

A control theoretic approach to achieve proportional fairness in 802.11e EDCA WLANs

Xiaomin Chen, Ibukunoluwa Akinyemi and Shuang-Hua Yang*
 Department of Computer Science, Loughborough University

Abstract—This paper considers proportional fairness amongst ACs in an EDCA WLAN redfor provision of distinct QoS requirements and priority parameters. A detailed theoretical analysis is provided to derive the optimal station attempt probability which leads to a proportional fair allocation of station throughputs. The desirable fairness can be achieved using a centralised adaptive control approach. This approach is based on multivariable state-space control theory and uses the Linear Quadratic Integral (LQI) controller to periodically update CW_{min} till the optimal fair point of operation. Performance evaluation demonstrates that the control approach has high accuracy performance and fast convergence speed for general network scenarios. To our knowledge this might be the first time that a closed-loop control system is designed for EDCA WLANs to achieve proportional fairness.

I. INTRODUCTION

Enhanced Distributed Channel Access (EDCA) was proposed in IEEE 802.11e-2005 standard to support QoS enhancement and service differentiation for WLAN applications [1]. It extends the basic Distributed Coordination Function (DCF) by classifying traffic flows into four different Access Categories (AC), namely voice, video, best-effort and background. Traffic with higher QoS requirements, e.g. shorter delay deadline, is assigned reda higher priority, and hence, on average, waits for less time before being sent to the channel. This mechanism is beneficial for high-priority traffic. Compared to the DCF, EDCA sacrifices the performance of low-priority traffic to some extent to provide QoS support for high-priority traffic. When the network is saturated with a large proportion of high-priority flows, an extremely unfair scenario will appear, in which the channel will be almost completely occupied by high-priority flows, e.g. VoIP or video streaming flows, however low-priority traffic, such as email or web browsing data, will suffer severe starvation.

Resource allocation in EDCA WLANs has therefore been the subject of considerable interest. The objective is to seek for a fair allocation of network resources (e.g. throughput, airtime and etc.) amongst different traffic types, and meanwhile, guarantee the specific QoS requirements and service differentiation. This paper considers proportional fair allocation of station throughputs amongst ACs redfor provision of distinct average delay deadlines and priority parameters. The 802.11e EDCA standard specifies four contention parameters to distinguish priority levels, which are minimum Contention Window (CW_{min}), maximum Contention Window (CW_{max}),

Arbitration Inter Frame Space (*AIFS*) and maximum Transmission Opportunity (*TXOP*). A set of default values for the four parameters are recommended in the standard for each physical (PHY) layer supported by 802.11e. As the default values do not take into account the varying WLAN conditions, and thus lead to suboptimal performance and no fairness guarantees, in this paper we find the optimal CW_{min} value that leads to proportional fair allocation of station throughputs while assuming *AIFS* and *TXOP* taking the recommended values and $CW_{max} = CW_{min}$. The optimal CW_{min} value corresponds to an optimal station attempt probability which is derived from the proportional fairness analysis.

In order to implement the derived proportional fair allocation in practice, a centralised adaptive approach which uses multivariable state-space control theory is then proposed. The WLAN is represented as a discrete multi-input multi-output (MIMO) linear time-invariant (LTI) state-space model. A state feedback control method, the Linear Quadratic Integral (LQI) control, is used to tune the CW_{min} value to drive the station attempt probability to the optimum so as to maintain a fair throughput allocation. We have demonstrated in simulations that the proposed control approach is adaptive to general network scenarios with high accuracy and fast convergence speed. To our knowledge this might be the first time that a closed-loop control system is designed for EDCA WLANs to achieve proportional fairness amongst ACs.

The remainder of this paper is organised as follows. red-Section II gives a comprehensive review on the state-of-the-art research on fairness and control theory approaches to solve network problems. Section III presents the theoretical analysis to derive the optimal station attempt probability which leads to proportional fair allocation of station throughputs given the constraints on the average delay deadlines for different ACs. Section IV describes a centralised adaptive control approach which can realise the proportional fair allocation derived from Section III in real networks. Section V evaluates the performances of the fairness algorithm and the proposed centralised control approach, and Section VI concludes the paper.

