Service differentiation without prioritization in 802.11 WLANs

Suong H. Nguyen, *Student Member, IEEE,* Hai L. Vu, *Senior Member, IEEE,* and Lachlan L. H. Andrew, *Senior Member, IEEE*

***Abstract*—Wireless LANs carry a mixture of traffic, with different delay and throughput requirements. The usual way to provide low-delay services is to give priority to such traffic. However this creates an incentive for throughput sensitive traffic also to use this service, which degrades overall network performance. We propose to allow applications to trade off delay for throughput, without giving preference to one class over another, by simultaneously scaling IEEE 802.11's $CW_{\min}$ and *TXOP limit* parameters.**

We provide a model of this scheme with two traffic classes, and show that increasing $CW_{\min}$ and *TXOP limit* in equal proportion reduces, but does not eliminate, the incentive for bulk data users to use the low-delay service. We show that subtracting a small constant from $CW_{\min}$ eliminates this incentive, while still giving improved performance to both classes.

***Index Terms*—802.11 EDCA WLANs, service differentiation.**

I. Introduction

Wireless networks carry a diverse mix of traffic, from voice with tight delay constraints to bulk file downloads with only long-term throughput requirements. Efficient use of the network requires services tailored to each of these traffic classes. The traditional approach to providing quality of service (QoS) is to prioritize real-time traffic at the expense of data traffic, as done by the default parameter setting of the IEEE 802.11e EDCA standard [1]. This creates an incentive for data applications to use the class intended for real-time users to gain higher share of resources. This can degrade network performance drastically [8] and QoS differentiation no longer occurs when all data users use the highest priority class [6]. To cope with this, policing mechanisms have been proposed [7], which increase the complexity of the network.

As an alternative to prioritization, we propose a simple scheme providing better service for both throughput-sensitive and delay-sensitive traffic and encouraging applications to use the service designed for them. The approach is based on the IEEE's standard 802.11e [1] for carrier sense multiple access with collision avoidance (CSMA/CA) using exponential backoff. Note that unlike [2], [3], we are not combatting deliberately malicious users who violate the protocol, but rather addressing application writers who optimize their code based on measured performance using all the available services.

The authors are with the Centre for Advanced Internet Architectures (CAIA), Faculty of Information and Communication Technologies, Swinburne University of Technology, Hawthorn, Melbourne VIC 3122, Australia. E-mail: {hsnguyen, hvu, landrew}@swin.edu.au.

TABLE I
DEFAULT MAC PARAMETERS IN 802.11E EDCA

AC	CW_{min}	CW_{max}	AIFSN	*TXOP limit* (for DSSS PHY)
Data (AC_BE)	31	1023	3	1 packet
Real-time (AC_VO)	7	15	2	3.264 ms

To support QoS, 802.11e uses Enhanced Distributed Channel Access (EDCA), in which the access point (AP) selects four Access Categories (ACs) which stations can use. The ACs use different values of four MAC parameters: $CW_{\min}$, $CW_{\max}$, *TXOP limit* and $AIFS$. In particular, CW_{min} controls how long a station waits before transmission and *TXOP limit* controls how much it can transmit per channel access. Note that applications cannot choose arbitrary combinations of these parameters, but only those permitted by the AP.

The 802.11e EDCA standard recommends four particular combinations of parameters. The parameters intended for throughput-sensitive bulk data and delay-sensitive voice are shown in Table I, taken from Table 7-37 of [1]. In particular, the AC for real-time service is given higher priority than that for data service; each parameter is set to a more aggressive value. This priority-based QoS provision works well provided the high-priority class is only used by low-throughput real-time traffic. However, when applications are optimized based on performance measurements, the priority-based approach creates an incentive for all users to use the class AC_VO, resulting in no QoS differentiation.

Instead, we seek to provide service differentiation without priority, by choosing ACs such that some parameters are less aggressive whenever others are more aggressive. The aim of service differentiation without priority is to provide "different but fair" service for different types of traffic, by allowing users to choose different points on a throughput-delay tradeoff curve. Our proposal consists of choosing the combinations of parameters allowed by the AP. As such, it does not require any additional mechanisms such as fair queueing or traffic policing.

Fair differentiated service has already been proposed in [10], [11], [12], [13] for wired networks. This has not been widely deployed, because it requires signals to be sent from the application to the core network. In contrast, for wireless links connected directly to the host running the application, no protocol changes are needed. Prior work in the wireless context can be divided into rewarding schemes [5], [6] and pricing schemes [8], [9]. The first approach uses 802.11e's contention-free period (CFP) to provide extra throughput to the data class, which is problematic because current wireless NICs do not implement the CFP. The pricing approach requires either micropayments of monetary prices, which make implementation difficult, or must impose prices through some other form of service degradation such as packet drops, which seems counter-productive.

In this paper, we will go through two steps to design our scheme.

First, in Section II, we propose a scheme called "proportional tradeoff" which provides better service for both traffic types, by judicious choice of $CW_{\min}$ and *TXOP limit* for the two ACs. To gain insight into the scheme, we propose a model of 802.11 WLANs in Section III with these two types of traffic from which the advantage of the scheme is shown in Section IV.

Second, we use a game theoretic model presented in Section III to investigate the incentives provided by the proportional tradeoff scheme. We show that our proposed scheme does not eliminate data users' incentive to use the real-time class, despite providing better service for data users than the default EDCA parameters. Hence the proposed scheme is modified in Section V by reducing the data class's $CW_{\min}$ slightly to give throughput-sensitive applications the incentive to use the bulk-data service. This modification is called "proportional incentive adjusted" or "PIA". It is shown analytically that this gives the bulk-data class higher throughput than the real-time class under all traffic loads. It is also shown by simulation to give better delay performance to the real-time class than the parameter settings without service differentiation.

II. Proposed proportional tradeoff scheme

Our aim is to provide service differentiation for data traffic and real-time traffic, without the prioritization implied by standard 802.11e parameters. To this end, we now propose a mechanism which improves service for both types of traffic by increasing $CW_{\min}$ and *TXOP limit*. We do not use the AIFS parameter because it provides load-dependent prioritization which is not appropriate to achieve a "fair" service differentiation.

In particular, we define two service classes: "bulk data" abbreviated by B and "real-time" abbreviated by R. Users of class B can transmit more packets per channel access but less often while users of class R transmit more often but only one packet per channel access. To achieve this, class B has higher *TXOP limit* but commensurately higher $CW_{\min}$. This is similar to the method used in [4] to ensure fairness. Let W_t and W_d be the values of $CW_{\min}$ used by classes B and R respectively. Then

$$W_t = \eta W_d, \quad \eta > 1 \tag{1}$$

where η is the number of packets sent by class B within its *TXOP limit*. Class B is designed for throughput-sensitive data users, and class R is for delay-sensitive real-time users.

The logic behind this scheme is that real-time traffic requires low delay and often has only one packet to send at a time but the packet needs to be sent as soon as possible; hence, it always use class R. Whereas, data traffic requires high throughput; hence, it does not matter that they wait a little longer, if they are then able to transmit more once they gain access to the channel.

We will show that when data users use class B, this scheme actually improves service for both types of traffic. This benefit comes from the reduction of collision probability in the network due to the lower attempt probability of data sources.

Note that MAC parameter setting in our scheme is per flow, not per node, because each EDCA station can support different flows with different types of traffic.

III. Model

In this section, we present models (called wireless and game models) to study respectively the performance of our proportional tradeoff scheme given user choices, and the user choices it induces. Note that these models apply to arbitrary MAC parameters, not only those satisfying (1). This allows them to be used to study the incentive-adjusted scheme in Section V.

Consider a network consisting of $N_u \geq 0$ non-saturated sources with arrival rate λ, and $N_s \geq 1$ saturated sources in an infrastructure topology. The non-saturated sources represent the real-time users while data users are modeled as saturated sources. For modeling simplicity, we assume each station transmits packets of only one source, although this is not required by the scheme itself. Each unsaturated station uses class R (sends a burst of one packet per channel access) while each saturated station can use either class R or class B (sends a burst of $\eta \geq 1$ packets per channel access).

A. Wireless model

We now present, without justification, the model derived in [14]. Note that this model has been validated extensively in [14], showing its accuracy under a wide range of scenarios.

The model uses the following assumptions:

- Channel conditions are ideal (no channel errors, hidden terminals or capture effect) and EDCA operates in basic mode (no RTS/CTS).
- Unsaturated stations always do backoff before a transmission. This standard assumption (see, e.g., [15], [16]) is reasonable in the presence of saturated stations.
- Sources perform binary exponential backoff until they successfully transmit a packet (there is no retransmission limit and $CW_{\max} = \infty$). This assumption is made for notational and computational simplicity; however, simulations show that qualitative results from this

model still hold when these two backoff parameters are truncated as in the standard.

The model uses the following notation:

- Subscripts s, t, d, and u denote any saturated user, a saturated user using class B, a saturated user using class R, and an unsaturated user using class R, respectively.
- N_k ($k \in \{s,t,d,u\}$) denotes the number of users of type k. Note that $N_s = N_t + N_d$.
- W_k ($k \in \{t,d,u\}$) is the minimum contention window of users of type k. Recall that a saturated user using R transmits only one packet per channel access and has $W_d = W_u$.
- η is the *TXOP limit* of a user of type t, which is the number of packets it can send per channel access[1]; l_{sat} and l_{nonsat} are the payload length of a packet from saturated and unsaturated sources, respectively. It is assumed that $l_{nonsat} < l_{sat}$.
- τ_k ($k \in \{t,d,u\}$) is the probability that a user of type k attempts to transmit in a given slot.
- p_k ($k \in \{t,d,u\}$) is the collision probability of a packet from a user of type k.
- Y is the (random) virtual slot time, which is the time interval between the starts of two consecutive decrements of the backoff counter. In particular, it is σ if the slot is idle, T_u if the slot contains transmissions of only non-saturated stations, T_d if the slot contains collision involving at least one saturated station or a successful transmission of a saturated station using class R, and T_t if the slot contains a successful transmission of a saturated station using class B.

Note that T_u, T_d and T_t are given as follows

$$\begin{aligned} T_u &= T_{difs} + T_{nonsat} + T_{sifs} + T_{ack} \\ T_t &= T_{difs} + \eta(T_{sat} + T_{ack}) + (2\eta - 1)T_{sifs} \\ T_d &= T_{difs} + T_{sat} + T_{sifs} + T_{ack} \end{aligned} \tag{2}$$

where T_{difs}, T_{sifs} and T_{ack} are the duration of DIFS, SIFS and transmission of an ACK packet, respectively, and T_{sat} and T_{nonsat} are the transmission time of a packet from saturated and non-saturated sources, respectively.

The inputs to our model are N_t, N_d, N_u, W_t, $W_u = W_d$, η, λ, l_{sat} and l_{nonsat}. The outputs are attempt probabilities τ_k, collision probabilities p_k ($k \in \{t,d,u\}$), and the throughput of saturated sources.

Central to the model is a set of fixed-point equations, where the attempt probabilities of saturated and unsaturated sources are expressed in terms of the collision probabilities of saturated and unsaturated sources, and vice versa. The fixed point equations are as follows.

