INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Admission Control and Tuning of Medium Access Parameters in 802.11e Networks

Juliana Freitag *Nelson L.S. da Fonseca*
José Ferreira Rezende

Technical Report - IC-05-33 - Relatório Técnico

December - 2005 - Dezembro

Admission Control and Tuning of Medium Access Parameters in 802.11e Networks

Juliana Freitag* Nelson L. S. da Fonseca* José F. de Rezende†

Abstract

The IEEE 802.11e standard is intended to support applications with QoS requirements. However, such provisioning cannot be achieved when the network load is high. In this paper, the effectiveness of two mechanisms for measurement-based admission control in 802.11e networks is evaluated. In addition, a new control mechanism which dynamically tunes parameters of the 802.11e contention-based access method is introduced. The proposed mechanism aims at providing QoS as well as ameliorating the problem of delay asymmetry.

1 Introduction

The IEEE 802.11 standard for Wireless LANs [5] will play a significant role in the wireless Internet access scenario. Among its advantages it can be mentioned high transmission rate, low-cost and license-free technology. This standard is a strong candidate to make part of a wide-coverage access system, either integrated to third-generation cellular networks or simply composing a whole 802.11 solution for home networking and hotspots.

An extension to the standard, called 802.11e [6], has been proposed to offer Quality of Service (QoS) for applications, since the standard provides only best effort service. This extension incorporates a new coordination function, the Hybrid Coordination Function (HCF), which is composed of two medium access methods for the provisioning of service differentiation: the Enhanced Distributed Channel Access (EDCA) contention-based method and the HCF Controlled Channel Access (HCCA) non-contention-based method.

The EDCA method provides service differentiation using Access Categories (ACs). Each AC has a separate queue, with different values for the medium access parameters: minimum contention window (CW_{min}), maximum contention window (CW_{max}) and arbitrarion interframe space (AIFS). Prioritization is obtained by assigning smaller values for the parameters of high priority ACs. To enhance performance and improve channel utilization, QoS stations (QSTAs) can send frames in bursts during Transmission Opportunity intervals (TXOPs). During these intervals, a QSTA can send a sequence of frames without having to contend for the medium. Along with medium access parameters, the QoS access point (QAP) determines the limit of the TXOP interval for each AC.

*Institute of Computing, University of Campinas, 13081-970 Campinas, SP.
†COPPE, Federal University of Rio de Janeiro, 21945-970,Rio de Janeiro, RJ.

Although the 802.11e extension provides the tuning knobs for QoS control, it does not define how this support is to be provided under specific network conditions. This shortcoming has motivated the development of new mechanisms for monitoring and control of service levels. These mechanisms are of paramount importance to guarantee efficient use of available network resources and the maintenance of QoS levels of existing applications.

Admission control, one of these mechanisms, provides QoS guarantees by restricting the number of flows present in a domain. Various types of admission control mechanisms have been proposed for fixed networks ([7], [10], [15]). Parameter-based mechanisms are based on static values for traffic parameters, whereas measurement-based ones (MBAC) dynamically estimate traffic parameter values by actually measuring the traffic. It has been shown that MBAC mechanisms provide greater network utilization than parameter-based mechanisms [15].

Moreover, Breslau, Jamin and Shenker suggest that a large number of MBAC algorithms has nearly identical performance [13]. However, Qiu and Knightly [7] have shown that different MBAC schemes do indeed lead to different kinds of performance.

In 802.11 networks, the access point is responsible for forwarding all traffic to/from stations which makes it the network bottleneck. Since both QAP and QSTAs have the same probability of accessing the medium, the queues in the QAP can rapidly build up, increasing downlink delay and, consequently, making QoS provisioning unfeasible. A discrepancy between the downlink and uplink delays can lead to the acceptance of a low number of mobile stations into a 802.11e cell, due to the impact of downlink delay values on admission decisions. Thus, to cope with this discrepancy, it is necessary to adapt the access parameters of QSTAs and QAPs dynamically so that balanced delay values can be achieved.

This paper addresses the problem of Quality of Service provisioning by proposing the adoption of both admission control and dynamic tuning of medium access parameters. The contribution of this paper is twofold. It provides an evaluation of the effectiveness of measurement-based admission control for 802.11e networks, as well as an algorithm for ameliorating the discrepancy between uplink and downlink delays. Two mechanisms for admission control in 802.11e networks are compared: Time-Window/Measured Sum MBAC and Traffic Envelope MBAC. Results indicate that the Traffic Envelope MBAC provides a better performance than does the Measured Sum MBAC. Moreover, the scheme proposed for the dynamic tuning of medium access parameters does produce balanced delay values, thus facilitating provisioning of QoS for both uplink and downlink traffic, as well as the admission of a larger number of flows.

The paper is structured as follows. Section 2 presents basic concepts of admission control mechanisms, and Section 3 describes the admission control schemes enhanced to work on 802.11e networks. Section 4 presents the simulation methodology. Section 5 compares the performance of the two admission control schemes. Section 6 presents the motivation for the development of a scheme for dynamic tuning of 802.11e QoS parameters, while Section 7 describes a novel mechanism for such tuning and Section 8 evaluates the effectiveness of the tuning mechanism. Finally, Section 9 presents some conclusions.

2 Measurement-Based Admission Control

Measurement-based admission control schemes do not require analytical models of network traffic since the traffic is characterized by measures extracted from the true traffic itself, with admission decisions based on these measures. Such measurement-based schemes consist of three inter-related components: (1) the signaling protocol by which new flows are established, (2) the traffic measurement module and (3) the admission control or decision module which accepts or rejects requests for the establishment of new flows. The Time-Window/Measured Sum MBAC scheme, for example, uses separate traffic measurement procedures and decision algorithms (MS scheme), whereas the Traffic Envelope MBAC scheme incorporates both measurement and decision making in a single solution (TE scheme).

2.1 Time-Window/Measured Sum MBAC scheme

The Time-Window/Measured Sum MBAC scheme combines a time-window measurement method and a measured sum decision algorithm [15]. The Measured Sum decision algorithm admits a new flow with load ν_f if the load, after acceptance, occupies less than a pre-defined proportion of the channel capacity, i.e.,

$$\nu + \nu_f < \mu * C \tag{1}$$

where ν_f is the flow load, μ is a user-defined utilization target, ν is the measured load of existing traffic, and C is the channel capacity.

At high utilization levels, a measurement-based approach can fail since delay variations are exceedingly large. Thus, admission control algorithms must keep link utilization below a certain target level in order to avoid such fluctuations. In the present paper, the μ value is set according to the wireless link utilization.

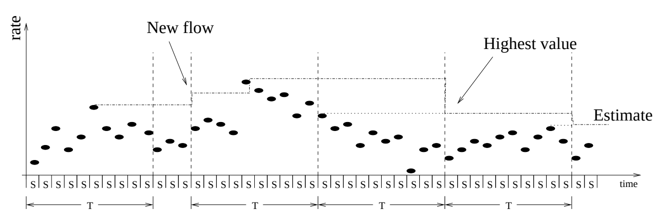

Figure 1: Time-window measurement of network load.

In the MS scheme, the average load is computed every S sampling period, as shown in Figure 1. The algorithm is as follows [15]: at the end of a measurement window T, the load estimate for the next window is the highest average load within the window. When a new flow is admitted, the estimate is increased based on the parameters of the new request. If a newly computed average is above the estimate, the estimate is immediately increased to the level of the new average. At the end of every window, the estimate is adjusted to the actual load measured in the previous window.

The window determines the duration of time for which the algorithm retains the history of the measurements. A short window makes the algorithm to react faster to traffic changes.

