Performance analysis of 802.11e EDCA WLANs with saturated and non-saturated sources

Suong H. Nguyen, Hai L. Vu, Lachlan L. H. Andrew
Centre for Advanced Internet Architectures, Technical Report 091013A
Swinburne University of Technology
Melbourne, Australia
hsnguyen@swin.edu.au, hvu@swin.edu.au, landrew@swin.edu.au

***Abstract*—In WLANs, realtime traffic has always been given higher priority than data traffic. This creates an incentive for data users to pretend to be real-time users, which benefits themselves and degrades the performance of the rest. Our report proposes a tradeoff mechanism which benefits both low-rate realtime traffic and greedy data traffic. To understand the interaction between these two types of traffic and estimate the optimal value of MAC parameters with the proposed tradeoff mechanism, we develop a simple model of an 802.11e EDCA WLAN with these two types of traffic based on mean-field approximation as in previous works. However, when a network carries some large packets and many small packets, the collision probability after a large packet is much larger than predicted by previous models and our simple model. Therefore, we propose another model which captures this effect.**

I. Introduction

In recent years, WLANs have become very popular and are widely deployed, due to the rapid increase in demand for Internet access at any time and any place through WiFi-enabled mobile devices such as laptops and personal digital assistants (PDAs).

Internet applications over WLANs consist of not only conventional applications such as email, file transfer or web surfing but also delay-sensitive ones such as voice and video. To provide different quality of service (QoS) for different types of traffic, an extension of IEEE 802.11 known as IEEE 802.11e [1] was built, which defines a contention-based medium access control (MAC) scheme called the Enhanced Distributed Channel Access (EDCA). EDCA provides service differentiation by giving different access priorities to different types of traffic, which also focuses on prioritizing real-time traffic at the expense of data traffic as traditionally. This creates an incentive for data users to pretend to be real-time users, which benefits themselves and degrades the performance of the rest.

In this report, we propose a trade-off mechanism which provides benefits for both data and low-rate real-time traffic by using two MAC parameters CW_{min} and *TXOP limit* of 802.11e flexibly. CW_{min} controls how long a station needs to wait before sending a packet and *TXOP limit* determines how many packets a station can send without contending again once it gains channel access. Another MAC parameter of service differentiation defined in 802.11e, Arbitration Inter-frame Space ($AIFS$) is not used in our mechanism because it provides priority aggressively. To understand the interaction between these two types of traffic and estimate the optimal value of MAC parameters with the proposed tradeoff mechanism, a model of WLANs with heterogeneous traffic, and *TXOP limit* and CW_{min} service differentiation is needed.

So far, there have been several works dealing with heterogeneous non-saturated traffic in one-class IEEE 802.11 DCF WLANs. Among those, [5] and [6] have proposed analytical models using Markov chain while [7] has proposed one using mean-value analysis method. However, these models do not include *TXOP limit* and CW_{min} service differentiation. In contrast, almost all previous works in IEEE 802.11e EDCA have modeled heterogeneous traffic with CW_{min} and $AIFS$ service differentiation, a few of which [8], [9] include *TXOP limit* differentiation. Nevertheless, most of them have dealt with either saturated or non-saturated traffic. Only a few such as [10] and [11] modeled an 802.11e EDCA network with both saturated and non-saturated traffic, which are based on Markov chain. However, these two models have not included *TXOP limit* differentiation. Furthermore, most of previous works ignore the residual service time in calculating MAC service time of a packet from non-saturated stations and assume that a packet from a non-saturated station always executes backoff process before the first transmission attempt, which is not consistent with medium access mechanism of 802.11 DCF and 802.11e EDCA.

In summary, previous models are either too complicated or do not capture some features needed in our scenario. Therefore, in this report, we develop a model for an 802.11e EDCA WLAN with low-rate realtime traffic and greedy data traffic based on renewal reward theorem [12] (hereafter called "basic model"). As the majority of existing analytical models to evaluate the performance of MAC protocol in WLANs, this model is also based on a fundamental assumption introduced in a seminal paper of Bianchi [2] which stated that, at each transmission attempt, and regardless of the number of retransmissions suffered, each packet of a source collides with constant and independent probability (hereafter called "mean-field approximation").

WLANs support a wide range of applications with a variety of packet sizes and this variability is set to increase in 802.11e WLANs which effectively allow very large packets controlled by *TXOP limit* parameter. This diversity leads to a new phenomenon: When a network carries some large packets and many small packets, the collision probability after a large packet is much larger than predicted by previous models and our basic model. The inaccuracy stems from the fact that packets may experience different collision probabilities at different times, i.e., the collision probability is not homogeneous across time slots in the system due to the existence of big packets. This makes the mean-field approximation inappropriate in estimating the collision probability of sources sending small packets. The collision probability is of importance because the energy consumption of the battery powered mobile devices depends on the number of packet transmissions, which is directly related to the collision probability. Therefore, we propose another analytical model that captures this effect (hereafter called "big-packet model").

The remainder of the report is organized as follows. Section II briefly introduces the main characteristics of the IEEE 802.11e EDCA. The proposed tradeoff mechanism is described in Section III followed by Section IV which investigates the impact of big packets on sources sending small packets. Section V presents our analytical model of an 802.11e EDCA WLAN with low-rate realtime traffic and greedy data traffic. Section VI evaluates the model and the tradeoff mechanism. Finally, Section VII concludes the report.