1) Fixed point model: First, the attempt probability of a saturated source of type $k \in \{t,d\}$ is

$$\tau_k = \frac{\mathbb{E}[\text{Attempts per burst}]}{\mathbb{E}[\text{Slots per burst}]} = \frac{\sum_{j=0}^{\infty} p_k^j}{\sum_{j=0}^{\infty}(\mathbb{E}[U_{kj}]+1)p_k^j}$$

where U_{kj} is a random variable representing the number of backoff slots in the jth backoff stage of a saturated station. U_{kj} is an integer uniformly distributed on $[0, 2^j W_k - 1]$, whence

$$\mathbb{E}[U_{kj}] = \frac{2^j W_k - 1}{2}.$$

Then,

$$\tau_k = \frac{2}{W_k \frac{1-p_k}{1-2p_k} + 1}, \quad k \in \{t,d\} \tag{3a}$$

Second, the attempt probability of an unsaturated source is

$$\begin{aligned} \tau_u &= \frac{\mathbb{E}[\text{Attempts per source per sec}]}{\mathbb{E}[\text{Number of system slots per sec}]} \\ &= \frac{\mathbb{E}[\text{Packets per source per sec}]\mathbb{E}[\text{Attempts per packet}]}{\mathbb{E}[\text{Number of system slots per sec}]} \\ &= \frac{\lambda \sum_{j=0}^{\infty} p_u^j}{(1/\mathbb{E}[Y])} \end{aligned}$$

where Y is the (random) virtual slot time. Thus

$$\tau_u = \lambda \mathbb{E}[Y] \frac{1}{1-p_u}. \tag{3b}$$

Finally, the collision probability of a station is probability at least one of the remaining stations transmits in a given slot, namely

$$p_k = 1 - \frac{(1-\tau_t)^{N_s - N_d}(1-\tau_d)^{N_d}(1-\tau_u)^{N_u}}{1-\tau_k}, \quad k \in \{t,d,u\} \tag{3c}$$

The fixed point is between the attempt probabilities in (3a)–(3b), and the collision probabilities in (3c), with the mean slot duration $\mathbb{E}[Y]$ given by

$$\mathbb{E}[Y] = a_\sigma \sigma + a_u T_u + a_d T_d + a_t T_t \tag{4a}$$
$$a_\sigma = (1-\tau_t)^{N_t}(1-\tau_d)^{N_d}(1-\tau_u)^{N_u} \tag{4b}$$
$$a_u = \left(1-(1-\tau_u)^{N_u}\right)(1-\tau_t)^{N_t}(1-\tau_d)^{N_d} \tag{4c}$$
$$a_t = N_t \tau_t (1-\tau_t)^{N_t - 1}(1-\tau_d)^{N_d}(1-\tau_u)^{N_u} \tag{4d}$$
$$a_d = 1 - (a_i + a_u + a_t) \tag{4e}$$

where a_σ, a_u, a_d, and a_t are the probabilities that a slot has the duration of σ, T_u, T_d, and T_t, respectively.

2) Throughput of data users: Two measures of throughput for saturated users are of interest here. In addition to the true throughput in packets/s, denoted S, some of our results apply to the more tractable measure of throughput in packets/slot, denoted C.

The throughput in packets/slot is given by

$$C_t = \tau_t(1-p_t)\eta \tag{5a}$$
$$C_d = \tau_d(1-p_d) \tag{5b}$$

where subscripts t and d denote saturated sources using class B and R, respectively.

The throughput in packets/s is given by

$$S_k = \frac{C_k}{\mathbb{E}[Y]}, \quad k \in \{t,d\}. \tag{6}$$

[1]This differs slightly from [1] in which *TXOP limit* is a duration.

B. Game model

1) Description: The above WLANs can be modeled as a game in which users are players. A player i can choose an action which is either to use class B or to use class R. Based on the actions of other players and its action, the player i will get a payoff, which is the throughput of a saturated user or the reciprocal of delay for an unsaturated station.

Using class R is a dominant strategy for unsaturated stations, since it reduces their delay regardless of what other stations do. For this reason, we will not treat unsaturated stations as players, but simply model their effect on the throughput obtained by the saturated users.

2) Model: A game of the wireless network described above is denoted by a quadruple $\langle \mathcal{P}, (A_i)_{i\in\mathcal{P}}, (u_i)_{i\in\mathcal{P}}, N_u \rangle$ where

- $\mathcal{P} = \{1, \dots, N_s\}$, the set of players, contains the saturated users.
- For every $i \in \mathcal{P}$, $A_i = \{B, R\}$ is the set of actions available to player i, where action B is to use MAC parameters $(CW_{\min}, TXOP) = (W_t, \eta)$, and action R is to use MAC parameters $(W_u, 1)$. Note that all the players have the same action space.
- For every $i \in \mathcal{P}$, the payoff $u_i(a)$ is the throughput of player i under action a in which each player l plays a_l. There are two forms of the game, corresponding to the two types of throughput described in Section III-A2:
  - Game 1: $u_i(a)$ is given by C_t of (5) if $a_i = B$, or C_d otherwise;
  - Game 2: $u_i(a)$ is given by S_t of (6) if $a_i = B$, or S_d otherwise.

  Note that the values of C_t, C_d, S_t and S_d depend implicitly on the entire action profile a.

Note that this is a symmetric game [17], since each player has the same opportunities, and for each player, the same actions yield the same payoffs.

Let us define the action profiles as follows

$$\begin{aligned} a_A &= (a_1 = B, a_{-1} = \text{all B}) \\ a_B &= (a_1 = R, a_{-1} = \text{all B}) \\ a_M &= (a_1 = B, a_{-1} = \text{all R}) \\ a_N &= (a_1 = R, a_{-1} = \text{all R}) \end{aligned}$$

where a_1 is the action of the player 1 while a_{-1} is the action profile of other data users except the user 1.

Besides, we also define A_D as a set of action profiles given as follows

$$A_D = \Big\{ a \in A | (a_1 = R \text{ and } a_2 = B) \Big\}$$

and a_D as a particular action profile in A_D.

The properties of a game can be investigated using the wireless model (3)–(6).

IV. Properties of proportional tradeoff scheme

In this section, we first use the MAC model (3)–(6) to derive some properties of the proportional tradeoff scheme. Then, we validate these results using ns-2 simulation.

Recall that under the proportional scheme, relationship of CW_{min} between class B and class R is given in (1).

A. Theoretical results

We will investigate properties of the proportional scheme, both when data users always use class B ("simple" users), and where they use whichever class gives them higher throughput ("measurement driven" users). Even though our results are proved for unbounded $CW_{\max}$ and number of retransmissions and some of them are for a network with only data users, we will show by simulation that they apply when these assumptions are lifted.

1) Simple users: Here we use the wireless model to study network performance under the proposed scheme.

The following theorem, proven for a network with only and arbitrary number of data users, states that with the proportional scheme ($\eta > 1$), data users will be receiving higher throughput than when there is no service differentiation ($\eta = 1$). This theorem is a special case of Theorem 6 and so is the proof.

Theorem 1: Under the wireless model based on (3)–(6) with $W_t = \eta W_u$ ($\eta \geq 1$, $W_u > 1$), $N_d = N_u = 0$, the throughput in packets/s of each data user increases with η.

The above theorem is based on the following lemma proven in Appendix A-A.

Lemma 1: Under the wireless model (3)–(4) with $N_d = N_u = 0$ and $W_u > 4$, the collision probability and attempt probability of each data user decrease with the increase of its CW_{min}.

This lemma suggests that under the proportional tradeoff mechanism where CW_{min} of data users is proportionally scaled with η, their collision probability reduces, which explains for the increase of their throughput as stated in Theorem 1.

To show the benefit of the proportional tradeoff scheme for both data and real-time users, we consider a simple network of mixed traffic. The following theorem shows that under this scheme, both data and real-time users can improve their service. This is proved in Appendix A-B. Note that to keep the algebra tractable, assuming that $W_k >> 1$, we simplify (3a) of the wireless model by

$$\tau_k = \frac{2}{W_k} \frac{1 - 2p_k}{1 - p_k}. \tag{7}$$

Theorem 2: Under the wireless model based on (3)–(5) with (3a) replaced by (7), $W_t = \eta W_u$, $N_u = N_t = 1$, $N_d = 0$, and $\lambda T_u \leq 1$, (T2-1) the throughput per slot of a saturated station increases with an infinitesimal increase of η at $\eta = 1$. Moreover, (T2-2) the collision probability of an unsaturated station decreases when η increases.

According to (T2-1), the benefit of the proposed scheme for saturated sources is observed locally around $\eta = 1$. This is quite a disappointing result due to the use of approximation (7); however, at larger η, we found by numerical results that the result still holds except when T_{difs} is very close to T_{sifs}, which is not the case in 802.11 standards. Besides, we also found that without that approximation, the result holds for any $\eta \geq 1$. However, the model without the approximation is algebra intractable.

Although the result in Theorem 1 is for scenarios with only data users and that in Theorem 2 is for a simple mixed-traffic scenario, we show by simulation that they hold for most scenarios with mixture of traffic types and higher number of stations of each type. In particular, simulation shows the fact that collision probability of unsaturated stations reduces will lead to the decrease of their mean delay except under light traffic load. Under light traffic load, when saturated stations transmit less often but longer bursts, the reduction of collision probability of an unsaturated station does not dominate the increase of the residual time which is an additional delay observed by an unsaturated station arriving at an empty queue and senses channel busy due to ongoing transmission from other stations.

2) Measurement driven users: When data users are measurement driven (using whichever class gives them higher throughput), we use the game model to analyze network performance. Under the proportional scheme, the action space of the game is

$$A0 = \{(\eta W_u, \eta), (W_u, 1)\}, \quad \text{for a given } \eta > 1$$

where $(\eta W_u, \eta)$ and $(W_u, 1)$, respectively, are MAC parameters of class B and R.

The main result in this section is that the proportional tradeoff scheme still creates an unfortunate incentive for a data user to choose class R, although this incentive is small. This will be shown through two theorems below by comparing the throughput of a data user choosing class R first with that of another data user choosing class B in the same network and then with its own throughput when it chooses class B.

We start with the following theorem, proven in Appendix A-D.

Theorem 3: Under the wireless model (3)–(6), in the game $\langle \mathcal{P}, (A0_i)_{i\in\mathcal{P}}, (S_i)_{i\in\mathcal{P}}, N_u\rangle$ with $W_i > 4$, a data user using R has higher throughput than other data users using B in the same network. In particular,

$$S_1(a_D) > S_2(a_D)$$

This theorem is based on the following lemma, proven in Appendix A-C.

Lemma 2: Under the wireless model (3)–(4), (4) with $N_d \geq 1$, when $W_t > W_d > 4$, data users using voice parameters will attempt to transmit more often than the rest using data parameters, or

$$\tau_d > \tau_t$$

The following theorem states that, when the actions of all but one user are the same, the remaining user is better off by using class R. The theorem is proven in Appendix A-F.

Theorem 4: Consider the wireless model based on (3)–(5) with (3a) replaced by (7), in the game $\langle \mathcal{P}, (A0_i)_{i\in\mathcal{P}}, (C_i)_{i\in\mathcal{P}}, 0\rangle$ with $W_i > 4$. For action profiles in which all users $i \neq 1 \neq j$ have the same action $a_i = a_j$, data user 1 gets higher throughput per slot when using class R than when using class B.

