Fairness Provision in the IEEE 802.11e Infrastructure Basic Service Set †

Inanc Inan, Feyza Keceli, and Ender Ayanoglu
Center for Pervasive Communications and Computing
Department of Electrical Engineering and Computer Science
The Henry Samueli School of Engineering
University of California, Irvine, 92697-2625
Email: {inan, fkeceli, ayanoglu}@uci.edu

Abstract

Most of the deployed IEEE 802.11e Wireless Local Area Networks (WLANs) use infrastructure Basic Service Set (BSS) in which an Access Point (AP) serves as a gateway between wired and wireless domains. We present the unfairness problem between the uplink and the downlink flows of any Access Category (AC) in the 802.11e Enhanced Distributed Channel Access (EDCA) when the default settings of the EDCA parameters are used. We propose a simple analytical model to calculate the EDCA parameter settings that achieve weighted fair resource allocation for all uplink and downlink flows. We also propose a simple model-assisted measurement-based dynamic EDCA parameter adaptation algorithm. Moreover, our dynamic solution addresses the differences in the transport layer and the Medium Access Control (MAC) layer interactions of User Datagram Protocol (UDP) and Transmission Control Protocol (TCP). We show that proposed Contention Window (CW) and Transmit Opportunity (TXOP) limit adaptation at the AP provides fair UDP and TCP access between uplink and downlink flows of the same AC while preserving prioritization among ACs.

I. INTRODUCTION

IEEE 802.11 Wireless Local Area Network (WLAN) is built around a Basic Service Set (BSS) [1]. While a number of stations may gather to form an independent BSS with no connectivity to the wired network, the common deployment is the infrastructure BSS which includes an Access Point (AP). The AP provides the connection to the wired network.

The IEEE 802.11 standard [1] defines Distributed Coordination Function (DCF) as a contention based Medium Access Control (MAC) mechanism. The 802.11e standard [2] updates the MAC layer of the

† This work is supported by the Center for Pervasive Communications and Computing, and by Natural Science Foundation under Grant No. 0434928. Any opinions, findings, and conclusions or recommendations expressed in this material are those of authors and do not necessarily reflect the view of the Natural Science Foundation.
former 802.11 standard for Quality-of-Service (QoS) provision. In particular, the Enhanced Distributed Channel Access (EDCA) function of 802.11e is an enhancement of the DCF. The EDCA scheme (similarly to DCF) uses Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) and slotted Binary Exponential Backoff (BEB) mechanism as the basic access method. The major enhancement to support QoS is that EDCA differentiates packets using different priorities and maps them to specific Access Categories (ACs) that use separate queues at a station. Each AC, within a station \((0 \leq i \leq 3)\) contends for the channel independently of the others. Levels of services are provided through different assignments of the AC-specific EDCA parameters; Contention Window (CW) sizes, Arbitration Interframe Space (AIFS) values, and Transmit Opportunity (TXOP) limits.

The DCF and the EDCA are defined such that each station in a BSS uses the same contention parameter set. Therefore, fair access can be achieved in the MAC layer for all the contending stations in terms of the average number of granted transmissions, over a sufficiently long interval. However, this does not translate into achieving fair share of bandwidth between uplink and downlink flows in the 802.11e infrastructure BSS. An AC of the AP which serves all downlink flows has the same access priority with the same AC of the stations that serve uplink flows. Therefore, an approximately equal number of accesses that an uplink AC may get is shared among all downlink flows in the same AC of the AP. This leads to the uplink/downlink unfairness problem in the WLAN where each individual downlink flow gets comparably lower bandwidth than each individual uplink flow gets at high load. This phenomenon will be described further in Section II-A.

We deal with weighted fair channel access between the uplink and the downlink flows of the same AC in the IEEE 802.11e infrastructure BSS. Using a simple analytical approach, we calculate the EDCA parameter settings that achieve a given utilization ratio between the uplink and the downlink transmissions. Comparing with simulation results, we noticed that sticking only with analytical results that are based on ideal condition assumptions may result in inaccuracies in a real WLAN scenario. Therefore, we also propose a simple model-assisted measurement-based dynamic EDCA parameter adaptation algorithm that provides weighted fair resource allocation in an arbitrary scenario.

Most of the data traffic in the Internet is carried by Transmission Control Protocol (TCP), while most of the real-time applications use User Datagram Protocol (UDP). UDP employs one-way unreliable communication. On the other hand, TCP defines reliable bi-directional communication where the forward link data rate depends on the rate of received Acknowledgment (ACK) packets in the backward link.
Another key contribution of this study is that our solution considers the effects of this difference on the design of the weighted fairness support algorithm.

II. BACKGROUND

In this section, we first present the uplink/downlink unfairness problem in the IEEE 802.11(e) WLAN at high traffic load. Next, we provide a brief review of the literature on this subject.

A. Problem Definition

In the 802.11e WLAN, at high load, a bandwidth asymmetry exists between contending upload and download flows which use the same AC. This is due to the fact that the MAC layer contention parameters are all equal for the AP and the stations. If \( n \) stations and an AP are always contending for the access to the wireless channel using the same AC, each host ends up having approximately \( \frac{1}{n+1} \) share of the total transmissions over a long time interval. This results in \( \frac{n}{n+1} \) of the transmissions to be in the uplink, while only \( \frac{1}{n+1} \) of the transmissions belonging to the downlink flows. This is the WLAN uplink/downlink unfairness problem stated previously. The uneven bandwidth share results in downlink flows experiencing significantly lower throughput and larger delay. The congestion at the AP may result in considerable packet loss depending on the size of interface buffers.