II. IEEE 802.11e Enhanced Distributed Coordination Function

Before introducing our model, let us recall the basic operation of EDCA. EDCA is designed to provide service differentiation to the original 802.11 DCF [1]. A thorough presentation of the main features of 802.11e is provided in [3], [4].

In each EDCA station, frames from different classes are mapped to one of four Access Categories (ACs). Each AC in a station behaves like a virtual station which has an independent queue and implements independent DCF. These ACs have different QoS by using different values of four MAC service differentiation parameters: $AIFS$, CW_{min}, CW_{max} and *TXOP limit*.

When an AC i has a new frame to send, it first senses the channel. If the channel is detected idle for a period equal to its $AIFS$ parameter ($AIFS_i$), it transmits the frame. Otherwise, if the channel is sensed busy within the $AIFS_i$ period, the AC i continues sensing the channel until it is detected idle for an $AIFS_i$. Then, the AC i starts the backoff process by initializing its backoff counter to a random number uniformly distributed between 0 and $(CW_i - 1)$ where CW_i is the contention window of the AC i which is initially set to CW^i_{min} and doubles after each unsuccessful transmission until it reaches CW^i_{max}. CW_i is reset to CW^i_{min} after each successful transmission or after the number of unsuccessful transmission of a frame reach a retry limit. The backoff counter is decreased by one at every idle slot time (each idle slot time (hereafter denoted as T_{slot}) is a constant defined by physical layer), frozen during channel activity period and resumed one slot time before the expiration of an $AIFS_i$ time which is after a channel activity period ends. When the backoff counter reaches zero, the AC i can gain channel access. However, if two or more ACs in a station have their backoff counters to reach zero simultaneously, a virtual collision occurs, which is solved by granting the channel access to the highest-priority AC while the others must invoke the backoff procedure with doubled CW_i and the same retransmission counter. After gaining channel access, the AC i is permitted to transmit several consecutive frames provided that it does not occupy the channel for a period of time longer than its *TXOP limit*.

To notify the transmitting AC that the transmitted frame is successfully received, an acknowledgment (ACK) is sent back from the receiver after a period of time equal to Short Inter-frame Space (SIFS) since it finishes receiving the frame. If an ACK is not received within a specified ACK timeout, the AC assumes that the transmitted frame is unsuccessfully received. Then, it either schedules a retransmission by starting a backoff process with doubled CW_i or CW^i_{max} (whichever is smaller) or drops the frame if the retransmission counter exceeds the retry limit. After an $AIFS_i$ time since receiving the ACK packet successfully, the transmitting AC will invoke backoff process with the contention window of CW^i_{min} which is called post-backoff process.

III. Proposed tradeoff mechanism

Unlike the traditional approach which gives realtime traffic higher priority than data traffic, we propose a tradeoff mechanism which provides benefit for both low-rate realtime and data traffic.

While low-rate realtime traffic requires small delay and usually has one packet to send at a time, data traffic cares about high throughput. Therefore, in this mechanism, low-rate realtime traffic always chooses the minimum *TXOP limit* and CW_{min} while data traffic can choose higher *TXOP limit* but then it must increase CW_{min} in proportion.

By using higher CW_{min}, data traffic attempts to transmit less often, which leads to the decrease of collision among data sources, and collision between data sources and realtime sources. As a result, by increasing *TXOP limit* in proportion with CW_{min}, the throughput of data traffic can be improved. Besides, the access delay of realtime traffic may also be improved. The efficiency of this tradeoff mechanism will be verified in Section VI.

IV. Impact of big packets

By allowing data traffic to use large *TXOP limit*, the above tradeoff mechanism can create large packets, the impact of which is investigated in this section.

A. Description of the impact

Consider a WLAN with a large number N_u of unsaturated sources sending small packets, each with rate λ, and one source sending big packets of size L_b and transmission duration T. In this scenario, it is possible for sufficient small packets to accumulate during the transmission of a large packet, that the collision probability of small packets is significantly under-estimated by the mean field approximation.

While a large packet is being sent, on average $N_u\lambda T$ new small packets will arrive to the system. These will all attempt to transmit within the short persistence time, which in 802.11 is uniformly distributed up to 32 slots. As a result, the longer the big packet is, the more small packets will attempt to transmit soon afterwards, and the higher the collision probability during that period.

Due to the effect of large packets, there exist high-contention and low-contention periods, which makes the contention level of slots not homogeneous. However, the mean-field approximation used in previous models assumes that the contention level is the same for every slot, which does not take into account the effect of large packets.

In systems such as 802.11, in which backoff intervals are measured in slots rather than absolute time, this effect primarily affects the first transmission attempt. On retransmission attempts, the sources are synchronized to the slot times, and are no more likely to transmit after a large (busy) slot than an idle slot. As a result, the collision probability of the first attempt is significantly larger than that of retransmission attempts.

B. When does this effect occur?

The effect described above is largest when the following conditions occur:

- $N_u\lambda T$ is large (at least comparable to 1), which implies that
- the ratio of big packets' size and small packets' size is reasonably large;
- The time interval between big packets is of the same order as the time to clear the backlog of unsaturated sources caused by a busy period;
- Stations sending small packets are very unsaturated, so that minimal queue builds up even during the big packet transmissions;
- The number of unsaturated stations is large.

Moreover, the impact is clearer when the arrival process of small packets at a source is quasi-periodic than when it is Poisson, because this maximizes the number of unsynchronized arrivals.

V. The analytical model

We now derive a model to evaluate the mechanism proposed in Section III, which captures the effect described in Section IV.