Although the throughput in Theorem 4 is in packets/slot, simulation demonstrates that this result still holds for packets/s.

Recall that an action profile is a *Nash equilibrium* if no player can get higher payoff by changing its action while others keep theirs unchanged [18]. From Theorem 4, the action profile with all data users using class R is the unique symmetric Nash equilibrium. Then, according to Theorem 1, the throughput of each data user at Nash equilibrium is less than that when all of data users use class B. Section V will consider an improved scheme that avoids that issue.

The proof of Theorem 4 is based on the following lemma.

Lemma 3: Consider the wireless model based on (3)–(5) with (3a) replaced by (7), in the game $\langle \mathcal{P}, (A0_i)_{i\in\mathcal{P}}, (C_i)_{i\in\mathcal{P}}, 0\rangle$ with $W_i > 4$. For action profiles in which all users $i \neq 1 \neq j$ have the same action $a_i = a_j$, data user 1 transmits more often when using class R than when using class B.

The result of this lemma explains for the increase of the throughput as stated in Theorem 4.

B. Simulation results and discussion

Recall that the properties of proportional tradeoff scheme in Sec. IV-A are proved for unbounded $CW_{\max}$ and number of retransmissions and some of them are for a network with only data users. In this section, we will use simulation to validate those in more general scenarios with both data and real-time users, and a limited number of retransmissions. The simulations used *ns-2.33* [19] with the EDCA package [20].

We simulated a network as described in Section III. The traffic type used for unsaturated sources is Poisson. Saturated sources receive CBR traffic at a rate faster than they can transmit. All stations use the user datagram protocol (UDP). The general MAC and physical layer parameters are set to the default values in IEEE 802.11b, as shown in Table II.

The MAC parameters specific to classes B and R used in the proportional scheme are given in Table III with $W_t = \eta W_u$.

1) Simple users ($N_d = 0$):

a) Scenario 1 ($N_t = N_u = 1$): Figs. 1 and 2 show the throughput of a data user and the collision probability of an unsaturated station, respectively. As can be seen, when η increases, the throughput increases and the collision probability decreases, which shows the benefit of the proportional scheme. This confirms the result of Theorem 2.

TABLE II
802.11B MAC AND PHY PARAMETERS

Parameter	Symbol	Value
Data bit rate	r_{data}	11 $Mbps$
Control bit rate	r_{ctrl}	1 $Mbps$
PHYS header	T_{phys}	192 μs
MAC header	l_{mac}	288 $bits$
UDP/IP header	l_{udpip}	160 $bits$
ACK packet	l_{ack}	112 $bits$
Slot time	σ	20 μs
SIFS	T_{sifs}	10 μs
Retry limit		7

TABLE III
MAC PARAMETERS OF CLASSES B AND R

Class	CW_{min}	CW_{max}	AIFSN	*TXOP limit* (packets)
B	W_t	$2^5 W_t$	2	η
R	W_u	$2^5 W_u$	2	1

b) Scenario 2 ($N_t > 1$, $N_u > 1$): To highlight the advantage of our scheme, we compare it with the default EDCA parameters (Table I) within the same scenarios (same N_s, N_u, λ, l_{sat}, l_{nonsat}).

The throughput of a data user, and mean delay and loss probability of a real-time user, under the proportional scheme are shown in Figs. 3 and 4 and Table IV respectively, as functions of η for different N_s. Note that in Figs. 3 and 4, the performance metric at each η is normalized by that at $\eta = 1$. Moreover, the performance metrics under the default parameter setting (Table I) with all data users using class AC_BE and real-time users using class AC_VO and under the default parameter setting with all users using class AC_VO are also shown for comparison. The performance metric of the default parameter setting in Figs. 3 and 4 is also normalized by that of the proportional scheme at $\eta = 1$.

From Fig. 3, the throughput of a data user increases with η, which suggests that the proportional scheme with $\eta > 1$ always provides better service for data users than the scheme with no service differentiation ($\eta = 1$). This corroborates the result of Theorem 1. Also our scheme provides significantly better throughput for data users than

TABLE IV
LOSS PROBABILITY OF A REAL-TIME USER

	η \ N_s	2	10	18
Proportional tradeoff	1	–	1.62e-4	8.17e-4
	2	–	1.9e-5	8.1e-5
	3	–	–	1.4e-5
	4	–	–	1.4e-5
	5	–	–	0.8e-5
	6	–	–	–
	7	–	–	–
Default setting with all data users using AC_BE		–	–	3.1e-5
Default setting with all data users using AC_VO		1.0e-4	7.37e-2	26.07e-2

("–" denotes no loss found during simulation time)

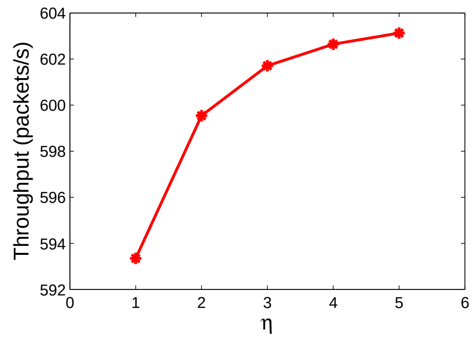

Fig. 1. Throughput of a data user as a function of class B's *TxOP limit* in packets (η). ($\lambda = 30$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $N_t = N_u = 1$, $N_d = 0$, $W_u = 32$, $W_t = \eta W_u$.)

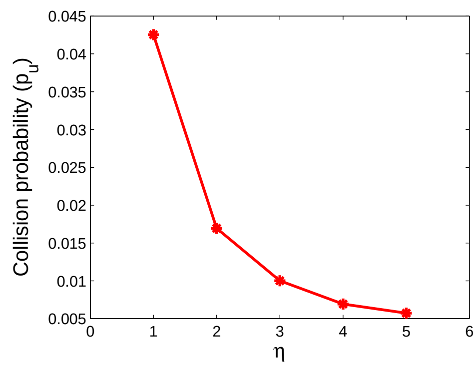

Fig. 2. Collision probability of a real-time user as a function of class B's *TxOP limit* in packets (η). ($\lambda = 30$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $N_t = N_u = 1$, $N_d = 0$, $W_u = 32$, $W_t = \eta W_u$.)

does the default parameter setting, either with all data users using class AC_BE or with all data users using class AC_VO.

Note that the benefit of the proposed scheme increases with the contention level in the network. In Fig. 4, when the load is high enough, our scheme with $\eta > 1$ provides significant improvement in terms of mean delay of real-time users in comparison with the case of no service differentiation ($\eta = 1$). At light load (e.g. $N_s = 2$), the improvement is negligible. This is acceptable because delay only matters at high load. Figure 4 also suggests that at each network load, there exists an optimal value of η at which mean delay is minimum (e.g. $\eta = 2$ at $N_s = 2$ and $\eta = 6$ at $N_s = 10$). This optimal η increases with the network load. Besides, Table IV shows that the loss probability of real-time traffic decreases with the increase of η. Also compared with the default parameter setting which prioritizes real-time traffic with all data users using class AC_BE, we expect the performance will be worse for real-time users under the proportional scheme as shown in Fig. 4 and Table IV. However, compared with the default parameter setting with all data users using class AC_VO, our proportional scheme provides much better service for real-time users. (The apparent exception for $N_s = 18$ is due to the much higher loss rate under the default scheme when all users use AC_VO; see also Fig. 10.)

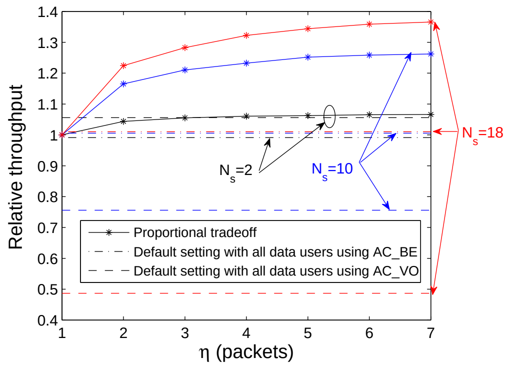

Fig. 3. Throughput of a data user as a function of class B's *TxOP limit* in packets (η), scaled by that of the proportional scheme at $\eta = 1$. ($\lambda = 20$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $N_u = 4$, $W_u = 32$, $W_t = \eta W_u$.)

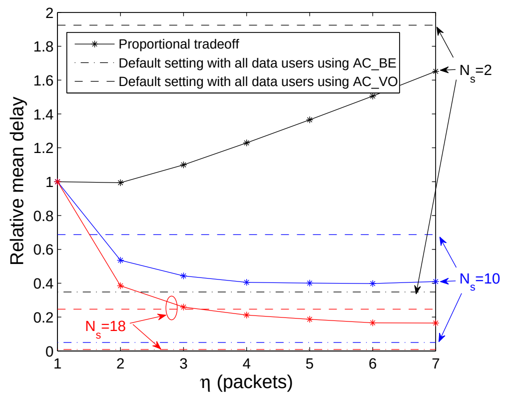

Fig. 4. Mean delay of a real-time user as a function of class B's *TxOP limit* in packets (η), scaled by that of the proportional scheme at $\eta = 1$. ($\lambda = 20$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $N_u = 4$, $W_u = 32$, $W_t = \eta W_u$.)

Although the optimal η in our proportional scheme depends on traffic load, the majority of the benefit is obtained by $\eta = 2$.

2) One measurement driven user ($N_d \leq 1$):

a) A data user using class R has higher throughput than another using class B: We simulated the network scenario with $\lambda = 20$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $N_u = 4$, $W_u = 32$, $W_t = \eta W_u$, η varied, $N_s = \{2, 10, 18\}$ consisting of $N_d = 1$ and $N_t = N_s - 1$. What we have found is that a data user using class R gains higher throughput than other data users using B, which confirms the result of Theorem 3.

b) A data user gets higher throughput when using class R than using class B: To see if a data user has incentive to use class R, we consider the case that $N_s - 1$ data users use class B while the other chooses whether to use action R (outcome a_B) or not (outcome a_A). The ratio of throughput in packet/s of the undecided data user in a_B and a_A are shown in Fig. 5 as functions of η for different N_s. For comparison, we also show the ratio of throughput of a data user when it uses class AC_VO and class AC_BE while other data users use class AC_BE under the default parameter setting.

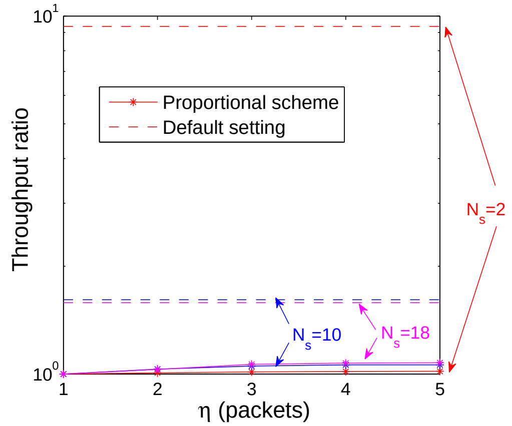

Fig. 5. Ratio of throughput in packets/s of a data user when it uses "real-time" class and "data" class while other data users use "data" class as a function of class B's *TxOP limit* in packets (η). ($\lambda = 20$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $N_u = 4$, $W_u = 32$, $W_t = \eta W_u$.)

