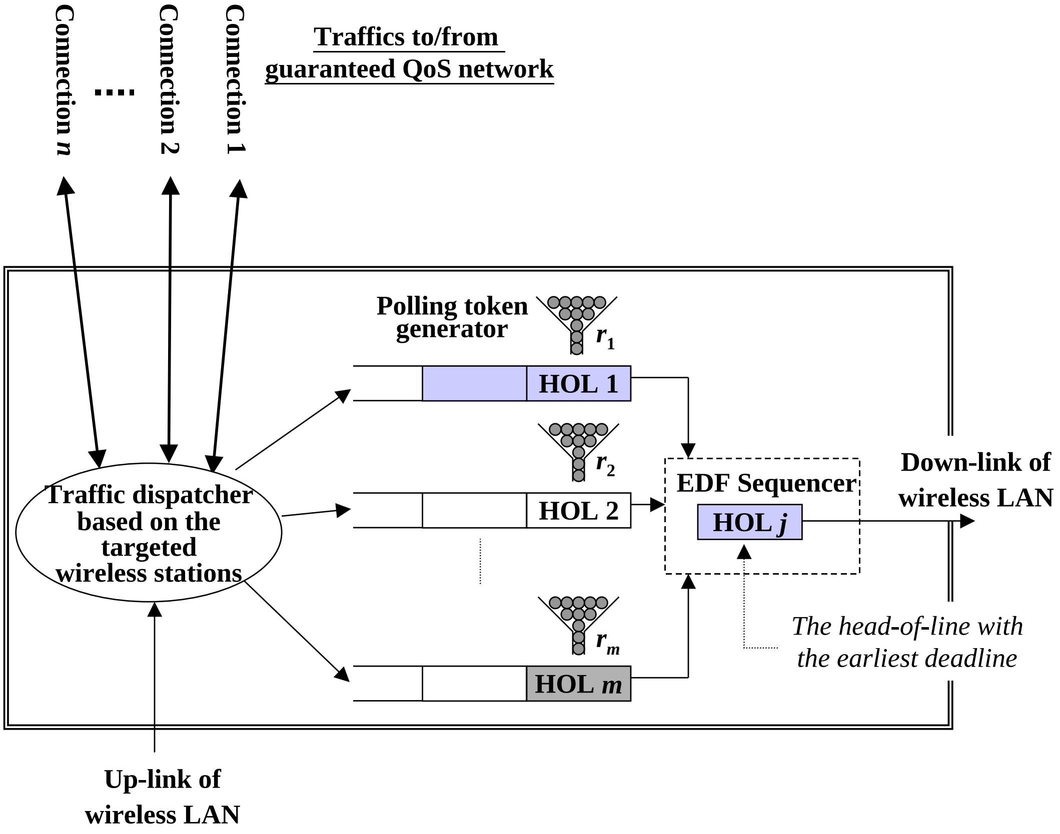


Fig. 7. Architecture of the earliest-deadline-first polling scheduler with token generators, where m represents the number of associated PCF-enabled wireless stations.

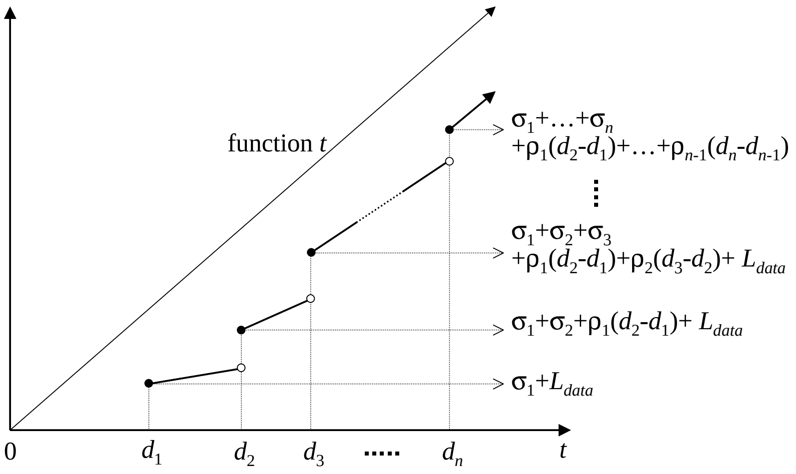


Fig. 8. Criterion for the schedulability of the nonpreemptive EDF scheduler under (σ, ρ, d) source models.

