Unitat Tècnica de Gestió de Tercer Cicle
UNIVERSITAT POLITÈCNICA DE CATALUNYA

UNIVERSITAT POLITÈCNICA DE CATALUNYA

entel

Toward All-IP Networks
IP and Wireless Networks Convergence

Ph.D. Thesis by

Xavier Pérez Costa

Thesis Advisor

Sebastià Sallent Ribes

Universitat Politècnica de Catalunya (UPC)
Department of Telematics Engineering

Campus Nord, Edifici C3
c/ Jordi Girona, 1-3
08034 Barcelona, Spain

February 16, 2005

5.7.1 Contention Window Computation

The difficulty of DWFQ relies in determining the CW values that leads to the desired bandwidth distribution of Equation 5.18. The approach we have chosen for the calculation of the CW is a dynamic one.

In order to be able to properly adjust the CWs, we introduce a variable L_i, the *label*, defined as

$$L_i = \frac{r_i}{W_i} \tag{5.19}$$

where r_i is the estimated bandwidth experienced by flow i and W_i is its *weight*. The estimated throughput, r_i, is updated every time a new packet is transmitted:

$$r_i^{new} = (1 - e^{-t_i/K})\frac{l_i}{t_i} + e^{-t_i/K} r_i^{old} \tag{5.20}$$

where l_i and t_i are the length and inter-arrival time of the transmitted packet, and K is a constant.

With the above definition of the label L_i, the resource distribution expressed in Equation 5.18 can be achieved by imposing the condition that the label L_i should have the same value for all the flows:

$$L_i = L \quad \forall i \tag{5.21}$$

Note that the actual value of L can vary in time (depending on the number of flows for example).

Equation 5.21 is fulfilled by using the following algorithm: having calculated its own label L_i, each station includes its label in the header of the packets it sends. For each observed packet, if the L_i in the packet's header is smaller than the L_i of the station, the station increases its CW by a small amount, while in the opposite case it decreases its CW by a small amount. In this way, the L_i of all flows tend towards a common value, L.

The above explanation describes the basics of the algorithm. However, in the adjustment of the CW, there are additional aspects that have to be taken into account:

- We do not want the CW to increase above the values defined by the 802.11 standard; as argued above, for the backward compatibility reasons the basic service (with a *weight* equal to 1) uses the CWs defined in the 802.11 standard, and any higher *weight* should receive a "better than average" kind of treatment and therefore the values of the CW should be lower.

- If the low sending rate of the application is the reason for transmitting below the desired rate, then the CW should obviously not be decreased. This can be detected by the fact that in this situation the transmission queue is empty.
- CWs should not be allowed to decrease in such a way that they negatively influence the overall performance of the network. If the channel is detected to be below its optimum limit of throughput due to too small values for the CWs (i.e. overload), the CW should be increased. This aspect will be discussed in Section 5.7.2.

The above considerations lead to the algorithm of Equation 5.22. This algorithm computes a value p which is used to scale the CW values defined in 802.11. Note that, besides this scaling of the CW, the backoff time computation algorithm is left as defined in the 802.11 standard (i.e. the Contention Window is doubled after every unsuccessful transmission attempt for a given number of times).

$$\begin{array}{c}\text{For each observed packet:}\\ \texttt{if}\ (L_{own} > L_{rcv})\ \texttt{then}\ p = (1+\Delta_1)p\\ \texttt{else if}\ (queue\_empty)\ \texttt{then}\ p = (1+\Delta_1)p\\ \texttt{else}\ p = (1-\Delta_1)p\\ p = min\{p, 1\}\\ CW = p \cdot CW_{802.11}\end{array} \tag{5.22}$$

where L_{own} is the label L_i calculated by the station, L_{rcv} is the label of the observed packet, and Δ_1 is computed as follows:

$$\Delta_1 = k\left|\frac{L_{own} - L_{rcv}}{L_{own} + L_{rcv}}\right| \tag{5.23}$$

where k is a constant equal to 0.01.

5.7.2 Overload

So far we have not discussed one important issue which is the *overload*. In fact, due to the nature of our protocol and in particular due to the dynamic way of adjustment of the size of the CW, a mechanism for controlling the overload is necessary.

As we can see it in the algorithm of Equation 5.22, each station adjusts its CW only on the basis of its own requirements. Such "selfishness" can easily be disastrous, due to the following side effect of the small CWs. We have been arguing so far that, the smaller the CW for a given station, the larger the throughput received by this station. The other bad consequence of such a procedure is that the more stations have small CWs, the bigger the probability of a collision. One can easily see that, for a big number of stations with a high *weight*, this can lead to an absolute blockage of the channel. Once all of the stations start decreasing their CWs in order to get the desired relative throughput, the number of collisions will start increasing, leading to even smaller CWs, and as a consequence, continuous collisions. A solution to this problem is to extend the algorithm of Equation 5.22 with the following condition:

$$\begin{gathered}
\text{For each observed packet:} \\
\texttt{if } (overload) \texttt{ then } p=(1+\Delta_2)p \\
\texttt{else if } (L_{own} > L_{rcv}) \texttt{ then } p=(1+\Delta_1)p \\
\texttt{else if } (queue_empty) \texttt{ then } p=(1+\Delta_1)p \\
\texttt{else } p=(1-\Delta_1)p \\
p=min\{p,1\} \\
CW=p\cdot CW_{802.11}
\end{gathered} \tag{5.24}$$

where Δ_2 is a constant equal to 0.25.

Let us now explain how we actually detect *overload.* As we have mentioned before, a big number of stations trying to transmit with a high *weight*, i.e. decreasing its CWs, leads to an increase of the number of collisions. If we now provide each station with a collision counter[1], which determines how many collisions a packet experiences before it is successfully transmitted, we can write the following simple condition determining overload

$$\texttt{if } (av\_nr\_coll > c) \texttt{ then } overload = true \tag{5.25}$$

where c is a constant that has to be properly adjusted. If c is too low, flows with high *weights* will not be allowed to decrease their CWs sufficiently, and as a consequence they will not be able to achieve the desired bandwidth distribution. On

[1]Note that in 802.11 collisions can only be detected through the lack of the Ack. However, a missing Ack can also be caused by other reasons different than a collision. In the simulations section we study the impact into our algorithm of having missing Acks due to errors in the channel (see Section 5.7.4).

the other hand, if c is too large, the number of collisions in the channel will be very high and the overall performance will be harmed. This constant, therefore, represents a tradeoff between the level of accurateness of the bandwidth distribution and the efficiency (i.e. total throughput) of the channel. These tradeoff has been studied via simulation (see Section 5.7.4), and an optimum value for c has been chosen according to simulation results.

The average number of collisions, (av_nr_coll), in Equation 5.25 is calculated after each successful transmission in the following way

$$av\_nr\_coll = (1 - t) \cdot num\_coll + t \cdot av\_nr\_coll \tag{5.26}$$

where in order to smoothen its behavior, we use some sort of memory, taking into account the last calculated value of av_nr_coll. The constant t is a small number (in our case $t = 0.25$) playing the role of a smoothening factor.

5.7.3 Multiple Flows per Node

In our discussion of DWFQ we assumed that only one flow exists at each node. In general, it is possible that each node may maintain multiple flows locally. In this case we modify the DWFQ protocol as described below (see Figure 5.20).

- A node i transmitting n flows with weights $W_1, \ldots, W_n$ uses the label

  $$L_i = \frac{r_i}{\sum_{j=1}^{n} W_j} \tag{5.27}$$

  where r_i is the estimated aggregated bandwidth experienced by node. This gives to the node the total bandwidth necessary for all its flows.

- Node i uses a weighted fair queuing scheduler with weights $W_1, \ldots, W_n$ to choose the next packet to transmit. This distributes the total bandwidth of the node among its flows proportionally to their weights $W_1, \ldots, W_n$.

5.7.4 Performance Evaluation & Discussion

To test the performance of the architecture presented in this section, we simulated it on a network consisting of a number of wireless terminals in a 2 Mbps Wireless LAN communicating with a fixed node through the Access Point (AP). These

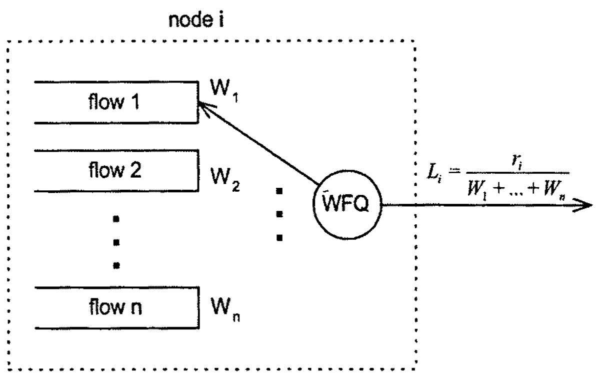

Figure 5.20: Multiple Flows per Node.

simulations were performed in ns-2 [42]. For this purpose, the algorithm of Equation 5.24 was inserted into the existing implementation of the 802.11 MAC DCF protocol in ns-2.

We chose to use the RTS/CTS mechanism in all cases. This mechanism, optional in the 802.11 standard, increases bandwidth efficiency in case of many collisions. Since our architecture may lead to larger number of collisions than the normal 802.11 MAC DCF, this mechanism can be especially beneficial in our case. Note, however, that the proposed DWFQ extension would also work without the RTS/CTS mechanism.

As already mentioned in Section 5.7, we assume that all nodes are sending just one flow, except for the downlink link in experiment 5.7.4.

Instantaneous Bandwidth Distribution

In DWFQ the desired bandwidth distribution is achieved by adjusting adaptively the CW of DWFQ stations according to the measured performance. Figure 5.21 shows this dynamic adjustment; the simulation corresponds to a scenario with a total number of 10 stations, 2 of them with a *weight* of 2 and the rest with a *weight* of 1. All stations are sending UDP CBR traffic with a packet size of 1000 bytes. It can be seen when comparing the instantaneous bandwidth of high priority and low priority stations that their ratio oscillates around the desired value.

Bandwidth Distribution as a function of the weight

With the proposed DWFQ extension, the throughput experienced by a station should be proportional to the *weight* assigned to its flow. Figure 5.22 shows the

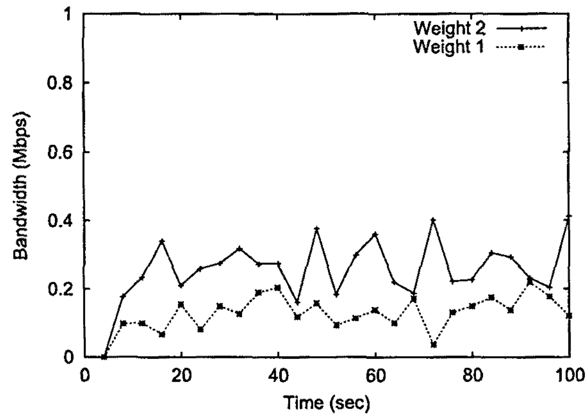

Figure 5.21: Instantaneous Bandwidth for Elastic traffic.

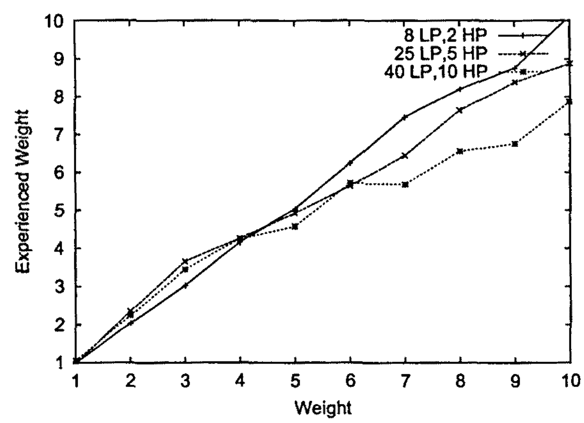

Figure 5.22: Bandwidth Distribution as a function of the *share*.

ratio between the throughput experienced by high priority (HP) and low priority (LP) stations (*experienced weight*) as a function of the *weight* assigned to the high priority stations when low priority stations have a *weight* equal to 1. Ideally, the resulting function should be the identity (i.e. a diagonal); that is, an *experienced weight* equal to the *weight*. In the figure it can be seen that the simulated results are quite close to the ideal. Only in the case of large weights and a large number of stations, the results obtained differ noticeably from the ideal case; however, not even in this case differences are too big (e.g. with 50 stations and a *weight* of 10, the *experienced weight* is 8).

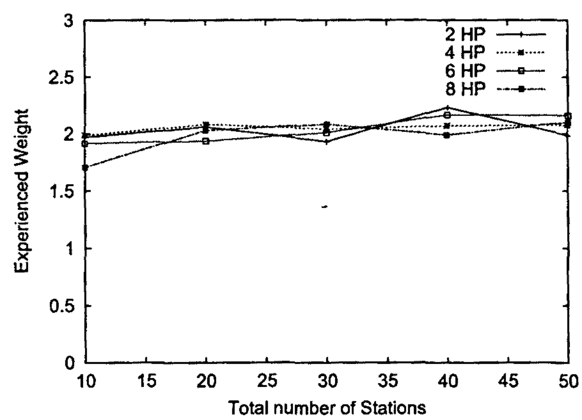

Figure 5.23: Bandwidth Distribution as a function of the number of stations.

Impact of the number of stations

The proposed algorithm for DWFQ relies on the experienced throughput estimated by each station. Note that the higher the number of stations, the lower the throughput received by each station. Since a low throughput is more difficult to estimate with exactitude than a high throughput, a large number of stations may negatively impact the performance of the algorithm. Figure 5.23 shows this impact when high priority stations have a *weight* of 2 and low priority ones have a *weight* of 1. Note that, in all cases, the experienced ratio between throughput of high and low priority stations keeps close to the desired value, which is 2. We conclude that the number of stations has a negligible impact on the *experienced weight*.

Impact of the parameter c

In Section 5.7.2 the constant c has been defined as the maximum average number of collisions allowed. This limit is needed in order to avoid loss of efficiency due to too small CWs. Since we are using the RTS/CTS mechanism, the number of collisions will never be bigger than 8 (according to the standard, a packet is dropped after 8 RTS tries). Therefore, the chosen value for c must be in the range of $0 < c < 8$.

In order to analyze the impact of c we chose to use a scenario with a large number of stations (100 stations), half of them with very high weights ($weight = 6$). This scenario leads to many stations with very small CW, and, therefore a high number of collisions, in such a way that collisions are controlled by the parameter c. Note that in a scenario without many collisions the impact of c would be almost

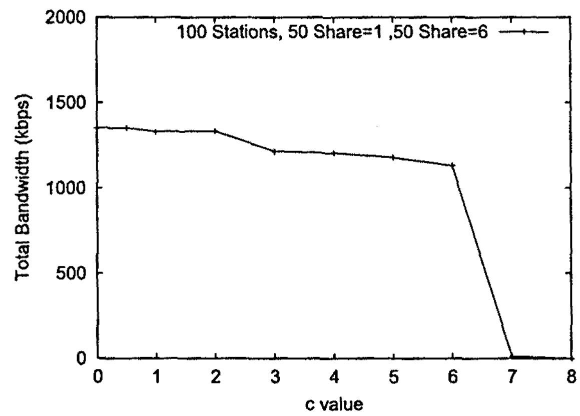

Figure 5.24: Throughput as a function of *c*.

null. Figures 5.24, 5.25 and 5.26 show the total throughput, the number of drops per successful packet and the *experienced weight* as a function of c in the above scenario. In these figures it can be seen that if the value of c is too high, the total throughput experienced is very low, and the percentage of losses very high. In the extreme case ($c > 7$) the throughput drops to 0 and the drops increase drastically. The reason for this is that with such values of c, CWs are allowed to decrease too much and the probability of collision gets too big. Note that in this case low priority stations totally starve and the differentiation tends to infinite, since the whole bandwidth is used by high priority stations.