The results may even be more catastrophic in the case of TCP flows. The TCP receiver returns TCP ACK packets to the TCP transmitter in order to confirm the successful reception of data packets. In the case of multiple uplink and downlink flows in the WLAN, returning TCP ACKs of upstream TCP data are queued at the AP together with the downstream TCP. When the bandwidth asymmetry in the forward and reverse path builds up the queue in the AP, the dropped packets impair the TCP flow and congestion control mechanisms which assume equal transmission rate both in the forward and reverse path [3].

TCP’s timeout mechanism initiates a retransmission of a data packet if it has not been acknowledged during a timeout duration. However, any received TCP ACK can cumulatively acknowledge all the data packets sent before the data packet for which the ACK is intended to. When the packet loss is severe in the AP buffer, downstream flows will experience frequent timeouts resulting in significantly low throughput. On the other hand, due to the cumulative property of TCP ACK mechanism, upstream flows with high congestion windows will not experience such frequent timeouts. In this case, it is a low probability that many consecutive TCP ACK losses occur for the same flow. Conversely, flows with low congestion window (fewer packets currently on flight) may experience frequent timeouts and decrease their congestion
windows even more. Therefore, a number of upstream flows may starve in terms of throughput while others enjoy a high throughput. This results in unfairness between the TCP upstream flows on top of the unfairness between the uplink and the downlink.

Fig. 3 shows the average throughput of individual flows for a scenario of 10 uplink UDP, 10 downlink UDP, 10 uplink TCP and 10 downlink TCP connections in an ns-2 simulation [4],[5]. Each connection is initiated by a separate station. All stations employ 54 Mbps data rate at the physical layer. The packet size is 1500 bytes for all flows. UDP flows are mapped to an AC with $CW_{\text{min}} = 31$ and $CW_{\text{max}} = 511$. TCP flows use an AC with $CW_{\text{min}} = 63$ and $CW_{\text{max}} = 1023$. For both ACs, AIFS N values are set to 2 and TXOP limits are 0. Other simulation parameters are as stated in Section IV. The results illustrate the throughput unfairness of the uplink and the downlink flows. The throughput unfairness between uplink TCP connections is also significant. Moreover, data packet losses at the AP buffer have almost shut down all downlink TCP connections.

B. Related Work

There are two groups of studies in the literature related to this work.

The first group works within the constraints of the default 802.11 contention parameters. In [6], the effect of the AP buffer size in the wireless channel bandwidth allocation for TCP is studied. The proposed solution of [6] is to manipulate advertised receiver windows of the TCP packets at the AP. Uplink/downlink fairness problem is studied in [7] using per-flow queueing. A simplified approach is proposed in [8] where two separate queues for TCP data and ACKs are used. In our previous work, we proposed using congestion control and filtering techniques at the MAC layer to solve the TCP uplink unfairness problem [9]. Two queue management strategies are proposed in [10] to improve TCP fairness. A rate-limiter approach is used in [11] which requires available instantaneous WLAN bandwidth estimation in both directions.

The second group proposes changes at the MAC layer access parameters to achieve improved fairness. Our work also falls into this category. AIFS and CW differentiation is proposed for improved fairness and channel utilization in [12]. A simulation-based analysis is carried out for a specific scenario consisting of TCP and audio flows both in the uplink and the downlink. An experimental study is carried out in [13] to decide on CW and TXOP values of the AP and the stations for a scenario with TCP uplink and downlink flows. Both solutions propose that individual uplink and downlink streams use separate ACs. No guidelines are provided on how to decide on the EDCA parameters that achieve fair resource allocation for an arbitrary scenario. Also, the interaction of TCP flow and congestion control mechanisms with
the MAC is not addressed. In [14], it is proposed that the AP accesses the channel in Point Interframe Space (PIFS) completion without any backoff when the interface queue size goes over a threshold. The use of TXOP is evaluated in [15] for temporal fairness provisioning among stations employing different data rates. Achieving weighted fairness between uplink and downlink in DCF is studied through mean backoff distribution adjustment in [16]. A mechanism that dynamically tunes CW and TXOP values in order to prevent delay asymmetry of realtime UDP flows is proposed in [17]. An adaptive priority control mechanism is employed in [18] to balance the uplink and downlink delay of VoIP traffic.

III. WEIGHTED FAIR ACCESS BETWEEN UPLINK AND DOWNLINK FLOWS

In this section, we first describe the simple analytical model we propose in order to find the AIFS, $CW_{min}$, and TXOP settings of the ACs that provide weighted fairness between uplink and downlink flows. Next, we propose a parameter adaptation algorithm which dynamically updates the analytically calculated CW and TXOP values of the AP regarding simple network measurements. As we will describe in Section III-D, our dynamic solution also addresses the effects of the slow-start phase of TCP.

Every beacon interval, the AP announces the values of the AC-specific EDCA parameters to the stations. The stations overwrite their EDCA parameter settings with the new values if any change is detected. Due to the specific design of the EDCA Parameter Set element in the beacon packet, the stations can only employ CW values that are integer powers of 2, i.e., the AP encodes the corresponding 4-bit fields of $CW_{min}$ and $CW_{max}$ in an exponent form. A key point which the studies in the literature have missed is that the CW settings of the ACs at the AP are not restricted to the powers of 2. The ACs at the AP may use any value and this value does not have to be equal to what is announced via beacons.

