A QoS-enabled MAC Architecture for Prioritized Service in IEEE 802.11 WLANs

Wei Liu[1] Wenjing Lou[2] Xiang Chen[1] Yuguang Fang[1]

[1]Wireless Networks Laboratory (WINET)
Department of Electrical and Computer Engineering
University of Florida, Gainesville, FL 32611
{liuw, xchen@ecel., fang@ece.}@ ufl.edu

[2]Department of Electrical and Computer Engineering
Worcester Polytechnic Institute, Worcester, MA 01609
wjlou@ece.wpi.edu

Abstract **–In this paper we propose a novel media access control (MAC) architecture to support differentiated service in IEEE 802.11 WLAN. By employing the MAC-core as base and different adaptors as add-ons in this architecture, the resulting MAC can provide prioritized services with different delays and throughputs. Our simulation studies show that the resulting MAC protocols based on this architecture can achieve low access delay and high throughput for high priority traffic while maintaining fairness.**

Index terms **—Quality of Service (QoS), MAC, Wireless LANs**

I. INTRODUCTION

WLANs have been widely used and deployed in recent years for its flexibility and non-wired capability. As the interest of the integrated multimedia services over high-speed networks including wireless LANs is steadily increasing, it is natural for us to expect that the WLANs support real-time applications with Quality of Service (QoS) guarantees as the same as its wired counterpart. However, the IEEE 802.11 medium access control (MAC) protocol [1], the most commonly used MAC in the current WLANs, was designed for asynchronous best-effort data services. Without proper modifications, it does not appear satisfactory to provide QoS guarantees for real-time applications. Since the channel bandwidth in the wireless environment is limited, the general strategy to support QoS is to set up some kind of priority scheme or differentiation mechanism, under which the delay-sensitive traffic can have higher priority to access the channel than the less time-critical traffic. In recent years, considerable research efforts have been expended in tackling the QoS problem in the MAC layer. Since IEEE 802.11 PCF mode is designed for real-time services, some efforts have been made on the PCF mode operations. In [2] [3], the authors proposed several polling schemes for the PCF mode, and in [4] the authors adopted some call admission control mechanisms to support real-time traffics with certain QoS guarantees. Though the PCF mode with centralized polling can provide some support for real-time service, we believe that the distributed MAC with QoS is more flexible and effective than the centralized MAC, as the dominant operational mode in IEEE 802.11 LANs is the DCF mode, i.e., the distributed MAC protocol. Another drawback of the PCF mode is that the PCF is an optional mechanism and is not supported by most current wireless cards. Some studies [5][6] also show that, compared with other schemes, the PCF mode performs poorly either alone or when incorporated with the DCF mode. Some researchers proposed some new distributed MAC protocols to address the QoS issue. The Balckbust in [7] is a novel mechanism that provides high priority for real-time traffic with good performance in terms of throughput and access delay. Unfortunately, it imposes special requirement on high priority traffic, which violates the promise that any enhancement to the standard must be fully compatible with the existing 802.11 standard. We argue that the DCF mode can be extended to support differentiated service for its easy implementation and high efficiency of medium sharing. From the IEEE 802.11 specification for DCF mode, we observe that there are several aspects we can make some modifications in order to support the differentiated service with QoS guarantees while at the same time maintaining the backward compatibility.

- *Minimum and Maximum Contention Window Size*: To avoid collision, the DCF mode randomly chooses the backoff time (*BT*) from the interval (0,*CW*), where *CW* is the contention window size and *CW* is in the interval of [*CWmin, CWmax*].
- *Backoff Increasing Factor*: It is a scaling factor for *CW*. In DCF mode, when a collision is detected (no acknowledgement is received), *CW* will be scaled by a backoff-increasing factor and new backoff time (*BT*) will be chosen again.
- *Interframe Space*: It is the time that the stations need to wait until it can send any frame. In DCF, DIFS is used before sending RTS/DATA frame and SIFS is used before sending ACK/CTS frame.
- *Backoff Time Distribution*: It is the distribution of how the *BT* is chosen from the interval (0,*CW*).
- *Frame Length*: it is the maximum frame length that the station is allowed to transmit each time.