Consider a wireless network with N_u unsaturated or non-greedy nodes (e.g. voices) with initial contention windows W_u, and N_t saturated nodes (e.g. TCP) with initial contention windows W_t. Packets sent by unsaturated nodes occupy the channel for time T_u, and those sent by the saturated nodes take time T_{ts} if they are successfully transmitted, or T_{tc} if they collide. These inputs can be related to physical 802.11 parameters by

(1a) – (1d):

$$T_i = \delta \quad (1a)$$
$$T_u = T_{difs} + T_{nonsat} + T_{sifs} + T_{ack} \quad (1b)$$
$$T_{ts} = T_{difs} + \eta(T_{sat} + T_{ack}) + (2\eta - 1)T_{sifs} \quad (1c)$$
$$T_{tc} = T_{difs} + T_{sat} + T_{sifs} + T_{ack} \quad (1d)$$

where δ is the length of an idle backoff slot; T_{difs}, T_{sifs} and T_{ack} are the duration of DIFS, SIFS and transmission time of an ACK packet, respectively; T_{sat} and T_{nonsat} are the transmission time of a packet from saturated and unsaturated sources, respectively; and η is the number of packets which can be sent by the saturated station during its *TXOP limit* once having access to the channel.

In our model, we make the following assumptions: (A1) Each saturated station always has packets available for transmission; (A2) Non-saturated stations have the same average packet arrival rate λ which is small enough so that their queue rarely builds up; (A3) Channel conditions are ideal (no channel errors, hidden terminals or capture effect); (A4) Stations perform binary exponential backoff until they successfully transmit a packet (no retransmission limit and no maximum contention window limit); (A5) EDCA operates in basic mode (no RTS/CTS) and all stations use the same AIFS which is equal to DIFS.

Let τ_t and τ_u be the probability that saturated stations and non-saturated stations, respectively, attempt to transmit in a given slot.

Let p_t and p_u be the collision probabilities experienced by packets from saturated and non-saturated stations, respectively.

Furthermore, p_{u1} and p_{u2} are the collision probability of a packet from non-saturated stations on its first attempt and retransmission attempts, respectively.

The inputs to our model are N_u, N_t, W_t, W_u, η, λ, T_{sat} and T_{nonsat}.

A. Fixed point model

The model evolves a set of fixed-point equations where the collision probability of a packet from saturated stations p_t and collision probability of a packet from non-saturated stations on its first attempt p_{u1} and retransmission attempts p_{u2} are expressed in terms of the attempt probability of saturated stations τ_t and attempt probability of non-saturated stations τ_u, with an opposing set of equations for the attempt probabilities expressed in terms of the collision probabilities.

$$\tau_t = \frac{\text{E[Number of attempts per burst]}}{\text{E[Number of slots per burst]}} = \frac{\sum_{i=0}^{\infty} p_t^i}{\sum_{i=0}^{\infty}(E[U_{ti}]+1)p_t^i} \approx \frac{2(1-2p_t)}{W_t(1-p_t)} \quad (2a)$$

$$\tau_u = \frac{\text{Number of attempts per non-saturated source}}{\text{Number of slots}} = \frac{\lambda\ g(p_{u1}, p_{u2})\ \eta}{S_t(\tau_t,\tau_u)\left(\sum_{i=0}^{\infty}(E[U_{ti}]+1)p_t^i\right)} \approx \frac{\lambda\ g(p_{u1}, p_{u2})\ \eta}{S_t(\tau_t,\tau_u)\left(\frac{W_t}{2(1-2p_t)}\right)} \quad (2b)$$

$$p_t = 1-(1-\tau_t)^{N_t-1}(1-\tau_u)^{N_u} \quad (2c)$$
$$p_{u1} = h(\tau_t,\tau_u) \quad (2d)$$
$$p_{u2} = 1-(1-\tau_t)^{N_t}(1-\tau_u)^{N_u-1} \quad (2e)$$

where U_{ti} is a R.V. representing the number of backoff slots in the i-th backoff stage of a packet from saturated sources which is uniformly distributed over $[0, 2^iW_t - 1]$ and $E[U_{ti}]$ denotes the mean of U_{ti} given by

$$E[U_{ti}] = \frac{2^iW_t - 1}{2} \approx 2^{i-1}W_t \quad (3)$$

and $g(p_{u1}, p_{u2})$ is the average number of attempts per packet from non-saturated stations; the expressions of $g(p_{u1}, p_{u2})$ and $h(\tau_t,\tau_u)$ depend on whether we treat p_{u1} and p_{u2} the same (see Section V-A1) or not (see Section V-A2); $S_t(\tau_t,\tau_u)$ is the throughput of each saturated station in packets/s, which analogously to [2] is given by:

$$S_t(\tau_t,\tau_u) = \frac{E[\text{payload successfully transmitted per slot}]}{E[\text{slot length}]} = \frac{a_s\ \eta}{E[Y]} \quad (4a)$$

where Y is a R.V. representing the duration of a backoff slot, the mean of which is given as following

$$E[Y] = a_iT_i + a_uT_u + a_{tc}T_{tc} + a_{ts}T_{ts} \quad (4b)$$
$$a_s = \tau_t(1-\tau_t)^{N_t-1}(1-\tau_u)^{N_u} \quad (4c)$$
$$a_i = (1-\tau_t)^{N_t}(1-\tau_u)^{N_u} \quad (4d)$$
$$a_u = \left(1-(1-\tau_u)^{N_u}\right)(1-\tau_t)^{N_t} \quad (4e)$$
$$a_{ts} = N_t\tau_t(1-\tau_t)^{N_t-1}(1-\tau_u)^{N_u} \quad (4f)$$
$$a_{tc} = 1-(a_i + a_u + a_{ts}) \quad (4g)$$

Note that a_s is the probability that the tagged saturated station successfully transmits a burst in a given slot, a_i

is the probability that no stations transmit in a given slot, a_u is the probability that only non-saturated stations transmit in a given slot, a_{ts} is the probability that a saturated station successfully transmits a burst in a given slot and a_{tc} is the probability that there is collision involving at least one saturated station in a given slot.

