Inter-Frame Space (IFS) Based Service Differentiation for IEEE 802.11 Wireless LANs

Chun-Ting Chou and Kang G. Shin

Real-Time Computing Laboratory
The University of Michigan
Ann Arbor, MI 48109-2122, U.S.A.
{choujt,kgshin}@umich.edu

Sai Shankar N.

Philips Research In USA
Briarcliff , NY 10510, USA
sai.shankar@philips.com

Abstract—**A novel channel access control for the IEEE 802.11 wireless standard is proposed to provide differentiated service in a wireless LAN. It relies on controlling the length of the Inter-Frame Space (IFS) interval, during which a station has to detect a quiet channel before decrementing its backoff time. We derive a simple model to mathematically analyze the service differentiation achievable with Arbitration IFS (AIFS), such as weight- or priority- proportional throughputs, and then verify its effectiveness via simulation. The simulation results confirmed the correctness of the proposed control scheme, and demonstrated its effectiveness. A small difference between stations' IFS values is shown to make substantial service differentiation between them. This, in turn, reduces the channel idle time which could be very long if the control relies on the commonly-used contention window size, rather than IFS. The proposed scheme, therefore, achieves higher channel utilization, especially when the total number of stations is not large.**

I. INTRODUCTION

The IEEE 802.11 standard [1] provides mobile stations simple but effective access to wired networks via a shared wireless medium. Each station uses carrier sense multiple access (CSMA) with collision avoidance (CA) to contend for the channel. In case of collision, the stations involved in the collision perform an exponential random backoff to minimize the possibility of subsequent collisions. With this channel access mechanism, all stations can have an equal share of system throughput in a distributed manner. Its simplicity, together with high transmission capacity, makes the 802.11 wireless LAN a good candidate for the next-generation wireless network that requires support for Quality of Service (QoS). The only problem with the current 802.11 wireless LAN is that all stations are treated in an egalitarian way. Thus, service differentiation cannot be provided without any further control.

The IEEE 802.11e standard [2] is proposed to solve this problem based on the original 802.11 medium access control (MAC). Stations under this new standard still use CSMA/CA with exponential random backoff to contend for the channel. However, according to the assigned priorities or weights, stations may be assigned different parameters for channel contention. The parameters that can be manipulated are minimum contention window size, maximum contention window size, retry limit, and AIFS time. For example, the authors of [3][4]

chose a station's contention window size to be inversely proportional to its assigned "weight" in order to approximate the weighted fairness. Differential treatment between delay-sensitive and best-effort traffic can also be achieved by some heuristic control over the contention window size [5]. In [6], the parameters used in stations' random backoff process, namely the minimum/maximum contention window size and the retry limit, are computed based on a Markovian model such that the transmission time acquired by individual stations can be finely controlled.

Even though the desired service differentiations are not same in these proposals, they all use the same idea: *by giving a less preferred station a larger contention window size, we force it to back off for a longer time before it can initiate a transmission*. The "preferred" stations will then have a better chance to acquire the channel and can improve their transmission performance. For example, if the minimum contention window size of one station is 32 and that of the other station is 64, the resultant throughput of the first station will be approximately twice as that of the second station [4]. The exact computation of a station's contention window size is more complicated [6], but it achieves service differentiation in an IEEE 802.11 wireless LAN. The only problem of controlling stations' contention window sizes is that, since some stations will use larger contention window sizes, a wireless LAN could be idle for a long time, especially when the number of stations in the wireless LAN is small, thus reducing the overall system throughput.

In this paper, we propose a new way of controlling stations' AIFS to achieve service differentiation. We will show that only a slight difference between stations' AIFS values is needed to achieve the differentiation achievable by controlling the contention window size. With appropriately-chosen stations' AIFS values, the desired service or QoS differentiation can be realized in the 802.11 wireless LAN. Moreover, since all stations can use the same (and smaller) contention window size, a higher system throughput can be achieved, especially when the number of stations is not large.