By properly assigning different *CWmin*, *CWmax, Backoff Increasing Factor* and *Frame Length*, defining different *Interframe Space* and *Backoff Time Distribution* function according to the traffic priority, we may provide QoS support for higher priority traffic. We notice that modification of even one aspect would lead to a new access mechanism. Some researches [8][9] only studied the performance of one of such aspects. In [10] the authors studied a combination of second

 0-7803-7974-8/03/$17.00 © 2003 IEEE

and the fourth aspects for Ad Hoc networks, but not for WLANs. In [11] IEEE 802.11e task group proposed the enhanced DCF (EDCF), which is the combination of the first and the third aspects. From the literature, each modified MAC protocol resulting from the above aspects is shown to be more effective by the respective designer. However, it is uncertain how these different aspects compare with each other on a unified basis. There are some related researches attempting to improve the channel utilization, throughput or efficiency. A fast collision resolution (FCR) scheme is recently proposed by us to improve the throughput [12]. Other researches [13][14] show improvement by taking into consideration channel states, energy consumption, security, and other factors. To our knowledge, no effort has been made to study all of the schemes under a unified architecture, or to investigate their pros and cons.

The rest of this paper is organized as follows. We first propose a QoS-enabled MAC architecture and place all the aforementioned aspects into the architecture in Section II. In Section III, we evaluate their performance on a unified basis. Moreover, the cooperation between different mechanisms is investigated. Finally, concluding remarks are given in Section IV.

II. QoS-enabled MAC Architecture

We assume that service differentiation mechanism is employed at upper layer, and traffics are assigned with different priorities according to different QoS requirements. We propose a new architecture to address the QoS in the MAC layer as shown in Figure 1. In this architecture, the MAC-core is the fundamental part where the basic functions of IEEE 802.11 are implemented. The plug-in adaptors are add-ons, in which different mechanism is defined and implemented. MAC core also coordinates the basic MAC functions and the mechanisms defined in the adaptors. One key feature of the architecture is its extensibility that the adaptor can be defined and implemented according to the QoS requirements, and more adaptors can be plugged in on demand.

Next we describe some adaptors used in the architecture as shown in Figure 1. Contention Window Size Adaptor (*CWA*) defines the minimum contention window size and the maximum contention window size for different priorities. Typically the high priority traffic may have smaller window size, so that the high priority traffic has better chance to have small backoff time (*BT*), leading to high probability to seize the medium over low priority traffic. Backoff Factor Adaptor (*BFA*) defines different backoff-increasing factors according to the priorities. Similarly, we often let high priority traffic have small values, so that once collision occurs, high priority traffics will choose smaller contention window sizes and have better chance to seize the medium than the low priority ones. Backoff Distribution Adaptor (*BDA*) defines different backoff time distributions for different priority services. By using different distributions, statistically high priority traffic may have better chance to choose small backoff time (*BT*) than the low priority traffic. Inter-frame Space Adaptor (*IFSA*) defines different inter-fame spaces for different priority services. Similar to the PCF mode using the shorter PIFS, high priority traffic may have smaller IFS than the low priority traffic, thus the high priority traffic can have greater chance to seize the medium. Frame Length Adaptor (*FLA*) defines the maximum frame length for different priority traffic. We may let high priority traffic have longer frame length than the low one, so that the high priority traffic may consume more bandwidth and have higher throughput than the low one. Collision Resolution Adaptor (*CRA*) is the adaptor that employs the mechanism such as FCR we proposed in [12] to quickly resolve the collision or avoid possible future collisions. With the fast collision resolution mechanism, we can reduce the number of collisions and the number of idle slots due to backoff, which in turn will improve the overall throughput. Different priority traffic may have different collision resolution mechanisms. In FCR, we demand all the stations in collisions and in the deferring state to increase their *CWs* and choose another backoff time (*BT*) once the start of a new busy period is sensed. Here, for example, when collision occurs, we may let the station in deferring mode with high priority traffic not take any action, while let those with low priority traffic carry out the FCR procedures. Thus, intuitively, those stations with high priority services may have better chance to seize the medium. In order to provide fairness among all stations, in FCR, a transmission limit is set, and after a number of consecutively successful transmissions, the station should choose *CWmax* as its contention window size and carry out the backoff procedures [12]. Channel State Adaptor (*CSA*) defines the characteristics or operations for different priorities in specific channel state. When the channel is good, we may allow the station that seize the medium to transmit longer, so that the number of slots wasted due to backoff is greatly reduced and the overall throughput improved in turn. Buffer State Adaptor (*BSA*) defines the characteristics or operations for different priority in specific buffer state. For example, in order to maintain some fairness between the different priorities, when the buffer of the low priority approaches to some predefined threshold, the station may allow the low priority service seize the channel for a while to avoid buffer overflow.