The sampling period (S) influences the measured burstiness. Short samples periods result in high averages, thus providing a more conservative admission control algorithm. According to [15], it is recommended the following rule of thumb be used: $\frac{T}{S} \leq 10$.

2.2 Traffic Envelope MBAC scheme

In the Traffic Envelope MBAC scheme [2], the admission decision for each flow is based solely on the aggregate measurements made at egress routers. The technique consists on measuring and controlling the aggregate arrival and service envelopes for each traffic class at the egress nodes.

To obtain the arrival envelope, the time when each packet enters the ingress node queues must be stored and made available to the egress node. Ingress nodes thus insert a timestamp containing the packet arrival time into the IP packet header. At the egress node, the service time when the packet leaves the node at the outgoing interface is also registered. This ensures that when calculating the envelopes, queuing delays at both edge nodes can be also accounted for.

The traffic is characterized via aggregate peak-rate envelopes as follows. Let $A[s, s+I_k]$ denote the flow arrivals in the interval $[s, s+I_k]$. The rate during this particular interval is given by $A[s, s+I_k]/I_k$, with the peak rate over any interval of length I_k given by $R_k = max_s A[s, s+I_k]/I_k$. The set of rates R_k which binds the flow rate over intervals of length I_k is defined by the peak-rate envelope.

Time is slotted with width $\tau = I_1$, which corresponds to the minimum interval of the measured rate envelope. The maximal rate envelope over the past T time slots from the current time t is defined as:

$$R_k^1 = \frac{1}{k\tau} \quad max_{t-T+k \leq s \leq t} \; A[(s-k+1)\tau, s\tau] \tag{2}$$

for $k = 1, ..., T$.

The aggregate maximal rate envelope over intervals of length $I_k = k\tau$ in the most recent $T\tau$ seconds is given by the envelope R_k^1, $k = 1, ..., T$. R_k^1 estimates the short scale burstiness, as well as the autocorrelation structure of the aggregate flow.

R_k^1 is measured every T time slots and $R_k^m \leftarrow R_k^{m-1}$ for $k = 1, ..., T$ and $m = 2, ..., M$. Thus, for each iteration, the oldest time window envelope is discarded and the most recent M window envelopes are retained. Using this information, the variance of the measured envelopes over the past M windows can be calculated using the following equation.

$$\sigma_k^2 = \frac{1}{M-1} \sum_{m=1}^{M} (R_k^m - \bar{R}_k)^2 \tag{3}$$

where $\bar{R}_k$ is the mean of the R_k^m's, $\sum_m \frac{R_k^m}{M}$.

The variability of the aggregate envelope is measured over $T.M$ time slots to characterize the variation of the peak-rate envelope during a longer time scale. The variance of the measured envelope is used to determine the confidence values of the schedulability condition.

The service envelope describes the minimum service received by a traffic class during a certain time period. The envelope is obtained by measuring the service received when the class is backlogged. The variability of the envelope can be used to quantify the confidence level of the predicted QoS values.

Let us consider a single class of traffic. Let the j^{th} packet arrival time be denoted by a_j and its departure time by d_j. The delay of this individual packet is $d_j - a_j$. The envelope describes the service received by the flow over the long intervals in which the class is backlogged. A flow is said to be backlogged if it has at least one packet to be transmitted (Figure 2). A traffic flow is continuously backlogged for k packet transmissions in the interval $[a_j, d_{j+k-1}]$ if

$$d_{j+m} > a_{j+m+1} \quad \text{for all } 0 \leq m \leq k-2 \tag{4}$$

for $k \geq 2$. It is important to note that all packet transmissions are backlogged for $k = 1$ in the interval $[a_j, d_j]$.

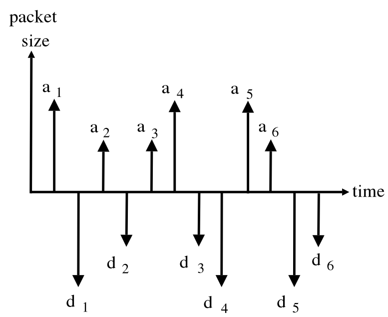

Figure 2: Example flow for service envelope computation.

Figure 2 illustrates an arrival departure sequence. After the departure of the first packet, the second packet arrives. The backlogging condition $k = 1$ holds for the first packet as well as for the second packet. The flow is backlogged for two consecutive packets since the fourth packet arrives into the system before the third packet departs.

The minimum service envelope mean and variance over the interval $[t - T\tau, t]$ at time t is measured as follows: the envelope is expressed as a vector of times $\overrightarrow{U}$ such that U_i is the maximum time required to service $i.L$ bits, where L is the number of bits in the minimum sized packet. Initially, $\overrightarrow{U} = 0$ and iteratively the final service envelope is computed considering all packets $1 \leq j \leq n$ in the window.

For the j^{th} packet, not only is the delay considered, but also longer backlogging intervals. Thus, the envelope is updated as

$$U_i = max(Ui, d_{j+k-1} - a_j) \tag{5}$$

where

$$i = \sum_{m=0}^{k-1} l_{j+m} \tag{6}$$

with l_{j+m} being the size of packet $j+m$ expressed in units L. For each packet, all $k \geq 1$ satisfying Inequality 4 are iteratively taken into account.

In Figure 2, two iterations are performed for packet 5, since two backlogging times need to be considered. For $k = 1$, $U_1 = max(U_1, d_5 - a_5)$ and for $k = 2$, $U_5 = max(U_5, d_6 - a_5)$, where the subscript of U_5 represents the combined sizes of packets 5 and 6.

The mean and variance of the service envelope is computed over successive windows, thus facilitating the computation of the confidence level of service provisioning.

Let us consider a system in which a traffic class has a measured maximum arrival envelope with mean $\bar{R}(t)$ and variance $\sigma^2(t)$. Moreover, let $\bar{S}(t)$ and $\psi^2(t)$ be the mean and variance, respectively, of the class measured minimum service envelope . The new flow with peak rate P is admissible with delay bound D if

$$t\bar{R}(t) + Pt - \bar{S}(t + D) + \alpha\sqrt{t^2\sigma^2(t) + \psi^2(t + D)} < 0 \tag{7}$$

where α is set according to the required violation probability [2]. Moreover, the following stability condition must be ensured.

$$\lim_{t\to\infty} R(t) < \frac{\bar{S}(t + D)}{t}. \tag{8}$$

If both requirements are satisfied, a message to admit the new flow is relayed by the admission agent and the user can start transmitting.

3 MBAC Schemes in 802.11e Networks

This section provides an explanation of how to adapt previous measurement-based admission control schemes to 802.11e networks.

3.1 Adaptation of the MS Scheme

In 802.11 networks, the Medium Access Control (MAC) protocol determines efficiency of sharing the limited communication bandwidth of the wireless channel [4]. The available capacity is influenced by several parameters, including the inter-frame duration, basic rate and data rate used by the QoS basic service set (QBSS), frame length and the backoff procedure. To consider the impact of these parameters, Inequality 1 should be replaced by:

$$v + v_f < B * \alpha * \mu \tag{9}$$

where B is the maximum data rate used by the QBSSs, and α is an estimation of the channel efficiency. α is computed by the ratio t_{MSDU}/t_m, where t_{MSDU} and t_m are the frame transmission time, without considering backoff procedure, and including it, respectively. t_{MSDU} is obtained by dividing the value of the MAC service data unit by the maximum transmission rate (B). The transmission time including backoff procedure is estimated by the following equation:

$$t_m = AIFS + t_{data} + SIFS + t_{ACK} + t_{backoff}(m) \tag{10}$$

where m is the number of active stations, and *AIFS* and *SIFS* are the inter-frame values defined by the standard, t_{data} and t_{ACK} are the transmission times for the data and the ACK frames, respectively; $t_{backoff}(m)$ is the average time consumed by the backoff procedure.