II. STATE OF THE ART

A. Fairness

Fairness has been the subject of a considerable body of literature on 802.11 WLANs [2]-[19]. The unfairness behaviors may be caused by a number of factors, e.g. hidden terminals, exposed terminals, capture, uplink/downlink unfairness, asymmetric radio conditions and multiple data rates and etc.,

* corresponding author S.H.Yang@lboro.ac.uk

which have been investigated in [2], [4], [5], [6], [8], [7], [11], [13]. There also exist distinct fairness criteria that are widely adopted in network resource allocation, such as time-based fairness, throughput-based fairness, proportional fairness, max-min fairness, weighted fairness and etc. (see for example [11], [18], [17], [19], [10], [9], [13], [14] and references therein). In this paper, we employ the proportional fairness criterion to deal with the unfairness amongst ACs with differentiated priority parameters.

The CSMA/CA scheduling used in 802.11 differs fundamentally from wired networks due to carrier sense deferral of the contention window countdown and the occurrence of colliding transmissions, both of which act to couple together the scheduling of station transmissions and lead to the rate region being nonconvex [20]. Therefore, well established utility fairness techniques from wired and TDMA networks cannot be directly applied to random access CSMA/CA wireless networks, and hence most studies on proportional fairness in 802.11 WLANs are confined to approximation approaches [2], [3], [13]. [18] corrects prior approximate studies and provides the first rigorous analysis of proportional fairness in 802.11 WLANs. It shows that there exists a unique proportional fair rate allocation and completely characterises the allocation in terms of the total air-time quantity. [17] extends the work in [18] by considering lossy links and BSC-based coding with delay deadline constraints. The optimal joint allocation of air-time and coding rate allows the throughput/loss/delay tradeoff amongst flows sharing network resources to be performed in a principled manner. The proportional fairness analysis in this paper builds upon the approaches used in [17] and [21].

In particular, the fairness issue in 802.11e EDCA WLANs has been given considerable attention. [11] derives a throughput allocation based on the proportional fair criterion in multirate 802.11e WLANs. It shows that in a proportional fair allocation high and low bit rate stations are assigned with the same share of channel time, and thus high bit rate stations obtain higher throughput. Two schemes that respectively involve Contention Window CW and Transmission Length TL (which is based on $TXOP$) are then proposed to achieve this allocation. [13] investigates the weighted proportional fairness in both single-rate and multi-rate 802.11e WLANs via test-bed experiments, and compares proportional fairness with time-based fairness in a multi-rate setting. Proportional fairness can be achieved by tailoring the CW_{min} , and the time-based fairness can be achieved by adjusting packet size or $TXOP$ limit. It concludes that in a multi-rate 802.11e WLAN proportional fairness with equal weights achieves higher performance than time-based fairness in terms of both aggregate utility and throughput. [11] and [13] deal with the unfairness behaviour arising in a WLAN due to asymmetric channel conditions, and [13] investigates the problem through test-bed evaluations relying on no theoretical analysis foundation. [12], [14], [16] address the unfairness problem existing amongst ACs. The priority-based service supported by 802.11e EDCA, while allowing differentiated service for flows of different priorities, cannot ensure service amount in proportion to their demands. [12] proposes a mechanism called Weighted Fair-EDCA (WF-EDCA) which uses Distributed Fair Scheduling

(DFS) in each backoff entity to provide weighted proportional fair service among different ACs. [14] proposes an algorithm to compute the optimal configuration of the EDCA differentiation parameters given a set of QoS requirements in terms of throughput and delay with multiple real-time and data ACs. A throughput and delay analysis is provided, based upon which the optimal configuration algorithm is derived by maximising the throughput using the weighted max-min fairness criterion. red[16] presents a new scheme which exploits differentiations of both inter frame space and contention window to achieve weighted fairness for two classes of services under EDCA mode in an 802.11e WLAN. Given the AIFSSs, the proposed scheme can properly set the corresponding CWs such that the ratio of the two classes' successful transmission probabilities can attain a pre-defined weighted-fairness goal. [15] proposes a dynamic contention window control scheme to achieve fairness between uplink and downlink TCP flows for the IEEE 802.11e EDCA-based WLANs while guaranteeing QoS requirements for real-time traffic. The proposed scheme first determines the minimum contention window size in the best-effort access category at APs, and then determines the minimum and maximum contention window sizes in higher priority access categories, such as voice and video, so as to guarantee QoS requirements for these real-time traffic. In this paper, we provide a comprehensive throughput and delay analysis that incorporates EDCA differentiation parameters for 802.11e EDCA WLANs and derive a throughput allocation to achieve proportional fairness amongst ACs.