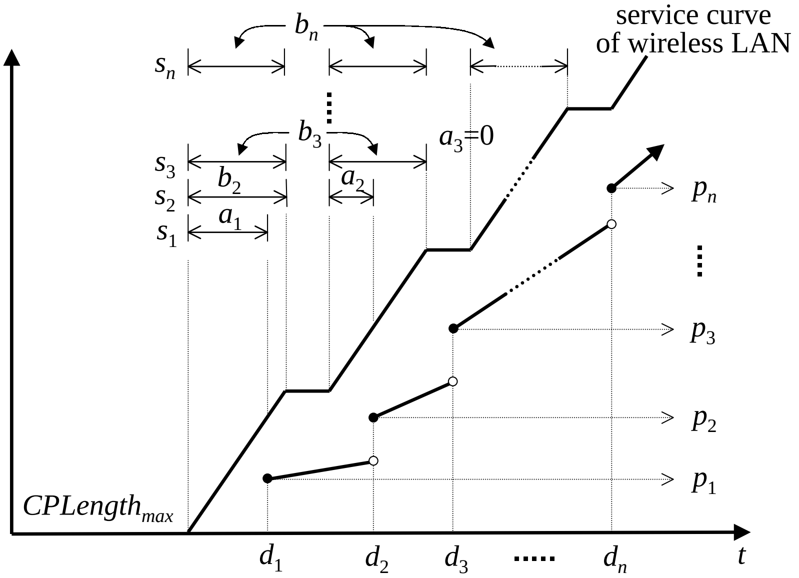


Fig. 9. The worst-case consideration at $t = d_j$, $1 \leq j \leq n$, for EDF/tg-Polling scheduler. (p_j = right-hand-side of condition (E2) for $1 \leq j < n$, and p_n = right- hand-side of condition (E3)).

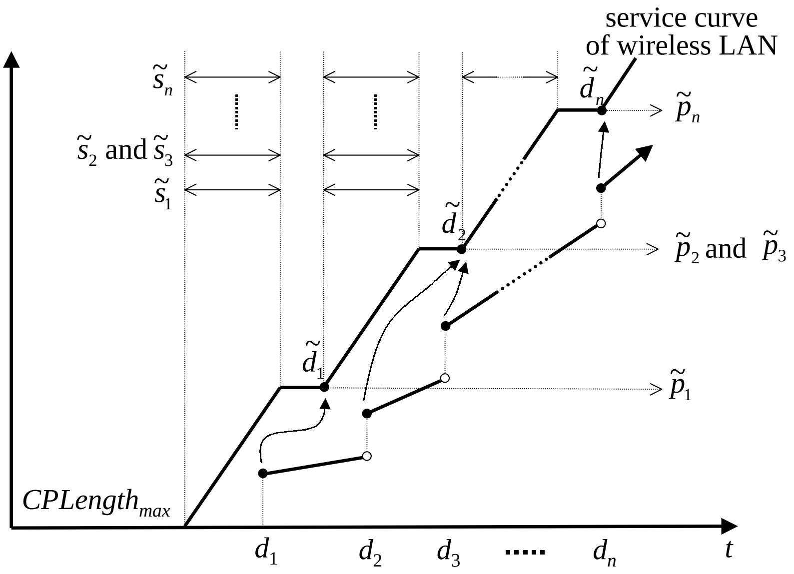


Fig. 10. The worst-case consideration at the end of the CP immediately following $t = d_j$, $1 \leq j \leq n$, for EDF/tg-Polling scheduler. ($\tilde{p}_j$ = right-hand-side of condition (E4) for $1 \leq j < n$, and $\tilde{p}_n$ = right-hand-side of condition (E5)).

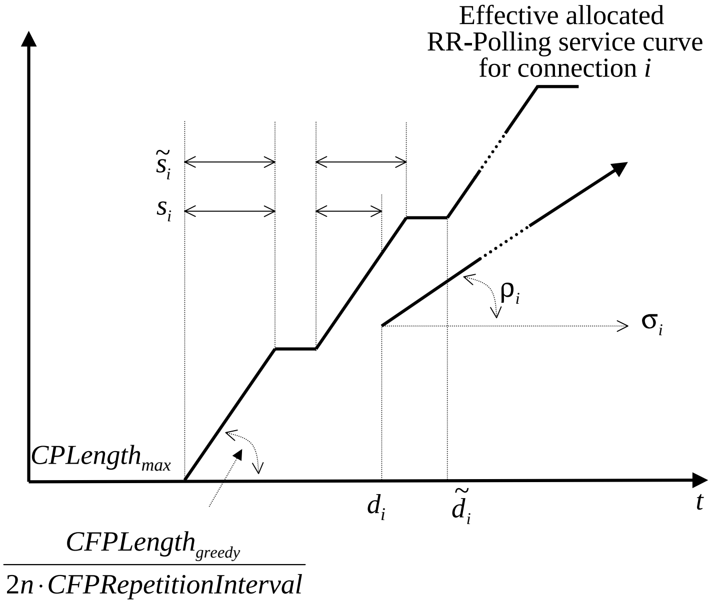


Fig. 11. The worst-case consideration for RR-Polling scheduler.