3.2 Adaptation of the Traffic Envelope Scheme

In 802.11e networks, the QAP is the only egress node, whereas wireless stations serve as ingress nodes for the Traffic Envelope (TE) scheme[1]. When a user wishes to initiate a new session, the wireless station sends a signal message to verify whether transmission is possible or not. The request handler, at the QAP then calls the admission control routine to determine whether this new flow can be admitted without interrupting the QoS of the already included flows. The network state is continually monitored at the egress node.

4 Simulation Experiments

Discrete event simulation experiments were carried out using the Network Simulator tool (*ns*) [8] to assess the effectiveness of the two schemes.

Each simulation experiment involved a random process of flow arrivals, with the flow admitted or rejected according to the decision of the algorithm. Rejected flows did not supply for a service request. Each flow that was accepted sent data packets throughout its randomly determined lifetime.

The topology of the simulated network consisted of a QAP wire-attached to a fixed node through a 100 Mbps link with a 2 ms delay. The QAP was located in the center of a 350 x 350 meter area, with the QSTAs uniformly distributed around it. The data rate in the wireless link was 11 Mbps, with a basic rate of 1Mbps.

Packets were generated according to voice, video or data source models. The voice model is an exponential "on/off" with a mean of 1.2 s and 1.8 s for "on" and "off" periods, respectively. The transmission rate during "on" periods was 64 Kbps, with a packet size of 256 bytes [12].

To simulate the conversational pattern for each voice connection, an exponential "on/off" flow from the wireless node to the fixed node (uplink) and other from the fixed node to the wireless node (downlink) were used. They were initiated with at most one second of difference.

MPEG encoded video traces of ARD News television programs [11] were used. The packets were 512 bytes long and were generated with an average rate of 720 Kbps and a peak rate of 3.4 Mbps.

Data sources generated packets according to an "on/off" model with Pareto distribution with the following parameters: packet size of 1024 bytes, mean burst duration of 250 ms, mean idle duration of 250 ms, and peak rate of 400 Kbps. The Pareto shape parameter was 1.9. This traffic generator produces highly bursty traffic which, when aggregated, forms a flow that has long range dependencies [17].

The delay requirements for voice, video and data traffic were 100 ms, 100 ms and 500 ms, respectively. The direction of each flow (uplink or downlink) was chosen randomly and the lifetime of the flows was distributed exponentially with a mean of 180 seconds for video traffic and of 300 seconds for voice and data traffic.

[1] When the traffic direction is from QAP to QSTAs, the QAP also functions as an ingress node.

UDP transport protocol with packet size equal to that used by the source to avoid fragmentation was used. For simplicity, no signaling protocol was considered, since in the simulation script it is possible to initiate a flow once it is accepted by the admission control. Nonetheless, such a protocol is necessary for a real implementation.

Both of the MBAC schemes require parameters which control the history retained by the estimation algorithm. Whenever possible, the parameter settings suggested in the original references were used [2, 15], although in some cases different values yielded an enhanced performance. One possible explanation for such differences would be the source models used.

The most relevant of these parameters is the measurement window size (T), which determines the lifetime of an estimate to be used for the admission of new flows. The miscalculation of the proper measurement window size can lead to resource underutilization [15]. Although large windows in the MS scheme result in a low number of delay violations, they lead to channel underutilization. The "optimal" value for T which would allow the fulfillment of QoS requirements for all three types of traffic, as well as resulting in high channel utilization was sought in the interval [1s, 500s] . Results of extensive simulations indicated that 400s would be the most appropriate measurement window size.

In the TE scheme, if the value of T is too small, the rate variation (σ_T) is too much, and the stability condition in Equation 8 may not be satisfied. If it is too large, the maximum measured rate ($\bar{R_k}$) tends to increase, since it will be chosen from a wider range of values, thus leading to a greater number of rejections (Equation 7) [7].

In this scheme, the "optimal" value for T was sought in the interval [0,005s, 5s] for the three types of traffic used in the simulation experiments. Values close to the lower bound of the interval led to unstable systems whereas those close to the upper bound resulted in low utilization. The results indicated that, independent of the traffic type, a value of 0.05 s led to the best performance.

5 Numerical Results

For each simulation experiment, the maximum number of flows accepted, the mean delay, the aggregated goodput and the blocking probability as a function of flow arrival rate was collected; the data from an initial warmup period were discarded. All experiments were repeated ten times using different seeds for the random number generator. The duration of each simulation was 600 seconds for experiments with a single media traffic and 3600 seconds for those with multimedia traffic.

5.1 Monomedia Traffic

The experiments with single media traffic used the following access parameters: AIFS = 2 and $CW_{min} = 7$. For the Traffic Envelopes MBAC, the value of α was 1.

Both MBAC schemes were compared to the contention access scheme of the 802.11e standard. Traffic Envelope MBAC, Measured Sum MBAC and the 802.11e standard are denoted by TE, MS and EDCA, respectively.

5.1.1 Voice Traffic

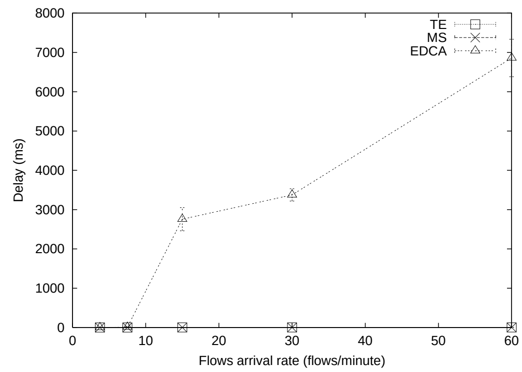

Figure 3: Mean delay for voice flows produced by TE, MS and EDCA.

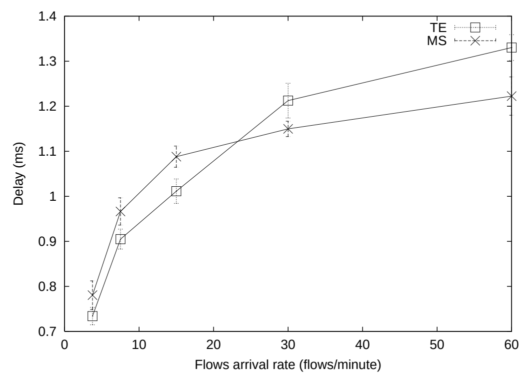

Figure 4: Mean delay for voice flows produced by TE and MS.

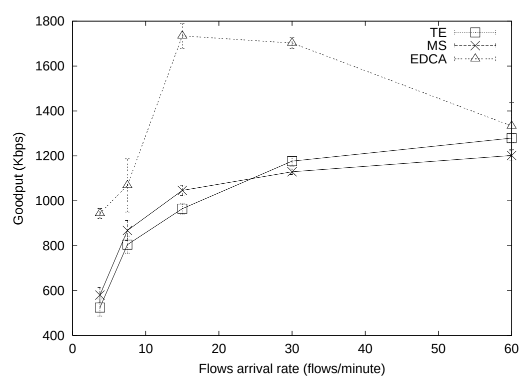

Figure 5: Goodput for voice flows produced by TE, MS and EDCA.