B. Control theory approach

Control theory has been applied to the area of communication networks in a wide range of aspects. For instance, [?] analyses a combined TCP and Active Queue Management (AQM) model from a control theoretic standpoint. It uses a nonlinear dynamic model of TCP to design a feedback control system depiction of AQM using the random early detection (RED) scheme. [23] introduces a control theoretical analysis of the closed-loop congestion control problem in packet networks. The control theoretical approach is used in a proportional rate controller, where packets are admitted into the network in accordance with network buffer occupancy. A Smith Predictor is used to deal with large propagation delays. [24] proposes a QoS-provisioning feedback control framework in order to achieve TCP uplink/downlink fairness and service differentiation. The medium access price (MAP) is delivered to TCP senders and the TCP senders adjust their sending rates to reduce congestion at the interface queue of the home gateway in an 802.11-based home network. [25] proposes a centralised adaptive control (CAC) approach to dynamically adjust the CW_{min} configuration of 802.11 WLANs with the goal of minimising the overall throughput performance. A proportional integrator (PI) controller is used to establish a closed-loop control system. [26] extends the work in [25] by considering real-time traffic in 802.11e WLANs. The CW_{min} configuration is adjusted in terms of minimising the average delay, which results in a better quality of experience (QoE) of the video traffic. [27] proposes a distributed adaptive control

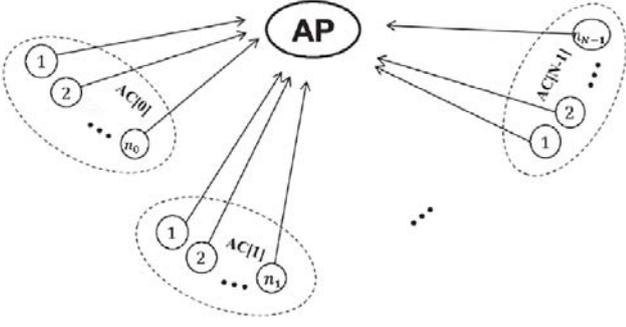


Fig. 1. Network model.

(DAC) algorithm based on the multivariable control theory. An independent PI controller is installed at each station in a WLAN and uses locally available information to drive the overall network performance to the optimum. In addition to the closed-loop control methods, [29] proposes an open-loop self-adaptive rate control approach for multi-priority WLANs. The approach is based on a biological competitive model which guides data flows to compete for network bandwidth in the way of a native ecosystem. The model parameters self-tune themselves to optimise the bandwidth utilization in an EDCA WLAN with multiple ACs. In this paper we employ the multivariable feedback control method to design a centralised adaptive control approach to achieve proportional fairness amongst different ACs. The WLAN is represented as a discrete MIMO LTI state-space model. The state feedback LQI controller is used to tune the CW_{min} value. To the best of our knowledge, this might be the first time that a closed-loop control system is used in 802.11 WLANs to deal with fairness issue.

III. PROPORTIONAL FAIRNESS IN MULTI-PRIORITY 802.11E WLANS

A. Network model

We consider a single-hop 802.11e EDCA WLAN with one AP and n client stations, as depicted in Fig. 1. The channel is assumed to be error-free for all supported PHY rates. Traffic flows are classified into N different ACs. We assume that each client carries flows of a single AC, so there are no virtual collisions in our setup. The number of stations in the i th AC is n_i . The total number of stations is thus $n = \sum_{i=0}^{N-1} n_i$. The analysis can be readily generalised to encompass situations where client stations have lossy links and carry more than one AC. But the current simplified error-free model is sufficient to capture performance features of differentiation settings in WLANs.