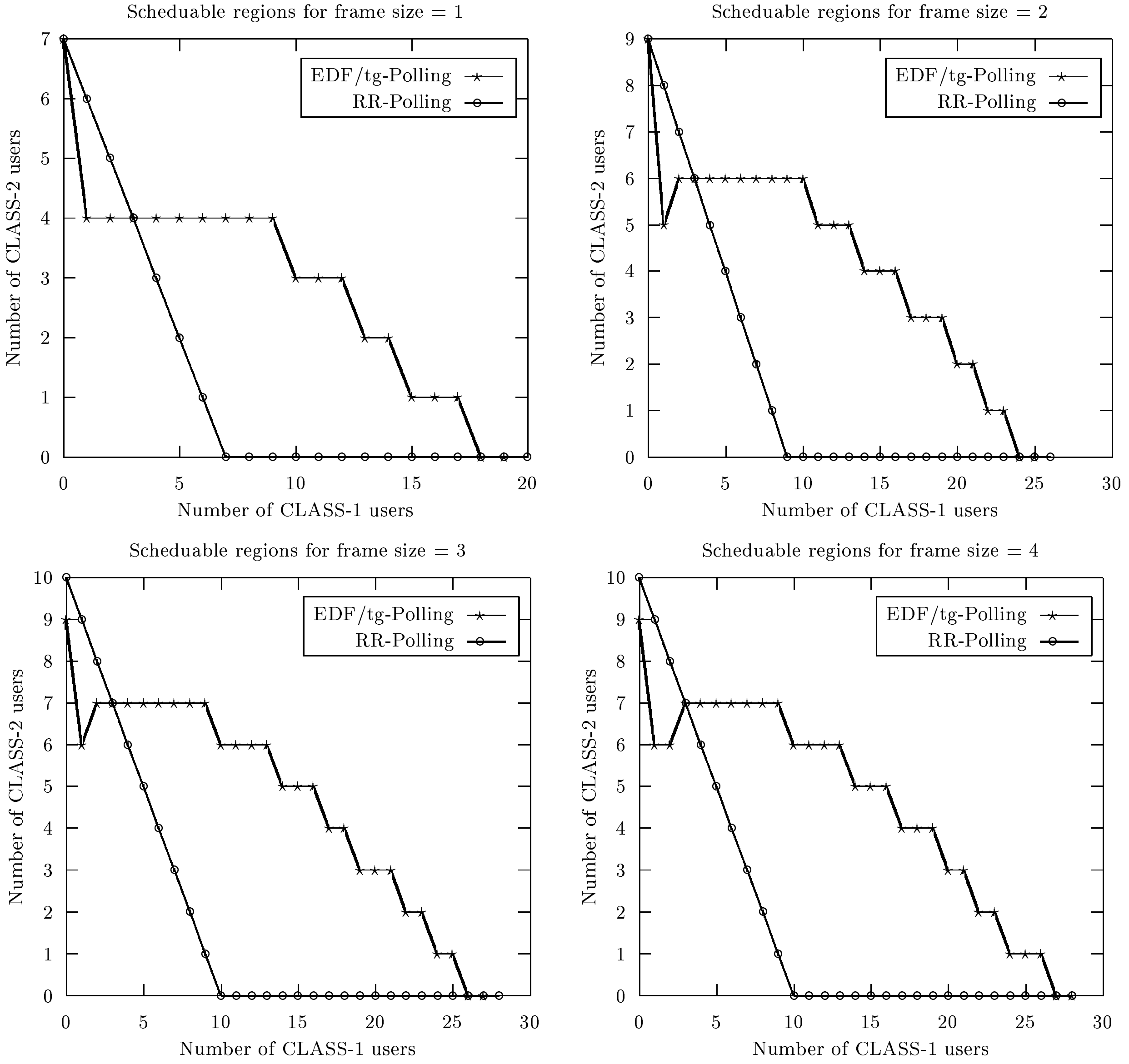


Fig. 12. (a) The schedulable regions for Modified-1 CFP setting with data frame size of 1, 2, 3 and 4 (ATM) cells.