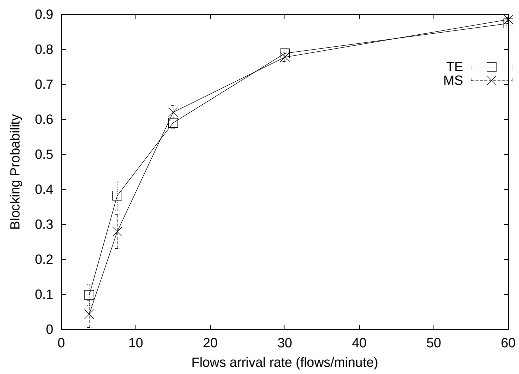

Figure 6: Blocking Probability for voice flows produced by TE and MS.

flows/ minute	Measured/Sum		Traffic Envelopes	
	uplink	downlink	uplink	downlink
3	1.96	1.85	1.82	1.68
7	2.58	2.57	2.41	2.31
15	3.01	3.04	2.89	2.91
30	3.24	3.38	3.21	3.41
60	3.36	3.63	3.58	4.29

Table 1: 95th Percentile delay for voice flows produced by TE and MS (ms).

Both the TE and MS schemes resulted in similar behaviour for voice traffic. Figure 3 shows the mean delay for the three schemes while Figure 4 shows the results of the two admission control schemes as a function of the mean arrival rate of flows.

Figure 3 illustrates the benefits of adopting admission control in 802.11e networks. Without admission control (EDCA), the required Quality of Service was maintained only for low mean arrival rates of new flows (less than 10 flows/minute). Moreover, the mean delay increases sharply as the arrival rate increases. Conversely, Figure 4 shows that under both MBAC schemes, it is possible to guarantee the established delay requirements. Actually, the mean delay is two orders of magnitude lower than the delay without admission control. Moreover, the mean delay does not undergo the sharp increase that it does when there is no admission control. Table 1 shows the 95th percentile delay of uplink and downlink voice traffic. This table indicates that there is no significant difference between the delays produced by the two MBAC schemes.

Figure 5 shows the mean goodput. When EDCA is used, high goodput values are obtained. However, they decrease as the flow arrival rate surpasses 15 flows/minute. The reason for this decrease is the high rate of collision, since in the absence of admission control, the number of flows in the network is high. For flow arrival rates of 15, 30 and 60 flows/minute, the maximum number of flows in the network when EDCA was used was 75, 140 and 230 respectively, whereas for the TE and MS schemes, the flows admitted were limited to from 25 to 29 flows. The high goodput under EDCA is obtained at cost of no QoS provisioning. Figure 6 illustrates the similarity of the blocking probability produced by the two MBACs schemes.

5.1.2 Video Traffic

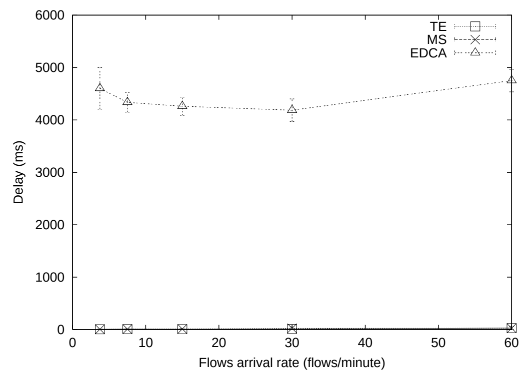

Figure 7: Mean delay for video flows produced by TE, MS and EDCA.

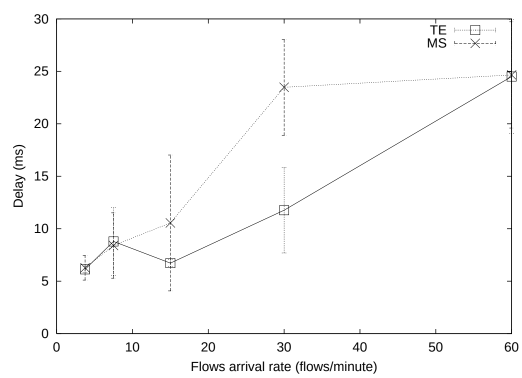

Figure 8: Mean delay for video flows produced by TE and MS.

Similarly to voice traffic Figures 7 and 8 show the mean delay as a function of the arrival rate for the three schemes and for the two MBAC schemes, respectively. As was the case for voice traffic, EDCA is not able to provide delay guarantees. Although the mean delay values were higher than 4000 ms for EDCA (Figure 7), under both MBAC schemes the delay values were less than 35 ms, as illustrated in Figure 8. The mean delay values given by TE are lower than those produced by MS. Moreover, TE produces higher goodput values than does MS, as shown in Figure 9. As in the experiments with voice traffic, when EDCA is used, the goodput decreases as the flow arrival rate increases due to the large number of flows allowed into the network (see Figure 10), with a consequent increase in the number of collisions. Although EDCA produces higher goodput than the other two MBAC schemes, the larger number of video sources in the cell considerably increases the delay.

The blocking probability is high for both schemes as this is necessary to guarantee QoS to all admitted users. As can be observed in Figure 11, TE provides the lowest

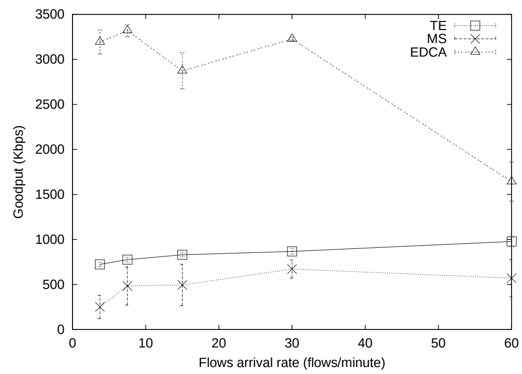

Figure 9: Goodput for video flows produced by TE, MS and EDCA.

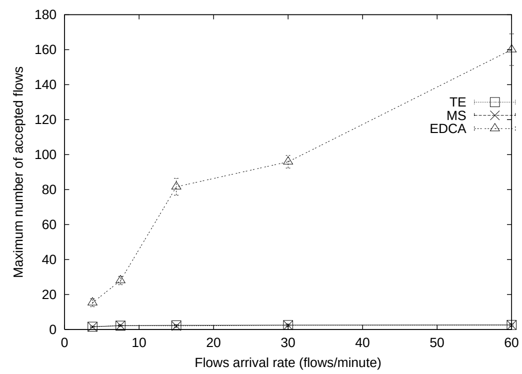

Figure 10: Maximum number of accepted video flows produced by TE, MS and EDCA.

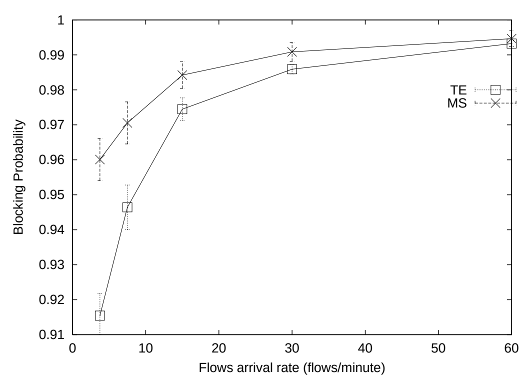

Figure 11: Blocking probability for video flows produced by TE and MS.

values for blocking probability, especially for low loads. The MS scheme produces blocking probability values greater than 0.95 even for extremely low arrival rates, which suggests that the MS underutilizes the available bandwidth due to its pessimistic admission decision making policies.

5.1.3 Data Traffic

The absence of admission control results in delay values much higher than the pre-defined requirement of 500 ms (Figure 12). When an admission control mechanism is used, however, the delay is less than 2 ms (Figure 13). The TE scheme produces a longer mean delay than does the MS scheme. This difference is due to the fact that the TE scheme accepts a larger number of users resulting in the consequent increase in delay with the arrival rate increases. Even so, the delay values are significantly lower than the required value.