B. Station throughput

We start with the analysis of station throughput under saturation conditions. For the i th AC, the following parameters are defined: CW_{min}^i is the minimum contention window; CW_{max}^i is the maximum contention window; t_i is the number of time slots in the period of $AIFS_i$, i.e. $AIFS_i = SIFS + t_i \times \sigma$

TABLE I
NOTATION

N	Number of ACs
n_i	Number of STAs in AC i
n	Total number of STAs
CW_{min}^i, W_i	Minimum contention window of AC i
CW_{max}^i	Maximum contention window of AC i
$AIFS_i$	Duration of AIFS of AC i
$SIFS$	Duration of SIFS
$EIFS$	Duration of EIFS
T_{RTS}	Duration of RTS
T_{CTS}	Duration of CTS
σ	Duration of a physical time slot
t_i	Number of time slots in $AIFS_i$
t_{min}	Minimum t value among all ACs
m_i	Number of packets in the TXOP burst of AC i
L	Packet size
r	PHY data rate
τ_i	Station attempt probability of AC i
P_{idle}	Slot idle probability
P_{succ}^i	Probability of a successful transmission by a station of AC i
P_{succ}	Probability of a successful transmission in a time slot
s_i	Station throughput of AC i
T_{col}	Duration of a collision
T_i^{succ}	Duration of a successful transmission from a station of AC i
T_i^o	Protocol overheads of a TXOP burst of AC i
T^o	Protocol overheads of a single packet within TXOP burst
M	Retry limit in exponential backoff algorithm
P_{col}^i	Conditional collision probability of AC i
P_{blk}^i	Blocking probability of AC i
D_i^{cd}	Expected countdown delay of AC i
D_i^{blk}	Expected blocking delay of AC i
D_i^{retx}	Expected retransmission delay of AC i
D_i^{succ}	Expected retransmission delay of AC i
D_i	Average delay of a TXOP burst of AC i
d_i	Delay deadline of a single packet of AC i
H	Plant linearisation matrix
$x(k)$	System state at instant k
$y(k)$	System output at instant k
$u(k)$	System input at instant k
$r(k)$	Controller input at instant k
K	LQI optimal gain matrix
Q	LQI state cost weighting matrix
R	LQI control cost weighting matrix

where $SIFS$ is the duration of the Short Interframe Space and σ represents the duration of a physical time slot; m_i is the number of packets transmitted in a TXOP burst. We assume that packets of all ACs have the same length of L bits, and are transmitted at the same PHY rate r Mbps under the assumption of error-free channels. Due to the use of TXOP bursting, the RTS/CTS exchange mechanism is used to make fast recovery from collisions. The rednotation used in this paper are listed in Table. I.

A MAC time slot may either be a PHY idle slot, a successful transmission or a colliding transmission. Let τ_i denote the probability that a station carrying a flow of AC i attempts to transmit in a time slot, so $0 < \tau_i < 1$. The probability that a time slot is idle is

$$P_{idle} = \prod_{i=0}^{N-1} (1 - \tau_i)^{n_i}$$

As the channel is assumed to be error-free, packet losses are only caused by collisions. The probability that a station with

a flow of AC i makes a successful transmission is then

$$P_i^{succ} = \tau_i (1 - \tau_i)^{n_i - 1} \prod_{j=0, j \neq i}^{N-1} (1 - \tau_j)^{n_j} = \frac{\tau_i}{1 - \tau_i} P^{idle}$$

The probability that a time slot is a successful transmission is then

$$P^{succ} = \sum_{i=0}^{N-1} n_i P_i^{succ} = P^{idle} \sum_{i=0}^{N-1} \frac{n_i \tau_i}{1 - \tau_i}$$