Observe in Fig. 5 that while other data users use "data" class, a data user can improve its throughput by using real-time class under both the proportional scheme and the default parameter setting. The throughput improvement under the default parameter setting is much higher than under the proportional scheme. This shows that the incentive to use real-time class of data users under the proportional scheme is greatly reduced from the default parameter setting, even though it is not entirely eliminated.

3) Many measurement driven users ($N_d \leq N_s$): The above shows that it is not a Nash equilibrium for all data users to use class B. To see if it is a Nash equilibrium for all to use class R, we consider the case that $N_s - 1$ data users use class R while the other chooses whether to use action R (outcome a_N) or not (outcome a_M). The ratio of throughput S in packets/s of the undecided data user, user 1, in these two outcomes are shown in Fig. 6 as a function of η for different N_s.

From Fig. 6, the throughput ratio is higher than 1 for $\eta > 1$, which means that while other data users use class R, a data user can improve its throughput by acting in the same way. This confirms that an action profile in which all data users choosing class R is a Nash equilibrium, as stated in Theorem 2. At the same time, we have found that the throughput achieved in this equilibrium is less than what could be obtained when all data users use class B.

However, this equilibrium gives a lower throughput than could be obtained when all data users use class B, as shown by the increase in relative throughput with η in Fig. 3. We next investigate a way to avoid this undesirable equilibrium.

V. Incentive Adjusted Scheme

Section IV-B showed that for networks with both data and voice users, our proportional scheme can improve service for both types of traffic relative to the scheme with no service differentiation, especially at high load.

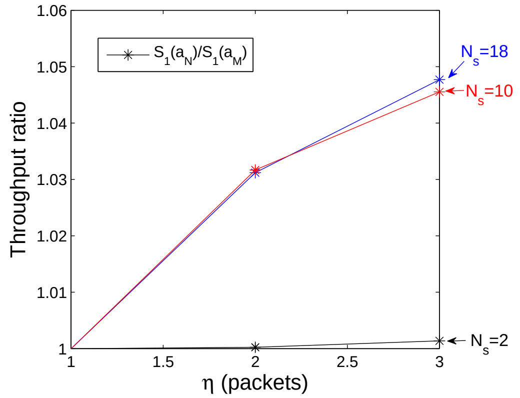

Fig. 6. Ratio of throughput in packets/s of the data user 1 when it uses class R ($S_1(a_N)$) and class B ($S_1(a_M)$) while other data users use class R as a function of class B's *TxOP limit* in packets (η). ($\lambda = 20$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $N_u = 4$, $W_u = 32$, $W_t = \eta W_u$.)

However, when a small fraction of data users chooses to use class R, their throughput can be slightly improved. Although the improvement is small, measurement-driven application design will still result in class R being chosen by throughput-sensitive applications. However, we will now show that our proposed scheme can easily be modified to eliminate this incentive issue. This is in contrast to priority-based schemes, which require explicit policing or pricing mechanisms.

A. Description of the incentive adjusted scheme, PIA

We modify the proportional scheme by reducing CW_{min} of class B by an amount $\epsilon > 0$ and hence providing higher benefit for users of class B. The reduction in $CW_{\min}$ for class B results in more throughput for a data user when it uses B compared with when it uses R, and thus data users have no incentive to use the real-time class. Recall that users can only select one of the access classes determined by the access point, and cannot choose arbitrary combinations of parameters.

In particular, under the PIA,

$$W_t = \eta W_u - \epsilon, \quad \eta > 1 \;\&\; \epsilon < (\eta - 1)W_u \tag{8}$$

Note that the performance of delay-sensitive users degrades as ϵ increases, and so we would like to use the smallest ϵ as possible.

B. Properties of the PIA

1) Theoretical results: We will investigate properties of the PIA, both when data users always use class B ("simple" users), and where they use whichever class gives them higher throughput ("measurement driven" users). Even though our results are proved for a network with only data users and for unbounded $CW_{\max}$ and number of retransmissions, we will show by simulation that they apply when these assumptions are lifted.

a) Simple users: Here we use the wireless model to study network performance under the PIA.

First, we define another variable θ which is given in terms of ϵ as follows

$$\theta = \frac{\epsilon}{\eta} \tag{9}$$

The following theorem states that the modified scheme provides better service for data users at higher η. The proof is given in Appendix B-A.

Theorem 5: Under the wireless model based on (3)–(6) with $W_t = \eta(W_u - \theta)$, $0 \leq \theta < (W_u - 1/\eta)$, $W_u > 1$, and $N_d = N_u = 0$, the throughput in packets/s of each data user is an increasing function of η for $\eta \geq 1$.

The following corollary comes from the result of the above theorem and Lemma 2.

Corollary 1: Under the wireless model based on (3)–(6) with $W_t = \eta(W_u - \theta)$, $W_u > 4$, and $N_d = N_u = 0$, the throughput in packets/s of each data user under the modified scheme ($\eta > 1$, $0 < \theta < W_u(1 - 1/\eta)$) is higher than that when there is no service differentiation ($\eta = 1$, $\theta = 0$). ■

b) Measurement driven users: When data users are measurement driven (using whichever class gives them higher throughput), we use the game model to analyze network performance. Specifically, under the incentive adjusted scheme PIA, the action space in the game model has the form

$$A1 = \{(\eta W_u - \epsilon, \eta), (W_u, 1)\}, \quad \text{for a given } \eta > 1 \;\& \;\; \epsilon < (\eta - 1)W_u$$

where $(\eta W_u - \epsilon, \eta)$ and $(W_u, 1)$, respectively, are MAC parameters of class B and R.

In this section, we first provide a lower bound on the value of ϵ, denoted by ϵ_0, such that bulk data users using class B get a higher throughput than those using class R, regardless of the network load. Then, we prove that when $\epsilon = \epsilon_0$, the action profile in which all data users use class B is the unique symmetric Nash equilibrium.

The following theorem, proven in Appendix B-B, provides a lower bound on the value of ϵ so that the throughput of a data user using class R is lower than that of another data user using class B in the same network. It's surprising that this lower bound only depends on η.

Theorem 6: Under the wireless model based on (3)–(6), in the game $\langle \mathcal{P}, (A1_i)_{i\in\mathcal{P}}, (S_i)_{i\in\mathcal{P}}, N_u \rangle$ with $W_i > 4$, when

$$\epsilon \geq \epsilon_0 = 4(\eta - 1) \tag{10}$$

then data users using B get higher throughput than those using R. In particular,

$$S_1(a_D) < S_2(a_D)$$

Then, under the modified scheme with $\epsilon = \epsilon_0$, using R is no longer a preferred choice for data users because they can gain higher throughput when using B. This is reflected in the following theorem proven in Appendix B-C.

Theorem 7: Consider the wireless model based on (3)–(5) with (3a) replaced by (7), in the game $\langle \mathcal{P}, (A0_i)_{i\in\mathcal{P}}, (C_i)_{i\in\mathcal{P}}, 0\rangle$ with $\epsilon = \epsilon_0$ and $W_i > 4$. For action profiles in which all users $i \neq 1 \neq j$ have the same action $a_i = a_j$, data user 1 gets higher throughput per slot when using class B than when using class R.

Although the throughput in Theorem 7 is in packets/slot, simulation demonstrates that this result still holds for packets/s.

From Theorem 7, the action profile with all data users using class B is the unique symmetric Nash equilibrium. Then, according to Corollary 1, the throughput of each data user at Nash equilibrium is greater than that when all of data users use class R.

Conjecture 1: When $\epsilon = \epsilon_0$, the modified scheme guarantees that data users have no incentive to use class R.

We do not have rigorous proof for this conjecture but we will show by simulation that it holds.

2) Simulation results and discussion: Recall that the properties of the PIA in Sec. V-B1 are proved for a network with only data users and for unbounded $CW_{\max}$ and number of retransmissions. In this section, we will use simulation to validate those in more general scenarios with both data and real-time users, and a limited number of retransmissions. The simulations used *ns-2.33* [19] with the EDCA package [20].

Note that the simulated network in this section is the same as one in Sec. IV-B.

To highlight the advantage of the incentive adjusted scheme PIA when all data users are measurement driven, we compare it with the case of no service differentiation ($\eta = 1$,$\epsilon = 0$) and the default EDCA parameter setting (Table I). Recall that under PIA, data users have no incentive to use real-time class. In contrast, all data users have incentive to use highest priority class in the default priority scheme.

In this section, the MAC parameters specific to classes B and R in our PIA scheme are also given in Table III with $W_t = \eta W_u - \epsilon$ and $\epsilon = \epsilon_0$.

a) Incentive compatibility: We will first verify that throughput-sensitive users have an incentive to choose class B, regardless of the class chosen by other users. To do this, we simulated a scenario with $N_s = 5$ data users, and plotted the results in Figure 7. Other parameters were $\lambda = 30$ pkts/s, $l_{sat} = 1040$B, $l_{nonsat} = 100$B, $N_u = 6$, $W_u = 32$, $W_t = \eta W_u - \epsilon_0$, $\eta = 2$, $\epsilon_0 = 4$.

Figure 7 shows the throughput a saturated user obtains by using either class B or class R, as a function of the number of competing data users who use class B. Regardless of the choices of the other users, a given saturated user gets a higher throughput by choosing the bulk-data class. If all data users choose class B, then the total throughput is maximal.

This was the objective in selecting ϵ_0. However, the result is stronger than is ensured by Theorem 7, since the latter only applied to the cases when either none or all of the competing data users used class R. Thus the simulation suggests that PIA is actually incentive compatible, and causes bulk data users to choose class B. This confirms Conjecture 1.

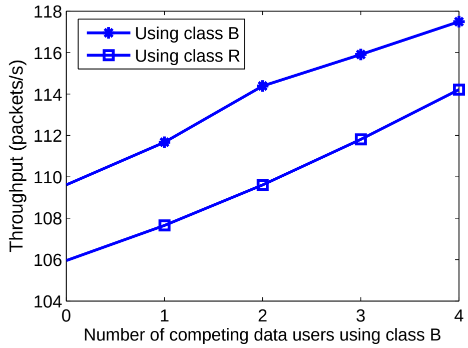

Fig. 7. Throughput of class-B and class-R data users as a function of the number of class-R data users N_d. ($\lambda = 30$ pkts/s, $l_{sat} = 1040$B, $l_{nonsat} = 100$B, $N_u = 6$, $N_s = 5$, $W_u = 32$, $W_t = \eta W_u - \epsilon_0$, $\eta = 2$.)

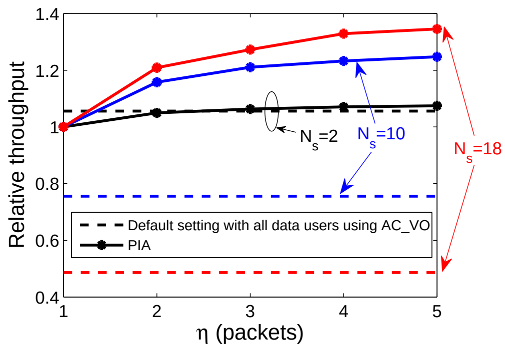

Fig. 8. Throughput of a data user as a function of class B's *TxOP limit* in packets (η), scaled by that of PIA at $\eta = 1$. ($\lambda = 20$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $N_u = 4$, $W_u = 32$, $W_t = \eta W_u - \epsilon_0$ for $\eta = \{1, 2, 3, 4, 5\}$.)

b) Comparison with the default EDCA parameters: We can now compare the performance of the proposed PIA with that of the default QoS classes, under the assumption that all users will use whatever class gives them the best performance. Under the default parameters, all users will use AC_VO, and under PIA bulk data users will use class B and real-time users will use class R.