Then, p_t, p_{u1}, p_{u2}, τ_t, τ_u and S_t can be determined by iteratively solving equations (2a) – (2e) using numerical techniques.

1) Traditional approach: Previous works have assumed that the collision probability of a packet from non-saturated stations is the same for every attempt. Using this assumption in (2), we have

$$h(\tau_t, \tau_u) = p_{u1} = p_{u2} = p_u \tag{5a}$$

where p_{u2} is determined by (2e), and $g(p_{u1}, p_{u2})$ is given by

$$g(p_{u1}, p_{u2}) = \frac{1}{1 - p_u} \tag{5b}$$

2) New approach: Unlike the traditional approach, we determine the collision probability of a packet from non-saturated stations on its first attempt and retransmission attempts separately. Then, we have

$$g(p_{u1}, p_{u2}) = 1 + \frac{p_{u1}}{1 - p_{u2}} \tag{6a}$$

and

$$h(\tau_t, \tau_u) = p_b \left(1 - (1-\tau_t)^{N_t}(1 - \frac{1}{W_u})^{N_{u1}}(1-\tau_{u2})^{N_{u2}}\right) \tag{6b}$$

where p_b is the probability that an arriving packet at non-saturated station finds the channel busy; N_{u1} is the average number of new packets from other non-saturated stations that come during W_u slots just before the transmission of the tagged packet and sense channel busy, which will contend with the tagged packet on its first attempt; N_{u2} is the average number of packets on their retransmission attempts from other non-saturated stations which will contend with the tagged packet on its first attempt; τ_{u2} is the probability that a non-saturated station attempts to retransmit in a given slot. Those are given as follows:

$$
\begin{aligned}
p_b &= 1 - \frac{a_i' T_i}{E[Y_u]} && \text{(6c)}\\
N_{u1} &= (N_u - 1)\lambda(2E[T_{res}] + p_b(W_u - 1)E[Y_u]) && \text{(6d)}\\
N_{u2} &= N_u - N_{u1} - 1 && \text{(6e)}\\
\tau_{u2} &= \frac{\text{Number of retransmission attempts per source}}{\text{Number of slots}}\\
&= \left(\frac{p_{u1}}{1 + p_{u1} - p_{u2}}\right)\tau_u && \text{(6f)}
\end{aligned}
$$

where $E[T_{res}]$ is the average residual service time of packets from other stations observed by an arriving packet at a non-saturated station and will later be given by (19) in Section V-B; Y_u is a R.V. representing the duration of a backoff slot experienced by a packet from a non-saturated station, the mean of which is calculated similar to (4b) where N_u is replaced with $N_u - 1$ in the expression of a_i, a_u, a_{ts} giving $a_i', a_u', a_{ts}', a_{tc}'$ as follows

$$
\begin{aligned}
E[Y_u] &= a_i'T_i + a_u'T_u + a_{tc}'T_{tc} + a_{ts}'T_{ts} && \text{(7a)}\\
a_i' &= (1-\tau_t)^{N_t}(1-\tau_u)^{N_u - 1} && \text{(7b)}\\
a_u' &= \left(1 - (1-\tau_u)^{N_u - 1}\right)(1-\tau_t)^{N_t} && \text{(7c)}\\
a_{ts}' &= N_t\tau_t(1-\tau_t)^{N_t - 1}(1-\tau_u)^{N_u - 1} && \text{(7d)}\\
a_{tc}' &= 1 - (a_i' + a_u' + a_{ts}') && \text{(7e)}
\end{aligned}
$$

Then, p_u can be determined by taking the weighted sum of p_{u1} and p_{u2} as following

$$p_u = \frac{1}{g(p_{u1}, p_{u2})}p_{u1} + \left(1 - \frac{1}{g(p_{u1}, p_{u2})}\right)p_{u2} \tag{8}$$

B. The delay model

The analysis of access delay in our model is based on [13].

1) Expression of access delay: Let D_u denote the access delay of each packet from non-saturated stations. We have

$$D_u = T_{difs} + A_u + T_{nonsat} \tag{9}$$

where A_u represents for the total backoff and collision time of a packet before it is successfully transmitted, which is given by

$$A_u = \begin{cases} A_{u0} & \text{w.p} \quad 1 - p_{u1} \\ A_{ui} & \text{w.p} \quad p_{u1}p_{u2}^{i-1}(1 - p_{u2})\ , \ i \geq 1 \end{cases} \tag{10}$$

where A_{ui} is the total backoff and collision time of a packet from non-saturated stations if it is successfully transmitted in i-th backoff stage, which is expressed as following

$$A_{ui} = \sum_{j=0}^{i} B_{uj} + \sum_{j=1}^{i} C_{uj} + X \tag{11}$$

where X is a R.V. given by

$$X = \begin{cases} -B_{u0} & \text{w.p} \quad 1 - p_b \\ T_{res} & \text{w.p} \quad p_b \end{cases} \tag{12}$$

and B_{uj} is the backoff time for the j-th backoff stage; C_{uj} is the duration of a collision involving the tagged packet. Those are given as follows

$$B_{uj} = \sum_{k=1}^{U_{uj}} Y_u \tag{13}$$

$$C_{uj} = \begin{cases} T_u & \text{w.p} \quad a'_u/(1-a'_i) \\ T_{tc} & \text{w.p} \quad 1-a'_u/(1-a'_i). \end{cases} \quad (14)$$

where U_{uj} is a R.V. representing the number of backoff slots in the j-th backoff stage of a packet from non-saturated sources, which is uniformly distributed over $[0, 2^j W_u - 1]$.