The throughput of a station carrying a flow of AC i is thus given by

$$s_i(\tau) = \frac{P_i^{succ} m_i L}{P^{idle} \sigma + \sum_{i=0}^{N-1} n_i P_i^{succ} T_i^{succ} + (1 - P^{idle} - P^{succ}) T^{col}}$$

in which $T^{col} = T_{RTS} + EIFS$ is the duration of a collision. $EIFS$ represents the duration of the Extended Interframe Space used in 802.11 WLANs, redwhich is given by $EIFS = T_{ACK} + SIFS + DIFS$. Note that T^{col} is the same for all ACs due to the use of RTS/CTS handshaking. T_i^{succ} is the duration of a successful transmission from a station sending traffic of AC i . It depends on the size of TXOP packet burst, and is thus given by

$$T_i^{succ} = T_i^o + m_i (T^{oo} + \frac{L}{r})$$

where $T_i^o = T_{RTS} + SIFS + T_{CTS} + AIFS_i$ is the protocol overheads associated with the transmission of a TXOP burst; T^{oo} is the protocol overheads associated with each packet transmission within a TXOP burst, i.e. $T^{oo} = T_{PHY_{hdr}} + 2SIFS + T_{ACK}$.

By working in terms of the quantity $\alpha_i = \frac{\tau_i}{1 - \tau_i}$, $\alpha_i > 0$ instead of τ_i , the station throughput is rewritten as

$$s_i = \frac{\alpha_i m_i L}{X \cdot T^{col}} \quad (1)$$

in which

$$X = \frac{\sigma}{T^{col}} + \sum_{i=0}^{N-1} n_i \left(\frac{T_i^{succ}}{T^{col}} - 1 \right) \alpha_i + \prod_{i=0}^{N-1} (1 + \alpha_i)^{n_i} - 1$$

C. Average delay

Next, we will calculate the average delay experienced by a TXOP burst of each AC. redWe start with the analysis for ordinary 802.11 MAC scheduling with binary exponential backoff algorithm and then move onto the scenario when $CW_{min} = CW_{max}$.

The delay is defined in this work as the duration since a station starts contending for the medium until the transmission is finished (either received successfully or dropped because of reaching the maximum retry limit). The calculation is based on the EDCA WLAN throughput model derived in [28]. The average delay consists of four expected delays, described as follows:

- *Expected countdown delay*: For each backoff stage j ($0 \leq j \leq M$, and M is the retry limit. We assume that $CW_{max}^i \geq 2^M CW_{min}^i$), the average countdown delay for AC i is $CW_{i,j} \sigma / 2$, in which $CW_{i,j} = 2^j CW_{min}^i$

is the contention window at the j th backoff stage. The expected delay associated with the backoff countdown process is then given by

$$D_i^{cd} = \sigma \times \left(\sum_{j=0}^M (P_i^{col})^j (1 - P_i^{col}) \sum_{h=0}^j \frac{CW_{i,h}}{2} + (P_i^{col})^{M+1} \sum_{h=0}^M \frac{CW_{i,h}}{2} \right)$$

where P_i^{col} is the conditional collision probability for the i th AC in the throughput model, i.e.

$$P_i^{col} = 1 - (1 - \tau_i)^{n_i - 1} \prod_{\substack{j=0 \\ j \neq i}}^{N-1} (1 - \tau_j)^{n_j} \quad (2)$$

- *Expected blocking delay*: During the countdown process, when a transmission is detected on the channel the backoff time counter is ‘‘frozen’’, and reactivated again after the channel is sensed idle for a certain period. A station is called blocked in our analysis when it senses (an) ongoing transmission(s) from some other station(s) during its countdown process. The blocking delay is the period during which a station is ‘‘frozen’’. The expected number of time slots in the backoff countdown process is D_i^{cd} / σ . At each time slot, a station could be blocked by either a successful transmission or a collision. For a station of AC i , the delay caused by a successful transmission from some other station is

$$D_i^{bs} = T_i^{succ} (n_i - 1) \tau_i (1 - \tau_i)^{n_i - 2} \prod_{\substack{j=0 \\ j \neq i}}^{N-1} (1 - \tau_j)^{n_j} + \sum_{\substack{j=0 \\ j \neq i}}^{N-1} T_j^{succ} n_j \tau_j (1 - \tau_j)^{n_j - 1} \prod_{\substack{k=0 \\ k \neq j, i}}^{N-1} (1 - \tau_k)^{n_k} (1 - \tau_i)^{n_i - 1}$$