We look at the scenarios where $\lambda = 20$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $N_u = 4$, $N_s = \{2, 10, 18\}$, $W_u = 32$, $W_t = \eta W_u - \epsilon_0$, $\epsilon_0 = 4(\eta - 1)$ and η is varied from 1 to 5. The case of no service differentiation ($\eta = 1$) is included for comparison.

The throughput of a saturated user under PIA is shown in Fig. 8 as functions of η for different numbers of saturated users, N_s. For comparison, the throughput under the default parameter setting (Table I) is also shown. The throughput is again normalized by that of PIA at $\eta = 1$.

The throughput increases faster with η under PIA than it did under the proportional scheme, which reflects the reduction in $CW_{\min}$. This shows that PIA provides better

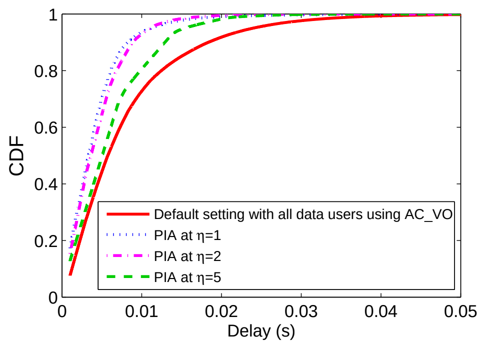

Fig. 9. Probability a packet of real-time users is successfully delivered as a function of delay. ($\lambda = 20$ pkts/s, $l_{sat} = 1040$B, $l_{nonsat} = 100$B, $N_u = 4$, $N_s = 2$,$W_u = 32$, $W_t = \eta W_u - \epsilon_0$ for $\eta = \{1, 2, 5\}$.)

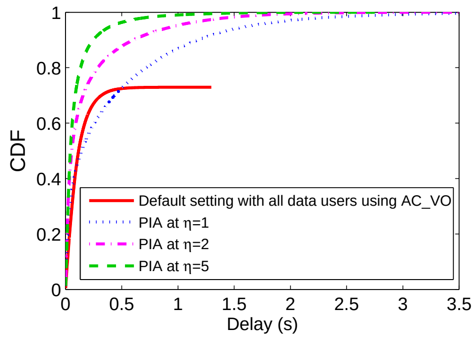

Fig. 10. Probability a packet of real-time users is successfully delivered as a function of delay. ($\lambda = 20$ pkts/s, $l_{sat} = 1040$B, $l_{nonsat} = 100$B, $N_u = 4$, $N_s = 18$,$W_u = 32$, $W_t = \eta W_u - \epsilon_0$ for $\eta = \{1, 2, 5\}$.)

service for data users than the proportional scheme without service differentiation ($\eta = 1$), especially at high load. This is in contrast to the default parameter setting with all data users using real-time class (AC_VO), for which the performance degrades rapidly at high load. For low load ($N_s = 2$), the default parameter setting performs better than the system without QoS differentiation ($\eta = 1$), because the more aggressive choice of $CW_{\min}$ is better matched to a small number of stations.

This improvement in throughput of PIA comes at the expense of increased delay for real-time users. Figs. 9 and 10 show the probability that a packet of a real-time user is successfully transmitted before a given delay, for different η and loads $N_s = 2$ and $N_s = 18$.

Figure 9 shows that PIA at both $\eta = 2$ and $\eta = 5$ gives a higher probability that a packet is successfully delivered at a given delay than the default parameter setting with all data users using the real-time class. This means that the average packet delay under PIA is smaller. In this lightly loaded case, $\eta = 2$ provides comparable service to $\eta = 1$ (no service differentiation), and $\eta = 5$ provides slight degradation, but less than that caused by the default prioritization.

In the heavily loaded case of Fig. 10, the cumulative distribution of delay for the default parameter setting never reaches 1, which indicates a high loss rate. In contrast, PIA has a low loss rate for all values of η tested, although some packets have very high delays. In this case, the benefit increases as η increases. Together with the result in Fig. 9, this implies that under PIA, the optimal η for real-time users increases with traffic load, as was observed for the proportional scheme. However, even using $\eta = 2$ for all loads appears to provide improvement over the default parameters.

In summary, although the optimal η in PIA depends on traffic load, it is clear that when $\eta = 2$, PIA provides better service for both types of traffic under typical scenarios. This implies that when designing a network with an unknown number of users, PIA can be implemented by simply setting $\eta = 2$ and $\epsilon = \epsilon_0 = 4$. Adaptive schemes that set η dependent on the estimated load are possible, but out of scope of this paper.

VI. Conclusion

It is important to be able to provide differentiated services, without giving all users the incentive to use a “highest priority” class. This paper has shown through both analysis and simulation that allowing users to adjust $CW_{\min}$ and *TXOP limit* in the same proportion provides service differentiation in WLANs. This scheme improves service for both data and real-time traffic, especially at high load. However, it still provides a slight incentive for data users to use real-time class’s parameters. This misalignment of incentives can be removed by increasing CW_{min} by a slightly smaller factor than the *TXOP limit*. For a given *TXOP limit* of data users, there always exists such a factor regardless of the traffic load. Our incentive adjusted scheme has many advantages over prior proposals: it improves service for both data and real-time traffic and provides the correct incentives for application optimizers, while allowing easy implementation by simply setting 802.11e MAC parameters.

Appendix A
Proofs of Properties of the Proportional Scheme

A. Proof of Lemma 1

Proof: When $N_u = 0$ and $N_d = 0$, (3) can be rewritten as follows

$$\tau_t = \frac{2}{W_t \frac{1-p_t}{1-2p_t} + 1} \equiv g_1(p_t) \quad (11a)$$

$$p_t = 1 - (1 - \tau_t)^{N_s - 1} \equiv g_2(p_t) \quad (11b)$$

From (11b), we have

$$\tau_t = 1 - (1 - p_t)^{1/(N_s - 1)} \quad (12)$$

The solution of (11) is the solution to $g_1(p_t) = g_2(p_t)$. First, for finite N_s, we have

$$g_1(0) = \frac{2}{W_t + 1} > g_2(0) = 0$$

$$g_1(1/2) = 0 < g_2(1/2) = 1 - (1/2)^{1/(N_s - 1)}$$

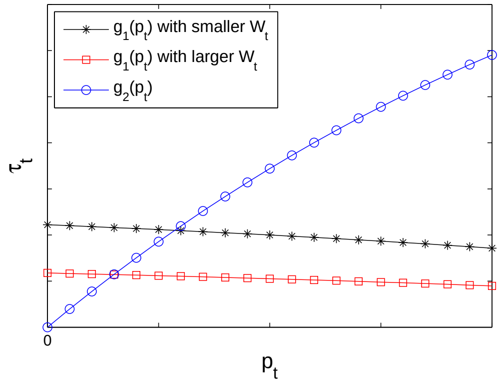

Fig. 11. Graphs of (11a) and (12) at different W_t.

This, together with the fact that $g_1(p_t)$ and $g_2(p_t)$ are continuous functions over $[0, \frac{1}{2}]$, implies that there exists solution to $g_1(p_t) = g_2(p_t)$.

Second, $g_2(p_t)$ is an increasing function of p_t and $g_1(p_t)$ is a decreasing function of p_t. Hence, it can be concluded that there is only one solution to $g_1(p_t) = g_2(p_t)$.

Define $g(p_t, W_t)$ by

$$g(p_t, W_t) = g_1(p_t) - g_2(p_t).$$

Let p_{t1} and p_{t2}, respectively, be the solution to $g(p_t, W_t) = 0$ at $W_t = W_{t1}$ and $W_t = W_{t2} > W_{t1}$.

It is clear that $g(p_t, W_t)$ is a decreasing function of W_t; hence, $g(p_{t2}, W_{t1}) > g(p_{t2}, W_{t2}) = g(p_{t1}, W_{t1}) = 0$. This, together with the fact that $g(p_t, W_t)$ is a decreasing function of p_t, implies that $p_{t2} < p_{t1}$.

From (11b), $p_{t2} < p_{t1}$ implies $\tau_{t2} < \tau_{t1}$. Therefore, when W_t increases, the collision probability and attempt probability of each data user decrease. This is illustrated in Fig. 11.

■

B. Proof of Theorem 2

Proof: Here we consider the simplest scenario: $N_u = N_t = 1$, $N_d = 0$. The throughput of saturated sources and the collision probability of unsaturated sources are used to evaluate the benefit of the proportional tradeoff scheme.

We will first prove Claim (T2-1) that under the proportional scheme (1), the throughput of saturated stations increases when η increases by a small amount around $\eta = 1$. Then, we will prove Claim (T2-2) that the collision probability of unsaturated stations decreases with the increase of η.

Although the scenario is simple, the proof is quite complicated because $\mathbb{E}[Y]$ depends in a complicated way on η.

1) Proof of Claim (T2-1) of Theorem 2: From (7) and (1), throughput per slot of saturated stations C_t in (5a) can be rewritten as follows

$$C_t = \frac{2}{W_u}(1 - 2p_t)$$

TABLE V
MATH EXPRESSION OF SYMBOLS IN THEOREM 2.

Symbol	Expression
K	$(2/W_u)(\eta - 4/W_u)(\sigma - T_u)$ $-(T_u - T_d)(4/W_u - \eta\tau_t)^2 + (8/W_u^2)(T_t - T_d)$
L	$(2/W_u)(\sigma - T_u)\tau_t(1-\tau_t)$ $+(2/W_u)(4/W_u - \tau_t)\tau_t(T_d - T_{difs})$
F	$1 + \lambda\Big((\sigma - T_u) + \frac{(4/W_u - \eta\tau_t)^2}{(2/W_u)\eta(1-\tau_t)^2}T_d$ $+ \frac{4/W_u - \eta\tau_t^2}{\eta(1-\tau_t)^2}(T_t - T_d)\Big)$
Z	$\frac{\lambda}{(1-\tau_t)^2}\Big(\frac{-\tau_t}{\eta}\Big(T_d + \frac{2(4/W_u - \eta\tau_t^2)(T_t - T_d)}{W_u(4/W_u - \eta\tau_t)^2}\Big)$ $+ \frac{(2/W_u)\tau_t(1-\tau_t)(4/W_u - \tau_t)(T_d - T_{difs})}{(4/W_u - \eta\tau_t)^2}\Big)$
H	$\frac{-(2/W_u)\tau_t}{(4/W_u - \eta\tau_t)^2} - \frac{\lambda L}{1-\tau_t}$
Q	$\frac{\eta(2/W_u)}{(4/W_u - \eta\tau_t)^2} + \frac{\lambda\mathbb{E}[Y]}{(1-\tau_t)^2} + \frac{\lambda K}{1-\tau_t}$

which increases with the decrease of p_t. Then, showing that saturated stations actually improve their throughput when η increases is equivalent to showing that p_t decreases with η. Thus, it is sufficient to show that $dp_t/d\eta < 0$.