2) Expression of mean access delay: From (9), mean access delay can be expressed as

$$E[D_u] = T_{difs} + E[A_u] + T_{nonsat} \quad (15)$$

where

$$\begin{aligned} E[A_u] \quad &\approx \quad \left(\frac{1-2p_{u2}+2p_{u1}}{2(1-2p_{u2})}\right) W_u E[Y_u] + \\ &+ \frac{p_{u1}}{1-p_{u2}} E[C_{uj}] + E[X] \end{aligned} \quad (16)$$

where $E[C_{uj}]$ can be calculated from (14) and $E[X]$ is given from (12) as following

$$\begin{aligned} E[X] &= -E[B_{u0}](1-p_b) + E[T_{res}]p_b \\ &\approx -(W_u/2)E[Y_u](1-p_b) + E[T_{res}]p_b \end{aligned} \quad (17)$$

where $E[B_{uj}]$ is given from (13) as follows

$$\begin{aligned} E[B_{uj}] &= E[U_{uj}]E[Y_u] \\ &= \frac{2^j W_u - 1}{2} E[Y_u] \approx 2^{j-1} W_u E[Y_u] \end{aligned} \quad (18)$$

Using approach in [14], $E[T_{res}]$ is given by

$$E[T_{res}] = \frac{E[Y'_u]}{2} + \frac{Var[Y'_u]}{2E[Y'_u]}. \quad (19)$$

where Y'_u is the duration of a busy period caused by transmissions of stations except the tagged non-saturated station, the distribution of which is given by

$$Y'_u = \begin{cases} T_u & \text{w.p} \quad a'_u/(1-a'_i) \\ T_{tc} & \text{w.p} \quad a'_{tc}/(1-a'_i) \\ T_{ts} & \text{w.p} \quad a'_{ts}/(1-a'_i) \end{cases} \quad (20)$$

VI. Numerical Evaluation and Discussion

This section has two objectives: (i) to validate our basic model (2), (4), (5) and big-packet model (2), (4), (6) by comparing these models with simulations; (ii) to verify the proposed tradeoff mechanism. The simulations were performed using *ns-2* simulator (version 2.33) [15], combined with an EDCA package [16].

Consider a network which consists of N_u non-saturated sources sending small packets and N_t saturated sources sending bursts of η packets. These stations will send packets to an access point in ideal channel conditions. The packet inter-arrival times of unsaturated sources are uniformly distributed in the range $1/\lambda \pm 10\%$, to model voice traffic with enough jitter to avoid phase effects. The rate was sufficiently low that queues rarely built up.

TABLE I
MAC and PHYS parameters for 802.11b system

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	T_{slot}	20 μs
SIFS	T_{sifs}	10 μs

Both saturated stations and non-saturated stations use the user datagram protocol (UDP). The MAC and physical layer parameters were the default values in IEEE 802.11b, as shown in Table 1.

Let l_{sat} and l_{nonsat} denote the payload length of a packet from saturated and unsaturated sources, respectively.

The transmission duration of each data packet in a burst from saturated stations, a data packet from non-saturated stations, and an ACK packet in our analytical model, respectively, are determined as follows:

$$T_{sat} = T_{phys} + \frac{l_{mac} + l_{udpip} + l_{sat}}{r_{data}}$$

$$T_{nonsat} = T_{phys} + \frac{l_{mac} + l_{udpip} + l_{nonsat}}{r_{data}}$$

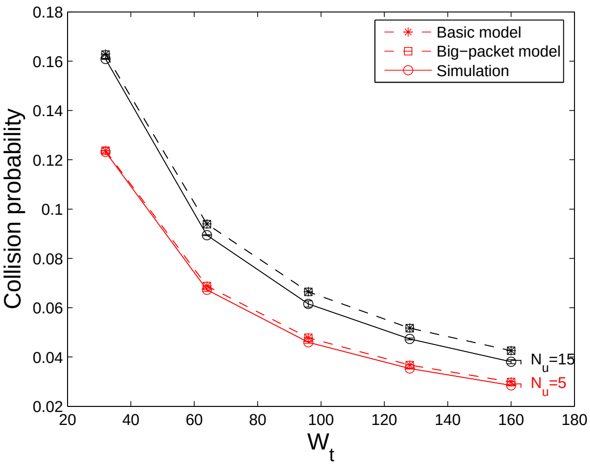
$$T_{ACK} = T_{phys} + \frac{l_{ACK}}{r_{ctrl}}$$

A. Validation of the basic model and the big-packet model

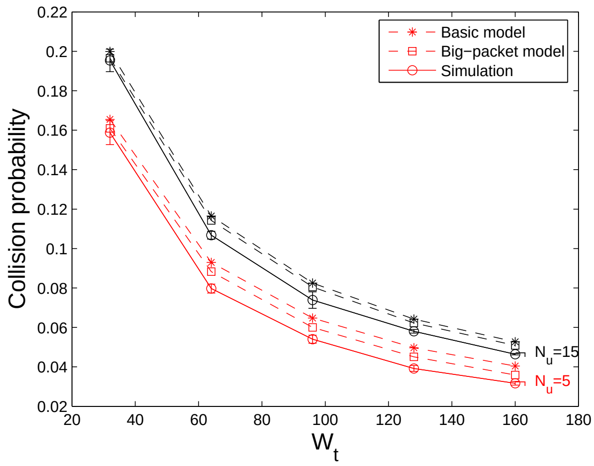
All simulation results are shown with 95% confidence intervals.