The blocking delay because of a collision is

$$D_i^{bc} = T^{col} \left(1 - (1 - \tau_i)^{n_i - 1} \prod_{\substack{j=0 \\ j \neq i}}^{N-1} (1 - \tau_j)^{n_j} - (n_i - 1) \tau_i (1 - \tau_i)^{n_i - 2} \prod_{\substack{j=0 \\ j \neq i}}^{N-1} (1 - \tau_j)^{n_j} - \sum_{\substack{j=0 \\ j \neq i}}^{N-1} n_j \tau_j (1 - \tau_j)^{n_j - 1} \prod_{\substack{k=0 \\ k \neq j, i}}^{N-1} (1 - \tau_k)^{n_k} (1 - \tau_i)^{n_i - 1} \right)$$

The expected blocking delay of AC i is thus

$$D_i^{blk} = \frac{D_i^{cd}}{\sigma} (D_i^{bs} + D_i^{bc})$$

- *Expected retransmission delay*: The expected retransmission delay for AC i is calculated by multiplying the expected number of retransmission attempts by the

collision duration, i.e.

$$D_i^{retx} = T^{col} \times \left(\sum_{j=0}^M j(P_i^{col})^j(1 - P_i^{col}) + (M+1)(P_i^{col})^{M+1} \right)$$

- *Expected successful transmission delay*: The expected successful transmission delay is the duration of a successful transmission multiplied by the probability that the transmission is not dropped, which is given by

$$D_i^{succ} = T_i^{succ}(1 - (P_i^{col})^{M+1})$$

Combining the above four delays, the average delay of a TXOP burst of AC i is therefore given by

$$D_i = D_i^{cd} + D_i^{blk} + D_i^{retx} + D_i^{succ}$$

The proposed approach in this paper works by finding the optimal contention window to achieve proportional fairness amongst ACs, so the exponential backoff algorithm is unnecessary in our setting and we simply set $CW_{max}^i = CW_{min}^i$, i.e. $M = 0$. To simplify notations, we hereafter refer to CW_{min}^i with W_i . The four expected delays then become:

$$D_i^{cd} = \sigma \frac{W_i}{2}$$

$$D_i^{blk} = \frac{W_i}{2}(D_i^{bs} + D_i^{bc})$$

$$D_i^{retx} = T^{col} P_i^{col}$$

$$D_i^{succ} = T_i^{succ}(1 - P_i^{col})$$

According to the throughput model in [28], when $CW_{max} = CW_{min}$ the station attempt probability under saturation conditions can be reduced to

$$\tau_i = \frac{2(1 - P_i^{blk})}{2(1 - P_i^{blk}) + W_i - 1} \quad (3)$$

in which

$$P_i^{blk} = 1 - \left[(1 - \tau_i)^{n_i - 1} \prod_{\substack{j=0 \\ j \neq i}}^{N-1} (1 - \tau_j)^{n_j} \right]^{t_i - t_{min} + 1}$$

is the probability that the backoff counter is suspended due to a busy channel during the period of $AIFS_i$. t_{min} is the minimum t value among all ACs.

Similarly by working in terms of the quantity $\alpha_i = \frac{\tau_i}{1 - \tau_i}$, the average delay experienced by a TXOP burst of AC i when $M = 0$ is then

$$D_i = \frac{W_i(\sigma + T^{col})}{2} + \frac{Y_i}{(1 + \alpha_i)^{n_i - 1}} (T_i^{succ} - T^{col}) + T^{col} + \frac{W_i Y_i}{2(1 + \alpha_i)^{n_i - 1}} (Z_i - T^{col} + (T_i^{succ} - T^{col})(n_i - 1)\alpha_i) \quad (4)$$

in which

$$Y_i = \sum_{j=0, j \neq i}^{N-1} (1 + \alpha_j)^{-n_j}$$

$$Z_i = \sum_{j=0, j \neq i}^{N-1} (T_j^{succ} - T^{col}) n_j \alpha_j$$

and

$$W_i = \frac{2}{\alpha_i} \left((1 + \alpha_i) \prod_{j=0}^{N-1} (1 + \alpha_j)^{-n_j} \right)^{t_i - t_{min} + 1} + 1$$