The rest of this paper is organized as follows. Section II gives a brief introduction of the IEEE 802.11 MAC for completeness. Section III describes the proposed control scheme and its application to control the stations' channel access. In Section IV, we present the numerical analysis results and compare them with the simulation results. Finally, conclusions are drawn in Section V.

II. The IEEE 802.11 MAC

The IEEE 802.11 standard supports two channel access mechanisms: Point-Coordinate-Function (PCF) and Distributed-Coordinate-Function (DCF). Since we are interested in achieving service differentiation in a distributed manner, we will only focus on the DCF mode.

A. DCF mode in the IEEE 802.11a/b standard

In the DCF mode, each station must contend for the channel using CSMA/CA. A station desiring to initiate transmission invokes the carrier-sense mechanism to determine whether the medium is busy or idle. If the medium is busy, the station defers the transmission until the medium is determined to be idle for a period of time equal to DCF Inter-Frame Space (DIFS). After this period of idle time, the station will wait for an additional time (i.e., the backoff time) before starting the transmission. This backoff time is determined by

$$BT = Random([0, CW]) \cdot aSlotTime,$$

where CW is the station's current contention window size and $aSlotTime$ is the duration of a time slot. If no medium activity is indicated for $aSlotTime$ seconds, the station will decrement its backoff time by $1 * aSlotTime$. If the medium is determined to be busy at any time during a backoff slot, the backoff procedure is suspended; that is, the backoff time will not be decremented for that slot. Transmission should commence whenever the backoff time of a station becomes zero.

In order to minimize collisions when multiple stations contend for the channel, each individual station should choose its CW as follows.

1) CW takes an initial value of $CW_{\min}$.
2) CW takes the next value in the series in Eq. (1) after making an unsuccessful attempt to transmit, until CW reaches its maximum value, $CW_{\max}$.
3) Once it reaches $CW_{\max}$, CW will remain there until it is reset.
4) CW will be reset to $CW_{\min}$ after (i) a successful transmission of a frame or (ii) the number of retransmission attempts reaches $retry\ limit$. (An IEEE 802.11 station should try to retransmit any unsuccessful frame up to $retry\ limit$ times before discarding that frame).

According to the IEEE 802.11b standard, the set of CW values should be a sequentially-ascending integer power of 2, minus 1, beginning with $CW_{\min}$ and continuing up to $CW_{\max}$:

$$\{CW = 2^j - 1 : j = K, K+1, \cdots, K+m\}. \quad (1)$$

Thus, $CW_{\min} = W_0 - 1 = 2^K - 1$, and $CW_{\max}$ is $2^m \cdot W_0 - 1$.

B. Enhanced DCF in the IEEE 802.11e standard

As mentioned earlier, the DCF does not support the concept of service differentiation or priorities. Basically, it is designed to provide an equal-probability channel access to all stations contending for the channel in a distributed manner. However, an equal access probability is not desirable among stations with different priority frames. Thus, the IEEE 802.11e standard offers a new channel access mode called *Hybrid Coordination*
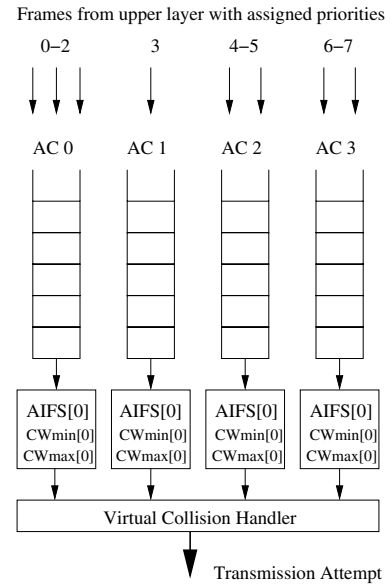

Fig. 1. Support for prioritized frames at the IEEE 802.11e MAC

Function (HCF). Details of the HCF can be found in [2]. Simply speaking, the HCF combines an enhanced DCF (EDCF) with the polling function. As its name suggests, EDCF is an enhanced version of the original DCF. The most important difference between DCF and EDCF is that EDCF is designed to provide differentiated channel accesses in a distributed manner for frames up to 8 different priorities (from 0 to 7). To achieve this, we have to manipulate the EDCF parameters: AIFS time, $CW_{\min}$ and $CW_{\max}$. The values of these EDCF parameters are announced by the access point (AP) via beacon frames. The AP can adapt these parameters dynamically, depending on network conditions. Basically, the smaller AIFS and $CW_{\min}$, the shorter the channel access delay for the corresponding priority, and hence, the more capacity share for a given traffic condition.