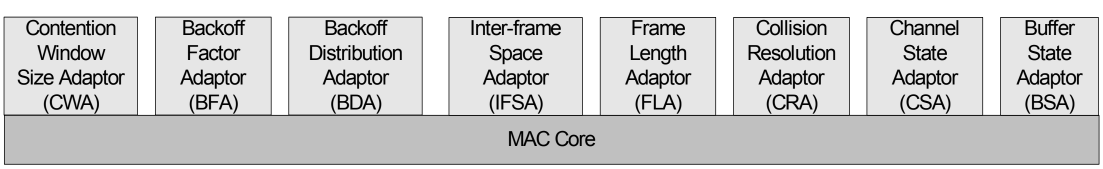

Figure 1. QoS-MAC Architecture

All adaptors are not independent, and may affect each other. Hence interactive design of deployed adaptors is necessary. By employing different adaptor settings or enabling/disabling some adaptors, we can fine-tune the MAC protocol and make it more effective. When there is no differentiation, $CW_{min}=15$; $CW_{max}=1023$; backoff increasing factor is 2; backoff distribution is chosen to be the uniform distribution; and the typical DIFS is used. The resulting MAC protocol will be the same as IEEE 802.11 for FHSS. The parameters in the various adaptors may have (but are not limited to) a fixed value; they can be some functions of the traffic load and other parameters of other adaptors. For example, the frame length may be a function of the channel state and buffer state.

In order to favor the high priority service, we may set smaller value for high priority service. We assume that the *j*th priority level is higher than *i*th priority level if $j>i$ .We may have $CWmin_j < CWmin_i$, $CWmax_j < CWmax_i$, $DIFS_j < DIFS_i$, and $BIF_j < BIF_i$. For the backoff time distribution, we may choose exponential distribution for the *j*th priority level and uniform distribution for the *i*th priority level, since the exponential distribution may have higher probability to have small value than the uniform distribution. All the above settings aim to give high priority service small backoff time, in this way the station with high priority service statistically seizes the medium with high probability.

III. Performance Evaluation

A. Simulation setup

To evaluate the performance of the proposed architecture, we use the OPENT to simulate the IEEE802.11 DCF functions. We modify the wireless LAN model in the OPENT into an extensible architecture by implementing the *CWA, BFA, IFSA, BDA and RCA***.** In order to study the effects of each adaptor to support prioritized service, we change the setting of adaptors to study the performance. In our simulations, we compare 6 different set of MAC protocols including the IEEE 802.11 MAC, which are described as follows:

- S_0: the IEEE 802.11 MAC. CWA ($CW_{min}=15, CW_{max}=1023$), BFA (*BIF=2*), BDA (*Uniform distribution*), IFSA (*DIFS*).
- S_1: IEEE 802.11+FCR. CWA ($CW_{min}=15, CW_{max}=1023$), BFA (*BIF=2*), BDA (*Uniform distribution*), IFSA (*DIFS*), CRA (*FCR*).
- S_2: IEEE 802.11+BD. *CWA* ($CW_{min}=15, CW_{max}=1023$), BFA (*BIF=2*), BDA (*distributions are defined as in Fig. 2*), IFSA (*DIFS*).
- S_3: IEEE 802.11+IFS. CWA ($CW_{min}=15, CW_{max}=1023$), BFA (*BIF=2*), BDA (*Uniform distribution*), IFSA ($DIFS_j=DIFS+slottime*(2-j), j=2, 1, 0$).
- S_4: IEEE 802.11+BIF. CWA ($CW_{min}=15, CW_{max}=1023$), BFA ($BIF_2=2, BIF_1=3, BIF_0=4$), BDA (*Uniform distribution*), IFSA (*DIFS*).
- S_5: IEEE 802.11+CW. CWA ($CW_{min2}=7$, $CW_{max2}=511$, $CW_{min1}=15$, $CW_{max1}=1023$, $CW_{min0}=32$, $CW_{max0}=2047$), BFA (*BIF=2*), BDA (*Uniform distribution*), IFSA (*DIFS*).
- S_6: IEEE 802.11 + IFS +BD+ FCR, CWA ($CW_{min}=15, CW_{max}=1023$), BFA (*BIF=2*), BDA (*distributions are defined as in Fig. 2*), IFSA ($DIFS_j=DIFS+slottime*(2-j), j=2,1,0$), and CRA (*FCR*).

According to the above definitions, each MAC, from S_1 to S_5, has only one difference from S_0 (IEEE 802.11 MAC) so that we can distinguish each individual adaptor's role in supporting the prioritized services. S_6 is a combination of S_1, S_2, and S_3. From S_6 we can see how different mechanisms work together to support prioritized service.

We simulate a WLAN consisting of 10 stations. At upper layer, we differentiate the incoming traffic (with arrival rate λ) with equal probability into three different priority levels ($P_2> >P_1>P_0$), where P_2 represents the highest priority. In our simulation, we use FHSS with the following basic MAC settings as shown in Table I.

TABLE I. SIMULATION PARAMETER SETTINGS

Parameters	**Values**
DIFS	128μs
SIFS	50μs
data rate	1Mbps
packet size	4096bits
slot time	28μs
maximum retry limit	7

B. Performance Results

First we compare the efficiency and the throughout of different schemes except S_6. Fig. 3.a shows the overall throughput and Fig. 3.b shows the overall efficiency. Here we define the efficiency as the ratio of the total number of packets successfully transmitted to the total number of packets received from the upper layer. Fig. 3c-3.e give the efficiency for the three priorities. We observe that the overall throughput