D. Proportional fair allocation

The 802.11e EDCA standard provides service differentiation by assigning different contention parameters to distinct ACs. Delay-sensitive traffic flows, such as voice over WLANs and streaming multimedia, are assigned with higher priorities. This mechanism has a significant cost for lower priority traffic flows as they can practically starve in dense network deployment. In this section we aim at finding the optimal $\alpha := [\alpha_i]_{i \in \{0, 1, \dots, N-1\}}$ to achieve fair allocation of station throughputs amongst ACs. Meanwhile we take into account the delay constraints for each AC. The utility function is defined as the sum of the log of station throughputs

$$\begin{aligned} \max_{\alpha} \quad & U(\alpha) := \sum_{i=0}^{N-1} n_i \log s_i(\alpha) \\ \text{s. t.} \quad & D_i(\alpha) \leq m_i d_i \quad 0 \leq i \leq N-1, \\ & \alpha_i > 0 \quad 0 \leq i \leq N-1. \end{aligned}$$

in which the station throughput is given by Eqn. (1); the average delay is given by Eqn. (4); d_i is the delay deadline for a single packet in a TXOP burst of AC i .

By plugging in the station throughput expression and removing the constant terms, the optimisation problem is simplified as

$$\begin{aligned} \max_{\alpha} \quad & U'(\alpha) := \sum_{i=0}^{N-1} n_i (\log \alpha_i - \log X) \\ \text{s. t.} \quad & D_i(\alpha) \leq m_i d_i \quad 0 \leq i \leq N-1, \\ & \alpha_i > 0 \quad 0 \leq i \leq N-1. \end{aligned}$$

1) *Non-convexity*: It can be verified by inspection of the second derivative that the objective function is not concave in α and hence the maximisation problem is not a standard convex optimisation task. We proceed by making the log transformation $\eta_i = \log \alpha_i$. The optimisation problem then becomes

$$\begin{aligned} \max_{\eta} \quad & U_1(\eta) := U'(e^{\eta}) \\ \text{s. t.} \quad & D_i(e^{\eta}) \leq m_i d_i \quad 0 \leq i \leq N-1. \end{aligned} \quad (5)$$

Lemma 1. $U_1(\eta)$ is concave in η .

The proof of Lemma 1 is included in the Appendix.

2) *Solving the optimisation with KKT conditions*: To solve this problem, we will use KarushKuhnTucker (KKT) conditions. The Lagrangian is

$$L = U_1(\eta) - \sum_{i=0}^{N-1} \mu_i (D_i(e^{\eta}) - m_i d_i)$$

where the Lagrange multiplier $\mu_i \geq 0$. Differentiating the Lagrangian with respect to η_i and setting it equal to zero yields

$$f_i(\boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{n}{X} \frac{\partial X}{\partial \eta_i} + \mu_i \frac{\partial D_i}{\partial \eta_i} + \sum_{\substack{j=0 \\ j \neq i}}^{N-1} \mu_j \frac{\partial D_j}{\partial \eta_i} - n_i = 0 \quad (6)$$