On the other hand, if the value of c is too low, we obtain a good total throughput and very low losses, but we do not achieve the desired differentiation. In the limit ($c = 0$) there is no differentiation at all and high priority stations get exactly the same throughput as low priority ones (i.e. *experienced weight* $= 1$). The reason for this is that, with such values of c, CWs are not allowed to decrease below the values for $weight = 1$ (i.e. the ones defined in the 802.11 standard), and, therefore, the DWFQ extension defined in this section is deactivated.

As a conclusion, c expresses a tradeoff between efficiency and differentiation, and it can be adjusted via administration depending on specific user preferences. In this study we have chosen to use an intermediate value: $c = 5$. With this value of c, a good level of differentiation is achieved while conserving a good overall efficiency.

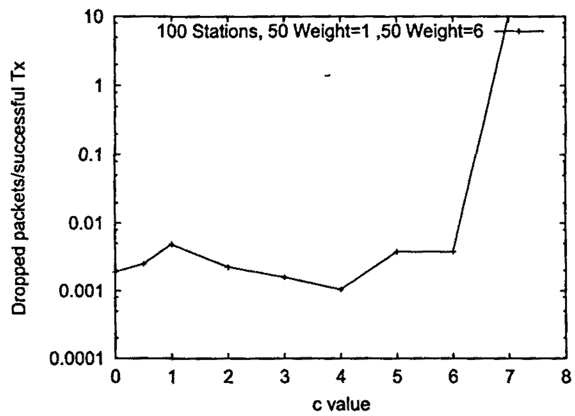

Figure 5.25: Drops as a function of c.

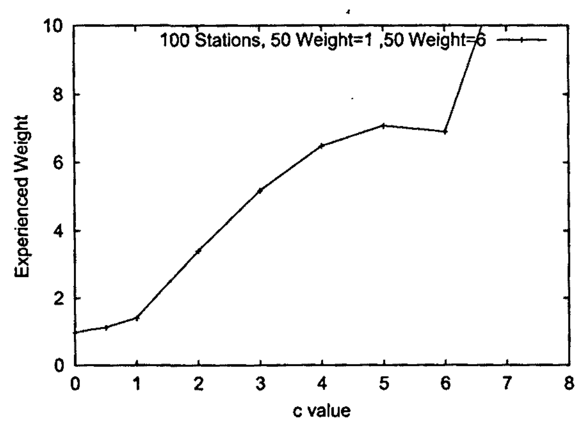

Figure 5.26: Experienced share as a function of c.

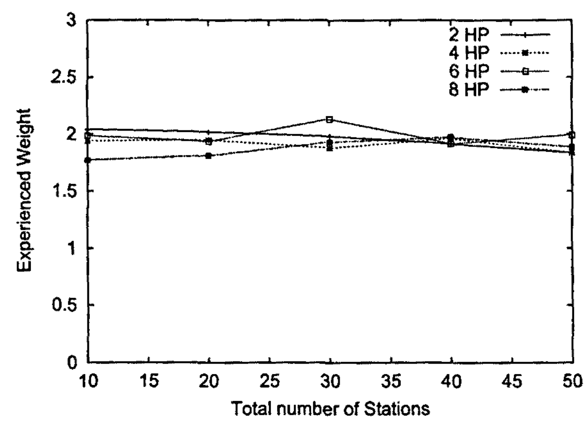

Figure 5.27: Impact of 802.11 terminals.

Backwards Compatibility

As explained in Section 5.7, stations conforming to the 802.11 standard receive in our architecture the same treatment as the stations with a *weight* equal to 1. Legacy 802.11 stations, however, do not carry in the header the L_i field, and this can have an impact into the overall performance of the DWFQ extension. This impact is studied in the simulation results shown in Figure 5.27. In this simulation, the number of stations with the DWFQ extension implemented is kept to 10, and the rest of the stations are legacy 802.11 terminals. Figure 5.27 shows the ratio between the throughput of high priority stations (*weight* 2) and low priority stations (*weight* 1) as the number of 802.11 terminals increases. It can be seen that this ratio is very close to the desired value, independent of the number of 802.11 terminals.

Channel utilization

Having DWFQ stations with a CW smaller than the CW defined in the current standard can impact the channel utilization. Figure 5.28 shows the channel utilization in the same scenario than the described for experiment 5.7.4, and compares it to the channel utilization with the current standard (i.e. with 0 high priority stations). It can be seen that the channel utilization keeps always close to the channel utilization of the current standard.

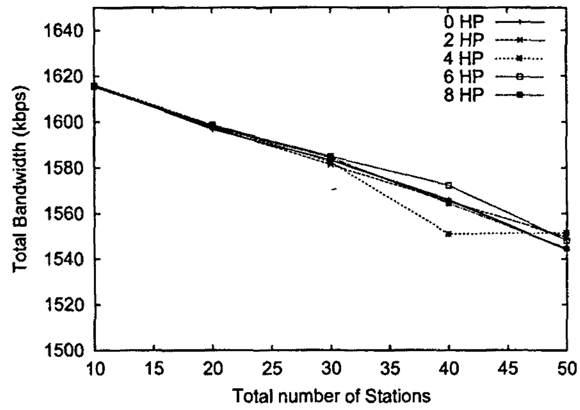

Figure 5.28: Channel utilization.

Packet drops

The algorithm proposed in this section for elastic traffic increases the aggressiveness of the MAC protocol, since it makes the CW smaller with respect to the current standard. This has an impact on the packet drops experienced at the MAC level, since after a certain number of unsuccessful retries the MAC protocol decides to drop the packet. The more aggressive we are, the higher is the probability for a retry to fail and, therefore, the probability of experiencing a packet drop.

We studied this impact in the simulations shown in Figure 5.29. It can be seen that packet drops at the MAC level increase with the total number of stations and decrease with the number of high priority stations. This is because the CW required to achieve the desired differentiation for a small number of high priority stations is very low, and therefore the probability of having 8 RTS/CTS collisions (i.e. a packet lost) is higher. However, the number of dropped packets always keeps very low: even in the worst case the percentage of packet drops is below 1%.

Channel Errors

Considering a non-ideal channel a not received ACK can be due to a channel error. As discussed in Section 5.7.2 we have introduced a collision counter which counts as a collision every sent packet (RTS) for which an ACK (CTS) has not been received. The effect of the channel errors in the collision counter would be the interpretation of a channel error as a collision. This could lead to assume falsely overload in the channel due to channel errors. The result would be an unnecessary increase of the CW, leading to a lower level of differentiation.

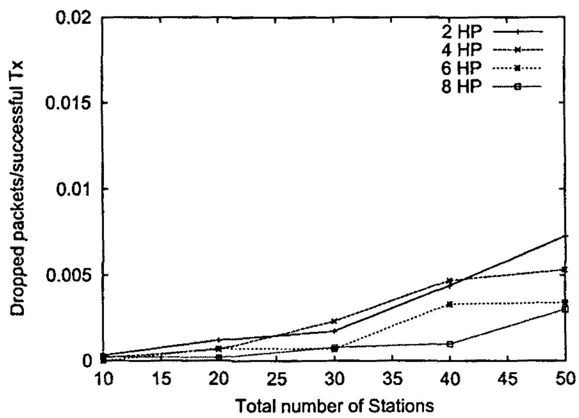

Figure 5.29: Packet drops.

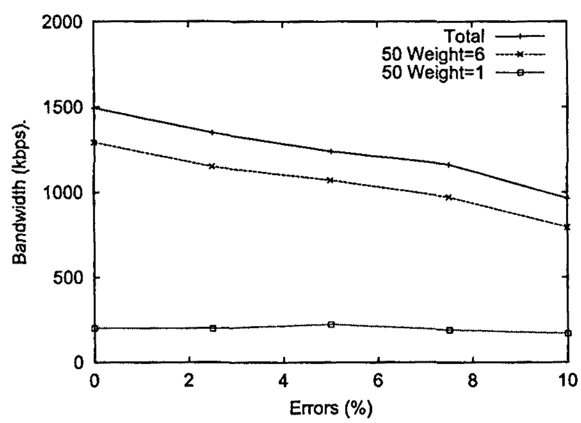

Figure 5.30: Level of differentiation as a function of the error rate

We have studied this impact under an extreme scenario as in Section 5.7.4 with a value of c equal to 5. We can observe from Figure 5.30 that the level of differentiation (*experienced weight*) is affected by the percentage of errors, as expected. Note that even in an extreme scenario with a high error percentage (10%) we still keep a reasonably high level of differentiation.

Hidden node impact

In order to study the impact of hidden nodes in DWFQ we simulated the scenario depicted in Figure 5.31. This scenario consists of three groups of stations (1,2,3) within the coverage area of the AP. Group 1 consists of one high priority station

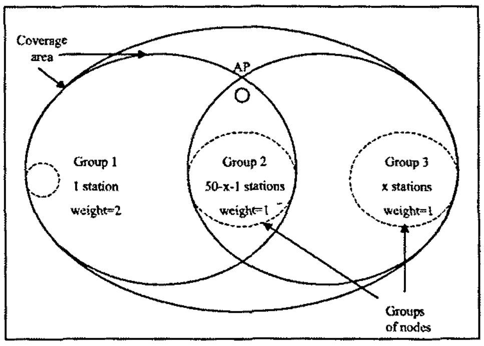

Figure 5.31: Simulation scenario

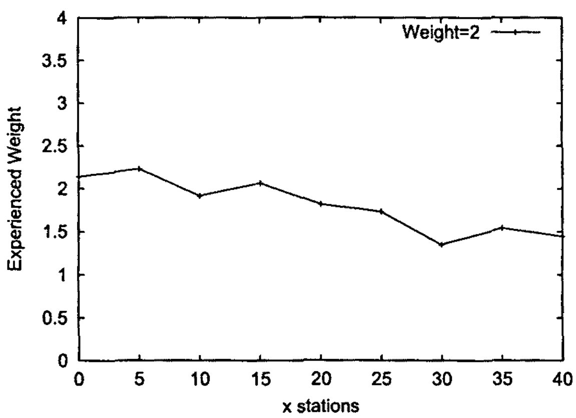

Figure 5.32: Hidden node

with *weight* equal to 2, and groups 2 and 3 consists of $50 - x - 1$ and x stations, respectively, all with a *weight* equal to 1. Groups 1 and 3 are hidden from each other.

Note that, since we are using the RTS/CTS mechanism, collisions of data packets due to the hidden node problem are avoided. However, the higher the number of stations hidden from the high priority station, the less accurate the computation of the CW in the high priority station will be. Figure 5.32 shows how this problem impacts the desired bandwidth distribution: the level of differentiation (*experienced weight*) decreases with the number of *hidden* stations. However, even in the extreme case when 80% of the stations are hidden ($x = 40$), we still keep a significant level of differentiation.

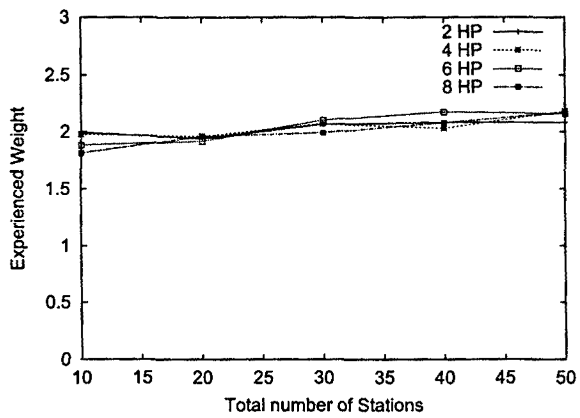

Figure 5.33: Sources UDP ON/OFF 1 ms.

Impact of bursty traffic

The simulations shown so far correspond to a constant traffic (UDP CBR sources). In order to gain a better understanding of the impact of different traffic sources to the performance of DWFQ, we have simulated it under bursty traffic (UDP ON/OFF sources).

In order to show the impact of different burst sizes, we performed two different simulations: one with a small burst (ON/OFF periods of 1 ms in average), and one with large bursts (ON/OFF periods of 500 ms in average). The simulation scenario was the same as the described in experiment 5.7.4.

Figure 5.33 shows the results when the ON/OFF periods are of 1 ms. Note that these results are very similar to the results of Figure 5.23 (CBR traffic), which means that short ON/OFF periods do not impact the performance of a station, as argued above. In Figure 5.34 it can be seen that the results for large ON/OFF periods are also very similar to the results of Figure 5.23, with a slightly higher oscillation.

TCP sources

Figure 5.35 shows the *weight* experienced by high priority stations for the scenario of experiment 5.7.4 with TCP sources. It can be seen that there are quite high oscillations in the *experienced weight* obtained, specially when the number of stations is high. This oscillation is due to the congestion control algorithm used in TCP. However, in average, the results obtained tend to the desired ones.

Note that, in contrast to the previous experiments, in this case we have down-link traffic consisting of TCP acknowledgments. Since this traffic consists of sev-

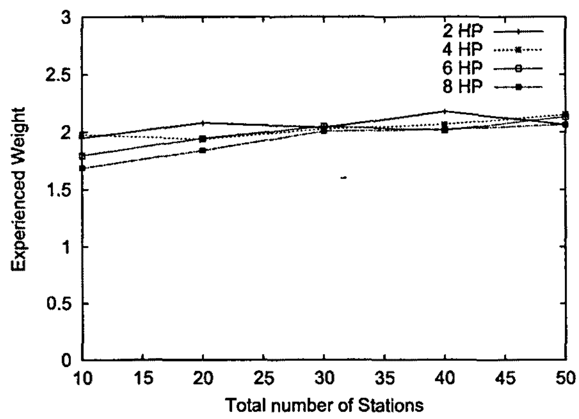

Figure 5.34: Sources UDP ON/OFF 500 ms.

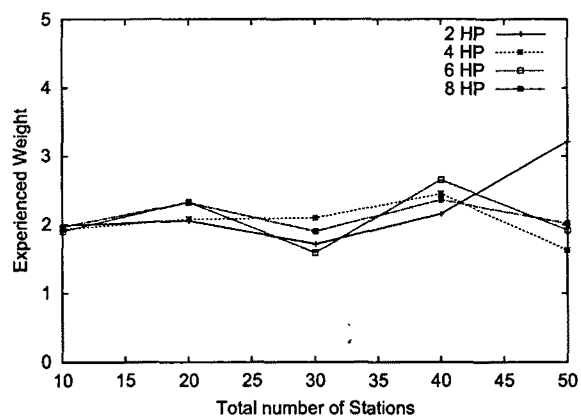

Figure 5.35: Differentiation with TCP Sources.

eral flows, we are in the multiple flows per node case. We used the solution explained in Section 5.7.3 to handle this case. We assigned to the flow i of TCP acknowledgments the same *weight* as the corresponding flow of TCP data packets. This is necessary in order to achieve the desired bandwidth distribution, since the TCP acknowledgments also impact the throughput of a TCP connection through the congestion control of TCP.