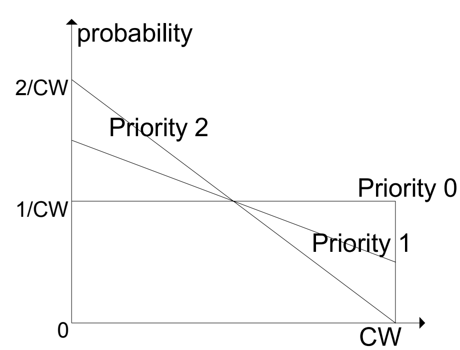

Figure 2. backoff distribution

and efficiency are almost identical for all the schemes except the S_1 with *CRA (FCR)*. One interesting observation for S_1 is that when $\lambda \leq 20$, the overall throughput and efficiency is less than those of other schemes; while $\lambda > 20$, S_1 has much better overall throughput and efficiency (This result matches our argument in [12]). This can be explained as follows. When λ is low, not so many collisions occur. Hence the FCR will unnecessarily incur lots of idle slots, leading to low channel utilization. When λ is high, collisions occur very frequently for IEEE 802.11 MAC. In this case, FCR exhibits its advantages in collision resolution because it emulates TDMA: one station consecutively uses the channel for a while and then gives the channel to another station. Thus the number of wasted slots is reduced and in turn, the channel utilization and fairness will be improved. In Fig 3.a, the overall throughput of S_1 is almost 80% of the total capacity (i.e., 0.8Mbps, which is close to the theoretical bound [15]), while for other schemes it is less than 70% (0.7Mbps). Generally speaking, all the efficiencies decrease with the increase of the arrival rate λ. The lower the priority, the faster the decreasing speed. Since all the schemes except S_1 cannot improve the overall throughput, in order to support high priority traffic, those schemes have no choice but to sacrifice the low priority traffic, especially when the network is overloaded. This can be observed from Fig. 3.e. All the schemes but S_1 can only transmit very few packets, if not nothing, for priority 0 traffic when λ is large. In contrast, this kind of starvation is not so serious for S_1 with FCR. From Fig. 3.c–3.e, S_3 shows the best efficiency for priority 2 traffic when λ is not too high. However, S_3 has the worst fairness. We can see that the throughput and the fairness for prioritized service cannot be simultaneously achieved. The fairness for FCR adaptor is much better than that of all the others when λ is high. In this sense, FCR scheme is also a fair collision resolution scheme.

Next, we study the MAC access delay and the end-to-end delay of different schemes except S_6. In general, for each scheme, the delays increase with the arrival rate λ. The lower the priority, the faster the increasing speed. Meanwhile, high priority traffic has small delay. From Fig. 3.f-3.k, we can see that *IFSA, CWA, BFA,* and *BDA* all play significant roles in supporting differentiated services with different MAC access delay and the end-to-end delay. S_2, S_3, S_4, and S_5 are capable of providing smaller MAC access delay and end-to-end delay for priority 2 traffic. Among them, S_3 should be the best one that has the smallest delays for high priority traffic, when λ is not too high. It is interesting to observe that, for S_3 at $\lambda=20$, the delays for priority 0 traffic are larger than those of all the other schemes. This is consistent with the above observation that S_3 has the worst fairness when it provides the high priority service with short delay. We also observe that some schemes have excessive, if not infinite, delays for priority 0 or priority 1 traffic when λ is very large. This matches the starvation phenomenon we have described before for the efficiency. For S_1, it is observed that it does not work well in terms of delays for small λ. This is due to the fact that, in the case of low collision probability, FCR always cause unnecessary backoffs, leading to long MAC access delays hence long end-to-end delays. When λ is very large, however, FCR works well and presents the best performance among all the schemes in terms of delays. This should be attributed to the fast collision resolution mechanism, which reduces the collisions and grants the station that has already successfully transmitted packets better chance to seize the channel.

From the above analysis, we can see that S_3 is the most promising one to support priority 2 traffic with small MAC access delay, small end-to-end delay, and greater efficiency, while S_1 has the best fairness performance among all. Based on this observation, we propose S_6, which combines S_1, S_2, *and* S_3 together, to investigate how different mechanisms work together to better support prioritized service. Seen from Fig. 3.a- 3.k, S_6 inherits the advantages from both *CRA* and *IFSA*. When λ is small, the difference between the schemes is negligible. When λ is large, say, 100 packets/sec, for priority 2 traffic, the small access delay due to *IFSA* is counteracted by FCR. But FCR's collision resolution mechanism reduces the number of retransmissions, which in turn greatly shortens the end-to-end delay. On the other hand, though fairness of FCR is weakened by the starvation properties of *IFSA*, S_6 still has better fairness property than S_3. The overall throughput of S_6 is also greater than that of S_1 or S_3. Therefore, S_6 performs best in our simulation study.