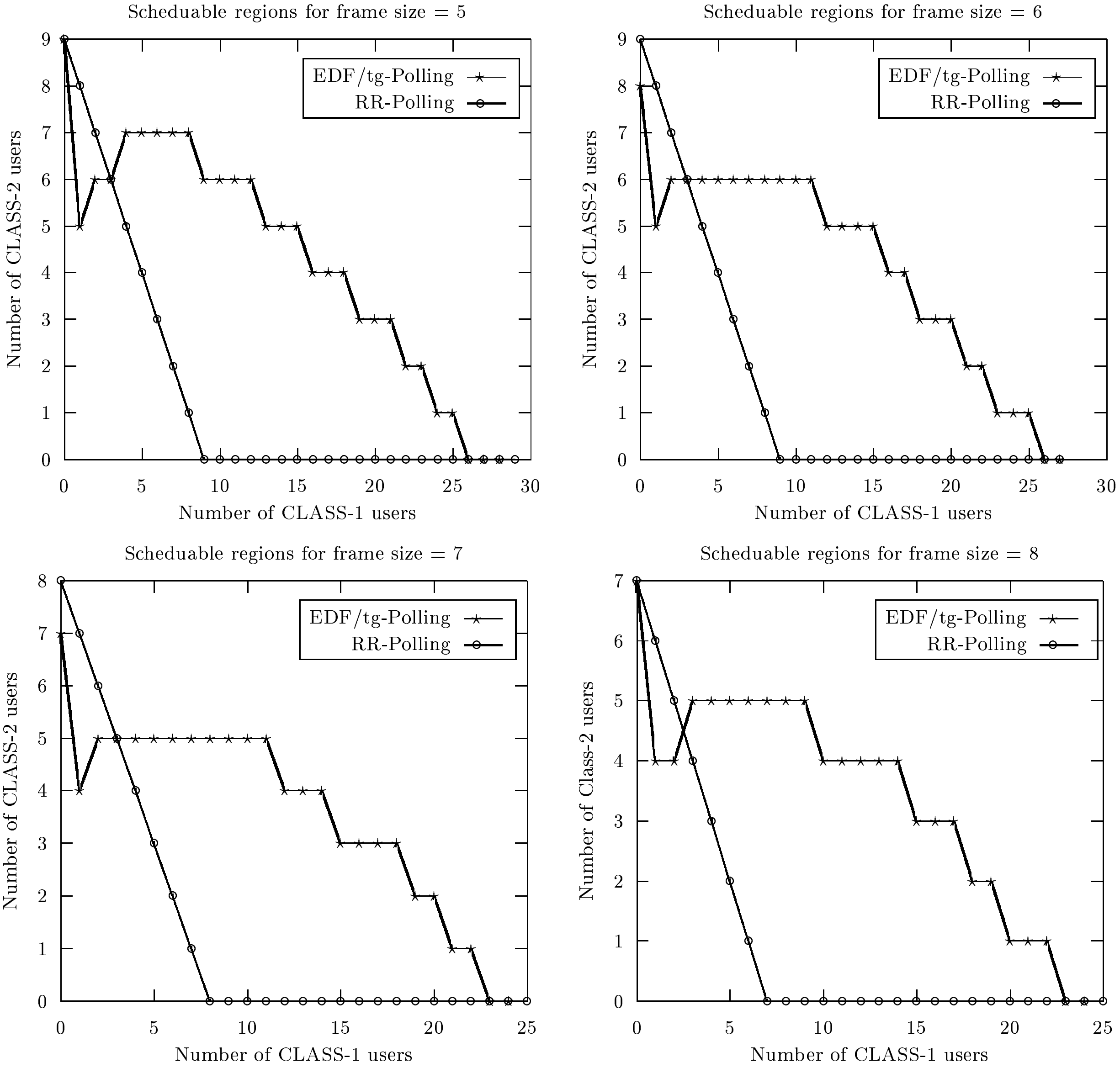
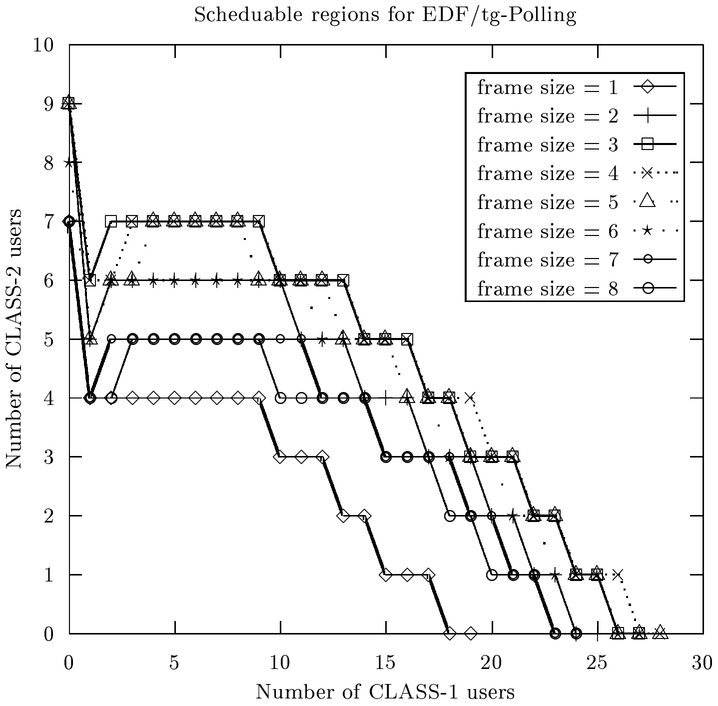


Fig. 12. (b) The schedulable regions for Modified-1 CFP setting with data frame size of 5, 6, 7 and 8 (ATM) cells.