By (3c), $p_t = \tau_u$ and $p_u = \tau_t$, whence by (3b),

$$p_t = \lambda E[Y]\frac{1}{1-\tau_t} \tag{13}$$

Taking derivative of (13) with respect to η,

$$\frac{dp_t}{d\eta} = \frac{\lambda}{(1-\tau_t)^2}\left(\frac{d\mathbb{E}[Y]}{d\eta}(1-\tau_t) + \mathbb{E}[Y]\frac{d\tau_t}{d\eta}\right) \tag{14}$$

Because $\mathbb{E}[Y]$ is a function of τ_t and τ_t is a function of η, to have a closed form expression of (14), we first express $d\mathbb{E}[Y]/d\eta$ as function of $d\tau_t/d\eta$ and then, express $d\tau_t/d\eta$ as function of $dp_t/d\eta$ as follows.

From (7),

$$p_t = 1 - \frac{2/W_u}{4/W_u - \eta\tau_t} \tag{15}$$

Substituting (15) into (4) and after some simple transformation steps, we have

$$\begin{aligned}\mathbb{E}[Y] =& (\sigma - T_u)(1-\tau_t)\frac{2/W_u}{4/W_u - \eta\tau_t} \\ &+ (T_u - T_d)(1-\tau_t) + T_d \\ &+ (T_t - T_d)\tau_t\frac{2/W_u}{4/W_u - \eta\tau_t}\end{aligned} \tag{16}$$

From (2), T_u and T_d are independent of η but T_t is a function of η. Hence,

$$\frac{d\mathbb{E}[Y]}{d\eta} = \frac{1}{(4/W_u - \eta\tau_t)^2}\left(K\frac{d\tau_t}{d\eta} + L\right) \tag{17}$$

where K and L are given in Table V.

Moreover, $d\tau_t/d\eta$ is expressed as a function of $dp_t/d\eta$ by taking derivative of (7) with respect to η as follows

$$\frac{d\tau_t}{d\eta} = -\frac{\tau_t}{\eta} - \frac{2}{W_u\eta}\frac{1}{(1-p_t)^2}\frac{dp_t}{d\eta} \tag{18}$$

Substituting (16), (17) with $d\tau_t/d\eta$ replaced by (18), and (18) into (14) gives

$$\frac{dp_t}{d\eta}F = Z \tag{19}$$

where F and Z are given in Table V. From (19), showing $dp_t/d\eta < 0$ is equivalent to showing $Z < 0$ and $F > 0$.

Since $\lambda T_u \leq 1$ by hypothesis, $T_t \geq T_d$, and $4/W_u - \eta\tau_t^2 > 4/W_u - \eta\tau_t > 0$ from (15), we have

$$F > 0$$

From Z in Table V,

$$\begin{aligned} Z <& \frac{\lambda}{(1-\tau_t)^2}\frac{\tau_t}{\eta(4/W_u-\eta\tau_t)^2}\bigg(\Big(-8/W_u^2 - \eta^2\tau_t^2 \\ &+ (8/W_u^2)\eta + (2/W_u)\eta(3-4/W_u)\tau_t\Big)T_d \\ &- (8/W_u^2 - (2/W_u)\eta\tau_t^2)T_t\bigg) \end{aligned} \tag{20}$$

At $\eta = 1$, $T_t = T_d$ and so (20) becomes

$$\begin{aligned} Z <& \frac{\lambda}{(1-\tau_t)^2}\frac{\tau_t T_d}{(4/W_u-\tau_t)^2}(4/W_u - \tau_t)\Big(-(2/W_u) \\ &+ \tau_t(1-2/W_u)\Big) \end{aligned} \tag{21}$$

From (21), to show $Z < 0$, it is sufficient to show that the last term on the right-hand side of (21) is less than 0. Denote that term by $h(\tau_t)$; hence,

$$h(\tau_t) = -(2/W_u) - \tau_t(1-2/W_u)$$

It is clear that $h(\tau_t)$ is an increasing function of τ_t over $[0, \frac{2}{W_u}]$. Hence, $h(\tau_t) \leq h(2/W_u) = -(2/W_u)^2$. This implies $Z < 0$ at $\eta = 1$.

The fact that $F > 0$ at $\eta \geq 1$ and $Z < 0$ at $\eta = 1$ imply $dp_t/d\eta < 0$ at $\eta = 1$. Hence, $dC_t/d\eta > 0$ at $\eta = 1$, which confirms Claim (T2-1).

2) Proof of Claim (T2-2) of Theorem 2: By (3c), $p_u = \tau_t$, whence showing p_u decreases with the increase of η is equivalent to showing τ_t decreases with η. Thus, it is sufficient to show that $d\tau_t/d\eta < 0$ as follows.

Taking derivative of (15) with respect to η,

$$\frac{dp_t}{d\eta} = \frac{-2/W_u}{(4/W_u-\eta\tau_t)^2}\left(\tau_t + \eta\frac{d\tau_t}{d\eta}\right) \tag{22}$$

Substituting (16), (17) and (22) into (14) and then after some simple transformation steps, we have

$$\frac{d\tau_t}{d\eta} = H/Q \tag{23}$$

where H and Q are given in Table V. This suggests that proving $d\tau_t/d\eta < 0$ is equivalent to proving $H < 0$ and $Q > 0$.

Substituting L into H, both given in Table V, gives

$$\begin{aligned} H =& -\frac{(2/W_u)\tau_t}{(4/W_u-\eta\tau_t)^2}\bigg(1 + \lambda(\sigma - T_u) \\ &+ (T_d - T_{difs})\lambda\frac{4/W_u-\tau_t}{1-\tau_t}\bigg) \end{aligned} \tag{24}$$

Since $\lambda T_u \leq 1$ by hypothesis, $T_d > T_{difs}$, and $4/W_u - \tau_t > 4/W_u - \eta\tau_t > 0$, it is clear from (24) that $H < 0$.

Substituting K into Q, both given in Table V, gives

$$\begin{aligned} Q =& \frac{1}{(1-\tau_t)^2(4/W_u-\eta\tau_t)^2}\bigg(\frac{2\eta}{W_u}(1-\tau_t)^2 \\ &\cdot\Big(1+\lambda(\sigma - T_u)\Big) + T_d(4/W_u - \eta\tau_t)^2 \\ &+ (2/W_u)(T_t - T_d)(4/W_u - \eta\tau_t^2)\bigg) \end{aligned} \tag{25}$$

Since $\lambda T_u \leq 1$ by hypothesis, $T_t \geq T_d$, and $4/W_u - \eta\tau_t^2 > 4/W_u - \eta\tau_t > 0$, it is clear from (25) that $Q > 0$. ∎

C. Proof of Lemma 2

Proof: From (3a), we have

$$1 - p_k = \frac{1}{2 - \frac{W_k}{2/\tau_k - 1}} \tag{26}$$

Moreover, dividing $1 - p_t$ from (3c) by $1 - p_d$ from (3c), we have

$$\frac{1-p_t}{1-p_d} = \frac{1-\tau_d}{1-\tau_t} \tag{27}$$

Then, by substituting $1 - p_t$ and $1 - p_d$ from (26) into (27) and after some simple transformation steps, we have

$$\frac{1-\tau_d}{2 - \frac{W_d}{2/\tau_d - 1}} = \frac{1-\tau_t}{2 - \frac{W_t}{2/\tau_t - 1}} \tag{28}$$

Let $g(\tau_k, W_k)$ be a function defined by

$$g(\tau_k, W_k) = \frac{1-\tau_k}{2 - \frac{W_k}{2/\tau_k - 1}} = \frac{1}{2}\frac{1-\tau_k}{1 - \frac{W_k}{2}\frac{1}{2-\tau_k}\tau_k} \tag{29}$$

Then, (28) becomes

$$g(\tau_t, W_t) = g(\tau_d, W_d) \tag{30}$$

Since $W_k > 4$ by hypothesis, the coefficient $\frac{W_k}{2}\frac{1}{2-\tau_k}$ of τ_k in the denominator of (29) is greater than 1. Hence, it can be easily shown that $g(\tau_k, W_k)$ is increasing in τ_k. Besides, $g(\tau_k, W_k)$ is also increasing in W_k. Therefore, from (30), $W_t > W_d$ implies that $\tau_t < \tau_d$. ∎

D. Proof of Theorem 3

Proof: Here we use the wireless model (3)–(6) to analyze the game.

Under the action profile a_D, we have

$$N_t \geq 1 \qquad N_d \geq 1$$

and

$$S_1(a_D) = S_d \qquad S_2(a_D) = S_t$$

Then, proving $S_1(a_D) > S_2(a_D)$ is equivalent to proving $S_d > S_t$, which is shown as follows.

Dividing S_d from (6) by S_t from (6),

$$\frac{S_d}{S_t} = \frac{\tau_d(1-p_d)}{\tau_t(1-p_t)\eta} \tag{31}$$

Substituting (27) into (31), we have

$$\frac{S_d}{S_t} = \frac{\tau_d(1-\tau_t)}{\tau_t(1-\tau_d)\eta} \tag{32}$$

From (32), showing $S_d > S_t$ is equivalent to showing the numerator of (32) greater than the denominator of (32) as follows.

First, we have

$$\begin{aligned}
&\tau_d(1-\tau_t) - \tau_t(1-\tau_d)\eta = \tau_d - \tau_t\eta + (\eta-1)\tau_t\tau_d \\
&= \tau_t\tau_d\left(\frac{1}{\tau_t} - \frac{\eta}{\tau_d} + \eta - 1\right) \\
&= \tau_t\tau_d\left(\frac{\eta W_d}{2}\left(\frac{1-p_t}{1-2p_t} - \frac{1-p_d}{1-2p_d}\right) + \frac{\eta-1}{2}\right)
\end{aligned} \tag{33}$$

In order for (33) to be greater than 0, it is sufficient to show that $p_t > p_d$. This is proven below.

Recall that under the action space $A0$ and by hypothesis,

$$W_t = \eta W_d > W_d > 4$$

which satisfies the conditions of Lemma 2. Hence, we have

$$\tau_d > \tau_t \tag{34}$$

From (27) and (34), the following holds

$$p_t > p_d \tag{35}$$

■

E. Proof of Lemma 3

Proof: To see how the attempt probability of the user 1 changes when its action changes from B to R, we investigate how the solution of the fixed point model changes accordingly.

Therefore, we first find the solution of the fixed point model and then, we prove that there exists a unique solution. Based on this solution, we next investigate how the solution changes with the action choice of user 1. Those will be presented as follows.

Denote the set of action profiles in which all users $i \neq 1 \neq j$ have the same action $a_i = a_j$ by

$$A_G = \left\{a \in A | (a_1 = B \text{ or } R, a_{-1} = X)\right\}$$

where X is an action profile of all data users $i \neq 1 \neq j$ in which $a_i = a_j$ known.

Let a_G denote a particular action profile in A_G.