Figure 1 shows how different-priority frames are handled at the MAC layer. Each frame from the higher layer arrives at the MAC along with specific priority. Then, each frame carries its priority value in the MAC frame header and is mapped into a separate queue as shown in Figure 1. Each queue, also called an *access category* (AC), behaves as a single enhanced DCF contending entity. Thus, each AC uses different AIFS, $CW_{\min}$, and $CW_{\max}$, instead of using the same parameters as in the original DCF, for the contention process to transmit a frame belonging to that AC. An 802.11e station implements four ACs. When there are more than one AC finishing the backoff at the same time, the collision is handled in a virtual manner. That is, the highest-priority frame among the colliding frames is chosen for transmission, and the others back off, as mentioned in the previous subsection.

III. The Control over AIFS

Figure 2 shows the channel access by stations in the EDCF mode. Since each station (or an EDCF entity) has a different AIFS, the station starts/resumes decrementing its backoff counter at a different time. Let AIFS[i] be the value of station i's AIFS. For example, station 1 may have a smaller AIFS
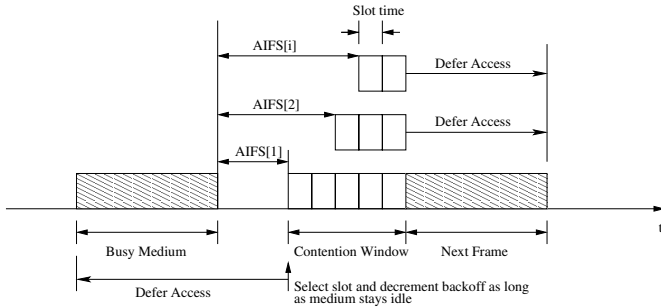
Fig. 2. Enhanced DCF mode in IEEE 802.11 wireless LANs

(a)

(b)

Fig. 3. Station-2's backoff decrement delay

than station 2 and let AIFS[2]−AIFS[1]=2. Thus, every time station 2 starts to decrement its backoff counter, station 1 has already decremented its backoff counter by 2 slot times. Let D be this "decrementing lag" of station 2 with respect to station 1. D is then a random variable with possible values 1 and 2. The reason why D could be less than 2 is that if station 1 chooses a backoff counter value less than 2 (say 1), then station 2 will have no chance to start/resume its backoff process before station 1 finishes its transmission. In this case, $D = 1$. With the definition of decrementing lag, the following relation can be found between any two consecutive collisions in which station 1 and station 2 are involved:

$$\sum_{i=1}^{n_1} BT_i^{(1)} = \sum_{j=1}^{n_2} BT_j^{(2)} + \sum_{h=1}^{n_1+n_2-1} D_h, \quad (2)$$

where $BT_i^{(j)}$ is the i-th backoff time chosen by station j for its transmission, and n_i represents the total number of times that station i acquires the access of channel. In the example above, the mean value of D should be very close to 2 because the probability of $BT_i^{(1)} = 1$ is very small ($\approx \frac{1}{CW}$).