As shown in Figure 14, TE provides higher goodput values than does MS, with the differences reaching the level of 1Mbps. Figure 15 plots the blocking probability. Note that the TE algorithm is less conservative than that of MS.

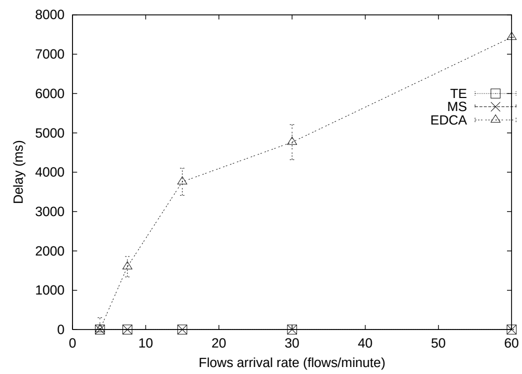

Figure 12: Mean delay for data flows produced by TE, MS and EDCA.

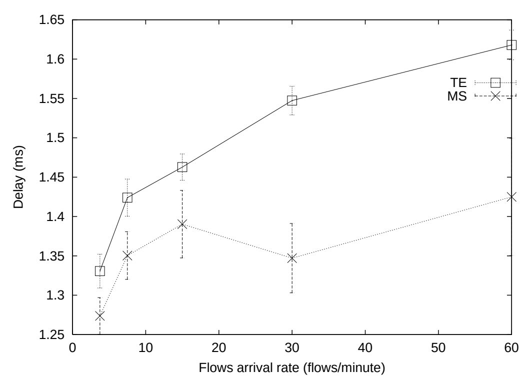

Figure 13: Mean delay for data flows produced by TE and MS.

In summary, results with monomedia traffic indicate the need to adopt admission control to provide an adequate QoS as well as the adequacy of MBAC approaches for this purpose. The TE scheme provides higher goodput and lower blocking probability values than does MS. Although it results in longer delays these results can be understood to reflect the fact that the TE scheme considers QoS requirements in traffic characterization and decision making. Such requirements are only indirectly considered by the MS scheme by

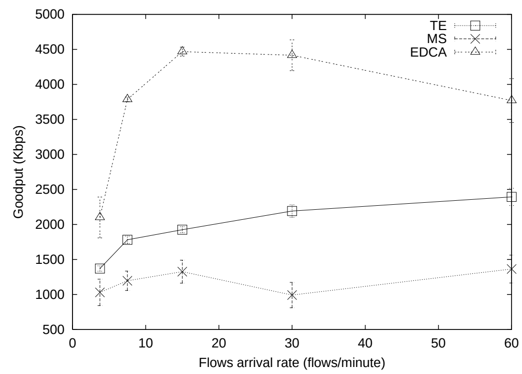

Figure 14: Goodput for data flows produced by TE, MS and EDCA.

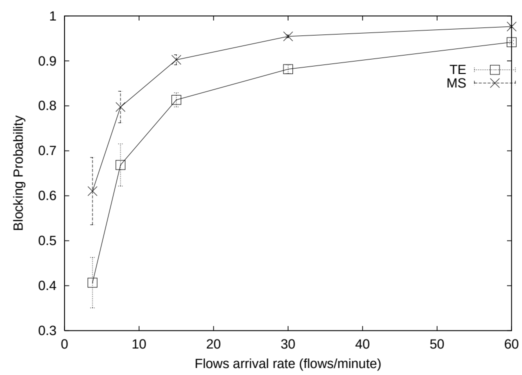

Figure 15: Blocking Probability for data flows produced by TE and MS.

considering the utilization level. The TE scheme can thus obtain higher utilization gains and, consequently, higher goodput and lower blocking probability values. Moreover, the MS scheme adopts stricter acceptance criterion than does TE [7].

5.2 Multimedia Traffic

To evaluate the MBAC approaches, simulation experiments considering multimedia traffic were also conducted. This section presents the experiments with the TE scheme only, since TE scheme produces more attractive results. In each simulation experiment, two media are considered. The load of one media is fixed and that of the other is varied so that the impact on performance of the media with the fixed load can be assessed. Different priority levels are assigned to each media; for the high priority class, the parameters values are AIFS = 2, $CW_{min} = 7$ and $\alpha = 1$; for the low priority class, the parameters are AIFS = 3, CW_{min} = 15 and $\alpha = 3$.

To evaluate the influence of video and data traffic on the service offered for voice traffic, the voice flow arrival rate was fixed at 7 flows/minute, the video and data flow arrival rates were varied between 3 and 60 flows/minute. Voice traffic was given high priority. Two scenarios were evaluated: voice and video traffic and voice and data traffic.

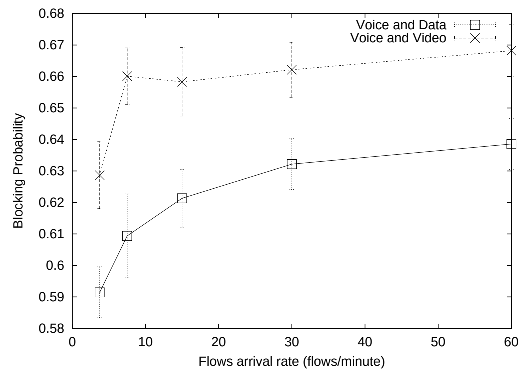

Figure 16: Blocking probability for voice traffic when its load is fixed and video/data load is varied.

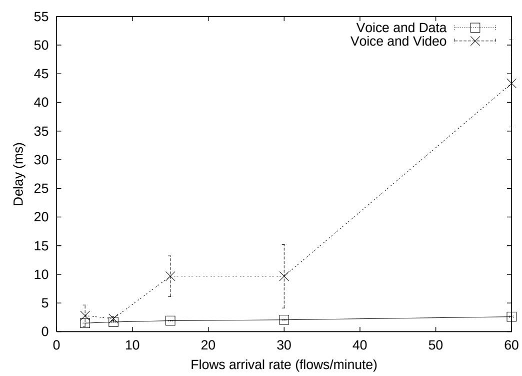

Figure 17: Mean delay for voice traffic when its load is fixed and video/data load is varied.

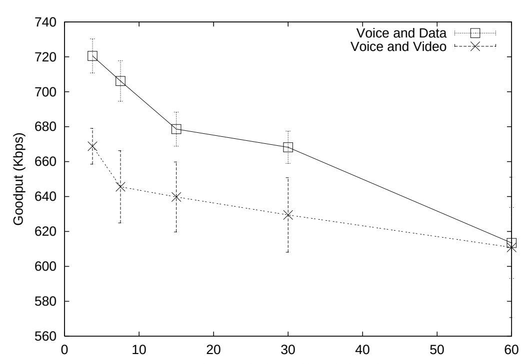

Figure 18: Goodput for voice traffic when its load is fixed and video/data load is varied.

video flows/minute	uplink	downlink
3	4.13	10.81
7	4.23	11.61
15	4.30	12.38
30	4.60	16.05
60	5.19	69.24

Table 2: 95th Percentile delay for voice traffic when its load is fixed and the video load is varied (ms).

Figure 16 plots the blocking probability of voice traffic as a function of the arrival rate of both data and video flows. This shows that the blocking probability is almost constant regardless of load increase. No change in order of magnitude was observed. Actually, at high loads an equilibrium point for blocking probability was achieved in both scenarios.

TE is able to provide the 100 ms delay required for voice traffic even when the arrival rate is high (Figure 17). Although the mean delay of voice traffic increases sharply under high video traffic loads, it is maintained at levels well below what is required. When the delay is analysed separately in uplink and downlink directions (see Table 2), it becomes evident that the increase observed is due to the increase in downlink traffic delay because the QAP is overloaded.