TCP vs UDP

When TCP and UDP flows compete with each other, the bandwidth distribution tends to favor UDP. This is because, in case of congestion, TCP backs off because

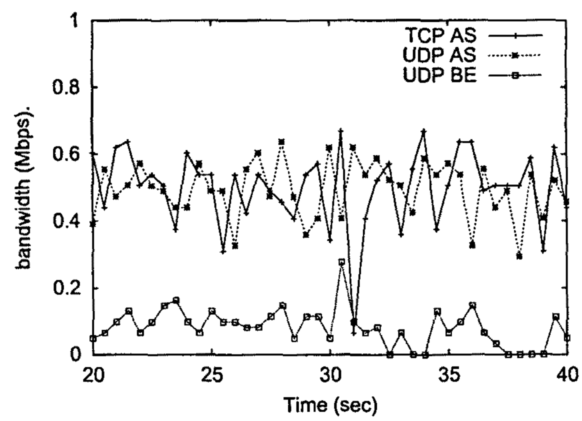

Figure 5.36: TCP vs UDP

of its congestion control mechanism, and UDP, without any kind of congestion control and therefore more aggressive, consumes the bandwidth left by TCP. An architecture for bandwidth allocation should overcome this different level of aggressiveness of the sources and provide all sources with their fair share of bandwidth independent of the congestion control algorithm they use.

To study the level of fairness between TCP and UDP achieved by DWFQ, we performed the following experiment: two high priority stations had a *weight* of 2, one sending an endless TCP flow and the other a UDP CBR flow. The remaining 8 stations had a *weight* of 1 (low priority) and were all sending UDP CBR traffic. Figure 5.36 shows the instantaneous bandwidth achieved by the TCP and UDP high priority sources and one UDP low priority source. It can be seen that, in average, the resulting bandwidth distribution is the desired.

From this experiment we conclude that DWFQ provides TCP with a fair treatment with respect to UDP. This is because the DWFQ algorithm adapts the CW to the aggressiveness of the source: a less aggressive source, like TCP, will see its CW reduced until it receives the committed throughput, while a more aggressive source, like UDP, will achieve its committed throughput with a larger CW.

5.7.5 Summary

In this section we have proposed the DWFQ architecture for providing weighted fair queuing in wireless LAN. DWFQ provides a flow with an average bandwidth proportional to its weight, but does not give any guarantees for individual packets (i.e. DWFQ can exhibit short-term unfairness).

The design goals of DWFQ have been to keep the MAC protocol fully distributed, to minimize the migration effort from the current standard, and to provide backward compatibility. We argue that a fully distributed MAC protocol is more efficient and flexible than a centralized one. We believe that the fact that DWFQ only requires minor changes in the computation of the CW facilitates the migration from the 802.11 standard. Finally, the computation of the CW has been designed in such a way that legacy 802.11 terminals receive the basic service in the proposed architecture.

The simulations performed show that DWFQ provides the desired bandwidth distribution among flows in a wide variety of scenarios. Because of the dynamic adaptation of the CW, this bandwidth distribution is independent of the level of aggressiveness of the sources and their willingness to transmit. Furthermore, simulation results show that DWFQ avoids harming channel utilization in case of overload.

5.8 Integration of 802.11e in Mobile Devices

In this section we evaluate the impact of using 802.11 power save mode in combination with the EDCA QoS mechanisms (EDCA+PSM). The objectives of the study are i) to determine whether, when both mechanisms interact, the required differentiation is still achieved, ii) quantify how significant the impact is and iii) evaluate the signaling load costs of the power save mode usage.

The analysis has been performed via simulation. We extended the 802.11b libraries provided by OPNET [80] to include the standard 802.11 power save mode described in Section 5.3 and a subset of the EDCA QoS mechanisms described in Section 5.2.

First releases of 802.11e-capable mobile devices, announced for dates before the completion of the 802.11e standard or the Wi-Fi^{TM} wireless multimedia certification start, should not be expected to implement the whole 802.11e functionality since the implementation will be based in the version of the draft available when the development started. In our study we focus in an 802.11e worst-case scenario, based on the currently available 802.11e chips, where neither Automatic Power Save Delivery, an 802.11e power saving enhancement, nor simultaneous decrement of the backoff counter for different ACs within a station is implemented. If the superposition of the mechanisms considered in our study results to be effective in such a scenario the same will apply to more favourable ones.

EDCA-TXOP durations are configured for all ACs to allow the transmission of one data frame after gaining access to the medium. The RTS/CTS mechanism has not been enabled to avoid its influence over the interaction being studied. PS-Polls are sent following the rules of the AC of the frames intended to be received to guarantee a better performance in the downlink of higher priority traffic. The listen interval configured for the power saving stations is 1. Frames of higher priority buffered at the AP AC queues are transmitted first. The length of the simulations performed is 120 seconds and at least 5 different seeds have been used to compute the values of each point in the graph.

5.8.1 Power Save Mode impact over the EDCA mechanisms

We study the impact of power save mode over the expected EDCA differentiation by analysing the differences in the performance when the three main EDCA differentiation mechanisms are varied, i.e., AIFS, CWmin and CWmax. The performance metrics used are average data throughput and MAC delay at 95% of the empirical cumulative distribution function (cdf). We define MAC delay as the delay experienced by a frame from the instant it arrives to the MAC layer and is placed in the power save mode queue (EDCA+PSM) or AC queue (EDCA) until it is successfully transmitted.

The simulation scenario chosen to illustrate the performance differences consists of a WLAN network composed of a high priority WLAN station (AC_VO), a low priority WLAN station (AC_BE) and an AP. The stations send/receive frames to/from their corresponding partners which are located in the wired domain and directly connected to the AP. The influence of the application traffic characteristics on the results has been avoided by guaranteeing that the WLAN stations always have a frame to transmit (125 bytes) and that the AP always has frames to transmit to both WLAN stations. Additionally, by doing this, the maximum achievable differentiation between both ACs can be observed. The beacon interval used in this first set of experiments is 20 ms.

AIFS

Figure 5.37 and 5.38 illustrate the effect of keeping an AIFS value of 2 for AC_VO and increasing the AIFS value for AC_BE from 2 to 34. The CWmin and CWmax values have been set to 31 to avoid the influence of the binary exponential backoff on the AIFS experiment.

The results in Figure 5.37 show that the variation of the AIFS parameter to provide a higher throughput to a certain AC is still effective even if it is combined with power save mode. However, significant differences can be observed in specific cases due to the reasons described below. AC_VO and AC_BE have, in the case of AIFS 2, exactly the same EDCA settings but the throughput achieved in the uplink (UL) by each AC of EDCA and EDCA+PSM is quite different. The

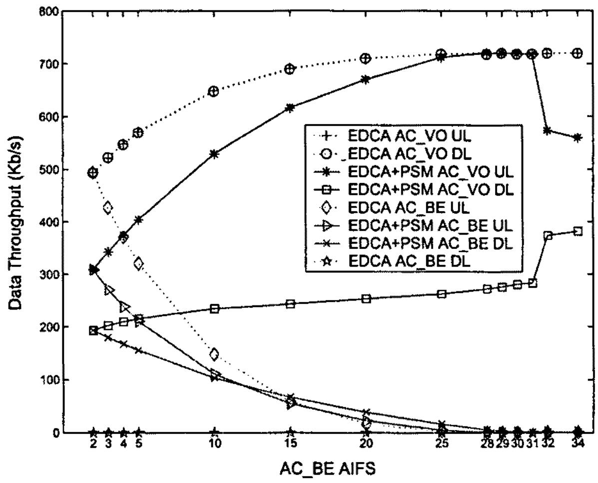

Figure 5.37: Impact of the AIFS variation of the AC_BE station on the average throughput

additional load of the EDCA+PSM stations which transmit PS-Polls in addition to the data frames is the reason for this lower throughput. The difference though gets reduced when the separation between the AIFS value of the AC_VO station and the AC_BE one increases. This happens because, in the EDCA+PSM case, an increase in the AC_BE AIFS value does not represent only a reduction in the throughput that the AC_BE can transmit but also on the number of AC_BE PS-Polls that are transmitted over the channel, i.e., leaving more bandwidth for the transmission of data frames in the channel.

In the downlink case (DL) very different behaviors are observed, as expected, since when the 802.11 power save mode is used the frames are buffered at the AP and only delivered if the corresponding PS-Poll is received. As a result, the downlink throughput achieved by AC_VO in the downlink under saturation conditions is significantly lower in the EDCA+PSM case than in the EDCA case. The downlink AC_BE throughput presents also a very particular result. In principle, considering strict priority of AC_VO over AC_BE, no AC_BE frames should be transmitted by the AP because in our scenario the AP always have at least one frame to transmit for each station. This is the case for EDCA but in the EDCA+PSM case, AC_BE frames are transmitted in the downlink since as explained in Section 5.3 only one packet *perstation* in power save mode can be buffered in the AC_VO or AC_BE queue of the AP until it is successfully transmitted. Therefore, in this experiment, after the successful transmission by the AP of an AC_VO packet, there can not be an AC_VO packet already waiting in the AC_VO queue so, if there is an AC_BE packet waiting, it will be transmitted.

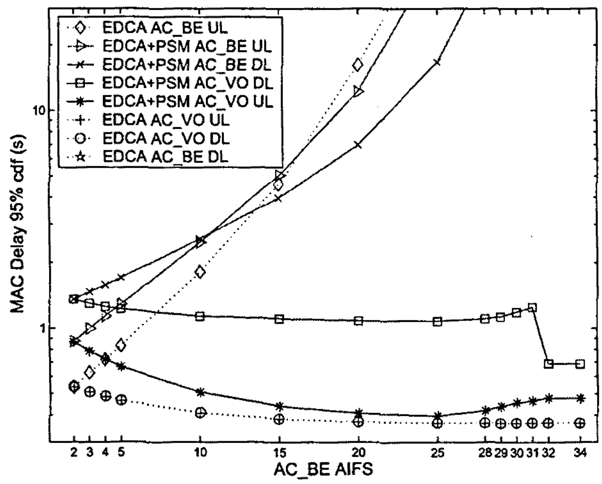

Figure 5.38: Impact of the AIFS variation of the AC_BE station on the MAC delay

Finally, there is a point in the EDCA+PSM case, AC_BE AIFS 31 in this experiment, where the difference between the AIFS of AC_VO and AC_BE is that large that no PS-Polls can be transmitted by the AC_BE stations. This results in the sharp change in the slope for AC_VO in the uplink and downlink throughput since in the downlink the high priority frames do not need to wait anymore for the transmission of the low priority ones so the throughput increases while the higher competition to access the channel and the larger number of PS-Polls to be sent yields a decrease for the AC_VO traffic in the uplink.

The delay experienced by frames at the MAC layer of a certain AC is expected to be inversely proportional to the throughput it achieves. Figure 5.38 depicts the MAC delay value at 95% of the cdf for the studied cases. As expected, the results correspond to the ones obtained for the throughput case. Note that some small differences in the slope of both figures occur because in the throughput case the points represent the average values while in the delay case the points represent the value at 95% of the cdf.

CWmin and CWmax

The impact of the 802.11 power save mode over the differentiation achieved by modifying the CWmin and CWmax values has been studied through two different experiments. In the CWmin analysis, the CWmax value of both ACs has been set to 1023, the CWmin value of AC_VO to 31 and the CWmin value of AC_BE has been varied from 31 until 1023. In the CWmax case, the CWmin value of both ACs has been set to 31, the CWmax value of the AC_VO to 31 and the CW-

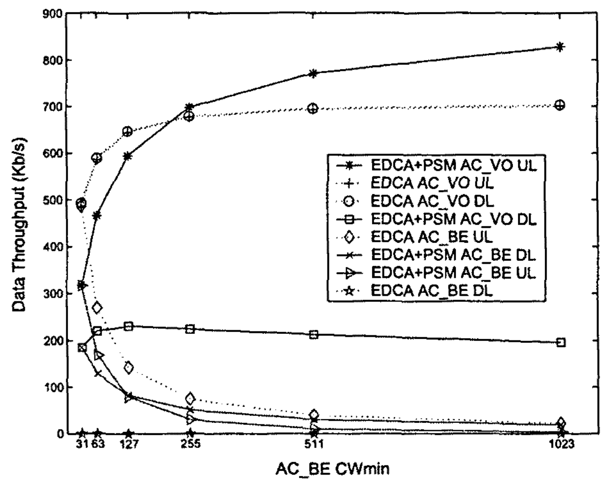

Figure 5.39: Impact of the CWmin variation of the AC_BE station on the average throughput

max value of the AC_BE has been varied from 31 until 1023. Since the CWmax importance increases according to the number of collisions, in this experiment results with 6 wireless stations of each AC are shown. Both experiments have been performed with an AIFS value of 2 for all ACs.

The results obtained in the CWmin experiment, see Figure 5.39, resemble the ones presented in the AIFS section. A similar reasoning for the performance differences between EDCA and EDCA+PSM applies here considering that now instead of providing a deterministic differentiation a statistical one is provided. This results in some major differences as compared to the AIFS case. No sharp change in the slope is observed for the EDCA+PSM average throughput of AC_VO in the uplink and downlink since in contrast to the AIFS experiment, the AC_BE average throughput value in the downlink does not reach zero. The reason for this is that while with AIFS we could arrive to a configuration where no AC_BE frame in the downlink could be transmitted, such a extreme case does not occur in the CWmin case since AC_BE traffic still has a small probability of being transmitted. Consequently, the EDCA+PSM AC_VO throughput in the downlink suffers a constant slight decrease from CWmin value 127 on due to having to wait for the lower priority frames transmission and the EDCA+PSM AC_VO throughput in the uplink benefits of the lower competition in the channel. The AC_BE EDCA+PSM throughput in the uplink is lower than in the downlink due to the additional PS-Polls sent by the station.

Figure 5.40 shows the throughput per station results in the CWmax case for 6 wireless stations of each AC. The results of this experiment for PSM present

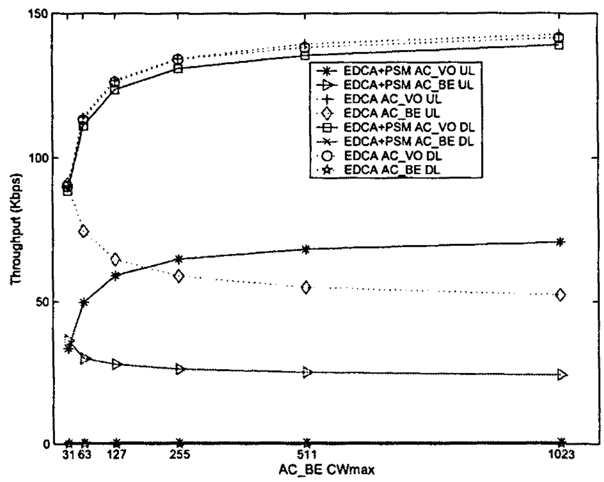

Figure 5.40: Impact of the CWmax variation of the AC_BE station on the average throughput

significant differences as compared to the CWmin case because now, even if the power save mode procedure to take frames out of the PSM queue is the same, in the scenario there is more than one AC_VO station so in the AP AC_VO buffer there is almost always a frame to be transmitted. As a consequence, the EDCA+PSM AC_BE downlink throughput is close to zero and the AC_VO one is similar to the EDCA case. In the uplink worse results than in the downlink are observed for the PSM AC_VO case because of sending PS-Polls. On the other hand, the AC_BE uplink throughput presents better results because a lower number of PS-Polls is sent.