Eq. (2) can be generalized for the case in which there are N different station classes. Here, we assume that stations in the same class have the same AIFS value. Let K_i be the number of class-i stations, and assume that class-1 stations have the smallest AIFS, class-2 stations have the second smallest AIFS, and so on. We also assume that all stations use the same CW_{min}, CW_{max}, and retry limit. In the steady state, Eq. (2) can be rewritten as

$$E[n_1]E[BT^{(1)}] =$$
$$(\sum_{j=1}^{N} K_j E[n_j] - E[N_{col}])E[D^{(k)}] + E[n_k]E[BT^{(k)}], \quad (3)$$

for $k = 2$ to $k = N$. $D^{(k)}$ is the "decrementing lag" of a class-k station as compared to a class-1 station, and $E[N_{col}]$ is the average number of collisions within the observed interval. Eq. (3) can be further rewritten as:

$$E[n_1]\frac{CW_{min}}{2} \approx$$
$$\sum_{j=1}^{N} K_j E[n_j]E[D^{(k)}] + E[n_k]\frac{CW_{min}}{2}. \quad (4)$$
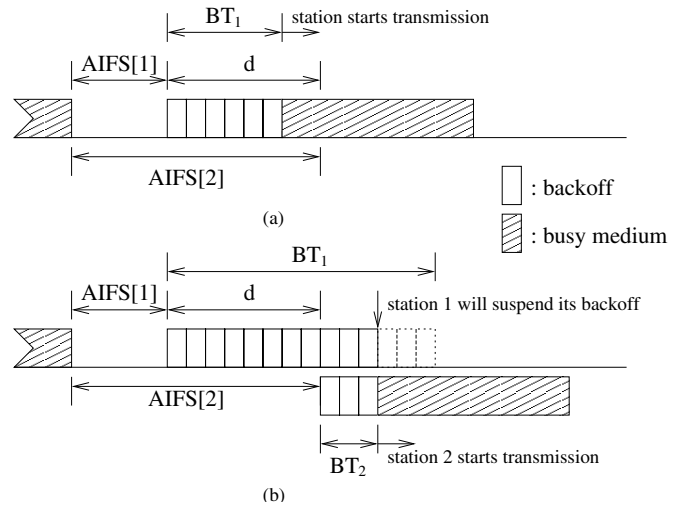
Here, we simply substitute $E[BT^{(i)}]$ by $\frac{CW_{min}}{2}$ and assume $\sum_{j=1}^{N} K_j E[n_j] \gg E[N_{col}]$. This is true when the total number of stations in a wireless LAN is not large. Later, we will show how to calculate $E[BT]$ when the number of stations is larger.

The ratio of each individual station's channel access to a class-N station's, $\frac{E[n_i]}{E[n_N]}$, can be obtained by solving the system of linear equations given in Eq. (4). The only problem left is how to calculate a class-k station's decrementing lag, $E[D^{(k)}]$. Once we solve it, we can choose the required values of AIFS for all stations such that the desired service differentiation , namely $\frac{E[n_i]}{E[n_N]}$, can be achieved.

Before giving the estimator of decrementing lag for the general case, we first consider a two-station case: one in class 1 and the other in class 2, in a wireless LAN. Let AIFS[2]−AIFS[1] = d be the difference of the stations' AIFS values. As shown in Figure 3-(a), if station 1 chooses its backoff time, BT_1, as any value between 1 to $d - 1$, the decrementing lag of station 2, $D^{(2)}$, will be equal to BT_1 because station 2 has not waited for AIFS[2] seconds to start decrementing its backoff. If BT_1 is larger than d, the computation of decrementing lag is a little more complicated but still can be approximated as follows:

- If $BT_1 - d < BT_2$, station 1 will win the current run of contention and thus $D^{(2)} = d$.
- Otherwise, station 2 will win the current run of contention. However, station 1's remaining backoff time may result in another $D^{(2)} < d$ in the next run if $BT_1 - d - BT_2 < d$, as illustrated in Figure 3-(b).