Under the action profile a_G, we have the following from (7)

$$\tau_1 = \frac{2(1-2p_1)}{W_1(1-p_1)} \tag{36a}$$

$$\tau_j = \frac{2(1-2p_j)}{W_j(1-p_j)}, \forall j \neq 1 \tag{36b}$$

where

$$W_1 = cW_j \tag{37}$$

in which c is different at different values of a_1. By hypothesis, $W_1 > 4$, $W_j > 4$. Hence, $cW_j > 4$ or $c > \frac{4}{W_j}$.

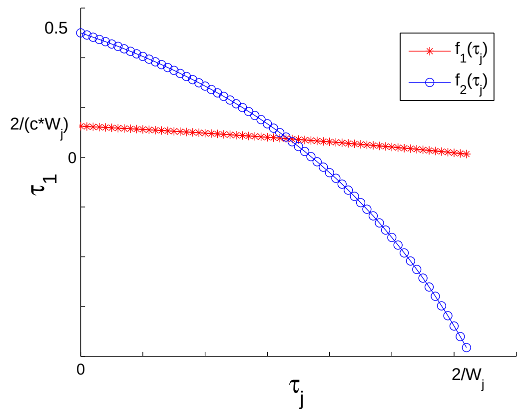

Fig. 12. $f_1(\tau_j)$ and $f_2(\tau_j)$

Moreover, since $N_u = 0$, from (3c), we have

$$p_1 = 1 - (1-\tau_j)^{N_s-1} \tag{38a}$$

$$p_j = 1 - (1-\tau_j)^{N_s-2}(1-\tau_1) \tag{38b}$$

Substituting (37) and (38a) into (36a) gives

$$\tau_1 = \frac{2}{cW_j}\left(2 - \frac{1}{(1-\tau_j)^{N_s-1}}\right) \equiv f_1(\tau_j) \tag{39}$$

From (36b),

$$p_j = 1 - \frac{2}{4 - W_j\tau_j} \tag{40}$$

Substituting (40) into (38b) and after some simple transformations, we have

$$\tau_1 = 1 - \frac{2}{(4-W_j\tau_j)(1-\tau_j)^{N_s-2}} \equiv f_2(\tau_j) \tag{41}$$

Then, the solution of the fixed point model is the positive solution to $f_1(\tau_j) = f_2(\tau_j)$. We now prove there exists positive solution to $f_1(\tau_j) = f_2(\tau_j)$ as follows.

It is clear that $f_1(\tau_j)$ and $f_2(\tau_j)$ are decreasing functions of τ_j on $[0, \frac{2}{W_j}]$. This is illustrated in Fig. 12.

Furthermore, since $W_j > 4$, at $\tau_j = 1 - 2^{-1/(N_s-1)} > 0$,

$$f_1(1 - 2^{-1/(N_s-1)}) = 0 \tag{42a}$$

$$f_2(1 - 2^{-1/(N_s-1)}) < 0 \tag{42b}$$

and at $\tau_j = 0$,

$$f_2(0) > f_1(0) > 0 \tag{43}$$

From (42), (43) and the continuity of $f_1(\tau_j)$ and $f_2(\tau_j)$, it can be concluded that there exists solution to $f_1(\tau_1) = f_2(\tau_2)$ and it is positive. Next we will prove that the solution is unique.

Define $f(\tau_j, c)$ as follows

$$f(\tau_j, c) = f_2(\tau_j) - f_1(\tau_j)$$

It can be easily shown that $f(\tau_j, c)$ is a decreasing function of τ_j. Together with the fact that $f_1(\tau_j)$ and $f_2(\tau_j)$ are decreasing functions, it can be concluded that the solution to $f_1(\tau_1) = f_2(\tau_2)$ is unique.

Now we know that the fixed point model has a unique solution. We will now investigate how the solution changes with the action choice of user 1 as follows. Note that when the user 1 changes its action, its CW_{min} (W_1) changes, causing the coefficient c in (37) changes.

Let τ_{j1} and τ_{j2} be the solution to $f(\tau_j, c) = 0$ at $c = c_1$ and $c = c_2 > c_1$, respectively.

It is clear that $f(\tau_j, c)$ is also an increasing function of c; hence, $f(\tau_{j1}, c_2) > f(\tau_{j1}, c_1) = f(\tau_{j2}, c_2) = 0$. This, together with the fact that $f(\tau_j, c)$ is a decreasing function of τ_j, implies that $\tau_{j1} < \tau_{j2}$. Therefore, when c increases or W_1 increases, τ_j increases and τ_1 decreases.

In particular, when a_1 changes from B to R, c decreases; hence, τ_j decreases and τ_1 increases. ■

F. Proof of Theorem 4

Proof: Let a_P and a_Q, respectively, be the action profiles defined as follows

$$a_P = (a_1 = B, a_{-1} = X)$$
$$a_Q = (a_1 = R, a_{-1} = X)$$

where X is an action profile of other data users except user 1 in which these data users have the same action known. Let j denote any player in $\mathcal{P}\backslash\{1\}$.

Based on (5), the successful transmission rate per slot of the data user in accordance with each action profiles a_P and a_Q, respectively, are given by

$$C_1(a_P) = \eta\tau_1(a_P)\big(1 - p_1(a_P)\big) \tag{44a}$$
$$C_1(a_Q) = \tau_1(a_Q)\big(1 - p_1(a_Q)\big) \tag{44b}$$

To show that $C_1(a_P) < C_1(a_Q)$, it's sufficient to show

$$\tau_1(a_Q) > \eta\tau_1(a_P) \tag{45a}$$
$$p_1(a_P) > p_1(a_Q) \tag{45b}$$

Those will be proven as follows.

First, note that the conditions of this theorem satisfy those of Lemma 3. Under the action space $A0$,

- when $a_{-1} =$ all B,
  - For the action profile a_P, $W_1(a_P) = W_j(a_P) = \eta W_u$; hence, c in (37) is equal to 1. Then, from (39), we have
$$\tau_1(a_P) = \frac{2}{\eta W_u}\left(2 - \frac{1}{\big(1-\tau_j(a_P)\big)^{N_s-1}}\right) \tag{46a}$$
  - For the action profile a_Q, $W_1(a_Q) = W_u$ and $W_j(a_Q) = \eta W_u$; hence, c in (37) is equal to $\frac{1}{\eta}$. Then, from (39), we have
$$\tau_1(a_Q) = \frac{2}{W_u}\left(2 - \frac{1}{\big(1-\tau_j(a_Q)\big)^{N_s-1}}\right) \tag{46b}$$
- when $a_{-1} =$ all R,
  - For the action profile a_P, $W_1(a_P) = \eta W_u$ and $W_j(a_P) = W_u$; hence, c in (37) is equal to η. Then, from (39), $\tau_1(a_P)$ is also given by (46a).
  - For the action profile a_Q, $W_1(a_Q) = W_j(a_Q) = W_u$; hence, c in (37) is equal to 1. Then, from (39), $\tau_1(a_Q)$ is also given by (46b).

From above, regardless of $a_{-1} =$ all R or all B, when a_1 changes from B in a_P to R in a_Q, c decreases. Then, from Lemma 3, for any player $j \neq 1$, we have

$$\tau_j(a_P) > \tau_j(a_Q) \tag{47}$$

Besides, $\tau_1(a_P)$ is given by (46a) and $\tau_1(a_Q)$ is given by (46b), regardless of $a_{-1} =$ all R or all B.

From (46) and (47), we obtain (45a) as follows

$$\begin{aligned}\tau_1(a_Q) &= \frac{2}{W_u}\left(2 - \frac{1}{(1-\tau_j(a_Q))^{N_s-1}}\right)\\ &= \eta\frac{2}{\eta W_u}\left(2 - \frac{1}{(1-\tau_j(a_Q))^{N_s-1}}\right)\\ &> \eta\frac{2}{\eta W_u}\left(2 - \frac{1}{(1-\tau_j(a_P))^{N_s-1}}\right) = \eta\tau_1(a_P)\end{aligned}$$

Moreover, based on (38a),

$$p_1(a_P) = 1 - \big(1-\tau_j(a_P)\big)^{N_s-1} \tag{48a}$$
$$p_1(a_Q) = 1 - \big(1-\tau_j(a_Q)\big)^{N_s-1} \tag{48b}$$

From (47) and (48), we have (45b). ■

Appendix B
Proofs of properties of the PIA

A. Proof of Theorem 5

Proof: First, note that Theorem 1 is a special case of Theorem 5 with $\epsilon = \theta = 0$. The proof of Theorem 5 is given as follows.

Consider two networks with $N_d = N_u = 0$, identical except that one has $\eta = \eta_1$ and the other has $\eta = \eta_2 > \eta_1$.

Let subscripts 1 and 2 refer to networks with $\eta = \eta_1$ and $\eta = \eta_2$, respectively; subscript i refer to a particular network ($i \in \{1, 2\}$).

Based on (4), $E[Y_i]$ is given by

$$\begin{aligned}E[Y_i] = T_d\Big(p_{ti} + (1-N_s)\tau_{ti}(1-p_{ti})\Big) +\\ + \sigma(1-\tau_{ti})(1-p_{ti}) + T_{ti}N_s\tau_{ti}(1-p_{ti})\end{aligned} \tag{49}$$

Substituting (49) into (6) and then after some simple transformation steps, we have

$$S_{ti} = \frac{1}{\sigma\Big(\frac{1-\tau_{ti}}{\eta_i\tau_{ti}}\Big) + \frac{T_d}{\eta_i}\Big(\frac{p_{ti}}{\tau_{ti}(1-p_{ti})} + (1-N_s)\Big) + T_{ti}\frac{N_s}{\eta_i}} \tag{50}$$

To show $S_{t1} < S_{t2}$, it's sufficient to show the denominator of S_{t1} from (50) is higher than that of S_{t2} from (50). This is equivalent to showing the following inequalities.

$$\sigma\Big(\frac{1-\tau_{t2}}{\eta_2\tau_{t2}}\Big) + T_{t2}\frac{N_s}{\eta_2} < \sigma\Big(\frac{1-\tau_{t1}}{\eta_1\tau_{t1}}\Big) + T_{t1}\frac{N_s}{\eta_1} \tag{51a}$$
$$\frac{p_{t2}}{\tau_{t2}(1-p_{t2})} < \frac{p_{t1}}{\tau_{t1}(1-p_{t1})} \tag{51b}$$

1) Proof of (51a): Because the conditions of Theorem 5 satisfy those of Lemma 1, we have the following from Lemma 1

$$\tau_{t1} > \tau_{t2} \qquad p_{t1} > p_{t2} \tag{52}$$

Note that $W_{ti} = \eta_i(W_u - \theta)$. Then, multiplying τ_{ti} from (3a) with η_i and after some simple transformations, we have

$$\tau_{ti}\eta_i = \frac{2}{(W_u - \theta)\frac{1-p_{ti}}{1-2p_{ti}} + \frac{1}{\eta_i}} \tag{53}$$

From (52) and (53), it's clear that $\tau_{ti}\eta_i$ increases with η_i. This is because when η_i increases, p_{ti} decreases by (52), which leads to the decrease of $\frac{1-p_{ti}}{1-2p_{ti}}$. Moreover, $1/\eta_i$ also decreases with the increase of η_i. Hence, we have

$$\eta_1\tau_{t1} < \eta_2\tau_{t2} \tag{54}$$

We have (51a) because

$$\sigma\Big(\frac{1-\tau_{t1}}{\eta_1\tau_{t1}}\Big) + T_{t1}\frac{N_s}{\eta_1} - \Big(\sigma\Big(\frac{1-\tau_{t2}}{\eta_2\tau_{t2}}\Big) + T_{t2}\frac{N_s}{\eta_2}\Big)$$
$$= \sigma\Big(\frac{1}{\eta_1\tau_{t1}} - \frac{1}{\eta_2\tau_{t2}}\Big) + \Big(\frac{1}{\eta_1} - \frac{1}{\eta_2}\Big)\sigma(2N_s - 1) > 0$$

The last inequality comes directly from (2) with $T_{difs} = T_{sifs} + 2\sigma$ [1] and (54).