1) Scenario 1 with small bursts: In this scenario, $N_t = 3$, $N_u = \{5, 15\}$, $\lambda = 15$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $W_u = 32$, $\eta = 2$, and W_t is varied.

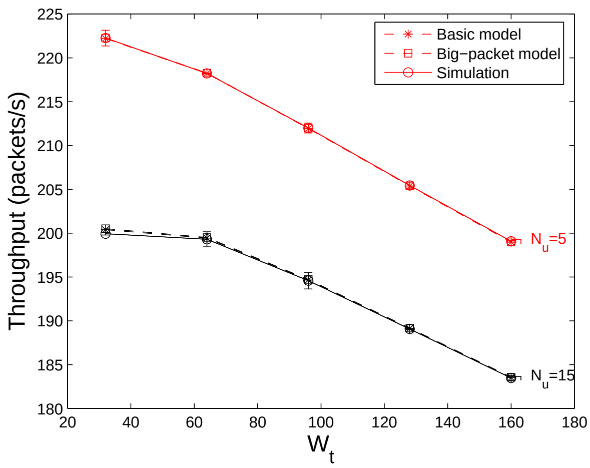
Collision probability of a packet from saturated stations, throughput of a saturated station, collision probability of a packet from non-saturated stations, and mean access delay of a packet from non-saturated stations, respectively, are shown in Figure 1 and 2 as a function of N_u and W_t. These figures show results determined from the big-packet model, the basic model and simulation.



(a) Collision probability of a packet from saturated sources

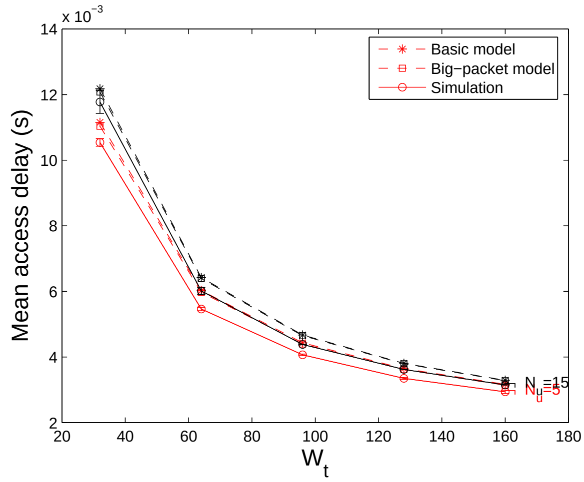


(a) Collision probability of a packet from non-saturated sources



(b) Throughput of a saturated source

Fig. 1. Collision probability and throughput of a saturated source. ($N_t = 3$, $N_u = \{5, 15\}$, $\lambda = 15$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $W_u = 32$, $\eta = 2$.)



(b) Mean access delay of a packet from non-saturated sources

Fig. 2. Collision probability and mean access delay of a packet from non-saturated sources. ($N_t = 3$, $N_u = \{5, 15\}$, $\lambda = 15$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $W_u = 32$, $\eta = 2$.)

These figures shows that the results determined from both the basic model and the big-packet model roughly match with the simulation.

The small deviation between two models in Figure 2(a) is because when determining the collision probability of a packet from non-saturated stations, the basic model makes approximation that a packet from non-saturated stations always comes and senses channel busy while the big-packet model takes into account the probability that a packet from non-saturated stations comes and senses channel busy. The deviation in collision probability between two models results in the deviation in mean access delay observed in Figure 2(b).

In scenarios where saturated stations send small bursts, the difference in collision probability of a packet from non-saturated stations between two models is small. A small change in the collision probability of a packet from non-saturated stations does not have much affect on the collision probability of a packet from saturated stations. This is illustrated in Figure 1 which shows indistinguishable difference in the collision probability and throughput of a saturated station between two models.