5.8.2 Costs of the Power Save Mode and EDCA interaction

The usage of the power save mode mechanism implies a decrease in the WLAN bandwidth available for the data transmission due to the additional signaling load and an increase in the downlink delay according to the beacon interval value. We define two different experiments to quantify the importance of this degradation. In both experiments two examples of EDCA configurations are considered based on the values proposed in the 802.11e draft: aggressive (Conf-A) and conservative (Conf-C) (see Table 5.2). The decrease of effective wireless channel bandwidth and the signaling load introduced is evaluated through an experiment where we increase the number of wireless stations (50% of each AC) until the saturation of the channel is reached. The beacon interval impact over the downlink delay is quantified, de-

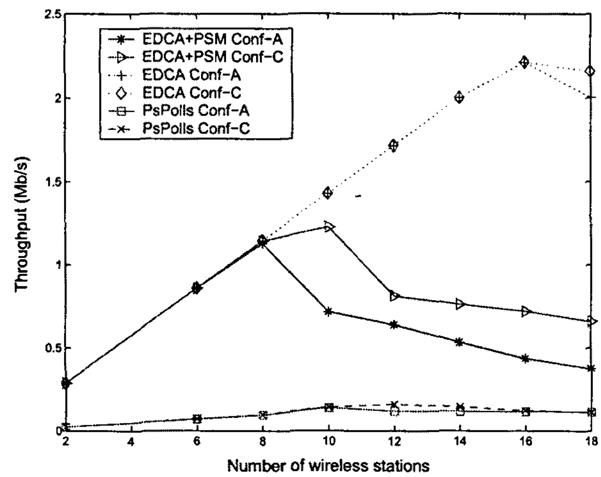

Figure 5.41: Impact of the number of stations on the wireless LAN data throughput

	AIFS	CWmin	CWmax
Conf-A, AC_VO	2	7	15
Conf-A, AC_BE	2	15	31
Conf-C, AC_VO	2	7	15
Conf-C, AC_BE	2	31	1023

Table 5.2: EDCA configuration for Conf-A and Conf-C

pending on the number of wireless stations, by increasing the beacon interval value from 20 to 100 ms. Each wireless station has its pair in the wired domain. Data traffic is generated at a constant bit rate (240 bytes every 30 ms) by the applications in the wired and wireless nodes.

Figure 5.41 shows the data and PS-Poll throughput sent over the wireless channel and received by its destination according to an increasing number of wireless stations. The effect of the additional overhead from power save mode on the achievable effective throughput is clearly noticeable demonstrating that it can not be neglected. When congestion is reached, a clear difference in the behavior for Conf-A and Conf-C can be observed where Conf-C because of its larger CW values for AC_BE obtains better results with respect to the maximum achievable data throughput since the number of collisions in the channel are lower. The signaling load introduced by the PS-Polls follows a similar behavior and represents in our scenario around 8% of the achievable data throughput when the channel is not congested.

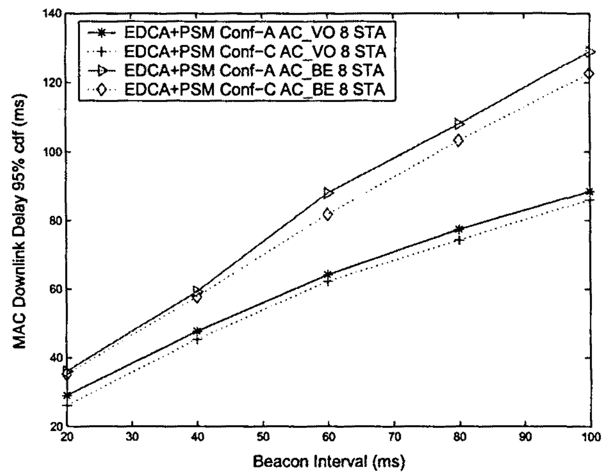

Figure 5.42: Impact of the beacon interval on the downlink delay

Based on the previous results we have chosen the case of 8 wireless stations to study the impact of the beacon interval as the case where the maximum differences between AC_VO and AC_BE will be observed before the wireless channel is congested. The results are depicted in Figure 5.42 where an almost linear increment according to an increasing beacon interval value is observed. As expected, the longer the beacon interval is the longer it takes to the stations to be informed about buffered packets at the AP to be retrieved. Note that because we compute the MAC downlink delay of the cdf value at 95% the effect of the More Data mechanism, which counteracts the beacon interval importance, is less noticeable than if the average would be used. The differences between AC_VO and AC_BE are due to the higher priority of the PS-Polls and in the transmission at the AP. Conf-C performs slightly better than Conf-A because we are close to the congestion case.

5.8.3 Power Saving Evaluation

In this section the effectiveness of the power save mode mechanism is studied in a scenario as the one used in Section 5.8.2. We compute the percentage of time spent in active mode by the stations in power save mode for both EDCA configurations depending on the beacon interval considered and the number of wireless stations (50% of each AC).

The results in Figure 5.43 clearly show the power consumption reduction for both ACs, at least around 40%, when the channel is not congested, i.e., number of

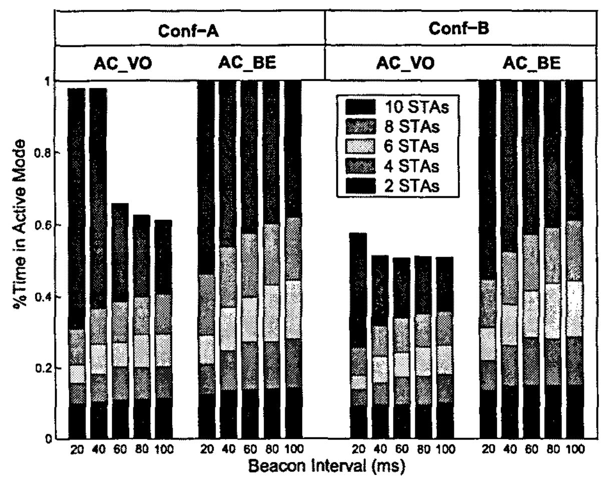

Figure 5.43: Impact of the beacon interval on the power saving efficiency

wireless stations 8 or below. AC_VO stations present a lower power consumption because they send and receive frames faster. Considering the frame generation rate of our scenario (30ms), we can clearly see that when the number of frames buffered in the AP is 2 or above, from the beacon interval case of 60ms on, the More Data mechanism tends to reduce the beacon interval impact since it randomizes the transmission time of the PS-Polls. For a lower beacon interval value, the longer the beacon interval the more PS-Polls collisions. The 10 stations case presents a percentage of time in active mode close to 1 for the AC_BE case since the channel is getting congested and a lower one for AC_VO. The aggressiveness of Conf-A that produces a longer time for resolving collisions when the channel is congested benefits from a longer beacon interval in the AC_VO case because of the randomizing effect of the More Data mechanism.

5.8.4 Summary

The upcoming mobile devices including 3G and WLAN access technologies introduce new technological challenges that need to be addressed. We identified the combination of the 802.11e QoS mechanisms with the 802.11 power save mode functionality as a key element toward the introduction of WLAN capabilities in battery limited devices as cellular phones. The performance of the superposition of both mechanisms though, requires to be studied to determine its suitability for the case considered.

A comprehensive overview of the 802.11 power save mode and a description of the latest functionality included in the 802.11e draft, comprising QoS and power save mode enhancements, has been provided followed by a description of the functionality to be expected in the worst case in the first versions of the devices. Based on this, we evaluated the performance of the combination of the EDCA QoS mechanisms and the 802.11 power save mode via OPNET simulations. Through this analysis a deep insight on the impact of power save mode over the EDCA QoS functionality and its causes was acquired. Therefore, the results of this study are twofold. First, we provided quantitative results of the performance differences to be expected when the power save mode is used and of the power saving efficiency. Second, we provided the reasoning behind the impact of the different parameters pointing out the elements to be taken into account when configuring both mechanisms to be used together.

The main conclusions that can be drawn from our results are i) the 802.11e mechanisms to provide different QoS are influenced by the 802.11 power save mode functionality but are still highly effective ii) an important impact is observed on the downlink delay, as expected, where special attention has to be paid to the beacon interval configuration iii) the additional overhead introduced by the power save mode decreases considerably the effective bandwith available for data transmission in the wireless LAN and iv) in devices where power saving is a critical issue, the significant increase in the power usage efficiency in no congestion conditions justifies the power save mode costs.

5.9 Mechanisms to Reduce the Downlink Delay Due to Power Save Mode

The standard power save mode functionality defined by IEEE 802.11 introduces a dependency between the Beacon Interval value (BI) and the data frames downlink delay (AP to station) since the process to retrieve frames from the AP only starts after receiving a beacon indicating that frames are buffered at the AP. This dependency implies that some BI configurations might result in downlink delays above the acceptable for some applications, e.g., VoIP, as we will show in Section 5.10.1. In order to remove the dependency between the BI and the downlink delay several modifications of the standard Power Save mode can be designed which can be divided in two main different groups: proactive and reactive.

- *Proactive* approaches remove the dependency between the BI and the downlink delay by forcing a station to send PS-Polls at certain time intervals independently of whether a beacon has been received indicating that there is traffic buffered at the AP for that station. By setting a certain PS-Poll send interval below the BI, a *soft* upper bound (statistical guarantee) of the downlink delay can be provided. However, depending on the application considered, this might result in PS-Polls being unnecessarily generated, e.g., web browsing where packets will be received at the AP in a bursty fashion rather than at a constant rate.

- *Reactive* approaches take advantage of the instant when a station awakes to send a data frame for sending a PS-Poll. In this case, the dependency with the BI is also eliminated but a new one with the uplink data frame transmission interval is introduced. Additionally, this mechanism can end up in some cases either in the transmission of a large number of PS-Polls not required by the application downlink rate, e.g., FTP uplink transfer, or in a large downlink delay when the application generates much more packets in the downlink direction than in the uplink.

Both approaches present a common problem when an AP receives a PS-Poll and no frame is buffered for the station that sent it. The problem comes from the fact of trying to reduce the downlink delay by generating PS-Polls without being totally sure that a data frame is buffered at the AP. According to the 802.11 standard Power Save mode description, PS-Polls sent by an 802.11 station to an AP can either be acknowledged by receiving the corresponding data frame or by an acknowledgment frame. However, in both cases, if no frame is buffered at the AP for the station, the station will remain awake expecting the reception of the data frame after sending the PS-Poll or after receiving the acknowledgment. As a result, the power saving efficiency will suffer a significant degradation. Note that this could not occur with standard power save mode since a PS-Poll is only sent after the indication that a frame is buffered at the AP.

We designed a simple solution for this problem that consists in using the More Data field included in the acknowledgment frames, which is currently not being used, to indicate to the station whether it should expect a data frame, i.e., remain awake, or not. By doing this, a station that sent a PS-Poll but has no frame buffered at the AP can go to doze mode immediately after receiving the acknowledgment, i.e., reducing in this case the time that a station remains awake to the minimum. In the rest of the document we will refer to an acknowledgment indicating that no frame is buffered for a certain station as a *No Data Acknowledgment (NDAck).*

The aforementioned proactive and reactive solutions are the first solution that one could think of but, as we will show in Section 5.10.1, they present similar problems with respect to the assumption of a certain application traffic pattern. Because of their *static* nature, when the traffic pattern assumed does not match with the one currently generated by the application, either unnecessary PS-Polls are generated or a large downlink delay is experienced. Different configurations could be defined depending on the downlink delay performance required and the expected traffic pattern, this would result though in an additional overhead for the user, forced to choose between different options, that should be avoided if possible.

In order to overcome the above mentioned issues we propose an adaptive power save mode algorithm (APSM) for the stations that using the information available at the MAC layer provides a soft upper bound of the MAC downlink delay by adapting to the downlink data frame interarrival time. In the following section the APSM algorithm designed is described. Due to the *dynamic* nature of our algorithm no application traffic pattern needs to be assumed.

5.10 Adaptive Power Save Mode Algorithm

The adaptive power save algorithm proposed has been designed to fulfill the following objectives:

1. Provide a soft upper bound of the MAC downlink delay according to common application requirements independently of the beacon interval value.
2. Keep the bound guarantee even in the case of more than one application per station (traffic mixed).
3. Guarantee a power saving efficiency similar or better than the one of standard power save mode.
4. Minimize the signaling load introduced in the channel.
5. Minimize the impact on the standard power save mode.

To achieve these objectives we have opted for an adaptive proactive approach instead of for a reactive one to have a solution independent of any uplink application characteristic or WLAN specific parameter, e.g., BI. The algorithm designed is based on the estimation of the current downlink data frame interarrival time to adapt the PS-Poll interval accordingly and in this way upper bound the MAC downlink delay to the frame arrival interval. We have chosen this approach because in general applications can cope with an end-to-end delay well above their frame generation interval, e.g., a delay sensitive application as VoIP, which codecs usually generate frames every 10-30ms, can deal with an overall delay of 150-300ms. Considering that in most of the networks including WLAN access the main contributor to the overall delay is usually the MAC layer, upper bounding the MAC downlink delay to the downlink frame interarrival time should satisfy the most stringent application requirements.

The APSM algorithm is started by a station after the reception of a beacon indicating that frames are buffered for that station and is stopped when the number of NDAck received in a row reach the n_NDAck_max value. Note that ASPM becomes equivalent to standard power save mode (except from the transmission of n_NDAck_max PS-Polls) when the downlink data frame interarrival time is so low that due to NDAcks we always disable the APSM mechanism before a new data frame arrives. The n_NDAck_max value represents a trade-off between certainty that the communication with a station has stopped and signaling load and power consumption savings. A higher value provides more guarantees that the algorithm does not go back to standard power save mode due to a very short inactive period of an application. On the other hand, a lower value guarantees a faster reaction resulting in lower signaling load and higher power saving, e.g., taking advantage of the silence periods of a VoIP source.

Once the APSM algorithm is started, a station continuosly sends PS-Polls to the AP at a certain PS-Poll interval ($pspoll\_interval$) starting from the configurable predefined value $pspoll\_interval\_init$ and converging to the downlink data frame interarrival time following the procedure described in Algorithm 1. The $pspoll\_interval\_init$ value should be chosen considering that the algorithm might require several iterations until is finally adapted and that the speed to increase the PS-Poll interval value is faster than the speed to decrease it. Therefore, if the usage of delay sensitive applications is expected, it is better to start from a small value, e.g., 10-50 ms. During the phase where the APSM algorithm is active, stations still follow the standard procedure of the More Data mechanism when receiving a data frame but do not awake to check the information of the beacon regarding buffered data frames or react to them. The More Data mechanism is still necessary to absorbe disadjustments when the PS-Poll interval estimated by the APSM algorithm is not yet adapted.