The goodput of voice traffic decreases with the load increase (Figure 18). However this decrease (about 20 Kbps) is not significant at all given the high access priority of voice traffic.

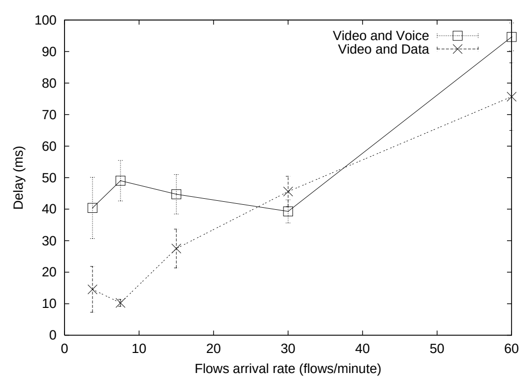

Figure 19: Mean delay for video traffic when its load is fixed and voice/data load is varied.

In the second experiment, the arrival rate for video flows is fixed at 7 flows/minute. Two scenarios were evaluated: video and voice traffic, with low priority video traffic, and video and data traffic, with high priority video traffic.

In both scenarios, the blocking probability, the maximum number of flows admitted and the goodput results are similar to those obtained in the monomedia experiment with video traffic only. It is therefore possible to conclude that the TE scheme is able to exploit the excess capacity available and admit additional voice/data flows, while maintaining the same number of video flows as in the experiments with video only.

Although the delay of video flows is affected by the presence of other media traffic, this does not surpass the 100 ms delay bound, as can be seen in Figure 19.

In the final set of experiments, data traffic had a fixed flow arrival rate (7 flows/minute) and low priority. The scenarios considered were: data and voice traffic and data and video traffic.

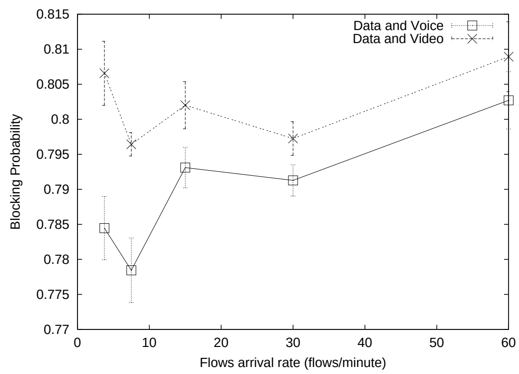

Figure 20: Blocking probability for data traffic when its load is fixed and voice/video load is varied.

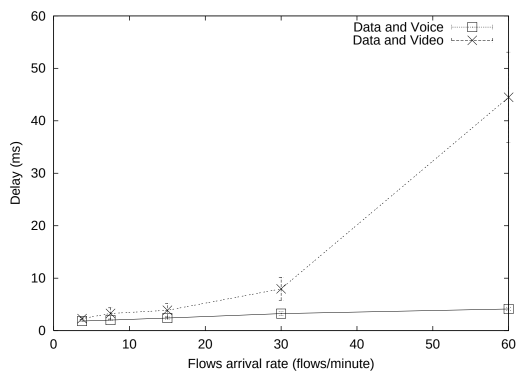

Figure 21: Mean delay for data traffic when its load is fixed and voice/video load is varied.

Figure 20 shows that the blocking probability of data traffic undergoes only slight variations (about 2%) as the voice and video traffic load increases. Furthermore, the delay is almost constant in the scenario with voice traffic, whereas in the scenario with video traffic it increases with an increase in load. However, as can be seen in Figure 21, the 500 ms delay requirement is met in both scenarios. Although the goodput of data traffic decreases when the load of high priority traffic increases (Figure 22), there is no starvation of the data traffic.

In summary, the results obtained with multimedia experiments show that the TE scheme is able to provide the QoS requirements for a traffic class, regardless of the behavior of

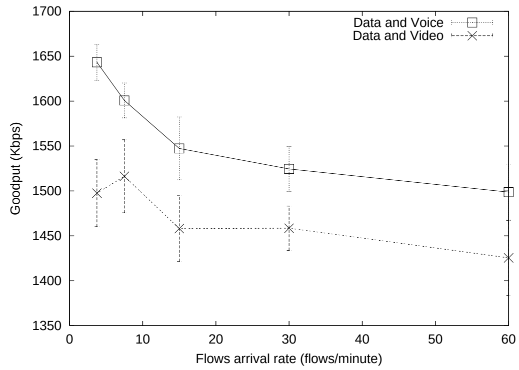

Figure 22: Goodput for data traffic when its load is fixed and voice/video load is varied.

other existing traffic classes. This is due to the ability of the algorithm to exploit the gains obtained from inter-class resource sharing [2].

6 Need for mechanism to cope with delay asymmetry

In 802.11e networks, the QAP is responsible for delivering traffic in both directions, thus competing with several QSTAs. The probability of accessing the medium is the same for the QAP and QSTAs in the same service class, thus leading to a more rapid increase in the QAP queue than in those at the QSTAs, resulting in long delays for downlink traffic.

Figure 23 illustrates the delay asymmetry phenomenon in infrastructured 802.11 networks[2]. This figure shows that the downlink delay is the major factor in the mean delay. Moreover, while the downlink component increases sharply with the increase of flow arrival rate, the uplink delay remains at levels lower than what is required. Table 3 shows the 95th percentile delay for the two traffic directions. Notice that 95th percentile for the downlink is three to four orders of magnitude greater than that for the uplink.

In summary, the downlink delay is the major factor responsible for network delay. Ameliorating the QAP bottleneck, will make it possible to obtain lower delay levels.

Recently, efforts have been made to deal with the delay asymmetry problem. In [1], it is shown that the asymmetry decreases when HCCA is used in conjunction with EDCA. Based on these results, it was suggested that the asymmetry problem be solved by increasing the amount of HCF contention-free transmission as much as possible. However, in [14] it was demonstrated that the HCCA scheduling algorithm is only efficient for flows with strict Constant Bit Rate (CBR) characteristics.

Another way of coping with the delay asymmetry problem is to adopt an additional mechanism to balance the utilization of the network between the QAP and the QSTAs.

Simulation results presented by Kong *et al* [16] show that if only a small number of stations are contending for the medium, a low CW_{min} value allows a prioritized access to the medium for these stations. Otherwise, it is necessary to adopt an adaptive scheme

[2]The scenario used in this experiment is presented in Section 8.

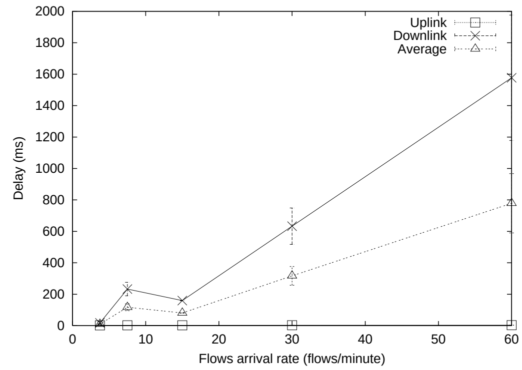

Figure 23: Delay of voice flows

flows/minute	uplink	downlink
3	5.08	66.40
7	6.34	654.93
15	7.79	3263.87
30	8.81	4611.89
60	10.26	8808.79

Table 3: 95th Percentile delay for voice traffic (ms)

to adjust the CW_{min} values according to the number of stations and the current network load [9]. It is shown that, although the differentiation provided by the AIFS can be very effective, it may reveal negative effects when the network is heavily loaded, and may result in the starvation of the lower priority traffic [16]. Moreover, in [3] it was demonstrated that the correct setting of the TXOP parameter not only increases WLAN capacity but also provides temporal fairness.