Thus, the average $D^{(2)}$ given $BT_1 \geq d$ can be calculated by

$$E[D^{(2)}|BT_1 \geq d] = \frac{(CW - (d-1)) \cdot d}{CW} + \frac{\sum_{i=1}^{(d-1)} i}{CW} \quad (5)$$

given that the stations choose their own backoff times uniformly within $[0, CW]$. Finally, combining the cases (a) and (b) in Figure 3, the average value of $D^{(2)}$ can be calculated as

$$E[D^{(2)}] = d - [\frac{d(d-1)}{CW_{min}} - \frac{d(d-1)^2}{2CW_{min}}], \quad (6)$$

since both stations are assumed to have all other parameters to be the same. If there are more than one class-1 station, $E[D^{(2)}]$ can be calculated by using the concept of union bound [7]:

$$E[D^{(2)}] \approx d - [\frac{d(d-1)}{CW_{min}} - \frac{d(d-1)^2}{2CW_{min}}] * K_1. \quad (7)$$

Finally, in view of the fact that only the stations with smaller AIFSs can contribute to the decrementing delay of the stations with larger AIFSs, we can get an estimate of $E[D^{(k)}]$ according to Eq. (7). Let $d_i^{(k)}$=AIFS[k]-AIFS[i] for $i = 1$ to $i = k - 1$. Then, the decrementing delay of a class-k station can be estimated by

$$E[D^{(k)}] = d_1^{(k)} - \sum_{i=1}^{k-1}[\frac{d_i^{(k)}(d_i^{(k)}-1)}{CW_{\min}} - \frac{d_i^{(k)}(d_i^{(k)}-1)^2}{2CW_{\min}}] * K_i. \quad (8)$$

It should be noted that even though some simplifications have been made in order to obtain Eq. (8), we will show later that it matches the simulation results very well. Finally, with Eqs. (4) and (8), we can compute the AIFS values necessary for the desired service differentiation in a wireless LAN that supports multiple station classes/priorities.

Our control scheme has many immediate applications in an IEEE 802.11 wireless LAN. For example, it can realize the weighted fairness, in terms of throughput between different classes/flows, once the ratio of throughputs is determined by the admission control of the IEEE 802.11e wireless LAN. Furthermore, we can also guarantee the frame delay as long as the station's traffic is regulated by some traffic shaping mechanism such as a leaky bucket. It can also be used to provide the weighted fair share of system airtime in an IEEE 802.11e wireless LAN [6] which supports multiple physical transmission rates (i.e., so-called link-adaptation in [8]).

IV. NUMERICAL ANALYSIS AND SIMULATIONS

We first show the accuracy of our estimator in Eq. (8) for 2 different cases, in which the wireless LAN supports 2 and 4 different station classes, respectively. We will then show how service differentiation, such as weighted shares of system throughput or transmission time, can be achieved by means of the proposed control mechanism. In all of the following simulations, we assume that all stations use the same parameters except the value of AIFS. The other EDCF parameters, $CW_{min} = 63$, $CW_{max} = 1023$, and $retry\ limit = 7$ are fixed in all of the simulations.[1] The differences between the AIFS values in different classes, in number of slot times, are given in each set of simulation.

A. The decrementing lag

Table I shows the average decrementing lag of class-2 stations, $E[D^{(2)}]$, given that there are only two classes and AIFS[2]−AIFS[1]=4. As mentioned earlier, the decrementing lag of class-2 stations does not change with the number of

[1]We choose $CW_{\min} = 63$ because the "legacy" IEEE 802.11b standard uses 31 as its default value.

		The number of class 2 stations K_2			
		1	2	3	Estimator
The number of class 1 stations K_1	1	3.82	3.82	3.82	3.82
	2	3.66	3.66	3.67	3.63
	3	3.52	3.52	3.51	3.45
	4	3.40	3.4	3.39	3.26
	5	3.29	3.29	3.29	3.15*