The information collected at the MAC layer to infer the actual downlink data frame interarrival time is: number of data frames received during a certain period (n_fr_rcvd), data frames received with the More Data bit set to true (MD) and acknowledgments indicating that no frames are buffered at the AP for a certain station ($NDAck$). Based on this information, we have designed the APSM algorithm described in Algorithm 1 which adapts the PS-Poll sending interval to the downlink data frame interarrival time observed at the MAC layer[2].

The reasoning used to design the algorithm is as follows. Ideally, if the PS-Poll interval configured would perfectly match the downlink data frame interarrival

[2]Note that in Algorithm 1 we show only the part of the algorithm that takes care of modifying the PS-Poll interval to be used.

Algorithm 1 Adaptive algorithm to bound the MAC Downlink delay

On receiving an ACK frame:
if $NDAck$ **then**
  if $n\_NDAck == n\_NDAck\_max$ **then**
    Stop the APSM algorithm
  else
    $pspoll\_interval = pspoll\_interval(1 + \frac{1}{n\_fr\_rcvd+1})$
    $n\_NDAck \leftarrow n\_NDAck + 1$
    $n\_fr\_rcvd \leftarrow 0$
    $update\_next\_MD \leftarrow false$
    $previous\_MD \leftarrow false$
  end if
else
  $n\_NDAck \leftarrow 0$
end if

On receiving a DATA frame:
if MD **then**
  if $update\_next\_MD$ **then**
    $pspoll\_interval = pspoll\_interval(1 - \frac{1}{k(n\_fr\_rcvd+1)})$
    cancel pending PS-Poll and schedule a new one
    $update\_next\_MD \leftarrow false$
    $n\_fr\_rcvd \leftarrow 0$
  end if
  if $previous\_MD == false$ **then**
    $n\_fr\_rcvd \leftarrow 0$
  end if
  $previous\_MD \leftarrow true$
else
  if $previous\_MD$ **then**
    if $n\_fr\_rcvd > 1$ **then**
      if $n\_MD\_burst > j$ **then**
        $pspoll\_interval = \frac{pspoll\_interval}{n\_fr\_rcvd+1}$
        cancel pending PS-Poll and schedule a new one
      else
        $n\_MD\_burst \leftarrow n\_MD\_burst + 1$
        $update\_next\_MD \leftarrow true$
      end if
    else
      $update\_next\_MD \leftarrow true$
      $previous\_MD \leftarrow false$
      $n\_MD\_burst \leftarrow 0$
    end if
  end if
  $n\_MD\_burst \leftarrow 0$
end if
$n\_fr\_rcvd \leftarrow n\_fr\_rcvd + 1$

time, when a PS-Poll is sent to the AP a *single* data frame would be buffered at the AP waiting for a PS-Poll arrival. When the PS-Poll interval is not yet *adapted* though, what happens is that if the PS-Poll interval is above the downlink data frame interarrival time the station will recover some packets using the More Data mechanism while if the PS-Poll interval is below the downlink data frame interarrival time, then the station will receive some NDAcks. Therefore, our algorithm is based on the usage of the data frames received with the More Data bit set to true (MD) and the reception of no data acknowledgments (NDAcks) as indicators of whether our PS-Poll interval estimation is above the desired value or below, respectively. Additionally, when the decision of modifying the current PS-Poll interval is taken, the required change is calculated taking into account the number of data frames received during the period comprised from the last moment this decision was taken until the current one (n_fr_rcvd).

Since one of our objectives is to minimize the signaling load introduced by the algorithm, the way of deciding when a change in the estimation is required and how fast this change should be is different depending on whether the estimation is above the actual downlink frame interarrival time or below. To achieve our target the PS-Poll interval used should always be equal or above the downlink frame interarrival time to avoid sending unnecessary PS-Polls. Therefore, we have introduced a factor $k \in [1, \infty]$ used to control the difference in the estimation value change and we modify the PS-Poll interval value for each NDAck received while in the MD case we change the estimation if we previously received another one. By introducing this asymmetry we achieve firstly a faster adaptation of the algorithm when the estimation is below the desired value and secondly the reduction of the danger of bouncing around the actual value once the estimation is above and close to the desired one. To avoid the problem that could occur due to our slower decreasing speed when a significant decrease of the PS-Poll interval would be necessary to adapt to the current downlink data frame interval, we included a mechanism (n_MD_burst) that detects whether a considerable change has occurred based on the configurable threshold j and reacts accordingly.

We deliberately restricted the algorithm design to use only MAC layer information in order to keep it transparent to the upper and lower layers easing thus its implementation. The complexity of the mechanism has been kept low by minimizing the parameters to be configured ($pspoll\_interval\_init$, n_NDAck_max, k and j) and by distributing the computation load between the stations instead of centralizing it at the AP.

Further improvements could be achieved based on additional assumptions. If we would consider only applications transmitting at a constant interval, by estimating the exact time at which these packets are actually generated we could further reduce the downlink delay. This would additionally assume that the variance in the delay introduced by the wireless channel is small compared to the generation interval. We have chosen not to rely in such strong assumptions since our aim is to design a generic enhancement of the 802.11 power save mode that covers the case not only of constant bit rate traffic sources but also of variable bit rates and interarrival times and specially that is suited for the case of a mix of traffic for a station. We consider the case of mixed traffic specially relevant since we assume that in the future mobile devices could be used for a voice/video communication while at the same time run an e-mail or web browsing application. Therefore, our algorithm has been designed to absorbe this additional bursty traffic by rapidly reducing the PS-Poll interval for keeping the soft upper bound downlink delay guarantee.

5.10.1 Performance Evaluation & Discussion

In this section we evaluate the performance of the different enhancement approaches described in Sections 5.9 and 5.10 as compared to standard 802.11 power save mode. The analysis has been performed via simulation. We extended the 802.11b libraries provided by OPNET 10.0 [80] to include the 802.11 standard power save mode and the potential enhancements described in Sections 5.9 and 5.10.

For illustration purposes we have configured the static *proactive* mechanism used in our simulations to send a PS-Poll every 30ms and the *reactive* mechanism to generate a PS-Poll after the uplink transmission of each data frame (ratio 1). The configuration chosen for our proposed APSM algorithm is $pspoll\_interval\_init$ of 10 ms, n_NDAck_max equal to 3, k to 2 and j to 1. The implementation of all enhancement approaches include the NDAck mechanism. With respect to the standard power save mode, the listen interval used is 1.

In the following sections, unless otherwise indicated, the number of wireless stations used for the simulations is 5. The value has been chosen large enough to guarantee that in our experiments the effect of collisions in the channel is taken into account but small enough to avoid saturation. Each wireless station has its corresponding pair in the wired domain. Communication occurs always between wireless-wired pairs through the AP. The length of the simulations performed is 120 seconds. The values in the graphs have been computed using the statistics obtained for each of the single wireless stations and using at least 10 different seeds.

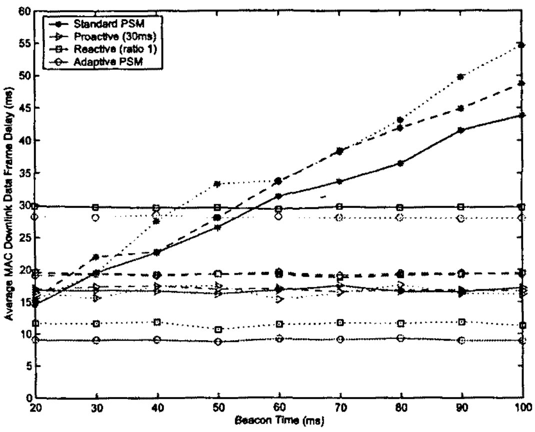

Figure 5.44: Impact of the beacon interval on the MAC downlink delay of data frames

Impact of the Beacon Interval

The purpose of the enhancements described in Sections 5.9 and 5.10 is to remove the dependency of the average MAC downlink data frame delay with respect to the beacon interval value. In this experiment we evaluate the effectiveness of the proposals computing the average data frame downlink delay for a beacon interval value varying from 20 to 100ms.

Figure 5.44 shows the results for a *symmetric application* sending and receiving every 40ms (dashed line), an *uplink asymmetric application* sending in the uplink (station to AP) every 20ms and receiving every 60ms (dotted line) and a *downlink asymmetric application* sending every 60ms and receiving every 20ms (solid line), all of them generate traffic at a 64kbps rate. The results are as expected, while the downlink delay of data frames increases according to the beacon interval value for standard power save mode, the one of the enhancement approaches remains constant around a value corresponding to half of the PS-Poll generation rate since the arrival time of the data frames to the AP within a PS-Poll interval is uniformly distributed. Note that the standard power save mode downlink delay does not increase linearly according to the beacon interval increase in the symmetric and asymmetric uplink case. The reason for that is the More Data mechanism effect that when it comes into play reduces the medium access delay since the probability of collision is lower than directly after receiving a beacon.

In the rest of the experiments the beacon interval used is 100ms.

Dynamic Adaptation of the APSM algorithm

A design goal of our APSM algorithm is to adapt the downlink delay of the data frames according to the characteristics of the downlink traffic generated by the application. In order to accomplish this, each station predicts the downlink data frame interarrival time at the MAC layer to generate PS-Polls at a rate that provides a soft upper bound guarantee for the downlink data frame delay.

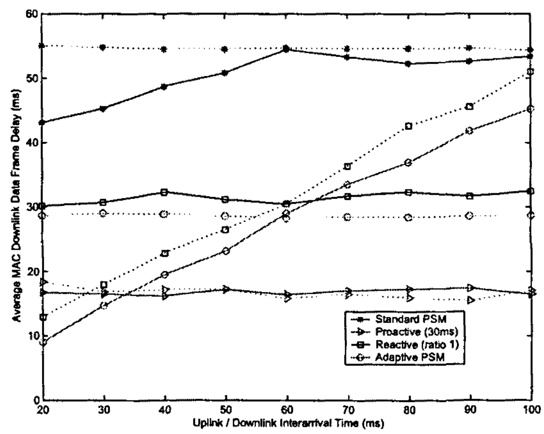

Figure 5.45: Impact of the traffic characteristics on the MAC downlink delay

Figure 5.45 presents the results regarding the average MAC downlink delay for an experiment where, in the case of the uplink asymmetric application (dotted line), we fix the downlink frame interarrival time to 60ms and increase the uplink from 20 to 100ms, and in the case of the downlink asymmetric application (solid line), we fix the uplink frame interarrival time to 60ms and increase the downlink from 20 to 100ms. The results show that the APSM algorithm achieves the desired objective of upper bounding the downlink data frame delay according to the downlink frame interarrival time. For the uplink asymmetric application case the downlink delay keeps constant around the 30ms value since the downlink frame interarrival time is 60ms and in the downlink asymmetric case the downlink delay increases according to an increasing downlink frame interarrival time.

With respect to the static proactive scheme, the downlink delay is constant for all cases due to the fact that, independently of the frame interarrival time in the downlink, the probability of a frame to arrive within a certain point between two consecutive PS-Polls is uniformly distributed. The same applies to the downlink asymmetric case of the reactive scheme with the only difference that now the PS-Poll generation rate is 60ms instead of 30ms. On the other hand, the downlink delay increases according to the uplink frame interarrival time for the reactive mechanism because the PS-Poll interval increases in the same way. Regarding the standard power save mode performance, in the asymmetric uplink case the average delay is constant since a change in the uplink rate does not influence the mechanism. However, in the asymmetric downlink case, the average delay decreases for a downlink rate of 60 to 20 ms due to the fact that occasionally the More Data phase lasts longer than the arrival of new frames.

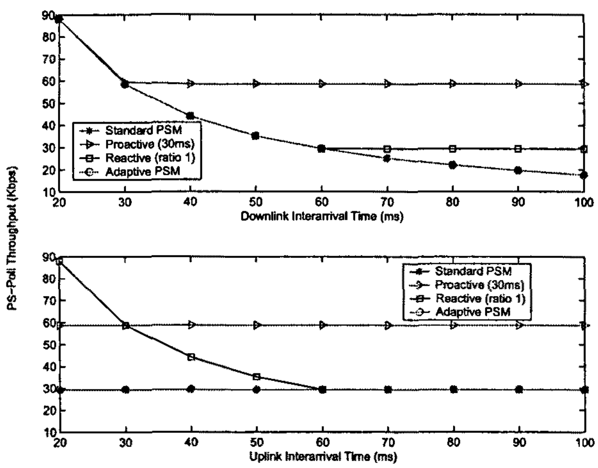

Figure 5.46: Impact of the traffic characteristics on the PS-Polls signaling load

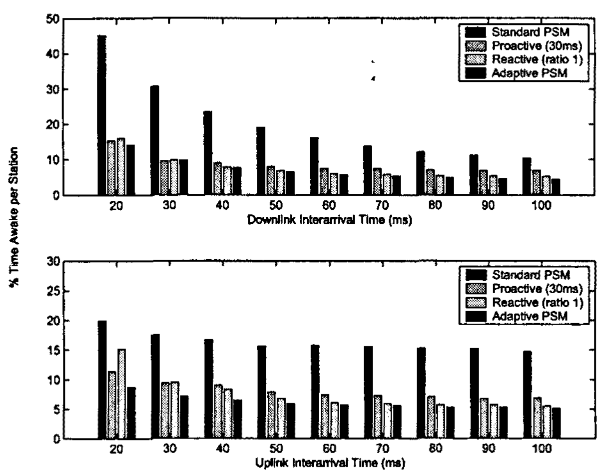

Figure 5.47: Impact of traffic characteristics on the power saving efficiency

In Section 5.10 we noted that another objective of our algorithm was to minimize the signaling load introduced in the wireless channel. Using the same experiment already described we analyze the PS-Poll signaling load introduced by

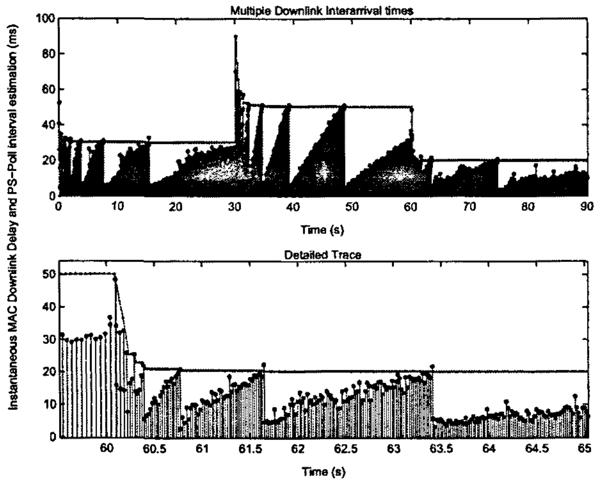

Figure 5.48: Dynamic adaptation of the APSM algorithm to the traffic characteristics

our algorithm and present the results in Figure 5.46. Obviously, the ideal would be to achieve with our APSM algorithm, which reduces the downlink delay, the same signaling load results than standard power save mode which does not send any unnecessary PS-Polls. In the figure we see that for both cases, asymmetric uplink and downlink applications, APSM introduces a PS-Poll signaling load almost perfectly matching the one of standard power save mode. We can conclude then that the additional signaling load introduced by the APSM algorithm during the adaptation phase as compared to standard PSM is negligible while it results in a significant performance improvement in the downlink delay.