The following section introduces a novel mechanism for introducing prioritized service by adjusting medium access parameters, thus solving the delay asymmetry problem and maintaining balanced delay values.

7 Dynamic Tuning of Medium Access Parameters

This section presents a mechanism for coping with delay asymmetry in 802.11e networks by dynamically adjusting the TXOP and the CW_{min} values. The TXOP parameter controls the duration of the period a station has the right to access the medium after a successful contention. The station has the right to send multiple frames as long as the total transmission time does not exceed the maximum TXOP duration. Thus, the adjustment of this parameter can improve the throughput.

The main cause for delay asymmetry is related to the fact that the QAP is responsible for delivering all the downlink traffic. Thus, at the QAP, classes must be assigned throughput proportional to the number of flows they furnish. With the mechanism introduced here, the QAP throughput is adjusted as follows: if the TXOP value of the i^{th} class allows the sending of q frames and if this value is sufficient for a QSTA to provide the required QoS, then, the TXOP value must be adjusted to send $k * q$ frames, where k is the number of flows of the i^{th} class at the QAP. In this way, downlink throughput is similar to uplink throughput.

Moreover, the TXOP value must be adjusted according to the throughput demands of the specific class, rather than the priority of that class, since class priorities are determined by delay requirements rather than throughput demands.

Considering these arguments, the mechanism proposed in this section adjusts the TXOP parameter as follows: the TXOP value for the class with the lowest packet arrival rate is set to zero, so that this class can send only a single frame when accessing the channel. For the other classes, the TXOP value is set proportionally to the ratio between the number of packets that have arrived in their queues and the number of packets that have arrived in the queue of the class with the lowest packet arrival rate. The mechanism adjusts the TXOP value at the QAP according to the number of downlink flows handled by each class, which can be provided by the admission control mechanism. CW_{min} is adjusted by considering the number of stations with active flows in each class to reduce the occurrence of collision. Class differentiation is provided by both CW_{min} and AIFS parameters.

For each execution, the mechanism checks whether the CW_{min} values are appropriate. After that, it performs the adjustment of the TXOP value in conformity with the load imposed by the applications. After both operations, the QAP announces a new set of QoS parameters.

ALGORITHM *adjustParameters* /* n is the number of classes */
1. for each class i, i from 0 to $(n-1)$
2. adjustContention(i)
3. adjustClassesTXOP( )
4. for each class i, i from 0 to $(n-1)$
5. $TXOP_{AP}(i) = TXOP_{AP}(i) * num\_flowsAP(i)$
6. sendQoSParameterSet( )

Figure 24: Main module for dynamic adjustment of per class access parameter values

Figure 24 shows the main module of the tuning mechanism. This module calls the adjustContention procedure (Figure 25) for each class, beginning with that with the highest priority level. This procedure checks whether the size of minimum contention windows should be increased in order to reduce the number of transmission attempts. To perform this check, the procedure compares $CW_{min}(i)$ with the number of QSTAs with accepted i^{th} class flows. It also checks whether $CW_{min}(i)$ can be reduced, by comparing this value to the number of QSTAs, which is identified by subtracting the number of downlink flows from the number of accepted flows in each class. It is assumed that the QAP has information about the number of downlink flows, and that each QSTA has only one active flow.

If the CW_{min} value of the i^{th} class is to be adjusted, the adjustment is made from class

ALGORITHM *adjustContention(i)*

if $stations(i) > CW_{min}(i)$
  for each class j, j from i to $(n-1)$
    $CW_{min}(i) = CW_{min}(i) * 2 + 1$
if $stations(i) < CW_{min}(i)/2$
  for each class j, j from i to $(n-1)$
    $CW_{min}(j) = (CW_{min}(j) - 1)/2$
for each class j, j from i to $(n-1)$
  if $CW_{min}(j) < CWMinimum(j)$
    $CW_{min}(j) = CWMinimum(j)$

Figure 25: Adjustment of minimum contention window size

ALGORITHM *adjustClassesTXOP*

$minimum = \infty$
for each class i, i from 0 to $(n-1)$
  $numPackets(i) = \lfloor RealLoad(i)/PacketsSize(i) \rfloor$
  if $numPackets(i) < minimum$
    $minimum = numPackets(i)$
for each class i, i from 0 to $(n-1)$
  if $(numPackets(i) = minimum)$
    $TXOP_{AP}(i) = 0$
    $TXOP_{STA}(i) = 0$
  else
    $TXOP_{AP}(i) = TxTime(i) * \lfloor numPackets(i)/minimum \rfloor$
    $TXOP_{STA}(i) = TxTime(i) * \lfloor numPackets(i)/minimum \rfloor$

Figure 26: Adjustment of the TXOP value

i to class $(n-1)$. In this way, differentiation based on CW_{min} values is maintained. Finally, the procedure checks whether or not the CW_{min} values are below a given minimum. If this happens, the CW_{min} value is set to the minimum value allowed.

After this, the main module calls the adjustClassesTXOP procedure (Figure 26). This procedure divides the real load per class by the average packet size in order to identify which class has received the lowest number of packets in the most recent monitoring interval. The class with the minimum number of packets received has its TXOP value set to zero; and all the other classes have their TXOP values set to the product of packet transmission time and the ratio between the number of packets received by that class and the minimum number of packets received. Information about both the real load per class and the packet size can be obtained by monitoring the transmission queues of all classes at the QAP.

The main module adjusts the TXOP value for each class at the QAP. The TXOP is set as the product of the value assigned by the adjustClassesTXOP procedure and the number of downlink flows accepted for the class. Finally, the access point informs the stations of the new QoS parameter set.

The mechanism for dynamically adjusting the TXOP and the CW_{min} values is periodically executed according to a pre-defined interval. Given that the 802.11e standard does not determine an interval for the QAP to announce the medium access parameters, the "ideal" interval was searched via extensive simulation experiments.

In the search for a solution to the delay asymmetry problem an algorithm involving the adjustment of AIFS and CW_{min} values was also pursued. However, this alternate algorithm is not effective. In the alternate algorithm to balance the delay, each type of traffic was mapped into different classes both at the access point and at the stations. The QAP had higher access priorities than do the QSTAs access priorities. For example, voice traffic carried by class 0 in a 802.11e network is carried by class 0 at the access point, but by class 1 at the stations. Class 0 has lower AIFS and CW_{min} values than class 1. The parameters provide class differentiation and their values are adjusted to give differentiated throughput to the classes.

Simulation experiments realized with this mechanism indicated that giving higher priority for the access point than for the stations is not effective. For small differences in AIFS and CW_{min} values, non-significant differences in throughput are obtained. For large differences, the delay of low priority class degrades considerably. Thus, it is not possible to use AIFS and CW_{min} for both class differentiation and for tuning per class throughput.

8 Effectiveness of the tuning Mechanism

The results presented in this section evaluate the proposed tuning mechanism. The simulated scenario[3] consisted of an infrastructured 802.11e network with real time (voice and video) and best effort (data) traffic. Flows were admitted or rejected according to the decisions of the TE admission control algorithm.

Figures 27 and 28 plot, respectively, the voice traffic and the video traffic delay, and Table 4 presents the 95th percentile delay for voice traffic. Note that the mechanism for dynamically adjusting TXOP and CW_{min} values was able to provide similar delays for uplink and downlink traffic regardless of the network load conditions. The 95th percentile of both uplink and downlink traffic is now of the same order of magnitude. Moreover, the recommended delay value of 100 ms is now achieved.

flows/minute	uplink	downlink
3	7.37	5.78
7	8.80	6.50
15	12.55	8.96
30	16.75	11.83
60	27.46	17.94

Table 4: 95th Percentile delay for voice traffic (ms)

[3]The scenario used in this experiment is presented in Section 8.