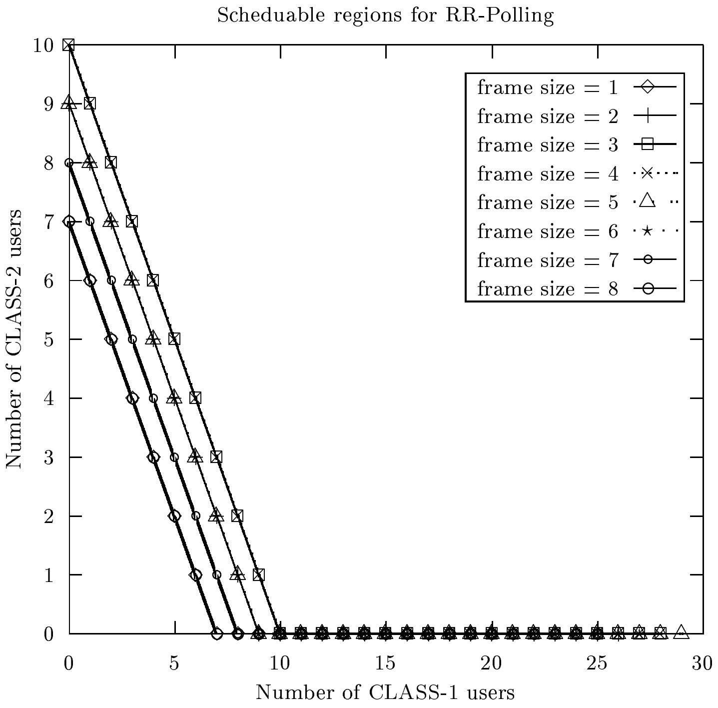
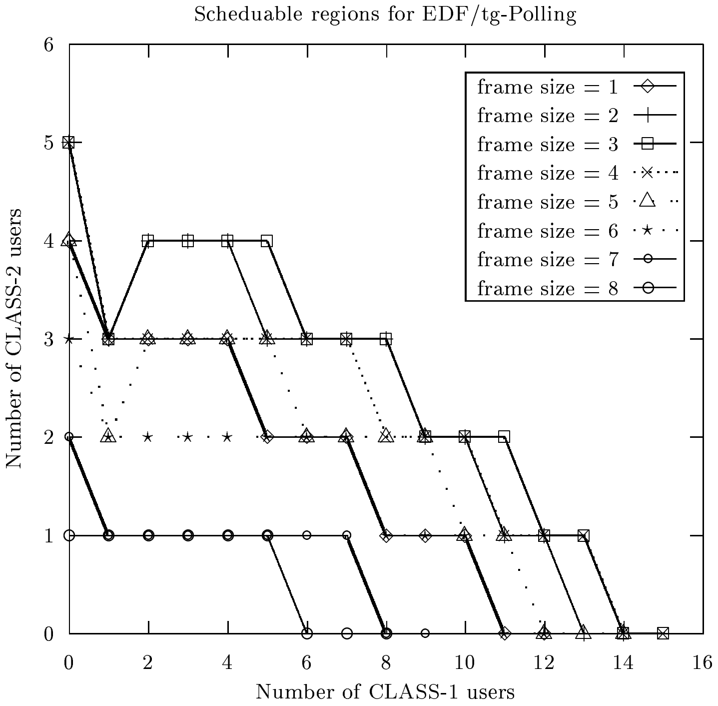


Fig. 13. The effect of frame sizes in schedulable regions for the Modified-1 CFP setting.



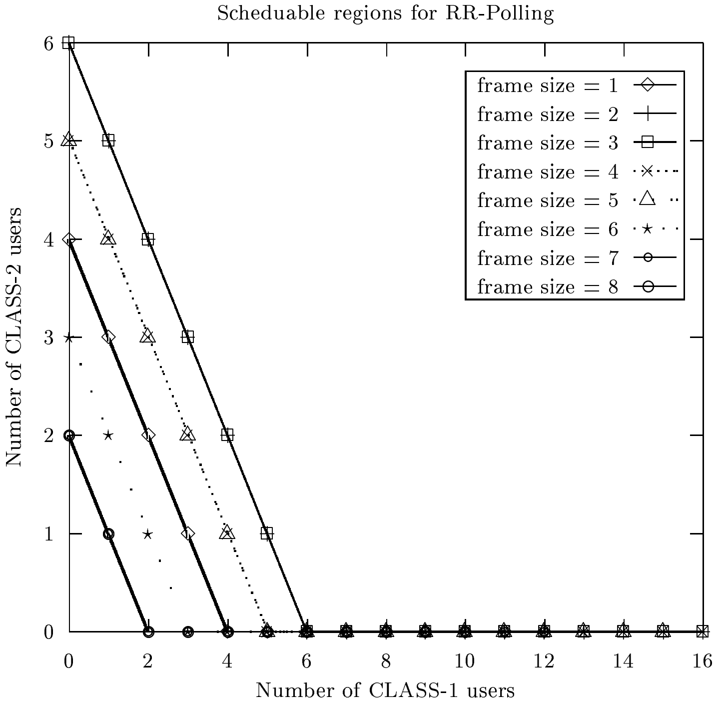
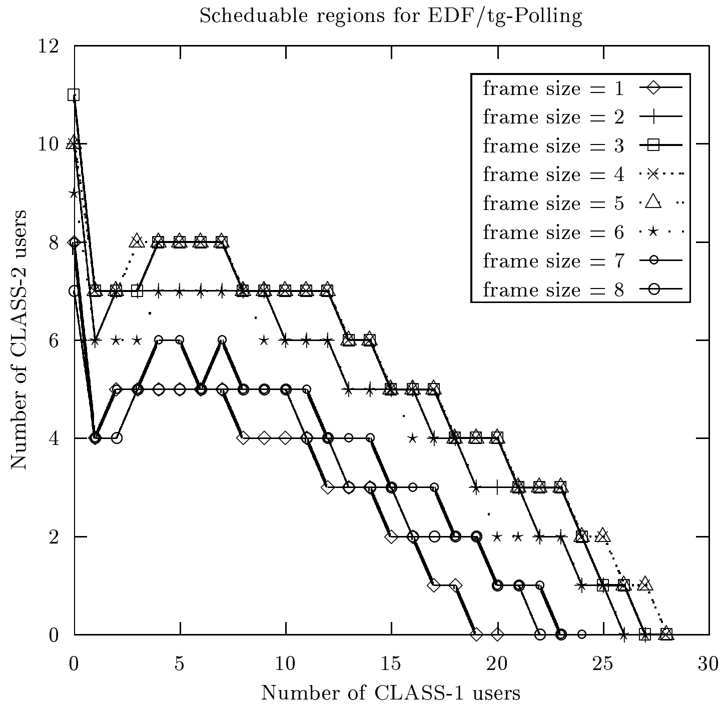


Fig. 14. The effect of frame sizes in schedulable regions for the Modified-2 CFP setting.