IV. CONCLUSIONS

In this paper we propose a QoS-based MAC architecture based on which we compare the performance of various MAC schemes (S_1 to S_6) in support of prioritized services. From the simulation study, it is observed that scheme S_1 to S_5 are all capable of supporting high priority services with low delay when the traffic arrival rate is not too large. S_3 is the most promising MAC in terms of delay. S_2 is also a good choice to support high priority with low delay. Since S_2 sacrifices the low priority traffic to a moderate degree, the fairness of S_2 is better than S_3. Compared with S_2 and S_3, the MAC schemes S_4 and S_5 do not have many distinguishing features. When the arrival rate is very large, S_1 is the best one in terms of delay and overall efficiency. As far as fairness is concerned, S_1 is the best and S_3 is the worst. All the schemes but S_1 cannot improve the throughput when the network is overloaded. More importantly, through our proposed scheme S_6, we can better understand how different adaptors can work closely to provide better performance. The superior performance of S_6 suggests a way to design an adaptive MAC protocol based on the proposed unified architecture. For example, at low traffic arrival rate, we may adopt S_2 (*BD*) or S_3 (*IFS*) to support high priority service with small delays, and at the same time, provide good fairness among different services. At high traffic arrival rate, it is better to enable FCR adaptor, which can provide low delays and good fairness for supporting prioritized services while avoiding the starvation of the low