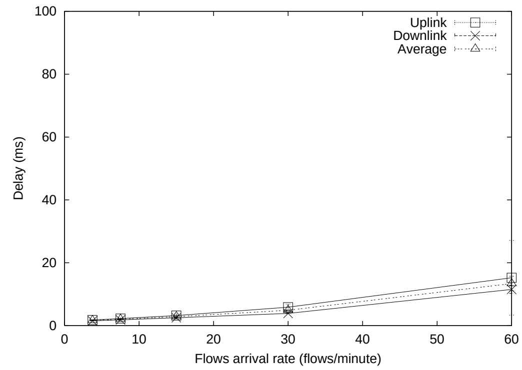

Figure 27: Delay of voice flows

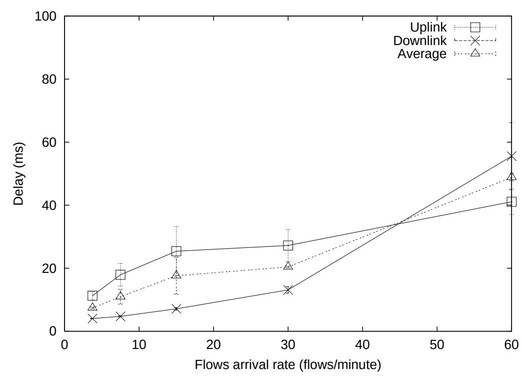

Figure 28: Delay of video flows

This example illustrates that using the proposed mechanism makes it possible to maintain the QoS required by real time applications. Furthermore, it is important to note that the delays are balanced without starving the best effort traffic, and without decreasing network utilization.

Figures 29 and 30 show, respectively, the goodput for both voice traffic and data traffic achieved by using the admission control mechanism both with and without the tuning mechanism. For voice traffic, the goodput is higher whereas it is lower for data traffic when the mechanism is used than when it is not. Overall, the tuning mechanism does not penalize the network aggregated goodput. In fact, the resources are allocated according to the classes needs in order to provide the required delay.

In summary, by jointly using an admission control mechanism and a mechanism for dynamic tuning of MAC parameters, it is possible to provide QoS for both uplink and downlink traffic, even under high load. Furthermore, the parameters can be adjusted so that the AIFS provides class differentiation, with the TXOP parameter balancing network utilization and the CW_{min} parameter reducing the rate collision.

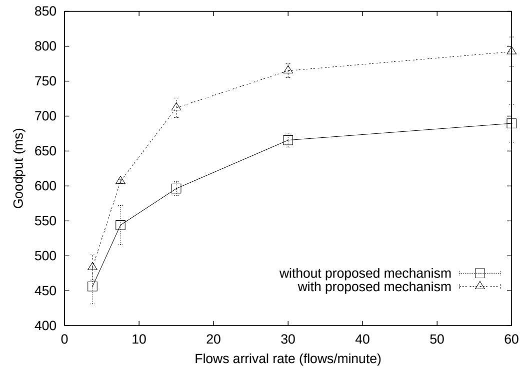

Figure 29: Goodput of voice flows

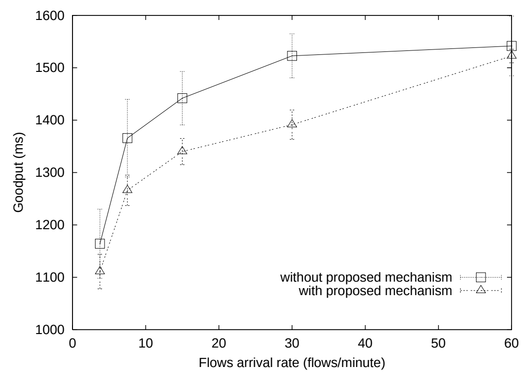

Figure 30: Goodput of data flows

9 Conclusions

In this paper, the need for admission control in 802.11e networks was assessed. Two measurement-based admission control algorithms for IEEE 802.11e networks were compared. The algorithms were evaluated according to four criteria: maximum number of admitted flows, blocking probability, mean delay and goodput. Both algorithms effectively provide the requested service for all types of traffic which encourages the use of MBAC schemes in 802.11e networks, specially for its efficiency and realistic assessment of network conditions. Of the two types tested the traffic envelope MBAC scheme produced a better performance than did the measured sum MBAC scheme.

This paper also introduced a novel mechanism for dynamically tuning the values of TXOP and CW_{min} so that QoS can be maintained and the asymmetry between downlink and uplink delay ameliorated. Simulation results indicate that when this mechanism is used in conjunction with an admission control mechanism comparable delay values are obtained by both downlink and uplink flows, thus improving the service provided for delay sensitive applications.

References

[1] A. Grilo and M. Nunes, *Performance evaluation of IEEE 802.11e*. In Proceedings of PIMRC, pp. 511-517, 2002.

[2] C. Cetinkaya, V. Kanodia, and E. Knightly, *Scalable services via egress admission control*. IEEE Transactions on Multimedia, vol. 3, pp. 69-81, 2001.

[3] D. Pong and T. Moors, *Fairness and Capacity Trade-off in IEEE 802.11 WLANs*. In Proceedings of IEEE Conference on Local Computer Networks (LCN), pp. 310-317, 2004.

[4] F. Cali, M. Conti, and E. Gregori, *Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit*. IEEE/ACM Transactions on Networking, vol. 8, pp. 785-799, 2000.

[5] Institute of Electrical and Electronics Engineers, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, 1999.

[6] Institute of Electrical and Electronics Engineers, *Medium Access Control (MAC) enhancements for quality of service (QoS)*. IEEE Draft for Standard 802.11e/D6, 2003.

[7] J. Qiu and E. Knightly, *Measurement-based admission control with aggregate traffic envelopes*. IEEE/ACM Transactions on Networking, vol. 9, pp. 199-210, 2001.

[8] K. Fall and K. Varadhan, *The NS Manual*. Technical Report, UC Berkeley, The VINT Project, 2002.

[9] L. Romdhani, Q. Ni, and T. Turletti, *AEDCF: enhanced service differentiation for IEEE 802.11 wireless ad-hoc networks*. In Proceedings of the IEEE Wireless Communications and Networking Conference, pp. 1373-1378, 2003.

[10] M. Grossglauser and D. Tse, *A framework for robust measurement-based admission control*. Computer Communications Review, vol. 27, pp. 237-248,1997.

[11] MPEG-4 and H.263 Video Traces for Network Performance Evaluation. URL: http://trace.eas.asu.edu, 2000.

[12] P. Brady, *A technique for investigating on-off patterns of speech*. Bell Systems Technical Journal, vol. 44, pp. 1-22, 1965.

[13] L. Breslau, S. Jamin, and S. Shenker, *Comments on the performance of measurement-based admission control algorithms*. In Proceedings of IEEE Infocom, pp. 1233-1242, 2000.

[14] P. Ansel, Q. Ni, and T. Turletti, *FHCF: an efficient scheduling scheme for IEEE 802.11e*. ACM/Kluwer Journal on Mobile Networks and Applications, in press.

[15] S. Jamin, P. B. Danzig, S. J. Shenker, and L. Zhang, *A measurement-based admission control algorithm for integrated services packet networks.* IEEE/ACM Transactions on Networking, vol. 5, pp. 56-70, 1997.

[16] Z. Kong, D. H. K. Tsang, B. Bensaou, and D. Gao, *Performance analysis of IEEE 802.11e contention-based channel access.* IEEE Journal on Selected Areas in Communications, vol. 22, pp.2095-2106, 2004.

[17] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, *Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level.* IEEE/ACM Transactions on Networking, vol. 5, pp. 71-86, 1997.