In the case of the proactive approach, the signaling load introduced is always a constant value except for the case of the asymmetric downlink application when it generates frames at 20 ms since some additional PS-Polls are sent due to the More Data mechanism. The reactive mechanism shows a very similar behaviour to the standard power save mode or APSM until an uplink/downlink interarrival time of 60ms because for both kind of applications the same PS-Poll rate is generated either due to the More Data mechanism that sends the additional PS-Polls required in the asymmetric downlink application case or due to the unnecessary PS-Polls sent in the asymmetric uplink application case. From the 60ms interarrival time point on the signaling load keeps constant in the asymmetric uplink case due to the More Data mechanism.

Even though the main aim of the APSM algorithm designed is to reduce the downlink delay introduced by the standard power save mode to a value acceptable for the QoS required by the applications, it is equally important to guarantee that our APSM algorithm does not degradate the power saving efficiency as compared to standard power save mode. Figure 5.47 represents the percentage of time that a station is awake in average for the same experiment previously described. The results show that APSM is always the best performing power saving mechanism. Standard power save mode is always the worst one since the transmissions of the PS-Polls of the stations are always concentrated around the time that follows the reception of the beacon so the probability of collision is higher than for the enhancement approaches that do not introduce this synchronization. The proactive and reactive mechanisms present a power saving efficiency closer or further to the one achieved by APSM depending on whether for a particular case the assumption of the downlink frame rate is better or worse.

Two important characteristics of an adaptive algorithm which are related with each other are the speed at which the algorithm adapts to changes and the stability once it is adapted. This trade-off has been studied by modifying during a simulation the downlink frame interarrival time from 30ms to 50ms and then to 20ms. In Figure 5.48 we can observe the dynamic adaptation during the simulation time of the PS-Poll interval estimation (solid line) and the instantaneous MAC downlink delay of the data frames (deltas) for one of the stations. In the detailed part of the graph, each point in the solid line represents the transmission of a PS-Poll and each delta the arrival of a data frame in the downlink direction with an amplitude corresponding to the MAC downlink delay experienced. As it can be observed, the APSM algorithm properly estimates the corresponding downlink frame interarrival time and adapts significantly fast to the rate modifications. The 'ramp up' behaviour occurs due to the algorithm design that prioritizes to be above the actual downlink frame rate to minimize the unnecessary PS-Polls sent by the station resulting in ramps that decrease their slope based on the data frames received with the More Data bit set.

Impact of a Realistic Scenario

In the previous section we have analyzed the performance of the APSM algorithm considering applications generating traffic at fixed time intervals and at a fixed bit rate in order to facilitate the interpretation of the results obtained. Once the expected behaviour of the algorithm has been validated, we focus in a more general scenario where realistic applications do not generate traffic constantly, e.g., G.711 audio codec with silence suppression (data rate 64kbps, frame rate 30ms), or do not generate traffic at a constant bit rate, e.g., MPEG-4 streaming of the movie Jurassic Park (average rate 150kbps, frames generated every 40ms [20]).

Figure 5.49 shows the proper adaptation of the PS-Poll rate estimated by the APSM algorithm during the active periods of the VoIP source (solid line where

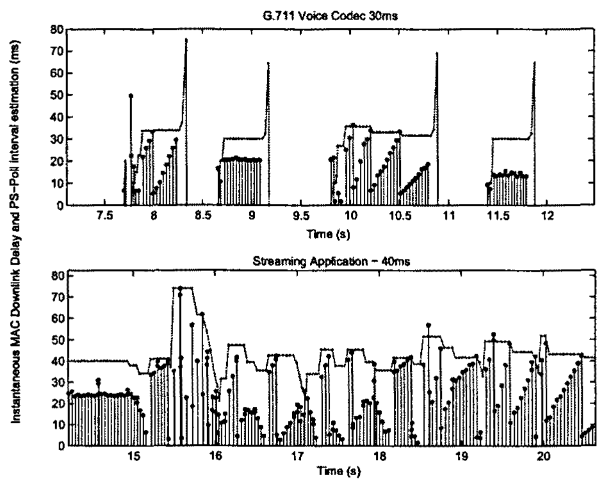

Figure 5.49: Dynamic adaptation of the APSM algorithm to the downlink frame interarrival time of real applications

each point represents a PS-Poll transmission) and the detection of the silence periods going back to doze mode. As a result, the soft upper bound downlink delay guarantee is provided (amplitude of deltas).

The adaptation to the variable data rate generated by the streaming application is depicted also in Figure 5.49. In this case, the variable bit rate of the application results in a variable downlink frame interarrival time, e.g., due to segmentation of large packets. The PS-Poll rate estimation successfully adapts to these changes while keeping around the expected average value of 40ms.

Regarding the effect of channel congestion over our algorithm, we study the mechanism robustness by increasing the number of wireless and corresponding wired stations in our scenario. In Figure 5.50 and 5.51 it can be appreciated that the APSM algorithm manages to improve the downlink delay and decrease the power consumption even in congestion conditions.

Impact of Traffic Mix

One of the main design objectives for our algorithm, as described in Section 5.10, is to keep the delay bound guarantee even in the case of more than one application per station (traffic mixed). In this experiment we analyze the effect over the APSM algorithm of adding Web traffic to the VoIP application used in the previous experiment for the case of 5 wireless stations. In this case we do not use silence suppression to facilitate the recognition of the web traffic burst over the voice traffic. The configuration used to emulate web traffic is page interarrival

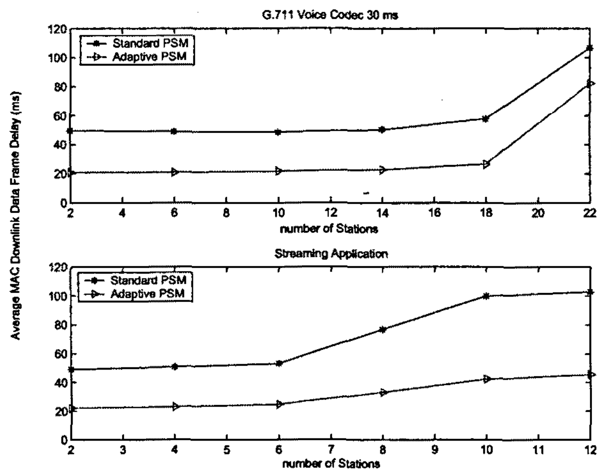

Figure 5.50: Impact of congestion in the channel on the MAC downlink delay

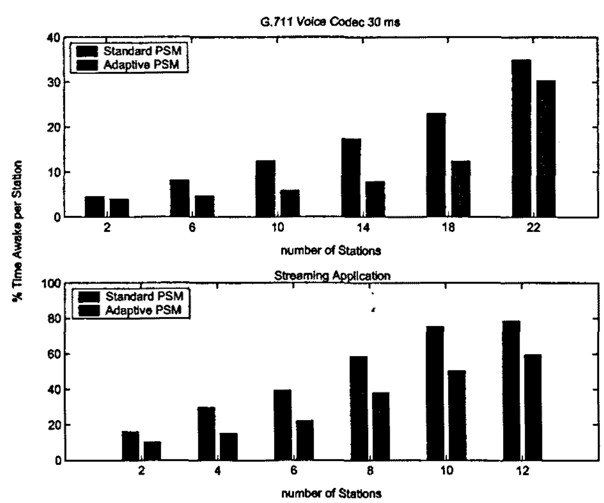

Figure 5.51: Impact of congestion in the channel on the power saving efficiency

time exponentially distributed with mean 10s and page size 10 KB plus 5 images of a size uniformly distributed between 0.5 and 2 KB.

As shown in Figure 5.52, the PS-Poll interval estimation of the algorithm (solid line) reacts as desired to the additional load decreasing the PS-Poll interval and therefore keeping the soft upper bound downlink delay guarantee of data frames (amplitude of deltas).

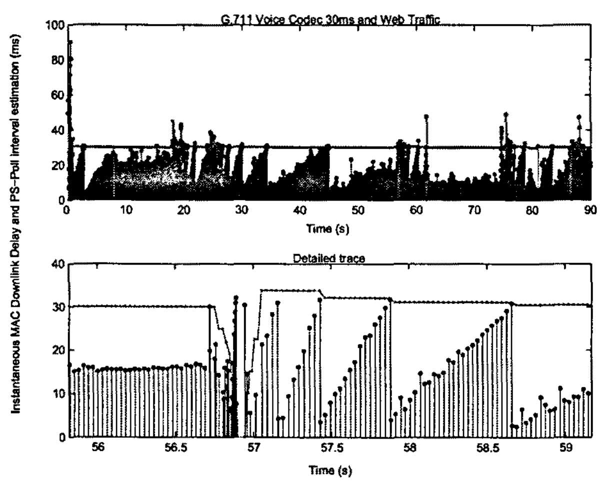

Figure 5.52: Dynamic adaptation of the APSM algorithm to the resulting downlink frame interarrival time produced by a voice and web application

5.10.2 Summary

The upcoming mobile devices with built-in wireless LAN technology require the usage of a power saving mechanism to ensure a reasonable battery duration. 802.11 provides a power save mode that reduces the power consumption but the MAC downlink delay introduced for data frames might result in an unacceptable QoS for some applications, e.g., VoIP.

In this document, we propose an adaptive power save mode algorithm that provides a soft upper bound of the data frames downlink delay (statistical guarantee) according to the downlink frame interarrival time. The decision for taking this approach is based in the observation that in general applications can cope with an end-to-end delay well above their frame generation rate. For instance, a delay sensitive application as VoIP, which codecs usually generate frames every 10-30ms, can deal with an overall delay of 150-300ms. Since in most of the networks including WLAN access the main contributor to the overall delay is usually the MAC layer, upper bounding the MAC downlink delay to the downlink frame interarrival time should satisfy the most stringent application requirements.

The power saving mechanism designed minimizes the impact over the standard power save mode to facilitate its implementation by re-using some of its mechanisms, e.g., beacon indication for starting the APSM algorithm or the More Data mechanism to correct disadjustments, by introducing a prediction algorithm of the PS-Poll rate which requires very simple operations and by using only MAC layer information.

The performance of our proposal has been analyzed and compared to 802.11

standard power save mode, a static proactive mechanism and a reactive one with respect to the resulting data frame MAC downlink delay, power saving efficiency and required signaling load. The study included the validation of the proper behaviour of APSM under congestion conditions, when used with realistic traffic sources as VoIP with silence suppression or streaming and also when used with traffic mixed from different applications for the same station. The results show that i) as expected, our algorithm fulfills its objective of providing an upper bound of the downlink delay according to the downlink frame interarrival time ii) the additional signaling load required in comparison to standard power save mode is negligible and iii) the power saving efficiency is increased because of the desynchronization of the PS-Polls transmissions.

5.11 Summary of Main Results

In the beginning of the chapter we pointed out the defficiencies of the 802.11 MAC layer to support QoS, provided an overview of the research activity performed to extend the MAC layer functionality and summarized the QoS mechanisms being developed by the 802.11e Working Group.

With our work we contributed to this effort by proposing three extensions aimed to provide delay guarantees 'DIME-EF' [114] and bandwidth guarantees, absolute 'ARME' or relative 'DWFQ' [115–118]. Additionally, we contributed to the 802.11e standardization effort [119] and studied the capabilities of EDCA in order to provide bandwidth guarantees [120].

DIME (DIffServ MAC Extension) is an extension of the 802.11 MAC protocol that provides support for Differentiated Services. The proposed extension consists of two optional modules: the Expedited Forwarding (EF) and the Assured Forwarding (AF). The Expedited Forwarding extension (DIME-EF) reuses the Interframe space of the Point Coordination Function of the IEEE 802.11 standard in a distributed manner, while the Assured Forwarding extension (ARME) relies on the Distributed Coordination Function with a modified algorithm for the computation of the Contention Window. Backwards compatibility for legacy 802.11 terminals is achieved by handling them as Best Effort users. We have shown that DIME-EF can meet the low delay requirements of real-time services when admission control is performed and that ARME provides bandwidth assurance to AF terminals in normal circumstances, while the leftover bandwidth is shared equally between Best Effort and AF. Furthermore, starving of Best Effort terminals is avoided in case of overload by trading off the bandwidth assurance of AF.

DWFQ (Distributed Weighted Fair Queing) is an extension of the 802.11 MAC layer that provides support for Weighted Fair Queuing. Our study has shown that

DWFQ achieves the desired bandwidth distribution among flows in a wide variety of scenarios. Because of the dynamic adaptation of the Contention Window, this bandwidth distribution is independent of the level of aggressiveness of the sources and their willingness to transmit. Furthermore, simulation results show that DWFQ avoids harming channel utilization in case of overload.

In addition to the study of mechanisms to extend the 802.11 MAC layer to provide QoS guarantees we analyzed the implications of introducing Wireless LAN capabilities in battery limited mobile devices as for example cellular phones [121]. The combination of the 802.11e QoS mechanisms with the 802.11 power save mode functionality was identified as a key element toward the introduction of WLAN functionality in such devices. The performance of the superposition of both mechanisms though, required to be studied to determine its suitability for the case considered. The main conclusions that can be drawn from our results are i) the 802.11e QoS mechanisms are influenced by the 802.11 power save mode functionality but are still highly effective ii) an important impact is observed on the downlink delay, as expected, where special attention has to be paid to the beacon interval configuration iii) the additional overhead introduced by the power save mode decreases considerably the effective bandwith available for data transmission in the wireless LAN and iv) in devices where power saving is a critical issue, the significant increase in the power usage efficiency in no congestion conditions justifies the power save mode costs.

Based on the results obtained regarding the impact of the 802.11 power save mode over the 802.11e QoS mechanisms we designed an Adaptive Power Save Mode (APSM) algorithm that provides a soft upper bound of the data frames downlink delay according to the downlink frame interarrival time while minimizing the battery consumption [122]. The performance of our proposal has been analyzed and compared to 802.11 standard power save mode, a static proactive mechanism and a reactive one with respect to the resulting data frame MAC downlink delay, power saving efficiency and required signaling load. The study included the validation of the proper behaviour of APSM under congestion conditions, when used with realistic traffic sources as VoIP with silence suppression or streaming and also when used with traffic mixed from different applications for the same station. The results show that i) as expected, our algorithm fulfills its objective of providing an upper bound of the downlink delay according to the downlink frame interarrival time ii) the additional signaling load required in comparison to standard power save mode is negligible and iii) the power saving efficiency is increased because of the desynchronization of the PS-Polls transmissions. This proposal has resulted in three patent applications [123–125].

Part of the work presented in this chapter has been used for the design and configuration of the NEC 3G/WLAN mobile terminal N900iL.

Chapter 6

Summary & Conclusions

During the last decades we have assisted to an increasingly faster evolution of communication networks pushed by the continuosly expanding number of users and their usage in different environments for a wider range of purposes. This evolution has been driven mainly by the high success of the Internet for wired networks based on packet switched technology (IP) and of 2G cellular systems for wireless networks based on circuit-switched technology. The fact that users have become used to the commodities of these communication networks is leading a convergence trend that should conclude with what is commonly referred to as *All-IP* networks. The path toward All-IP networks however, is still long and requires of the convergence of IP and wireless networks in order to fulfill the objective of providing in future mobile devices seamless services across heterogeneous networks.