A. Analytical Model

Fair access between uplink and downlink flows using the same AC can be provided by assigning different EDCA parameters for the AP and the stations. This results in two Traffic Classes (TCs) using the same AC. While uplink flows constitute the first TC, downlink flows constitute the second TC. In the analysis, we will treat the case with one AC (thus 2 TCs), since we address the weighted fairness problem between the uplink and downlink flows that are mapped to the same AC. Moreover, we only formulate the situation when there is only one TC per station, therefore no internal collisions can occur. Note that, this does not cause any loss of generality, since the analysis can be extended for larger number of ACs or TCs as in [19], and larger number of ACs per station as in [20],[21].
Our analysis considers the fact that the difference in AIFS creates the so-called contention zones as shown in Fig. [19],[22],[23],[21]. First, we calculate the average collision probability of each TC according to the long term occupancy of AIFS and backoff slots in saturation. The average collision probability of a TC is a function of transmission probabilities of all TCs. Next, we formulate the average transmission probability for each TC, which is a function of average collision probability of the same TC. This results in a set of nonlinear equations which can be solved numerically.

We define $p_{c_{i,x}}$ as the probability that TC$_i$ experiences a collision given that it has observed the medium idle for $AIFS_x$ and transmits in the current slot (note $AIFS_x \geq AIFS_i$ should hold). For notational simplicity, let uplink flows belong to TC$_0$ and downlink flows belong to TC$_1$. Let $d_i = AIFSN_i - AIFSN_{min}$ where $AIFSN_{min} = \min(AIFSN_0, AIFSN_1)$ and $AIFSN_i = SIFS + AIFSN_i \cdot T_{slot}$. Following the slot homogeneity assumption of [24], assume that each TC$_i$ transmits with constant probability, $\tau_i$. Also, let the total number of TC$_i$ in the BSS be $N_i$ (note that $N_1 = 1$). Then,

$$p_{c_{i,x}} = 1 - \frac{\prod_{i': d_{i'} \leq d_x} (1 - \tau_{i'})^{N_{i'}}}{(1 - \tau_i)}.$$ (1)

We use the Markov chain shown in Fig. 2 to find the long term occupancy of contention zones. Each state represents the $n^{th}$ backoff slot after completion of the $AIFSN_{min}$ idle interval following a transmission period. The Markov analysis uses the fact that a backoff slot is reached if no transmission occurs in the previous slot. Moreover, the number of states is limited by the maximum idle time between two successive transmissions which is $W_{min} = \min(CW_{i,max})$ for a saturated scenario. The probability that at least one transmission occurs in a backoff slot in contention zone $x$ is

$$p^{tr}_x = 1 - \prod_{i': d_{i'} \leq d_x} (1 - \tau_{i'})^{N_{i'}}.$$ (2)

The long term occupancy of the backoff slots $b'_n$ in Fig. 2 can be obtained from the steady-state solution. Then, the average collision probability $p_{c_i}$ is found by weighing zone specific collision probabilities $p_{c_{i,x}}$ according to the long term occupancy of contention zones (thus backoff slots)

$$p_{c_i} = \frac{\sum_{n=d_i+1}^{W_{min}} p_{c_{i,x}} b'_n}{\sum_{n=d_i+1}^{W_{min}} b'_n}.$$ (3)

where $x = \max \left( y \mid d_y = \max_{z} (d_z \mid d_z \leq n) \right)$ which shows $x$ is assigned the highest index value within a set of TCs that have $AIFSN$ smaller than equal to $n + AIFSN_{min}$. 


Given $p_{ci}$, we can calculate the expected number of backoff slots $E_i[t_{bo}]$ that TC$_i$ waits before attempting a transmission. Let $W_{i,k} = 2^{\min(k,m_i)}(CW_{i,min} + 1) - 1$ be the CW size of TC$_i$ at backoff stage $k$ where $CW_{i,max} = 2^{m_i}(CW_{i,min} + 1) - 1$, $0 \leq m_i < r_i$. Note that, when the retry limit $r_i$ is reached, any packet is discarded.

\[
E_i[t_{bo}] = \sum_{n=0}^{\infty} (p_{ri})^n \sum_{k=1}^{r} p_{ci}^{k-1}(1 - p_{ci}) \frac{W_{i,k}}{2} = \frac{1}{1 - p_{ci}} \sum_{k=1}^{r} p_{ci}^{k-1}(1 - p_{ci}) \frac{W_{i,k}}{2}.
\] (4)

Then as also shown in [23], the transmission probability of TC$_i$ can be calculated as

\[
\tau_i = \frac{1}{E_i[t_{bo}] + 1}.
\] (5)

The nonlinear system of equations (1)-(5) can be solved numerically to calculate average collision and transmission probabilities of each TC$_i$ for an arbitrary setting of EDCA parameters. We provide the validation of the proposed analytical model in [19].