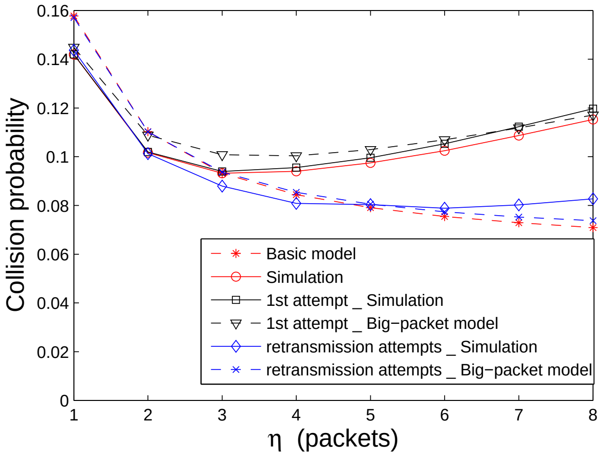
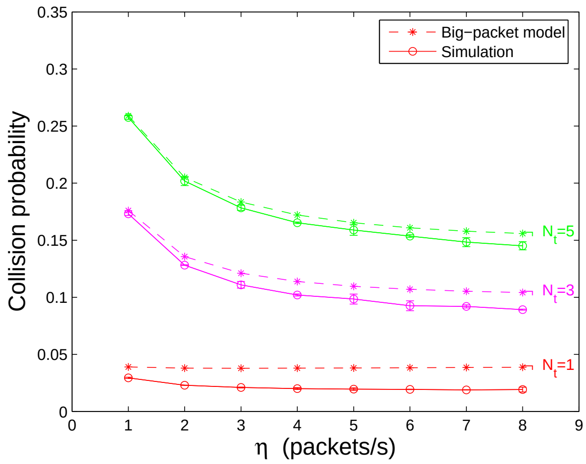
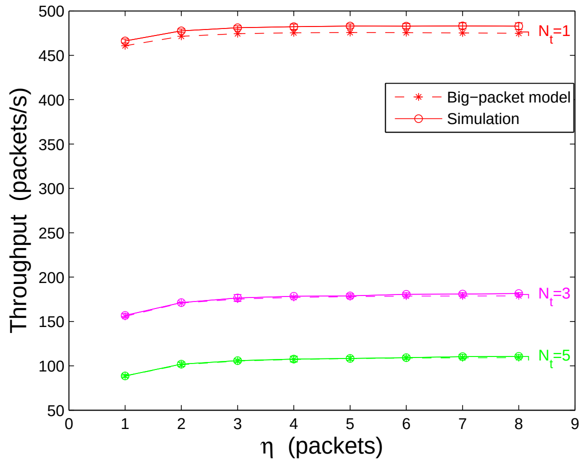


Fig. 3. Collision probability of a small packet from non-saturated sources. ($N_u = 10$, $N_t = 2$, $\lambda = 30$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $W_u = 32$, $W_t = \eta W_u$.)



(a) Collision probability of a packet from saturated sources



(b) Throughput of a saturated source

Fig. 4. Collision probability and throughput of a saturated source. ($N_t = \{1, 3, 5\}$, $N_u = 10$, $\lambda = 30$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $W_u = 32$, $W_t = \eta W_u$.)

2) Scenario 2 with a variety of burst sizes: In this scenario, $N_u = 10$, $N_t = 2$, $\lambda = 30$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $W_u = 32$, $W_t = \eta W_u$. This scenario is an example in which big packets have clear impact on the collision probability of small packets.

The collision probability of a small packet from non-saturated stations is shown in Figure 3 as a function of η. This figure shows the collision probability determined from the basic model and simulation, and the collision probability on the first attempt and retransmission attempts determined from the big-packet model and simulation. The basic model incorrectly predicts the collision probability to decrease monotonically, while the big-packet model can capture the right trend of the collision probability on both the first and retransmission attempts.

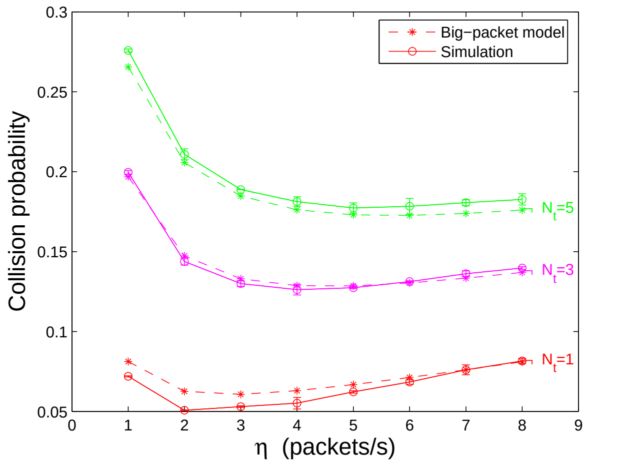
B. Verification of the proposed tradeoff mechanism

The efficiency of the tradeoff mechanism is examined in the following scenario: $N_t = \{1, 3, 5\}$, $N_u = 10$, $\lambda = 30$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $W_u = 32$, $W_t = \eta W_u$. According to the trade-off mechanism, saturated sources increase the spacing between their packets (W_t) in proportion to η.

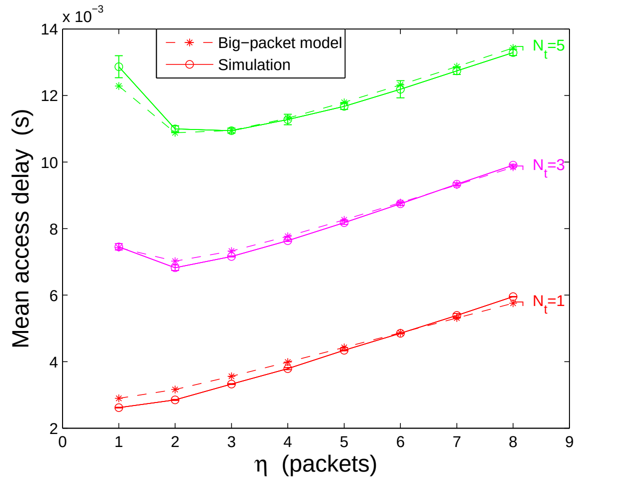
The collision probability of a packet from saturated stations is shown in Figure 4(a) as a function of η and N_t. It can be seen that when η and W_t increase, the collision probability decreases due to the increase of W_t. This explains for the increase of their throughput in Figure 4(b). Figure 4(b) shows that the tradeoff mechanism benefits greedy stations.