In this thesis the state of the art for IP networks and the two most predominant wireless access networks, UMTS and Wireless LANs, has been reviewed with respect to the enhancements required toward the objective of supporting services across different network technologies in a seamless manner. Three main areas of research were identified as key in the path toward All-IP networks: i) IP-based mobility management, ii) IP-based UMTS Radio Access Networks and iii) QoS for Wireless LANs.

Future All-IP networks require of an IP-based mobility management protocol to support mobility across heterogeneous networks. Several protocols are being proposed in the IETF for this purpose with Mobile IPv6, Hierarchical Mobile IPv6 and Fast Handovers for Mobile IPv6 being the ones with a higher acceptance. Our contribution in this area has been the thorough study of the performance enhancements provided by each protocol and their associated costs in realistic scenarios to support the design process of Mobile IPv6-based networks where the suitability of the different options has to be evaluated. In addition, we proposed a combination of Hierarchical Mobile IPv6 and Fast Handovers for Mobile IPv6, based on the description provided in the Hierarchical Mobile IPv6 draft, that aims to outperform both proposals by adding the advantages of each of them. Our proposal has been evaluated against the performance and costs of the other ones yielding in most of the cases better results with respect to the handoff delay and packet losses metrics while keeping constant the signaling load introduced outside the local domain.

During the evaluation of the IP-based mobility management protocols the mobility model influence over the performance results led to an additional line of study where the main characteristics of the Random Waypoint mobility model were analyzed. The random waypoint model is a commonly used mobility model for simulations of wireless communication networks. We gave a formal description of this model in terms of a discrete– time stochastic process which provides a deep understanding on the characteristics of the mobility created by its usage. The movement duration and the cell change rate enable us to make a statement about the 'degree of mobility' of a certain simulation scenario. Knowledge of the spatial node distribution is essential for all investigations in which the relative location of the mobile nodes is important. Finally, the direction distribution explains in an analytical manner the effect that nodes tend to move back to the middle of the system area. The results of our work are of practical value for performance analysis of communication networks to avoid misinterpretation of simulation results.

UMTS networks being deployed today, based on the Release'99 specifications, use ATM transport in the Radio Access Network (RAN). The specifications for future releases include options to support IP transport in the RAN, however, several challenges are introduced with regard to a significant increase in the transport protocol overhead and the provision of QoS guarantees. These challenges have to be addressed to ensure an efficient usage of the scarce RAN resources. Our work focused first in reducing the resources needed in the air interface for the expected most relevant application in future 3G networks, Voice over IP (VoIP). We designed a radio access bearer for VoIP, when Robust Header Compression (RoHC) is used, that reduces in about 50% the resources required in the air interface. Then, a multiplexing scheme for the reduction of the overhead in the wired part of the RAN that allows QoS differentiated scheduling was described and evaluated. The results showed an increase in the efficiency of the RAN resources usage between 100% and 260%, in our scenarios, when combined with RoHC. Finally, we proposed and evaluated a QoS differentiated scheduling mechanism based on Earliest-Deadline-First that fulfills the RAN specific synchronization requirements while providing the QoS differentiation required by the applications. The method, which main novelty is to perform the scheduling at the RNC instead of at intermediate routers, results in a very significant improvement with respect to packets discarded at the Node-Bs due to late arrival.

The IEEE 802.11 standard presents relevant defficiencies in order to support applications with QoS requirements. The integration of the Wireless LAN technology in future All-IP networks heavily depends on the success of the upcoming 802.11e standard which extends the 802.11 MAC layer to provide QoS guarantees. With our work in this area we contributed to the 802.11e research standardization effort and designed three extensions of the 802.11 MAC layer to provide delay guarantees (DIME-EF) and bandwidth guarantees, absolute (ARME) or relative (DWFQ). DIME (DIffServ MAC Extension) is an extension of the 802.11 MAC protocol that provides support for Differentiated Services. The proposed extension consists of two optional modules: the Expedited Forwarding (EF) and the Assured Forwarding (AF). We have shown that DIME-EF can meet the low delay requirements of real-time services when admission control is performed and that ARME provides bandwidth assurance to AF terminals in normal circumstances, while the leftover bandwidth is shared equally between Best Effort and AF.

DWFQ (Distributed Weighted Fair Queing) is an extension of the 802.11 MAC layer that provides support for Weighted Fair Queuing. Our study has shown that DWFQ achieves the desired bandwidth distribution among flows in a wide variety of scenarios.

Moreover, we analyzed the implications of introducing Wireless LAN capabilities in battery limited mobile devices as for example cellular phones. The combination of the 802.11e QoS mechanisms with the 802.11 power save mode functionality was identified as a key element toward the introduction of WLAN functionality in such devices. The performance of the superposition of both mechanisms though, required to be studied to determine its suitability for the case considered. The main conclusions drawn from our results were that the 802.11e QoS mechanisms are influenced by the 802.11 power save mode functionality but are still highly effective and that in devices where power saving is a critical issue, the significant increase in the power usage efficiency in no congestion conditions justifies the power save mode costs.

Based on the significant impact observed of the 802.11 power save mode over the 802.11e QoS mechanisms we designed and evaluated a solution that provides a soft upper bound of the data frames downlink delay according to the downlink frame interarrival time while minimizing the battery consumption. The results of the proposed Adaptive Power Save Mode (APSM) algorithm showed that our algorithm fulfills its objective of upper bounding the downlink delay and that the power saving efficiency is increased. Part of this work has been used for the design and configuration of the NEC 3G/WLAN mobile terminal N900iL.

The work realized in this thesis contributes to the convergence of IP and wireless networks toward All-IP networks. Significant research effort though is still needed regarding the interaction of QoS and security protocols to ensure a secure, reliable and fast handoff user experience across administrative domains.

Chapter 7

Acknowledgments

This thesis is the result during the past years of intensive and exciting work combining the PhD studies at Universitat Politècnica de Catalunya and my job at NEC Network Laboratories. I sincerely thank my PhD advisor Sebastià Sallent for his guidance and constant support during the whole process, and for providing me with the opportunity of realizing the PhD while working in Germany. Special thanks go also to Heinrich Stuttgen, the General Manager of NEC Network Laboratories, for helping me in the development of my career.

I deeply thank my ex-colleague Hannes Hartenstein for encouraging me to start the PhD, providing uncountable wise advices and his collaboration in many research activities.

A warm thank to my ex-colleague Albert Banchs for introducing me in the research path and motivating me to continuosly evolve resulting in remarkable years of joint research writing several papers.

Thanks to Marc Torrent and Daniel Camps for their oustanding master theses and for very interesting discussions. Special thanks as well to all students I had the pleasure to work with, Kjell Heinze, Wasan Pattara and Anirban Chakarbarti, for their many contributions.

I further thank Christian Bettstetter for his collaboration in the study of the properties of mobility models and Markus Radimirsch with whom Albert and me designed some of the QoS extensions for wireless LANs.

I want to thank as well all the friends and colleagues which made of this period of my life an experience that I will never forget.

Finally, this thesis would not have been possible without the help of three persons. I thank my mother and my sister for their wholehearted support and believe in me throughout my entire life. Melanie, I thank you for your unconditional love and support which always brought me light in the dark moments.

Xavier Pérez Costa

List of Figures

Bibliography

[1] C.Perkins (Ed.), "IP Mobility Support for IPv4," RFC 3344, August 2002.

[2] D.B. Johnson, C.Perkins, and Jari Arkko (Eds.), "Mobility Support in IPv6," RFC 3775, June 2004.

[3] Andrew T. Campbell et al., "Comparison of IP Micro-mobility Protocols," IEEE Wireless Communications Magazine, Vol. 9, February 2002.

[4] H.Soliman, C.Castelluccia, K.El-Malki, and L.Bellier, "Hierarchical MIPv6 mobility management (HMIPv6)," Internet Draft, work in progress, June 2003.

[5] G.Tsirtsis, A.Yegin, C.Perkins, G.Dommety, K.El-Malki, and M.Khalil, "Fast Handovers for Mobile IPv6," Internet Draft, work in progress, March 2003.

[6] C.E.Perkins and K-Y.Wang, "Optimized Smooth Handoffs in Mobile IP," IEEE Symposium on Computers and Communications (ISCC), 1999.

[7] H.Yokota et al., "Link Layer Assisted Mobile IP Fast Handoff Method over Wireless LAN Networks," ACM International Conference on Mobile Computing and Networking, September 2002.

[8] C.Castelluccia, "HMIPv6: A Hierarchical Mobile IPv6 Proposal," ACM Mobile Computing and Communication Review (MC2R), April 2000.

[9] Nicola Montavont and Thomas Noel, "Handover Management for Mobile Nodes in IPv6 Networks," IEEE Wireless Communications Magazine, August 2002.

[10] R.Hsieh et al., "Performance analysis on Hierarchical Mobile IPv6 with Fast-handoff over End-to-End TCP," IEEE Global Telecommunications Conference (GLOBECOM), 2002.

[11] C.Perkins, "IP encapsulation within IP," RFC 2003, October 1996.

[12] T.Narten et al., "Neighbor Discovery for IP Version 6," RFC 2461, December 1998.

[13] H.Y Jung et al., "Fast Handover for Hierarchical MIPv6 (f-hmipv6)," Internet Draft, work in progress, June 2003.

[14] "Network Simulator (ns), version 2," http://www.isi.edu/nsnam/ns.

[15] C.Bettstetter, H.Hartenstein, and X.Pérez-Costa, "Stochastic Properties of the Random Waypoint Mobility Model," In ACM Kluwer Wireless Networks, special issue on Modeling & Analysis of Mobile Networks (WINET),vol. 10, no. 5, Sept 2004 (accepted Mar 2003, to be published).

[16] IEEE, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Standard 802.11, June 1999.

[17] C-N.Chuah et al., "QoS Provisioning Using a Clearing House Architecture," International Workshop on Quality of Service (IWQoS), June 2000.

[18] "Artificial conversational speech," ITU-T p.59, Recommendation, 1993.

[19] "H.263, video coding for low bitrate communication," ITU-T/SG15 Recommendation, 1996.

[20] F.H.P. Fitzek and M.Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation," IEEE Network,Vol. 15,No. 6,pages 40-54, November/December 2001.

[21] "Mobiwan: ns-2 extensions to study mobility in Wide-Area IPv6 Networks," http://www.inrialpes.fr/planete/mobiwan.

[22] W.Mendenhall et al., "Mathematical Statistics with Applications, p.368," 4th Ed.PWS-KENT, 1990.

[23] S. Thomson and T.Narten, "IPv6 Stateless Address Autoconfiguration," RFC 2462, December 1998.

[24] D.Tang and M.Baker, "Analysis of a Local-Area Wireless Network," In Proceedings of the ACM Intern. Conf. on Mobile Comp. and Netw. (MOBICOM), 2000.

[25] X.Pérez-Costa and H.Hartenstein, "A Simulation Study on the Performance of Mobile IPv6 in a WLAN-Based Cellular Network," Computer Networks, special issue on "Towards a New Internet Architecture", September 2002.

[26] X.Pérez-Costa, R.Schmitz, H.Hartenstein, and M.Liebsch, "A MIPv6, FMIPv6 and HMIPv6 Handover Latency Study: Analytical Approach," In Proc. IST Mobile & Wireless Telecommunications Summit, Thessaloniki, Greece, June 2002.

[27] X.Pérez-Costa and M.Torrent-Moreno, "A Performance Study of Hierarchical Mobile IPv6 from a System Perspective," In Proceedings of IEEE International Conference on Communications (ICC), May 2003.

[28] X.Pérez-Costa, M.Torrent-Moreno, and H.Hartenstein, "A Simulation Study on the Performance of Hierarchical Mobile IPv6," in Proceedings of the International Teletraffic Congress (ITC), Berlin, Germany, August 2003.

[29] M.Torrent-Moreno, X.Pérez-Costa, and S.Sallent-Ribes, "A Performance Study of Fast Handovers for Mobile IPv6," in Proceedings of IEEE Local Computer Networks (LCN), September 2003.

[30] X.Pérez-Costa, M.Torrent-Moreno, and H.Hartenstein, "A Performance Comparison of Mobile IPv6,Hierarchical Mobile IPv6, Fast Handovers for Mobile IPv6 and their Combination," ACM SIGMOBILE Mobile Computing and Communications Review (MC2R), Volume 7, Issue 4, pp. 5-19, October 2003.

[31] "Mobility and Differentiated Services in a Future IP Network," IST-2000-25394, 2000.

[32] R. A. Guérin, "Channel occupancy time distribution in a cellular radio system," *IEEE Trans. Vehicular Techn.*, vol. 36, no. 3, Aug. 1987.

[33] J. G. Markoulidakis, G. L. Lyberopoulos, D. F. Tsirkas, and E. D. Sykas, "Mobility modeling in third–generation mobile telecommunication systems," *IEEE Pers. Comm.*, Aug. 1997.

[34] M. M. Zonoozi and P. Dassanayake, "User mobility modeling and characterization of mobility patterns," *IEEE Journal on Sel. Areas in Comm.*, vol. 15, no. 7, pp. 1239–1252, Sept. 1997.

[35] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang, "A group mobility model for ad hoc wireless networks," in *Proc. ACM Intern. Workshop on Modeling, Analysis, and Sim. of Wireless and Mobile Systems (MSWiM)*, Seattle, WA, 1999.

[36] C. Bettstetter, "Smooth is better than sharp: A random mobility model for simulation of wireless networks," in *Proc. ACM Intern. Workshop on Modeling, Analysis, and Sim. of Wireless and Mobile Systems (MSWiM)*, Rome, Italy, July 2001.

[37] D. Helbing, "Traffic and related self–driven many-particle systems," *Reviews of Modern Physics*, vol. 73, no. 4, pp. 1067–1141, Oct. 2001.

[38] D. Heidemann, "Ein mathematisches Modell des Verkehrsflusses," *Schriftenreihe des Instituts fr Verkehrswesen, Universitt Karlsruhe, Germany*, vol. 42, 1989.

[39] L. Briesemeister and G. Hommel, "Role–based multicast in highly mobile but sparsely connected ad hoc networks," in *Proc. Workshop on Mobile Ad Hoc Networking and Computing (MobiHoc)*, Boston, MA, Aug. 2000.

[40] E. Jugl, *Mobilittsmodellierung und Einflsse auf Systemparameter von Mobilfunksystemen*, Ph.D. thesis, TU Ilmenau, 2000.

[41] D. Lam, D. C. Cox, and J. Widom, "Teletraffic modeling for personal communication services," *IEEE Communications*, vol. 35, no. 2, pp. 79–87, Oct. 1997.

[42] "Network Simulator (ns), version 2," http://www-mash.cs.berkeley.edu/ns.

[43] "Glomosim," http://pcl.cs.ucla.edu/projects/glomosim/, Nov. 2002.

[44] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *In Proc. ACM Intern. Conf. on Mobile Comp. and Netw. (MOBICOM)*, Dallas, TX, Oct. 1998.