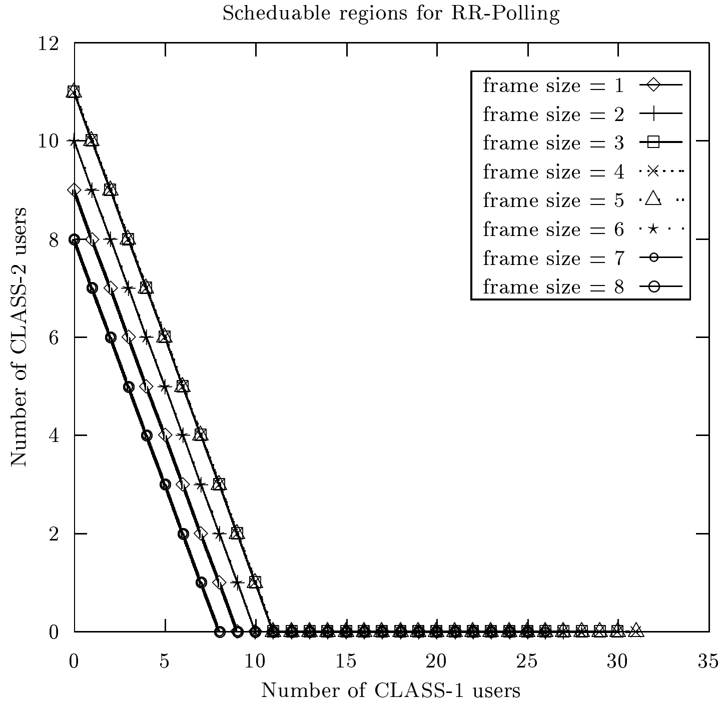


Fig. 15. The effect of frame sizes in schedulable regions for the Modified-3 CFP setting.