TABLE I

DECREMENTING LAG: $N = 2$ AND $AIFS[2] - AIFS[1] = 4$

class-2 stations, K_2, but only with the number of class-1 stations, K_1. In general, Eq. (7) gives a very good estimation of $E[D^{(i)}]$. It should be noted that the larger K_1, the larger the estimation error will be. The reason is because we used the union bound to derive Eq. (7). In Eq. (7), the second term accounts for the impact of class-1 stations' smaller AIFS on class-2 stations' decrementing lag. Since the union bound gives an upper bound of the second term in Eq. (7), the estimator will generate a smaller $E[D^{(2)}]$. Nevertheless, the largest estimation error is smaller than 5% (when $K_1 = 5$ and $K_2 = 3$). In this case, we actually use a better approximation for the mean window size, $E[BT]$, instead of using $\frac{CW_{\min}}{2}$ as in Eq. (4). Since the effect of exponential increase of random backoff time cannot be ignored when the number of stations is not small, we use

$$E[BT] \approx (1 - \frac{\sum_i K_i}{CW_{\min}})\frac{CW_{\min}}{2} + \frac{\sum_i K_i}{CW_{\min}}CW_{\min}, \quad (9)$$

to include the effects of collisions and the subsequent exponential increase of stations' backoff times. Here, $\frac{\sum_i K_i}{CW_{\min}}$ accounts for the collision probability and $CW_{\min}$ represents the average backoff time a station may choose after the first collision. We do not consider the effect of exponential increase of CW resulting from more than 2 consecutive collisions because they rarely occur.

Next we show the case in which there are four classes in the wireless LAN and AIFS[i]−AIFS[i − 1]=2 for $i = 2$ to 4. Again, when the number of stations is larger (e.g., ≥ 7), we used the better approximation given in Eq. (9) in order to get a more accurate estimation of the decrementing lag, $E[D^{(i)}]$. Even though deriving the estimator for the general case (i.e., Eq. (8)) needs some approximations, it is surprising that the estimation error is really small as shown in Table II. The largest estimation error occurs when $K_1 = 4$, $K_2 = 2$, $K_3 = 1$ and $K_4 = 1$, but it is still less than 10%.

B. The ratio of channel accesses

In this subsection, we will show how a small difference between stations' AIFSs is enough to provide differential treatment of different-class stations. Compared to the other proposals which may need to choose the contention window over a much larger range, our control scheme is more efficient, in terms of system throughput, especially when the total number of stations is not too large.

(K_1, K_2, K_3, K_4)	$E[D^{(2)}]$	$E[D^{(3)}]$	$E[D^{(4)}]$
(1, 1, 1, 1)	1.96	3.80	5.42
	1.969	3.817	5.35
(2, 1, 1, 1)	1.93	3.65	5.08
	1.937	3.63	4.91
(3, 1, 1, 1)	1.90	3.50	4.77
	1.906	3.45	4.57
(4, 1, 1, 1)	1.88	3.37	4.51
	1.89	3.31	4.17
(4, 2, 1, 1)	1.88	3.38	4.47
	1.89	3.29	4.04

* The value in the shaded columns are the simulation results.

TABLE II

DECREMENTING LAG: $N = 4$ AND $AIFS[i] - AIFS[i-1] = 2$.

		Class I			Class II		
I		STA 1	STA 2	STA 3	STA 4	STA 5	STA 6
	A	1.998	1.998	1.998	1.00	1.00	1.00
		1.96	2.02	1.99	1.01	1.02	1.00
	B	3.02	3.02	3.02	1.00	1.00	1.00
		3.07	3.06	3.03	1.01	1.00	1.02

	Class I		Class II		Class III	
II	STA 1	STA 2	STA 3	STA 4	STA 5	STA 6
	2.95	2.95	1.98	1.98	1.00	1.00
	3.06	3.08	2.01	1.97	1.00	1.00

	Class I		Class II		Class III		Class IV	
III	STA1	STA2	STA3	STA4	STA5	STA6	STA7	STA8
	4.23	4.23	3.08	3.08	2.11	2.11	1.00	1.00
	4.25	4.26	2.97	2.92	2.02	2.05	1.00	0.994

* The value in the shaded columns are the simulation results.