**B. Weighted Fairness between Uplink and Downlink Flows**

Let $\gamma_i$ be the probability that the transmitted packet belongs to an arbitrary user from TC$_i$ given that the transmission is successful. Also, let $p_{si,n}$ be the probability that a successfully transmitted packet at backoff slot $n$ belongs to AC$_i$. Then,

\[
\gamma_i = \sum_{n=d_i+1}^{W_{min}} b'_n \frac{p_{si,n}}{\sum_{j} p_{sj,n}}.
\] (6)

\[
p_{si,n} = \begin{cases} 
\frac{N_i \tau_i}{(1 - \tau_i)} \prod_{i':d_{i'} \leq n-1} (1 - \tau_{i'})^{N_{i'}}, & \text{if } n \geq d_i + 1 \\
0, & \text{if } n < d_i + 1.
\end{cases}
\] (7)

Let $U$ denote the utilization ratio between the downlink and the uplink transmissions of an AC. Let $N_{TXOP,i}$ denote the maximum number of packets that can fit in one TXOP of TC$_i$. Then, for our running example with one AC,

\[
U = \frac{\gamma_1 \cdot N_{TXOP,1}}{\gamma_0 \cdot N_{TXOP,0}}.
\] (8)

1) **Implementation of the Numerical Solution:** Without loss of generality, the EDCA parameters of the stations, $AIFS_0$, $CW_{min,0}$, and $N_{TXOP,0}$, are fixed at predetermined values. Then, the EDCA parameters of the TC at the AP, $AIFS_1$, $CW_{min,1}$, and $N_{TXOP,1}$, that achieve a required utilization ratio $U_r$ can be
calculated numerically as follows.

1) We assume AIFS differentiation is only used for the prioritization between the ACs not the TCs (thus $AIFS_0 = AIFS_1$).

2) When $AIFS_0 = AIFS_1$, after some algebra on (6)-(8),

$$U = \frac{\tau_1 \cdot (1 - \tau_0) \cdot N_{TXOP,1}}{\tau_0 \cdot (1 - \tau_1) \cdot N_{TXOP,0}}.$$  \hspace{1cm} (9)

Therefore, $\tau_1$ can be written in terms of $\tau_0$, $N_{TXOP,0}$, $N_{TXOP,1}$, and $U_r$. A numerical solution for $\tau_0$ and $\tau_1$ for given $U_r$ and a fixed value of $N_{TXOP,1}$ (initially, $N_{TXOP,1} = 1$) is obtained using (1)-(5).

3) $CW_{min,1}$ can be calculated as follows (the formula below is obtained using (8) and (9) in [25, Section IV-A]),

$$CW_{min,i} = \frac{2 - \tau_i}{\tau_i} \cdot \frac{(1 - p_{ci})(1 - 2p_{ci})(1 - p_{ci})}{(1 - p_{ci})^2(1 - (2p_{ci})^{m_i+1}) + 2^{m_i}p_{ci}^{m_i+1}(1 - 2p_{ci})(1 - p_{ci})(1 - p_{ci}^{m_i+1}) - 1} - 1.$$  \hspace{1cm} (10)

4) A simple controller block checks whether the prioritization among ACs are maintained or not for the new configuration. This block ensures that $CW_{min}$ of a low priority AC (at the AP or a station) is not smaller than $CW_{min}$ of a higher priority AC. Therefore, if analytically calculated $CW_{min,1}$ value does not satisfy the controller block requirements, $N_{TXOP,1}$ is doubled and the algorithm returns to step 3. The larger $N_{TXOP,1}$ is, the larger $CW_{min,1}$ will be.

5) If the calculated $CW_{min,1}$ is not an integer, it is rounded to the closest integer value.

A few remarks on the implementation are as follows.

- A numerical solution also exists when $AIFS_0$ and $AIFS_1$ are not equal, but the implementation differs since (9) does not hold. In such a case, $AIFS_1$ is also assigned an initial value as $N_{TXOP,1}$ and the nonlinear system of equations (1)-(8) is solved numerically. According to the controller block requirements on $CW_{min,1}$, the procedure may be repeated for updated values of $AIFS_1$ and $N_{TXOP,1}$.

- As previously mentioned, our formulation is valid for the situation when there is only one TC per station (including the AP). As an approximation, we assume that (1)-(8) still holds when there are multiple TCs at the AP. Indeed as only a few collisions are avoided when the internal collision procedure is run at the AP [26], the solution of (1)-(8) will be very close to an extension that exactly formulates the virtual collisions at the AP. In this case, if the AIFS values of TCs within an AC remains equal, it can be shown that (9) still holds for the TCs of the same AC. Therefore, we use
the implementation procedure previously stated for scenarios when larger number of ACs exist as long as there is one AC (or TC) per station and multiple TCs at the AP.

2) Proposed BEB Algorithm for non-integer CW values: As specified in [2], the initial value of $CW$ is set to the AC-specific $CW_{\text{min}}$. At each unsuccessful transmission, the value of $CW$ is doubled until the maximum AC-specific $CW_{\text{max}}$ limit is reached. The value of $CW$ is reset to the AC-specific $CW_{\text{min}}$ if the transmission is successful, or the retry limit is reached thus the packet is dropped.

The proposed analytical calculation for weighted fairness may decide a non-integer value of $CW_{\text{min},1}$ thus $W_{1,k}, k < r_1$. The simplest approach is rounding to the closest integer and employing the rounded value in the BEB.

Instead, we also propose the AP to choose integer $W'_{1,k}$ values from a probability distribution that satisfies $E[W'_{1,k}] = W_{1,k}$. For example, it is straightforward to show a simple discrete probability distribution such as $Pr(W'_{1,k} = \lfloor W_{1,k} \rfloor) = \lfloor W_{1,k} \rfloor - W_{1,k}$ and $Pr(W'_{1,k} = \lceil W_{1,k} \rceil) = W_{1,k} - \lfloor W_{1,k} \rfloor$ holds. According to the proposed algorithm, the EDCA function at the AP decides on the interval $(0, W'_{1,k})$ to select the backoff value regarding the given simple discrete probability distribution.