The collision probability of a packet from non-saturated stations is shown in Figure 5(a). When η and W_t increase, the collision probability initially decreases and then increases again. In terms of collisions of a tagged non-saturated station, there are collisions with saturated stations which go down due to the increase of W_t and collisions with other non-saturated stations which go up due to the increase of η. The dominance of the former causes the initial decrease of collision probability in Figure 5(a) and the increase afterwards is caused by the dominance of the latter.

Figure 5(b) shows the mean access delay of a packet from non-saturated stations as a function of η and N_t.



(a) Collision probability of a packet from non-saturated sources



(b) Mean access delay of a packet from non-saturated sources

Fig. 5. Collision probability and mean access delay of a packet from non-saturated sources. ($N_t = \{1, 3, 5\}$, $N_u = 10$, $\lambda = 30$ packets/s, $l_{sat} = 1040$ bytes, $l_{nonsat} = 100$ bytes, $W_u = 32$, $W_t = \eta W_u$.)

When η and W_t increase, for N_t greater than 1, the mean access delay first goes down and then goes up, which proves that the tradeoff mechanism also benefits non-saturated stations. The initial decrease of delay is caused by the decrease of the collision probability in Figure 5(a) and the increase afterwards is due to the increase of η. The optimal value of η which minimizes the mean delay may vary in different scenarios.

In summary, Figure 4(b) and 5(b) shows that the tradeoff mechanism can benefit both non-saturated stations and saturated stations. Although the optimal value of η may vary in different scenarios, in most cases, η of 2 provides significant improvement in throughput of saturated stations and mean delay of non-saturated stations.

As can be seen, the big-packet model can capture the trend of collision probabilities, mean delay, and throughput. Therefore, it can be used to estimate the optimal η in the tradeoff mechanism.

VII. Conclusion

Our report has shown that with the proposed tradeoff mechanism, both low-rate realtime traffic and greedy data traffic can be better off. Unlike traditional approach which gives realtime traffic higher priority than data traffic, this helps to reduce data traffic's incentive to pretend to be realtime traffic. To estimate the optimal value of MAC parameters with the proposed tradeoff mechanism, we propose a simple model of an 802.11e EDCA WLAN based on mean-field approximation as in previous works. However, when a network carries some large packets and many small packets, the collision probability after a large packet is much larger than predicted by previous models and our simple model. The collision probability is of importance because the energy consumption of the battery powered mobile devices depends on the number of packet transmissions, which is directly related to the collision probability. Therefore, we also propose another model which captures this effect.

References

[1] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements*, IEEE Standard 802.11e, 2005.

[2] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, 2000.

[3] G. Bianchi, I. Tinnirello and L. Scalia, "Understanding 802.11e Contention-based prioritization mechanisms and their coexistence with legacy 802.11 stations," *IEEE Network*, vol. 19, no. 4, pp. 28–34, 2005.

[4] N. Ramos, D. Panigrahi, and S. Dey, "Quality of Service Provisioning in 802.11e Networks: Challenges, Approaches, and Future Directions," *IEEE Network*, vol. 19, no. 4, pp. 14–20, 2005.

[5] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 Distributed Coordination Function in Nonsaturated Heterogeneous Conditions," *IEEE/ACM Transactions on networking*, vol. 15, no. 1, pp. 159–172, 2007.

[6] H. M. K. Alazemi, A. Margolis, J. Choi, R. Vijaykumar, and S. Roy, "Stochastic modelling and analysis of 802.11 DCF with heterogeneous non-saturated nodes," *Computer Communications 30*, pp. 3652-3661, 2007.

[7] X. Ling, L. X. Cai, J. W. Mark, and X. Shen, “Performance Analysis of IEEE 802.11 DCF with heterogeneous traffic,” *The 4th Annual IEEE Consumer Communications and Networking Conference (CCNC 2007)*, art. no. 4199105, pp. 49-53, 2007.

[8] I. Inan, F. Keceli, and E. Ayanoglu, “Modeling the 802.11e Enhanced Distributed Channel Access Function,” in *Proc. IEEE GLOBECOM’07*, pp. 2546–2551, 2007.

[9] J. Hu, G. Min, M. E. Woodward, and W. Jia, “A comprehensive analytical model for IEEE 802.11e QoS differentiation schemes under unsaturated traffic loads,” *IEEE International Conference on Communications 2008*, art. no. 4533088, pp. 241-245, 2008.

[10] P. E. Engelstad and O. N. Østerbø, “Non-saturation and saturation analysis of IEEE 802.11e EDCA with starvation prediction,” *Proceedings of the Eighth ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (ACM MSWiM 2005)*, pp. 224-233, 2005.

[11] P. Serrano, A. Banchs, and A. Azcorra, “A Throughput and Delay Model for IEEE 802.11e EDCA under non-saturation,” *Wireless Personal Communications*, vol. 43, pp. 467–479, 2007.

[12] V. Ramaiyan, A. Kumar, and E. Altman, “Fixed Point Analysis of Single Cell IEEE 802.11e WLANs: Uniqueness and Multistability,” *IEEE/ACM Transactions on networking*, vol. 16, no. 5, pp. 1080–1093, 2008.

[13] D. Xu, T. Sakurai and H. L. Vu, “An Access Delay Model for IEEE 802.11e EDCA,” *IEEE transactions on mobile computing*, vol. 8, no. 2, pp. 261–275, 2009.

[14] L. Kleinrock, ”Queueing systems,” John Wiley & Sons, Inc., NewYork, vol. 1, 1975.

[15] “The network simulator ns-2,” Available at http://www.isi.edu/nsnam/ns/.

[16] 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/.