TABLE III

THE RATIO OF STATIONS' THROUGHPUT: ESTIMATIONS VS. SIMULATIONS

We consider three different cases in which the wireless LAN supports 2 (case I), 3 (case II) and 4 (case III) classes of stations, respectively. In the case I-(A), we assume $K_1 = K_2 = 3$ and choose $AIFS[2] - AIFS[1] = 4$ based on Eqs. (4) and (8) such that each class-1 station can have twice the throughput of the class-2 station, while we choose $AIFS[2] - AIFS[1] = 7$ in the case I-(B) such that each class-1 station can have three times the throughput of the class-2 station. The simulation results in Table III show that the desired service differentiation can be achieved by our control scheme with a relatively small difference in AIFS values. If we control the stations' $CW_{\min}$ values instead of AIFS, all class-2 stations need to use $CW_{\min} = 127$ for setting A and $CW_{\min} = 255$ for setting B, given that class-1 stations use $CW_{\min} = 63$. In case II, we set AIFS[2]−AIFS[1]=3 and AIFS[3]−AIFS[3]=4 such that the ratio of stations' throughputs in each class is close to 3:2:1, given that there are 2 stations in each class. Even though the resulting ratio is not exactly the same as required, the largest deviation is only about 5%. Finally, we consider four classes with a desired throughput ratio close to 4:3:2:1, given that there are 2 stations in each class. Based on Eqs. (4) and (8), we need AIFS[2]−AIFS[1]=2, AIFS[3]−AIFS[2]=2, and AIFS[4]−AIFS[3]=3. Again, the simulation result shows that a small amount of difference among stations' AIFS still suffices to yield the desired service differentiation (with the largest deviation less than 6% as compared to the numerical results).

One should note that in all of these simulations, we only focus on the cases where the total number of stations is no more than 8, even though the control scheme is applicable to other cases. If the total number of stations is larger, the occurrence of long channel idleness due to large contention window size is very unlikely because there are always stations contending for the channel. Thus, controlling $CW_{\min}$ in [6] can achieve the required service differentiation without compromising the system throughput. In summary, the control over backoff process should adapt to the number of total stations in order to fully utilize the wireless channel while providing the needed service differentiation. *When the total number of stations is not large, we should control individual stations' AIFS values; otherwise, controlling $CW_{\min}$ is easier and also effective*.

V. CONCLUSIONS

In this paper, we proposed a novel control algorithm over AIFS time in the IEEE 802.11e wireless LAN to achieve fairness and service differentiation. With this control mechanism, we can provide QoS support , such as throughput differentiation, in an IEEE 802.11e wireless LAN or other wireless networks that use a similar MAC. The simulation results show the correctness and effectiveness of the proposed control. Since it only needs a small difference among stations' AIFS values, a higher channel utilization can be achieved, as compared to other schemes, especially when the number of stations is not large.

REFERENCES

[1] "IEEE 802.11b: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications", *IEEE Standard*, 1999.
[2] IEEE 802.11e/D3.0, *Draft Supplement to Part 11: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)*, May 2002.
[3] N. H. Vaidya, P. Bahl, and S. Gupta, "Distributed Fair Scheduling in a Wireless LAN", *ACM MobiCom'00*, pp. 167–178.
[4] D. Qiao and K. G. Shin, "Achieving Efficient Channel Utilization and Weighter Fairness for Data Communications in IEEE 802.11 WLAN under the DCF", *International Workshop on Quality of Service 2002*.
[5] M. Barry, A. T. Campbell and A. Veres, "Distributed Control Algorithms for Service differentiation in Wireless Packet Networks", *IEEE INFOCOM'01*, pp. 582–590.
[6] C.T. Chou, Sai Shankar N., and K.G. Shin, "Generalized Air Time Sharing : a new fairness concept and its sapplication to the IEEE 802.11 MAC-layer", *submitted for publication*.
[7] J. G. Proakis, "Digital Cimmunications", *Mcgraw-Hill*, 1995, pp. 263–264.
[8] Daji Qiao and Sunghyun Choi, "Goodput Enhancement of IEEE 802.11a Wireless LAN via Link Adaptation," in *Proc. IEEE ICC'01*, Helsenki, Finland, Jun. 2001.