Fig. 4 shows the downlink/uplink access ratio for increasing number of uplink and downlink flows. We assume equal $AIFS_N = 2$ for all the stations and the AP, and analytically calculate $CW_{\text{min},1}$ that achieves downlink/uplink access ratio of $U_r = 1$ when $CW_{\text{min},0} = 127$, $N_{TXOP,0} = 1$, and $N_{TXOP,1}$ is varied from 1 to 4. The performance of rounding the analytically calculated CW values is compared with the performance of the proposed BEB algorithm that uses the stated discrete probability distribution function. As the results imply, the proposed BEB algorithm maintains perfect weighted fairness while rounding the analytically calculated value may result in slight inaccuracies in terms of utilization ratio. As the number of uplink stations increase, $CW_{\text{min},1}$ that achieves $U_r = 1$ decreases. As Fig. 4 shows the effect of rounding is much more noticeable when $CW_{\text{min},1}$ is small. The effect of rounding becomes negligible as $N_{TXOP,1}$ (thus $CW_{\text{min},1}$) is increased.

C. Dynamic Parameter Adaptation

The IEEE 802.11 infrastructure BSS exhibits some non-ideal conditions which most of the analytical models ignore to maintain simplicity. For example,

- Accurate information on the instantaneous number of active flows may not always be available to the AP [27].
• If a station and the AP collide, the station’s transmission results in failure since the destination (the AP) is not in listen mode. However, there is some probability that the transmission of the AP results in success as a consequence of the capture effect depending on the spatial distribution and the power levels of the stations [28].

Such non-ideal conditions make finding the optimum EDCA setting analytically hard for any scenario. This also limits the use of proposed BEB algorithm for non-integer CW values. We propose a simple model-assisted measurement-based dynamic algorithm to adapt the analytically calculated $CW_{\min}$ values for such scenarios.

The AP carries out the dynamic adaptation for each AC every $\beta$ beacon intervals which is called an adaptation interval in the sequel. If it is detected as a new flow starting transmission or as an old flow becoming inactive at the last adaptation interval, the algorithm decides on new good EDCA parameters using the proposed analytical model which results in weighted fair resource allocation for the estimated number of uplink and downlink flows in ideal conditions. Otherwise, fine tuning on the $CW$ and the $TXOP$ values of the AC at the AP is carried out to make measured $U$ as close as to $U_r$.

We use a simple algorithm to estimate the number of active flows. More advanced approaches [27] can also be used. The AP counts the number of unique source and destination MAC addresses observed from incoming frames to estimate the number of uplink and downlink flows respectively. Let $n_u$ and $n_d$ denote the number of uplink and downlink flows labeled as active. If the AP receives a packet with the corresponding MAC address not on its list, it adds the new MAC address to the list and increments $n_u$ or $n_d$. If the AP does not receive any packet with the corresponding MAC address during the last adaptation interval, it deletes the MAC address from the list and decrements $n_u$ or $n_d$. Then, we define the required utilization ratio as

$$U_r = \frac{n_d}{n_u}. \quad (11)$$

If $U_r$ has been changed during the last adaptation interval, EDCA parameters are analytically calculated for $U = U_r$ and the fine tuning phase is skipped. Otherwise, solely fine tuning on $CW_{\min}$ is performed as follows. Every adaptation interval, the AP measures the number of successful uplink and downlink transmissions, $n_{tu}$ and $n_{td}$ respectively where $n_{td}/n_{tu}$ is the measured $U$ of the last adaptation interval. If $\frac{n_{td}}{n_{tu}} < (1 - \alpha) \cdot U_r$, then $CW_{\min,1}$ is decremented (where $0 \leq \alpha \leq 1$). Similarly, if $\frac{n_{td}}{n_{tu}} > (1 + \alpha) \cdot U_r$, then $CW_{\min,1}$ is incremented. Otherwise, no action is taken. Note that using steps equal in value to 1 in the $CW_{\min}$ adaptation is sufficient since the analytical calculation will provide a good initial guess.
D. TCP-MAC Interactions

TCP defines a reliable bi-directional communication where the forward link data rate depends on the rate of the received ACK packets in the backward link. This behavior of TCP constitutes the main difference between TCP and UDP access in the WLAN. The key observation is that, if we assume there are no packet losses in TCP connections (infinitely large interface buffers at the AP and the stations), the TCP access is fair irrespective of the EDCA parameter selection (which is not the case for UDP). This is due to the fact that the slow link limits the throughput for all TCP flows. However, when the buffer size at the AP (bottleneck) is limited, significant unfairness and low channel utilization is experienced as previously shown in Fig. 3. Therefore, for fair resource allocation and high channel utilization, packet losses at the AP buffer should be minimized. We configure our adaptation algorithm considering the TCP dynamics to achieve this objective.

None of the work in the literature on IEEE 802.11 MAC upload/download fairness considered the asymmetry in the forward and backward link packet rate during the slow-start phase of the TCP connections. During the slow-start phase, the packet rate in the forward link is twice the packet rate in the backward link. When the congestion avoidance phase is entered, the forward and the backward link packet rates become equal. When this asymmetry during slow-start is neglected, the download traffic is penalized with longer queueing delays. Depending on the buffer availability, significant packet loss may even occur during the slow-start. These may considerably affect the short-term fairness and the channel utilization.