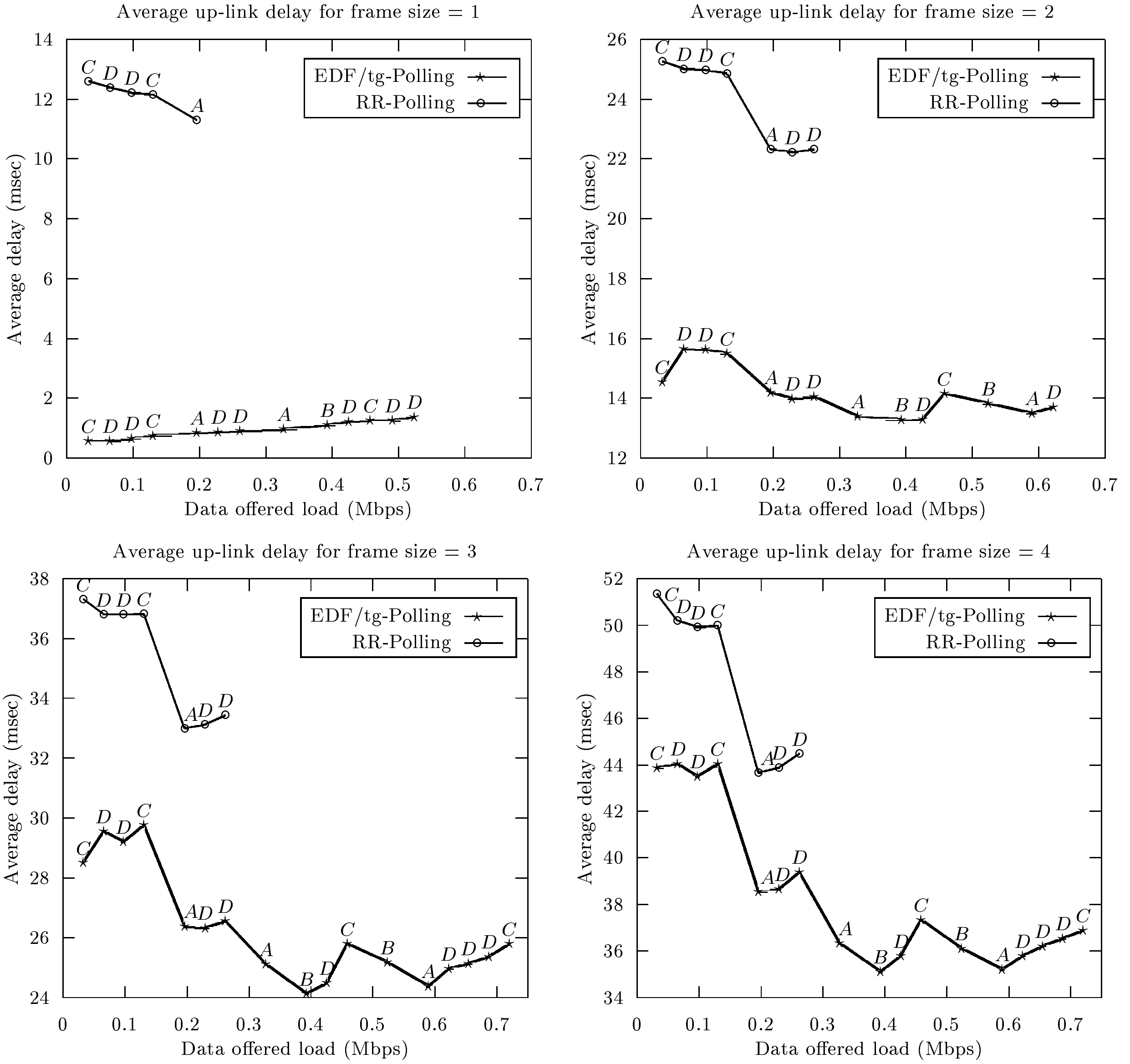


Fig. 16. (a) The average up-link delay for frame size = 1–4 data cells under the Modified-1 CFP setting.

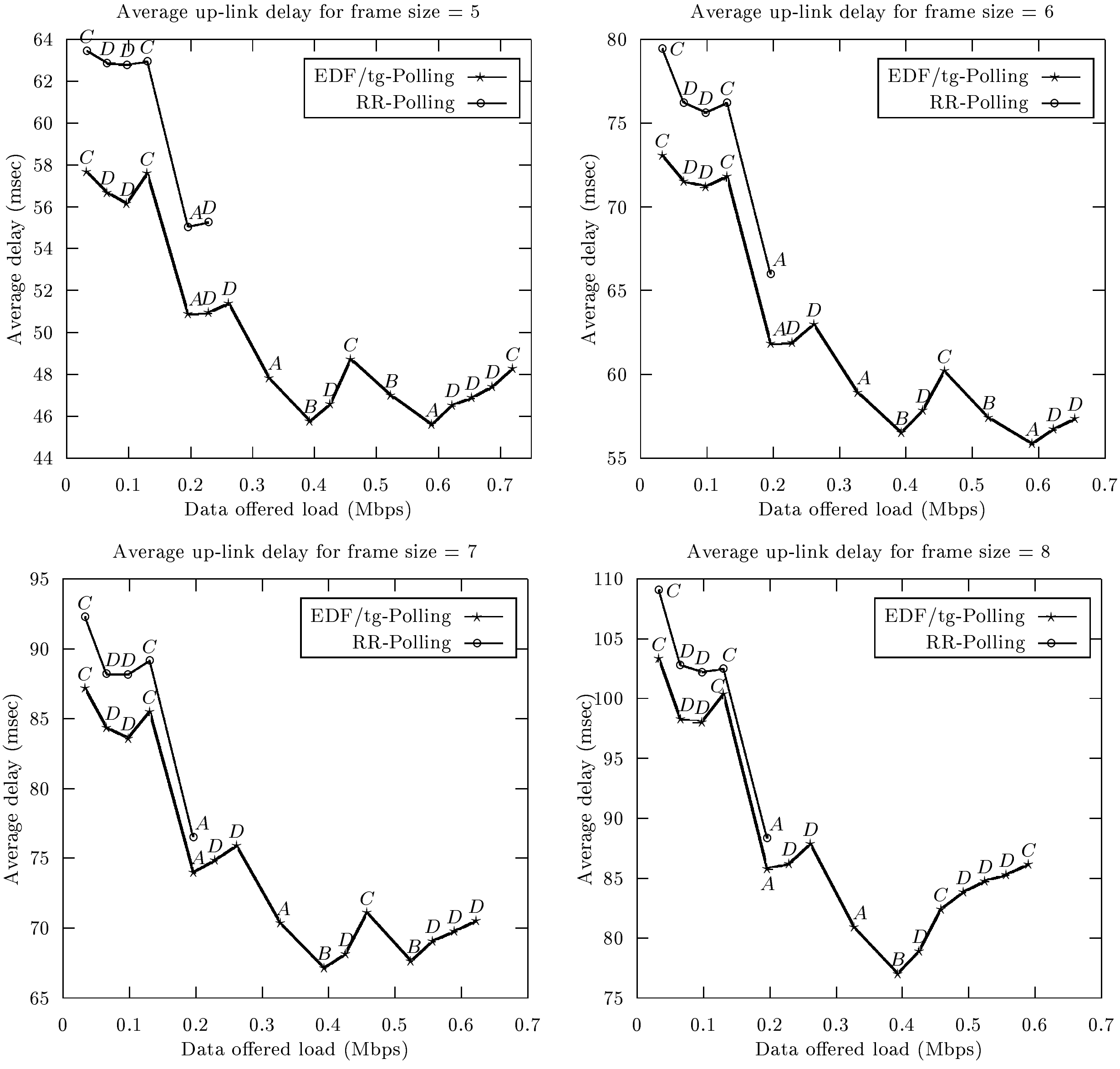


Fig. 16. (b) The average up-link delay for frame size = 5–8 data cells under the Modified-1 CFP setting.