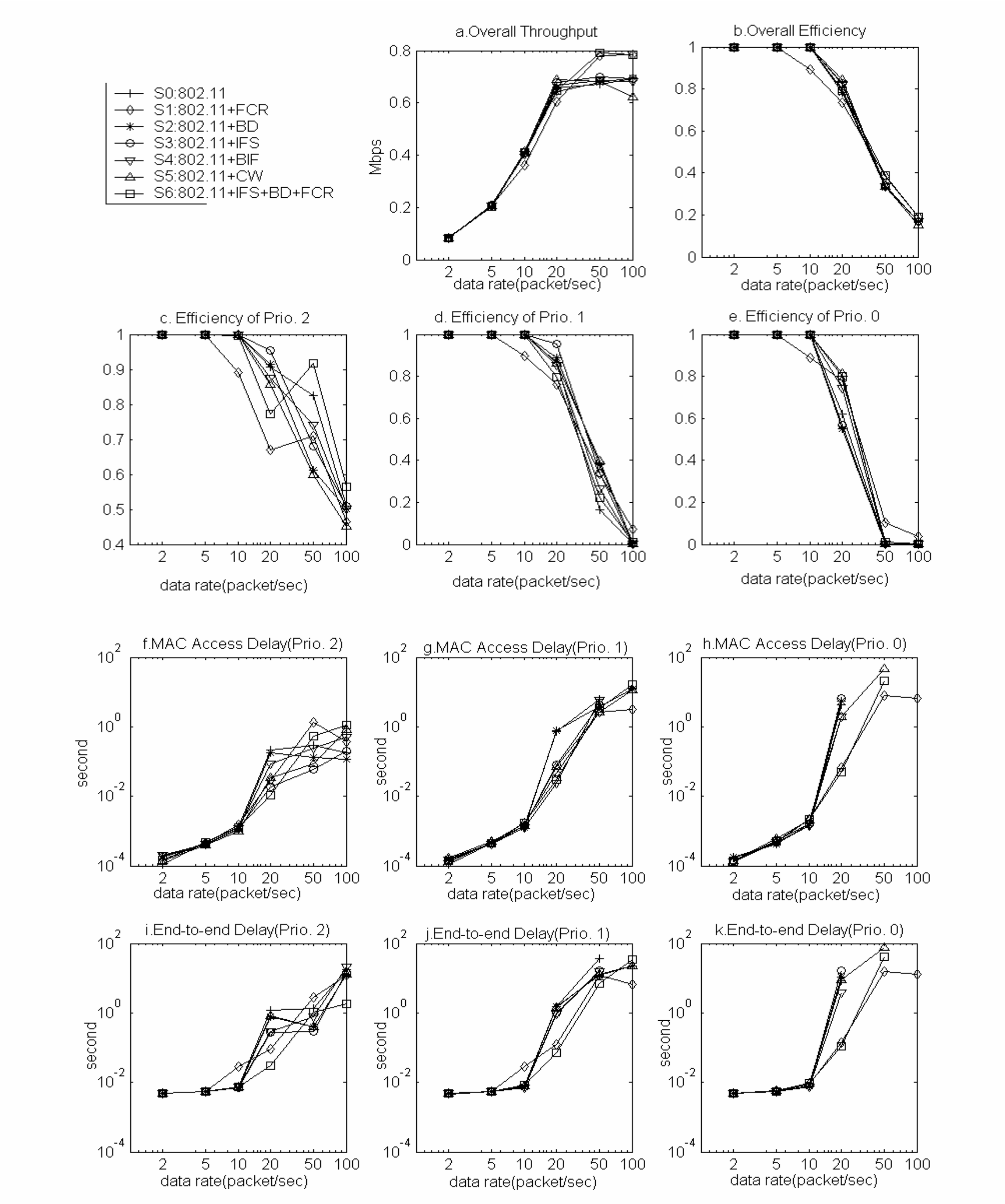

Figure 3. Simulation results

priority services. Since there may not be a universal solution that can be applied to all situations, a better way to provide QoS guarantees in WLANs is to dynamically change the setting of the adaptors and/or enable/disable some adaptor in the architecture in light of the instantaneous states of channels, station, and networks.

ACKNOWLEDGEMENT

This work was supported in part by the Office of Naval Research under grant N000140210464 (Young Investigator Award) and under grant N000140210554.

References

[1] *IEEE 802.11 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*,1997.

[2] J.-Y. Yeh and C. Chen, “Support of multimedia service with the IEEE 802.11 MAC protocol,” *IEEE ICC’ 2002*, pp.600-604, 2002.

[3] C. Coutras, S. Gupta and N. B. Shroff, “Scheduling of real-time traffic in IEEE 802.11 wireless LANs,” *Wireless Networks*, vol. 6, no. 6, pp457-466, 2000.

[4] M. Veeraraghavan, N. Cocker and T. Moors, “Support of voice service in IEEEE802.11 wireless LANs,” *Proc. of IEEE INFOCOM 2001*.

[5] A. Lindgren, A. Almquist and O. Schelen, “Evaluation of quality of service schemes for IEEE 802.11wireless LANs,” *Proc. of IEEE LCN’2001*.

[6] M. Barry, A.T. Campbell and A. Veres, “Distributed control algorithm for service differentiation in wireless packet networks,” *Proc. of IEEE INFOCOM 2001*.

[7] J.L. Sobrinho and A. S. Krishnakumar, “Real-time traffic over the IEEE 802.11 medium access control layer,” *Bell Labs Technical Journal*, pp.172-187, Autumn 1996.

[8] I. Aad and C. Castelluccia, “Differentiation mechanisms for IEEE 802.11,” *Proc. of IEEE INFOCOM 2001*.

[9] D-J. Deng and R-S. Chang, “A priority cheme for IEEE 802.11 DCF access method,” *IEICE Transaction on Communications*, E82-(1),January 1999.

[10] X. Pallot and L. E. Miller, “Implementing message priority policies over an 802.11 based mobile ad hoc network,” *IEEE MILCOM 2001*, Washington, October 2001.

[11] M. Benvenite, G. Chesson, M. Hoeben, A. Singla, h. Teunissen,M. Wentink, EDCF proposed draft text, *IEEE working doucument 802.11-01/131rl*,March 2001

[12] Y. Kwon, Y. Fang and H. Latchman, “Fast Collision Resolution (FCR) MAC algorithm for wireless local area networks,” *Proc. of IEEE GLOBECOM ’2002*.

[13] P. Lettieri and M.B. Srivastava, “Adaptive frame length control for improving wireless link throughput,range,and energy efficiency,” *Proc. of IEEE INFOCOM 1998*.

[14] P. Bhagwat, P. Bhattachrya, A. Krishna,and S. Tripathi, “Enahcing throughput over wireless LANs using channel state dependent packet scheduling,” *Proc. of IEEE INFOCOM 1996*.

[15] F. Cali, M. Conti and E. Gregori, “Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit,” *IEEE/ACM Transaction on Networking*, vol.8, no.6, Dce. 2000.