2) Proof of (51b): Let $g(\tau_t)$ be a function defined as follows

$$g(\tau_t) = \frac{p_t}{\tau_t(1-p_t)} \tag{55}$$

From (52), (51b) holds if $g(\tau_t)$ is an increasing function of τ_t, which is proven below.

Substituting (11b) into (55) gives

$$g(\tau_t) = \frac{1-(1-\tau_t)^{N_s-1}}{\tau_t(1-\tau_t)^{N_s-1}} = \sum_{i=0}^{N_s-2}(1-\tau_t)^{i-(N_s-1)}$$

Then, it is clear that $g(\tau_t)$ is a decreasing function of $(1-\tau_t) > 0$, or an increasing function of τ_t. ∎

B. Proof of Theorem 6

Proof: Here we use the wireless model (3)–(6) to analyze the game.

Under the action profile a_D, we have

$$N_t \geq 1 \qquad N_d \geq 1$$

and

$$S_1(a_D) = S_d \qquad S_2(a_D) = S_t$$

Then, $S_2(a_D) > S_1(a_D)$ is equivalent to $S_t > S_d$. Therefore, finding ϵ so that $S_2(a_D) > S_1(a_D)$ is equivalent to finding ϵ so that $S_t > S_d$, which is presented as follows.

Dividing S_d from (6) by S_t from (6),

$$\frac{S_d}{S_t} = \frac{\tau_d(1-p_d)}{\tau_t(1-p_t)\eta} \tag{56}$$

Under the action space $A1$ and by hypothesis,

$$W_t = \eta W_u - \epsilon > W_d = W_u > 4$$

which satisfies conditions of Lemma 2. Hence, based on Lemma 2 and (3c), we have

$$p_t > p_d \tag{57}$$

Dividing τ_t from (3a) by τ_d from (3a), we have

$$\frac{\tau_t}{\tau_d} = \frac{\dfrac{2}{W_t\frac{1-p_t}{1-2p_t}+1}}{\dfrac{2}{W_d\frac{1-p_d}{1-2p_d}+1}} \tag{58}$$

Substituting (57) into (58) gives

$$\frac{\tau_t}{\tau_d} > \frac{\dfrac{2}{W_t\frac{1-p_t}{1-2p_t}}}{\dfrac{2}{W_d\frac{1-p_d}{1-2p_d}}} \tag{59}$$

Substituting (59) into (56) gives

$$\frac{S_t}{S_d} > \frac{\eta W_d(1-2p_t)}{W_t(1-2p_d)} \tag{60}$$

From (27) and by substituting (57) into (58), we have

$$\frac{1-p_d}{1-p_t} = \frac{1-\tau_t}{1-\tau_d} < \frac{1-\dfrac{2}{W_t\frac{1-p_t}{1-2p_t}}}{1-\dfrac{2}{W_d\frac{1-p_d}{1-2p_d}}} \tag{61}$$

which is equivalent to the following, for $W_t > 4$ by hypothesis

$$p_t < \frac{\Big(W_d - 2 + (4-W_d)p_d\Big)W_t/W_d - W_t + 2}{4-W_t} \tag{62}$$

whence

$$\eta W_d(1-2p_t) > \frac{\eta W_d}{4-W_t}\Big(-W_t + 4W_t/W_d - 2(4-W_d)p_dW_t/W_d\Big) \tag{63}$$

From (60) and (63), when

$$W_t(1-2p_d) \leq \frac{\eta W_d}{4-W_t}\Big(-W_t + 4W_t/W_d - 2(4-W_d)p_dW_t/W_d\Big), \tag{64}$$

we have $S_t > S_d$.

After some simple transformations, (64) is equivalent to

$$\eta(4-W_d) - (4-W_t) \leq 0$$
$$\Leftrightarrow \eta(4-W_d) - \big(4-(\eta W_d - \epsilon)\big) \leq 0$$
$$\Leftrightarrow \epsilon \geq 4(\eta - 1) \tag{65}$$

∎

C. Proof of Theorem 7

Proof: Let a_P and a_Q, respectively, be the action profiles defined as follows

$$a_P = (a_1 = B, a_{-1} = X)$$
$$a_Q = (a_1 = R, a_{-1} = X)$$

where X is an action profile of other data users except user 1 in which these data users have the same action known. Let j denote any player in $\mathcal{P}\backslash\{1\}$.

First, note that the conditions of this theorem satisfy those of Lemma 3. Then, under the action space $A1$,

- when $a_{-1} =$ all B,
  - For the action profile a_P, $W_1(a_P) = W_j(a_P) = \eta W_u - 4(\eta - 1)$; hence, c in (37) is equal to 1.
  - For the action profile a_Q, $W_1(a_Q) = W_u$ and $W_j(a_Q) = \eta W_u - 4(\eta - 1)$; hence, c in (37) is equal to $\frac{W_u}{\eta W_u - 4(\eta-1)}$ which is less than 1.
- when $a_{-1} =$ all R,
  - For the action profile a_P, $W_1(a_P) = \eta W_u - 4(\eta - 1)$ and $W_j(a_P) = W_u$; hence, c in (37) is equal to $\eta - 4(\eta - 1)/W_u$ which is greater than 1.
  - For the action profile a_Q, $W_1(a_Q) = W_j(a_Q) = W_u$; hence, c in (37) is equal to 1.

From above, regardless of $a_{-1} =$ all R or all B, when a_1 changes from B in a_P to R in a_Q, c decreases. From Lemma 3, for any player $j \neq 1$, we have

$$\tau_j(a_P) > \tau_j(a_Q) \tag{66}$$

Because p_j from (40) is a decreasing function of τ_j, from (66), we have

$$p_j(a_P) < p_j(a_Q) \tag{67}$$

From (5), the successful transmission rate per slot of the data user j in accordance with each action profiles a_P and a_Q, respectively, are given by

$$C_j(a_P) = \eta\tau_j(a_P)\big(1 - p_j(a_P)\big) \tag{68a}$$
$$C_j(a_Q) = \eta\tau_j(a_Q)\big(1 - p_j(a_Q)\big) \tag{68b}$$

From (66), (67), and (68), it can be seen that

$$C_j(a_P) > C_j(a_Q) \tag{69}$$

Besides, under the action space $A1$, by applying the result in Theorem 6 with $a_D = a_Q$, we have

$$C_1(a_Q) < C_j(a_Q) \tag{70}$$

From (69) and (70), it is concluded that

$$C_1(a_P) = C_j(a_P) > C_1(a_Q) \tag{71}$$

■

ACKNOWLEDGMENT

This work was supported by Australian Research Council grants DP1095103 and FT0991594.

REFERENCES

[1] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11^{TM}, 2007.

[2] M. Cagalj, S. Ganeriwal, I. Aad, and J.P. Hubaux, "On selfish behavior in CSMA/CA networks," in *Proc. IEEE INFOCOM*, 2005.

[3] L. Galluccio, "A Game-Theoretic Approach to Prioritized Transmission in Wireless CSMA/CA Networks," in IEEE Vehicular Technology Conference, April 2009.

[4] N. H. Vaidya, A. Dugar, S. Gupta and P. Bahl, "Distributed fair scheduling in a wireless LAN," *IEEE Trans. Mobile Computing*, vol. 4, no. 6, pp. 616–629, Nov/Dec 2005.

[5] P. Nuggehalli, M. Sarkar, K. Kulkarni, and R. R. Rao, "Game-theoretic Analysis of QoS in Wireless MAC," in *Proc. INFOCOM* 2008.

[6] J. Price, P. Nuggehalli, and T. Javidi, "Incentive Compatible MAC-Layer QoS Design," In *Proc. CCNC* 2008.

[7] C.-T. Chou, S. N. Shankar and K. G. Shin, "Achieving per-stream QoS with distributed airtime allocation and admission control in IEEE 802.11e wireless LANs" in *Proc. INFOCOM* pp. 1584–1595, 2005.

[8] M. H. Cheung, A. H. Mohsenian-Rad, V. W. S. Wong, and R. Schober, "Random Access Protocols for WLANs based on mechanism design," in *Proc. IEEE Int. Conf. Commun.*, pp. 4892-4897, 2009.

[9] J. Price, P. Nuggehalli, and T. Javidi, "Pricing and QoS in Wireless Random Access Networks," in *Proc. GLOBECOM* 2008.

[10] P. Hurley and J.Y. Le Boudec, "ABE: Providing a low-delay service within best effort," *IEEE Network*, vol. 15, no. 3, pp. 60–69, 2001.

[11] M. Karsten, Y. Lin, and K. Larson, "Incentive-Compatible Differentiated Scheduling," in *Proc. HotNets IV*, 2005.

[12] B. Gaidioz and P. Primet, "EDS: A new scalable service differentiation architecture for internet," In *Proc. IEEE ISCC*, pp. 777–782, July 2002.

[13] V. Firoiu, X. Zhang, and Y. Guo, "Best Effort Differentiated Services: Trade-off Service Differentiation for Elastic Applications," In *Proc. IEEE ICT*, June 2001.

[14] S. H. Nguyen, H. L. Vu and L. L. H. Andrew, "Performance analysis of 802.11e with saturated and non-saturated sources, " (Online). Available ⟨http://caia.swin.edu.au/reports/091013A/CAIA-TR-091013A.pdf⟩.

[15] X. Ling, L. X. Cai, J. W. Mark, and X. Shen, "Performance Analysis of IEEE 802.11 DCF with heterogeneous traffic," in *Proc. IEEE Consumer Commun. Netw. Conf. (CCNC 2007)*, art. no. 4199105, pp. 49-53, 2007.

[16] B. Bellalta, C. Cano, M. Oliver, and M. Meo, "Modeling the IEEE 802.11e EDCA for MAC parameter optimization," in *Proc. IEEE Consumer Commun. Netw. Conf. (CCNC)*, vol. 1, pp. 390–394, 2006.

[17] P. K. Dutta, *Strategies and games: theory and practice*. The MIT Press, 1999.

[18] D. Fudenberg and J. Tirole, *Game Theory*. The MIT Press, 1991.

[19] "The network simulator ns-2," Available at ⟨http://www.isi.edu/nsnam/ns/⟩.

[20] S. Wietholter and C. Hoene, "An IEEE 802.11e EDCF and CFB simulation model for ns-2," Available at ⟨http://www.tkn.tu-berlin.de/research/802.11e_ns2/⟩.

[21] S. H. Nguyen, H. L. Vu and L. L. H. Andrew, "Service differentiation without prioritization in 802.11 WLANs," Available at ⟨http://caia.swin.edu.au/cv/snguyen/J2.pdf⟩.