Our solution is simple yet effective. Considering each TCP data and ACK streams of each connection as individual active flows, the parameter adaptation algorithm of Section III-C is used. Since TCP is fair irrespective of the EDCA parameter selection as long as there are no packet losses, fine tuning on \( CW_{\text{min}} \) is always skipped. Therefore, the AP does not have to measure \( n_{tu} \) and \( n_{td} \). On the other hand, fine tuning is carried out on TXOP assignments to overcome increased rate of downlink TCP data flows during slow-start. Since the forward to backward link packet rate ratio is 2 during the slow-start, the analytically calculated TXOP duration is multiplied by 2. Our approach is adapting the duration of the TXOP depending on the number of packets buffered at the interface queue. If the number of packets goes over a threshold value \( th \), doubled TXOPs are enabled until the number goes below the threshold again.

Assigning best-effort data flows a non-zero TXOP or a small \( CW_{\text{min}} \) may not be a favorable approach when multimedia flows coexist in the WLAN. The controller block located at the AP should check whether the QoS for admitted realtime flows is preserved or not in the WLAN with the \( CW_{\text{min}} \) and \( TXOP \) values.
calculated for uplink/downlink fairness.

IV. NUMERICAL AND SIMULATION RESULTS

We carried out simulations in ns-2 [4] in order to evaluate the performance of the proposed weighted fairness adaptation algorithm. For the simulations, we employ the IEEE 802.11e EDCA MAC simulation module for ns-2.28 [5].

We consider a network topology where each wireless station initiates a connection with a wired station where the WLAN traffic is relayed to the wired network through the AP. The stations are uniformly distributed on a circle and the AP is located at the center. The power thresholds are set so that every station can hear the other’s transmission. The data connections use either UDP or TCP NewReno. The UDP traffic uses a Constant Bit Rate (CBR) application. The TCP traffic uses a File Transfer Protocol (FTP) agent which models bulk data transfer. The default TCP NewReno parameters in ns-2 are used. The UDP traffic is mapped to a higher priority AC than the TCP traffic. All the stations are assumed to have 802.11g PHY using 54 Mbps and 6 Mbps as the data and basic rate respectively [29]. The packet size is 1500 bytes for all flows. The buffer size at the stations and the AP is set to 200 packets. We found $\beta = 5$, $\alpha = 0.5$, and $th = 50$ packets to be appropriate through extensive simulations.

Fig. 5 shows the average throughput of individual flows for a scenario of 10 uplink UDP, 10 downlink UDP, 10 uplink TCP and 10 downlink TCP connections (same scenario as in Fig. 3). At the stations, UDP flows are mapped to an AC with $CW_{min} = 31$ and $CW_{max} = 511$. TCP flows use an AC with $CW_{min} = 63$ and $CW_{max} = 1023$. For both ACs, AIFSN values are set to 2 and TXOP limits are 0. Unless otherwise stated, all data connections of the stations in other experiments use these ACs (thus these EDCA parameters). At the AP, we run the proposed algorithm designed for weighted fairness support in the downlink and uplink. Since the number of downlink and uplink flows are equal for both ACs, we define the downlink/uplink utilization requirement as $U_r = 1$. The analytical model decides on the $CW$ and the $TXOP$ that achieves $U_r = 1$. Fine tuning on $CW$ is carried out for the fairness of UDP flows. The $TXOP$ is adaptively doubled according to the proposed algorithm for TCP flows. The results illustrate that $U = 1$ is perfectly achieved in terms of throughput for both UDP and TCP flows.

We have tested the proposed algorithm for a range of network conditions.

a) Experiment 1: In the first set of experiments, we generate an equal number of TCP and UDP flows both in the uplink and downlink. Each flow starts at the same time and the simulation duration is 100 seconds. The wired link delay (denoted as Round Trip Time (RTT) in the titles of the figures) is
equal for all flows (30 ms). Fig. 6 shows the total throughput of TCP and UDP flows in each direction for the proposed algorithm. The results for the default 802.11e EDCA are also included for comparison. As the results depict, $U = 1$ is perfectly achieved in terms of average throughput for the proposed algorithm, while the default scheduler cannot maintain fair access. Fig. 7 shows the total throughput of TCP and UDP flows as well as the total system throughput for the proposed algorithm and the default case. The proposed algorithm can maintain more efficient channel utilization than the default EDCA while providing fair access. In Fig. 8, we present the performance in terms of fairness between individual TCP or UDP flows in the same direction for the proposed algorithm and the default EDCA. The performance metric we use is the widely used fairness index [30]. The fairness index, $f$, is defined as follows: if there are $n$ concurrent connections and the throughput achieved by connection $i$ is equal to $x_i$, $1 \leq i \leq n$, then

$$f = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \sum_{i=1}^{n} x_i^2}. \quad (12)$$

As the results imply, the proposed algorithm also provides fair access between UDP and TCP flows of the same direction. However, the default EDCA results in unfair resource allocation even between the TCP flows of the same direction. As we have described in Section II-A, the unfairness is more significant between TCP uplink flows. Although no unfair behavior is expected between UDP flows in the same direction, we have included these results in Fig. 8 for the sake of completeness.