[45] S. R. Das, C. E. Perkins, and E. M. Royer, "Performance comparison of two on-demand routing protocols for ad hoc networks," in *Proc. IEEE Infocom*, Tel Aviv, Israel, 2000.

[46] G. Holland and N. H. Vaidya, "Analysis of TCP performance over mobile ad hoc networks," in *In Proc. ACM Intern. Conf. on Mobile Comp. and Netw. (MOBICOM)*, Seattle, WA, Aug. 1999.

[47] D. M. Blough, G. Resta, and P. Santi, "A statistical analysis of the long-run node spatial distribution in mobile ad hoc networks," in *Proc. ACM Intern. Workshop on Modeling, Analysis, and Sim. of Wireless and Mobile Systems (MSWiM)*, Atlanta, GA, Sept. 2002.

[48] G. Resta and P. Santi, "An analysis of the node spatial distribution of the random waypoint model for ad hoc networks," in *Proc. ACM Workshop on Principles of Mobile Computing (POMC)*, Toulouse, France, Oct. 2002.

[49] J. Song and L. E. Miller, "Empirical analysis of the mobility factor for the random waypoint model," in *Proc. OPNETWORK*, Washington, DC, Aug. 2002.

[50] P. Santi and D. M. Blough, "An evaluation of connectivity in mobile wireless ad hoc networks," in *Proc. IEEE Intern. Conf. on Dependable Systems and Networks (DSN)*, Washington, DC, June 2002.

[51] T. Chu and I. Nikolaidis, "On the artifacts of random waypoint simulations," in *Proc. Intern. Workshop on Wired/Wireless Internet Comm. (WWIC)*, Las Vegas, NV, June 2002.

[52] C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," Tech. Rep., Technische Universität München, Institute of Communication Networks, Munich, Germany, Sept. 2002.

[53] E. M. Royer, P. M. Melliar-Smith, and L. E. Moser, "An analysis of the optimum node density for ad hoc mobile networks," in *Proc. IEEE Intern. Conf. on Comm. (ICC)*, Helsinki, Finland, 2001.

[54] C. Bettstetter, "Mobility modeling in wireless networks: Categorization, smooth movement, and border effects," *ACM Mobile Comp. and Comm. Review*, vol. 5, no. 3, July 2001.

[55] B. Ghosh, "Random distances within a rectangle and between two rectangles," *Bull. Calcutta Math. Soc.*, vol. 43, pp. 17–24, 1951.

[56] J. M. Hammersley, "The distribution of distance in a hypersphere," *Annals of Math. Statistics*, vol. 21, pp. 447–452, 1950.

[57] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*, McGraw–Hill, 4th edition, 2002.

[58] E. Weisstein, Ed., *Eric Weisstein's World of Mathematics*, chapter Probability and Statistics: Statistical Distributions, Wolfram Research, online edition, Dec. 2002.

[59] C. Bettstetter and C. Wagner, "The spatial node distribution of the random waypoint mobility model," in *Proc. German Workshop on Mobile Ad Hoc Networks (WMAN)*, Ulm, Germany, Mar. 2002.

[60] J. Li, J. Jannotti, D. De Couto, D. Karger, and R. Morris, "A scalable location service for geographic ad hoc routing," in *In Proc. ACM Int. Conf. on Mobile Comp. and Netw. (MOBICOM)*, Boston, MA, 2000.

[61] Chr.Bettstetter, H.Hartenstein, and X.Pérez-Costa, "Stochastic Properties of the Random Waypoint Mobility Model," ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM), 2002.

[62] X.Pérez-Costa, H.Hartenstein, and C.Bettstetter, "Towards a Mobility Metric for Reproducible and Comparable Results in Ad Hoc Networks Research," In Proc. ACM Intern. Conf. on Mobile Comp. and Netw. (MOBICOM) poster session, San Diego, USA, September 2003.

[63] X.Pérez-Costa, H.Hartenstein, and C.Bettstetter, "Towards a Mobility Metric for Reproducible and Comparable Results in Ad Hoc Networks Research," ACM SIGMOBILE Mobile Computing and Communications Review (MC2R), special feature on MobiCom 2003 posters, Volume 7, Issue 4, pp. 58-60, October 2003.

[64] "3GPP, IP Transport in the UTRAN Work Task Technical Report," 3GPP TR 25.933 V5.3.0, 2003.

[65] 3GPP TR 25.933, "IP Transport in UTRAN Technical Specification Group Radio Access Network," version 5.4.0, December 2003.

[66] "IP in the RAN as a Transport Option in 3rd Generation Mobile Systems," Mobile, Wireless Internet Forum, MTR-006, v2.0.0, 2001.

[67] K. Venken, D. De Vleeschauwer, and J. De Vriendt, "Dimensioning the Backbone Network for an IP-based UTRAN," in *Proceedings of the 2001 International Conference on Third Generation Wireless and Beyond*, San Francisco, CA, May 2001.

[68] K. Venken, D. De Vleeschauwer, and J. De Vriendt, "Designing a Diffserv-capable IP-Backbone for the UTRAN," in *Proceedings of the Second International Conference on 3G Mobile Communication Technologies*, March 2001.

[69] S. K. Kasera ad R. Ramjee, S. Thuel, and X. Wang, "Congestion Control Policies for IP-based CDMA Radio Access Networks," in Proceedings of INFOCOM, March 2003.

[70] 3GPP TR 34.128, "Common test environments for User Equipment (UE) conformance testing," version 4.7.0, June 2003.

[71] 3GPP TS 26.074, "AMR Speech codec speech processing functions, AMR Speech Codec Test Sequences," version 5.0.0, June 2002.

[72] S. Casner and V. Jabobson, "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links," RFC 2508, Februrary 1999.

[73] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," RFC 1889, January 1996.

[74] 3GPP TR 26.236, "Packet switched conversational multimedia applications; Transport protocols," version 5.3.0, June 2003.

[75] 3GPP TR 26.937, " Transparent end-to-end packet switched streaming service (PSS); Real-time Transport Protocol (RTP) usage model," version 1.5.0, June 2003.

[76] C. Bormann et al, "RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed," RFC 3095, July 2001.

[77] "Arrows Project Web page," `http://www.arrows-ist.upc.es`.

[78] J. J. Olmos and J. Perez-Romero, "Simulation tools presentation," `http://www.arrows-ist.upc.es/publications/workshop/UTRA-FDDSimulators.pdf`.

[79] 3GPP TR 25.942, "RF system scenarios," version 6.0.0, January 2003.

[80] "OPNET Technologies, Inc.," `http://www.opnet.com`.

[81] S. Shenker, C. Partridge, and R. Guerin, "Specification of Guaranteed Quality of Service," RFC 2212, September 1997.

[82] "Acticom," http://www.acticom.de.

[83] K.Venken, I.Gomez Vinagre, and J.De Vriendt, "Analysis of the Evolution to an IP-based UMTS Terrestrial Radio Access Network," IEEE Wireless Communications Magazine, October 2003.

[84] C. Fox R. Pazhyannur, I. Ali, "PPP Multiplexing," IETF RFC 3153, August 2001.

[85] "3GPP, Group Services and System Aspects; Services and Service Capabilities," 3GPP TS 22.105 V5.2.0, June 2002.

[86] "3GPP, Synchronisation in UTRAN stage 2," 3GPP TS 25.402 V5.2.0, 2003.

[87] Grenville Armitage, *Quality of Service in IP Networks*, Macmillan Technical Publishing, Chapter 3, pp.97-98, 2000.

[88] E.Biton and A.Orda, "QoS Provision with EDF Scheduling, Stochastic Burstiness and Stochastic Guarantees," in Proceedings of the International Teletraffic Congress (ITC), Berlin, Germany, August 2003.

[89] 3GPP TR 25.853, "Delay Budget within the Access Stratum," version 4.0.0, March 2001.

[90] X.Pérez-Costa, A.Banchs, J.Noguera, and S.Sallent-Ribes, "Optimal Radio Acess Bearer Configuration for Voice over IP in 3G UMTS Networks," In Proceedings of the 5th European Wireless, Barcelona, February 2004.

[91] X.Pérez-Costa, K.Heinze, S.Sallent-Ribes, and A.Banchs, "Study of Efficiency issues in an IP-based UTRAN," Submitted, January 2005.

[92] X.Pérez-Costa and A.Banchs, "Differentiated Downlink Packet Scheduling for an IP-based UTRAN," NEC-E Invention Disclosure Nr. NLE-13-03, July 2003.

[93] T. Nandagopal, S. Lu, and V. Bharghavan, "A Unified Architecture for the Design and Evaluation of Wireless Fair Queuing Algorithms," in *In Proc. ACM Intern. Conf. on Mobile Comp. and Netw. (MOBICOM)*, Seattle, WA, August 1999.

[94] S. Lu, V. Bharghavan, and R. Srikant, "Fair Scheduling in Wireless Packet Networks," in *In Proceedings of ACM SIGCOMM*, Cannes, France, August 1997.

[95] M. Barry, A. Veres, and A. T. Campbell, "Distributed Control Algorithms for Service Differentiation in Wireless Packet Networks," in *Proceedings of INFOCOM*, Anchorage, Alaska, April 2001.

[96] A. Ayyagari, Y. Bernet, and T. Moore, "IEEE 802.11 Quality of Service - White Paper," IEEE 802.11-00/028.

[97] A. Imad and C. Castelluccia, "Differentiation Mechanisms for IEEE 802.11," in *Proceedings of INFOCOM*, Anchorage, Alaska, April 2001.

[98] J.L Sobrinho and A.S. Krishnakumar, "Real-Time Traffic over the IEEE 802.11 Medium Access Control Layer," *Bell Labs Technical Journal*, Autumn 1996.

[99] C.-S. Hsu Y.-C. Tseng and T.-Y. Hsieh, "Power-saving protocols for ieee 802.11-based multi-hop ad hoc networks," in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies*, 2002.

[100] R.Zheng and R.Kravets, "On-demand power management for ad hoc networks," in *Proceedings of IEEE INFOCOM*, March 2003.

[101] R.Krashinsky and H.Balakrishnan, "Minimizing energy for wireless web access with bounded slowdown," in *Proceedings of the eighth Annual International Conference on Mobile Computing and Networking (MOBICOM)*, september 2002.

[102] E.Shih, P.Bahl, and M.J.Sinclair, "Wake on wireless: An event driven energy saving strategy for battery operated devices," in *Proceedings of the eighth Annual International Conference on Mobile Computing and Networking (MOBICOM)*, september 2002.

[103] F.A. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part I - Carrier Sense Multiple-Access Modes and their throughput delay characteristics," *IEEE Trans. on Communications*, vol. 23, no. 12, pp. 1417–1433, 1975.

[104] F.A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II - the Hidden Terminal Problem in Carrier Sense Multiple-Access Modes and the Busy-Tone Solution," *IEEE Trans. on Communications*, vol. 23, no. 12, pp. 1417–1433, 1975.

[105] C. Fullmer and J. J. Garcia-Luna-Aceves, "Floor acquisition multiple access (fama) for packet radio networks," Computer Communication Review,

vol. 25, (no. 4),(ACM SIGCOMM '95, Cambridge, MA, USA, 28 aug.-1 Sept. 1995) ACM, 1995.

[106] M. A. Visser and M. E. Zarki, "Voice and Data transmission over an 802.11 Wireless network," in *Proceedings of PIMRC*, Toronto, Canada, September 1995.

[107] "Wireless Fidelity (Wi-Fi) Alliance," http://www.wi-fi.com.

[108] S.Mangold, S.Choi, G.R.Hiertz, O.Klein, and B.Walke, "Analysis of IEEE 802.11e for QoS Support in Wireless LANs," IEEE Wireless Communications Magazine, December 2003.

[109] ETSI, "Broadband radio access networks (BRAN); HIgh PERformance Radio Local Area Network (HIPERLAN) Type 1; Functional Specification," European Norm 300 652 (V1.2.1), ETSI, 1998.

[110] S. Chevrel et al., "Analysis and Optimisiation of the HIPERLAN Channel Access Contention Scheme," Wireless Personal Communications 4, pp. 27-39, Kluwer Acadamic Publishers, 1997.

[111] J. Wozniak, "Performance Analysis of IEEE 802.11 and ETSI HIPERLAN Networks in the Presence of Hidden Stations," *Systems-Science*, vol. 24, no. 14, pp. 81–96, 1998.

[112] D. D. Clark and W. Fang, "Explicit Allocation of Best Effort Packet Delivery Services," *IEEE/ACM Transactions on Networking*, vol. 6, no. 4, pp. 362–373, August 1998.

[113] J. Ibanez and K. Nichols, "Preliminary Simulation Evaluation of an Assured Service," draft-ibanez-diffserv-assured-eval-00.txt, Internet draft, August 1998.

[114] A. Banchs, M. Radimirsch, and X. Pérez, "Assured and Expedited Forwarding Extensions for IEEE 802.11 Wireless LAN," In Proceedings of the tenth IEEE/IFIP International Workshop on Quality of Service (IWQoS 2002), Miami, USA, May 2002.

[115] A. Banchs, X. Pérez, M. Radimirsch, and S. Sallent, "Service Differentiation Extensions for IEEE 802.11," in *LANMAN*, Boulder, CO, March 2001.

[116] A. Banchs, X. Pérez, M. Radimirsch, and H. Stuettgen, "Service Differentiation Extensions for Elastic and Real-Time Traffic in 802.11 Wireless

LAN," in *High Performance Switching and Routing (HPSR)*, Dallas, Texas, May 2001.

[117] A. Banchs and X. Pérez, "Providing Throughput Guarantees in IEEE 802.11 Wireless LAN IEEE 802.11 Wireless LAN," in *In Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 2002)*, Orlando, FL, March 2002.

[118] A. Banchs and X. Pérez, "Distributed Weighted Fair Queuing in 802.11 Wireless LAN," in *In proceedings of the IEEE International Conference on Communications (ICC2002)*, New York, USA, April 2002.

[119] A.Banchs, X.Pérez, W.Pokorski, and M.Radimirsch, "A Proposal for Wireless MAC Multimedia Extensions," IEEE 802.11-00/205, July 2000.

[120] A.Banchs, X.Pérez-Costa, and D.Qiao, "Providing Throughput Guarantees in IEEE 802.11e Wireless LANs," in Proceedings of the International Teletraffic Congress (ITC), Berlin, Germany, August 2003.

[121] X.Pérez-Costa, D.Camps-Mur, and T.Sashihara, "Analysis of the Integration of IEEE 802.11e Capabilities in Battery Limited Mobile Devices," IEEE Wireless Communications Magazine, to be published.

[122] X.Pérez-Costa and D.Camps-Mur, "APSM: Bounding the Downlink Delay for 802.11 Power Save Mode," In Proceedings of IEEE International Conference on Communications (ICC) (accepted Dec. 2004, to be published).

[123] X.Pérez-Costa, "Adaptive Power Save Mode: Extension of IEEE 802.11 Power Save Mode ," NEC-E Invention Disclosure Nr.NLE-25-04.

[124] X.Pérez-Costa, "No Data Acknowledgment (NDACK)," NEC-E Invention Disclosure Nr.NLE-27-04.

[125] X.Pérez-Costa and D.Camps-Mur, "Access Category Extension of the Access Point Power Save Mode Queue," NEC-E Invention Disclosure Nr.NLE-29-04.