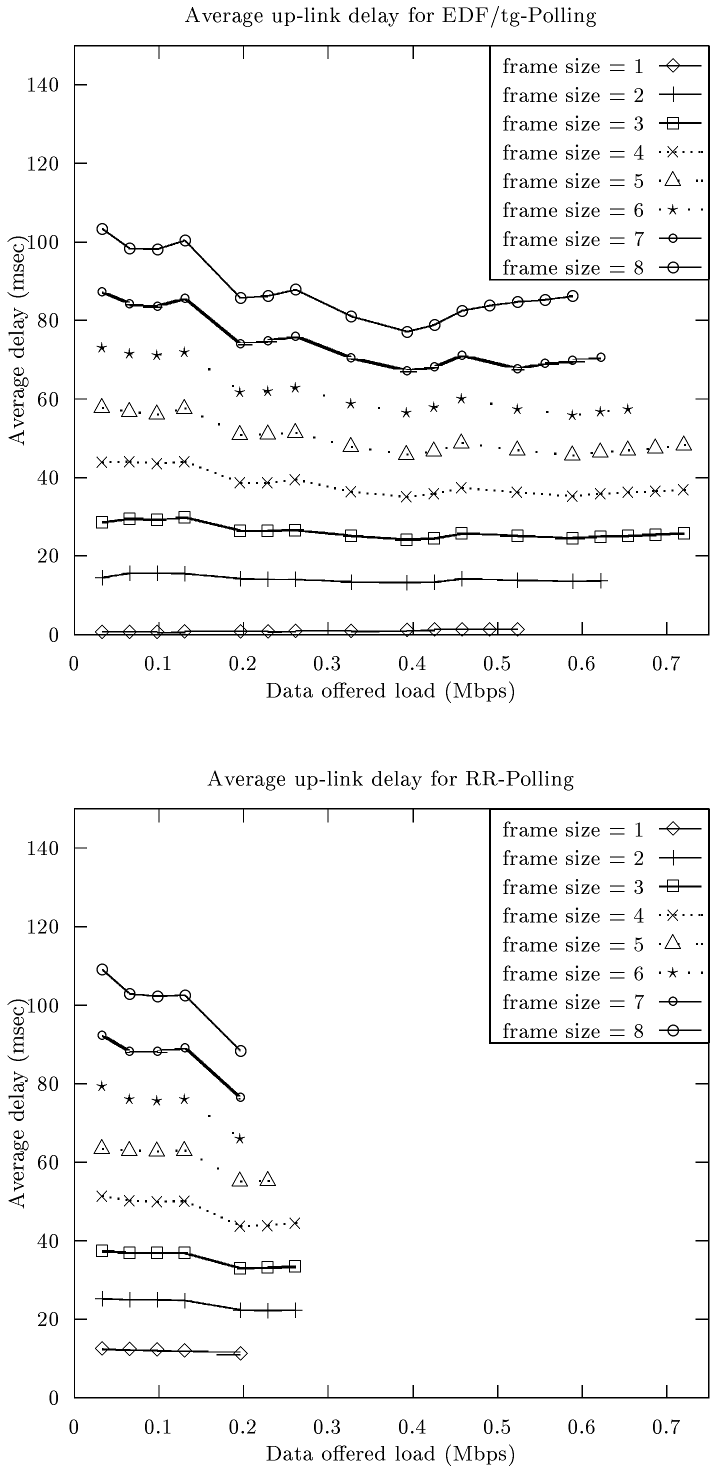


Fig. 17. Summary of Figure 16 in terms of the frame sizes.

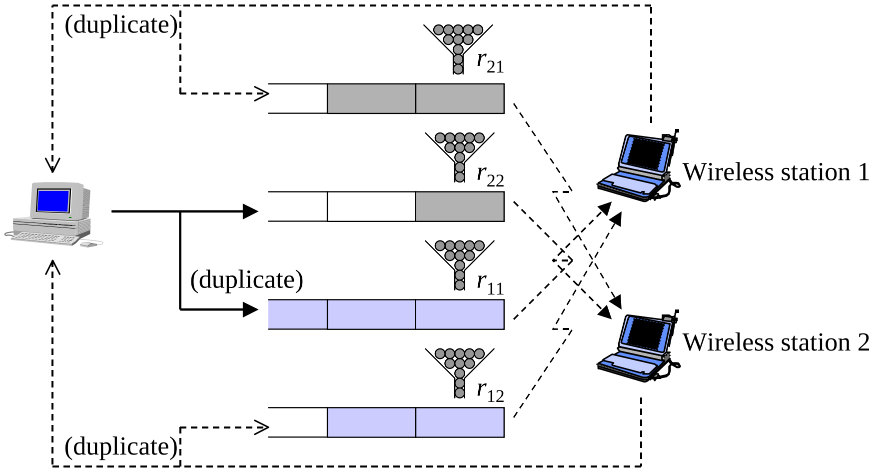


Fig. 18. An example of implementing a three-way call by means of three two-way calls. The duplication tasks can be implemented over the traffic dispatcher.