b) Experiment 2: We have repeated the simulation set of experiment 1 when the wired link delay is varied for TCP flows. The wired link delay of the first TCP connection is set to 24 ms and each newly generated TCP connection is assigned 4 ms larger wired link delay than the previous one. Therefore, the second TCP connection has 28 ms wired link delay, the third one has 32 ms wired link delay and so on. This holds for both uplink and downlink connections. UDP wired link delay is constant for each connection. Fig. 9 shows the average throughput of each TCP and UDP flow in each direction. Fig. 10 shows the total throughput of TCP and UDP flows as well as the total system throughput. Fig. 11 shows the performance in terms of fairness between individual TCP or UDP flows in the same direction. As the results show, the performance of the proposed algorithm in terms of fair resource allocation is independent of the duration of the wired link delay. High channel utilization and perfect fairness is maintained. On the other hand, as the comparison of Fig. 8 and Fig. 11 imply, the performance of default EDCA depends on the duration of the wired link delay. In the case of varying wired link delays, the unfairness between individual TCP flows both in the downlink and uplink is even worse.
c) **Experiment 3:** In the third set of experiments, we also generate an equal number of TCP and UDP flows both in the uplink and downlink. In this scenario, each uplink or downlink flow starts at different times and the simulation duration is 300 seconds. The wired link delay is equal for all flows. The first downlink UDP connection starts at $t = 5$ s. The first uplink UDP connection starts at $t = 10$ s. The first uplink TCP connection starts at $t = 7$ s. The first downlink TCP connection starts at $t = 12$ s. Then, a new flow of the same type arrives every 10 s. No other flow arrives after 200 s. Fig. 12 and Fig. 13 show the instantaneous UDP and TCP throughput of individual uplink and downlink flows respectively for default EDCA. The unfairness between uplink and downlink for both UDP and TCP and the unfairness between individual TCP flows both in the uplink and downlink are evident. Fig. 14 and Fig. 15 show the instantaneous UDP and TCP throughput of individual uplink and downlink flows respectively when the proposed algorithm is enabled. As the results imply, the proposed algorithm adaptively updates EDCA parameters and always maintains instantaneous $U_r$ (as calculated in (11)).

**d) Experiment 4:** We have repeated the simulation set of experiment 3 when the wired link delay is varied for TCP flows. We set different wired link delays using the way as previously stated. Fig. 16 and Fig. 17 show the instantaneous UDP and TCP throughput of individual uplink and downlink flows respectively for default EDCA. The unfairness between individual TCP flows both in the uplink and downlink are more pronounced when compared with the equal wired link delay scenario. Fig. 18 and Fig. 19 show the instantaneous UDP and TCP throughput of individual uplink and downlink flows respectively for the proposed algorithm. Since the proposed algorithm adaptively updates EDCA parameters, it maintains fair resource allocation. The downlink flows does not starve in terms of throughput.

**e) Experiment 5:** We have repeated the simulation set of experiment 3 when half of the TCP flows model short flows. The flow generation times follow the rules of experiment 3. The simulation duration is 450 s. No other flow arrives after 300 s. The short and long TCP flows are alternatively initiated both in the downlink and uplink. The short TCP flows consist of 31 packets and leave the system after all the data is transferred. Fig. 20 shows the total transmission duration for individual short TCP flows for the proposed algorithm and the default EDCA. Note that flow indices from 1 to 15 represent uplink TCP flows while flow indices from 16 to 30 represent downlink TCP flows. The file transfers with short durations can be completed in a considerably shorter time when the proposed algorithm is used. At high load, short flows experience significantly long delays and connection timeouts when default constant EDCA parameter selection is used.
f) Experiment 6: We have repeated the simulation set of experiment 5 when the wired link delay is varied for TCP flows. We set different wired link delays using the way as previously stated. Fig. 21 shows the total transmission duration for individual short TCP flows for the proposed algorithm and the default EDCA. The comparison of Fig. 20 and Fig. 21 reveals that the proposed algorithm performance in terms of short TCP flow completion time is independent of varying wired link delays among the flows.

g) Experiment 7: In another set of experiments, we consider three types of traffic sources; audio, video, and data. The audio traffic model implements a Voice-over-IP (VoIP) application as a Constant Bit Rate (CBR) traffic profile at 24 kbps. The constant audio packet size is 60 bytes. Although not presented here, similar results and discussion hold when the silence suppression scheme is used and the audio traffic exhibits on-off traffic characteristics. For the video source models, we have used traces of real H.263 video streams [31]. The mean and maximum video payload size is 2419 bytes and 3112 bytes respectively. The mean video data rate is 255 kbps. The audio flows are mapped to an AC with $CW_{\text{min}} = 7$ and $CW_{\text{max}} = 15$. The video flows use an AC with $CW_{\text{min}} = 15$ and $CW_{\text{max}} = 31$. For both ACs, AIFSN values are set to 2 and TXOP limits are 0. Fig. 22 shows the average throughput of uplink and downlink data flows when there are 5 voice and 5 video flows both in the uplink and downlink (a total of 20 flows with QoS requirements). Similarly, Fig. 23 shows the average throughput of uplink and downlink data flows when there are 10 voice and 10 video flows both in the uplink and downlink. We also compare the results with the proposed algorithm of [13]. As the results of default EDCA and [13] imply, sticking with constant EDCA parameters for any number of flows does not result in fair access no matter which EDCA parameter setting is used. On the other hand, the proposed adaptive algorithm effectively manages fair resource allocation for any number of stations. Note that we have not included the average throughput of the flows with QoS requirements in Fig. 22 and Fig. 23 since all audio and video flows get necessary bandwidth to serve offered load with zero packet loss rate. Fig. 24 compares the average delay of each QoS flow in each direction for default EDCA and the proposed algorithm when there are a total of 20 flows with QoS requirements. Similarly, Fig. 25 compares the average delay of each QoS flow in each direction for default EDCA and the proposed algorithm when there are a total of 40 flows with QoS requirements. As the results show, the QoS flows experience slightly larger delays when the proposed algorithm is used (due to smaller CW and larger TXOP assignment for data flows). On the other hand, the delay increase is well within the limits of QoS requirements. Moreover, fair resource allocation for data flows is provided.
V. CONCLUSIONS

We have proposed a model-assisted measurement-based dynamic EDCA parameter adaptation algorithm that achieves a predetermined utilization ratio between uplink and downlink flows of the same AC while keeping the prioritization among ACs. The key contribution is that depending on simple network measures, the proposed algorithm dynamically adapts the EDCA parameters calculated via a proposed analytical model. Another key insight is that the proposed algorithm differentiates the way of adaptation between UDP and TCP flows regarding their characteristics.

The proposed algorithm is fully compliant with the 802.11e standard. We propose AP to use any CW value, not necessarily exponents of 2. Our observation is that the 802.11e standard does not restrict the CW settings of the ACs at the AP to be the powers of 2, while the CW setting of the ACs at the STA should be powers of 2 due to the definition of specific fields in the beacon packet. Our approach provides the AP the freedom of satisfying any required utilization ratio through fine tuning on CW settings.

Via simulations, it is shown that fair resource allocation between uplink and downlink flows of an AC can be maintained in a wide-range of scenarios when the proposed model-assisted measurement-based dynamic EDCA parameter adaptation algorithm is used. The performance of the proposed algorithm in terms of fair resource allocation is shown to be independent of the duration of the round trip time of a connection. Short flows experience significantly low delays and no connection timeouts. Therefore, we conclude that the proposed method also provides short-term fairness. The QoS requirements of existing audio and video flows in the 802.11e WLAN are maintained. Our results also show that sticking with constant EDCA parameters at any scenario does not result in fair access no matter which EDCA parameter setting is used.

REFERENCES


[29] IEEE Standard 802.11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: Further Higher Data Rate Extension in the 2.4 GHz Band, IEEE 802.11g Std., 2003.


Fig. 1. EDCA backoff after busy medium.
Fig. 2. Transition through backoff slots in different contention zones for the example given in Fig.[1]
Fig. 3. Total throughput of 10 uplink UDP (indices 1-10), 10 downlink UDP (indices 11-20), 10 uplink TCP (indices 21-30) and 10 downlink TCP (indices 31-40 flows) when the AP and the stations use equal EDCA parameters.
Fig. 4. The downlink/uplink access ratio for increasing number of uplink and downlink flows.
Fig. 5. Total throughput of 10 uplink UDP (indices 1-10), 10 downlink UDP (indices 11-20), 10 uplink TCP (indices 21-30) and 10 downlink TCP (indices 31-40 flows) when the AP uses the proposed adaptation algorithm to achieve $U_r = 1$. 
Fig. 6. The total throughput of TCP and UDP flows in each direction (experiment 1).
Fig. 7. The total throughput of TCP and UDP flows as well as the total system throughput (experiment 1).
Fig. 8. Fairness index of individual TCP or UDP flows in the same direction (experiment 1).
Fig. 9. The total throughput of TCP and UDP flows in each direction (experiment 2).
Fig. 10. The total throughput of TCP and UDP flows as well as the total system throughput (experiment 2).
Fig. 11. Fairness index of individual TCP or UDP flows in the same direction (experiment 2).
Fig. 12. The instantaneous UDP throughput of individual uplink and downlink flows for default EDCA (experiment 3).
Fig. 13. The instantaneous TCP throughput of individual uplink and downlink flows for default EDCA (experiment 3).
Fig. 14. The instantaneous UDP throughput of individual uplink and downlink flows for the proposed algorithm (experiment 3).
Fig. 15. The instantaneous TCP throughput of individual uplink and downlink flows for the proposed algorithm (experiment 3).
Fig. 16. The instantaneous UDP throughput of individual uplink and downlink flows for default EDCA (experiment 4).
Fig. 17. The instantaneous TCP throughput of individual uplink and downlink flows for default EDCA (experiment 4).
Fig. 18. The instantaneous UDP throughput of individual uplink and downlink flows for the proposed algorithm (experiment 4).
Fig. 19. The instantaneous TCP throughput of individual uplink and downlink flows for the proposed algorithm (experiment 4).
Fig. 20. The total transmission duration for individual short TCP flows (experiment 5).
Fig. 21. The total transmission duration for individual short TCP flows (experiment 6).
Fig. 22. The average throughput of uplink and downlink data flows when there are 5 voice and 5 video flows both in the uplink and downlink (experiment 7).
Fig. 23. The average throughput of uplink and downlink data flows when there are 10 voice and 10 video flows both in the uplink and downlink (experiment 7).
Equal RTT for each TCP flow, number of QoS Flows: 4*5

Fig. 24. The average delay of each QoS flow in each direction when there are a total of 20 flows with QoS requirements (experiment 7).
Fig. 25. The average delay of each QoS flow in each direction when there are a total of 40 flows with QoS requirements (experiment 7).