$$i \in \{0, 1, \dots, N-1\}$$

in which

$$\begin{aligned} \frac{\partial X}{\partial \eta_i} &= \left(\frac{T_i^{succ}}{T^{col}} - 1 \right) n_i \alpha_i + n_i \prod_{j=0}^{N-1} (1 + \alpha_j)^{n_j} \frac{\alpha_i}{1 + \alpha_i} \\ \frac{\partial D_i}{\partial \eta_i} &= \left(\frac{Y_i \alpha_i ((T_i^{succ} - T^{col})(n_i - 1) \alpha_i + Z_i - T^{col})}{2(1 + \alpha_i)^{n_i - 1}} \right. \\ &\quad \left. + \frac{(\sigma + T^{col}) \alpha_i}{2} \right) \cdot \frac{\partial W_i}{\partial \alpha_i} + \frac{Y_i W_i \alpha_i}{2} (1 + \alpha_i)^{-n_i} (n_i - 1) \\ &\quad \left((T_i^{succ} - T^{col})((2 - n_i) \alpha_i + 1) - Z_i + T^{col} \right) \\ &\quad + Y_i \alpha_i (T_i^{succ} - T^{col}) (1 - n_i) (1 + \alpha_i)^{-n_i} \\ \frac{\partial D_j}{\partial \eta_i} &= \left(\frac{Y_j \alpha_i ((T_j^{succ} - T^{col})(n_j - 1) \alpha_j + Z_j - T^{col})}{2(1 + \alpha_j)^{n_j - 1}} \right. \\ &\quad \left. + \frac{(\sigma + T^{col}) \alpha_i}{2} \right) \cdot \frac{\partial W_j}{\partial \alpha_i} + \alpha_i (1 + \alpha_j)^{1 - n_j} \left((T_j^{succ} - T^{col}) \right. \\ &\quad \left. \left(\frac{W_j}{2} (n_j - 1) \alpha_j + 1 \right) + \frac{W_j}{2} (Z_j - T^{col}) \right) \frac{\partial Y_j}{\partial \alpha_i} \\ &\quad + \frac{Y_j W_j \alpha_i}{2} (1 + \alpha_j)^{1 - n_j} \frac{\partial Z_j}{\partial \alpha_i} \end{aligned}$$

The derivatives $\frac{\partial W_i}{\partial \alpha_i}$, $\frac{\partial W_j}{\partial \alpha_i}$, $\frac{\partial Y_j}{\partial \alpha_i}$ and $\frac{\partial Z_j}{\partial \alpha_i}$ are respectively given by

$$\begin{aligned} \frac{\partial W_i}{\partial \alpha_i} &= -\frac{2}{\alpha_i} \left((1 + \alpha_i) \prod_{j=0}^{N-1} (1 + \alpha_j)^{-n_j} \right)^{t_i - t_{min} + 1} \\ &\quad \left(\frac{1}{\alpha_i} + \frac{n_i - 1}{1 + \alpha_i} (t_i - t_{min} + 1) \right) \\ \frac{\partial W_j}{\partial \alpha_i} &= -\frac{2n_i}{\alpha_j (1 + \alpha_i)} (t_j - t_{min} + 1) \left((1 + \alpha_j) \right. \\ &\quad \left. \prod_{k=0}^{N-1} (1 + \alpha_k)^{-n_k} \right)^{t_j - t_{min} + 1} \\ \frac{\partial Y_j}{\partial \alpha_i} &= -n_i (1 + \alpha_i)^{-(n_i + 1)} \\ \frac{\partial Z_j}{\partial \alpha_i} &= (T_i^{succ} - T^{col}) n_i \end{aligned}$$

3) *Subgradient algorithm for optimal $\boldsymbol{\alpha}$* : Given the values of the Lagrange multipliers $\boldsymbol{\mu}^*$, the solution to Eqn. (6) specifies the optimal $\boldsymbol{\alpha}$. To complete the solution to the optimisation it therefore remains to calculate the optimal multipliers $\boldsymbol{\mu}^*$. These cannot be obtained in closed form since their values reflect the network topology. We proceed in a centralised manner by using a standard sub-gradient approach. The dual problem for the primal problem defined in Eqn. (5) is given by

$$\min_{\boldsymbol{\mu} \geq 0} g(\boldsymbol{\mu})$$

where the dual function $g(\boldsymbol{\mu})$ is given by

$$\begin{aligned} g(\boldsymbol{\mu}) &= \max_{\boldsymbol{\alpha}} U'(\boldsymbol{\alpha}) + \sum_{i=0}^{N-1} \mu_i (m_i d_i - D_i(\boldsymbol{\alpha})) \\ &= U'(\boldsymbol{\alpha}^*(\boldsymbol{\mu})) + \sum_{i=0}^{N-1} \mu_i (m_i d_i - D_i(\boldsymbol{\alpha}^*(\boldsymbol{\mu}))) \end{aligned}$$

For any $\boldsymbol{\alpha}$,

$$g(\boldsymbol{\mu}) \geq U'(\boldsymbol{\alpha}) + \sum_{i=0}^{N-1} \mu_i (m_i d_i - D_i(\boldsymbol{\alpha}))$$

and in particular, the dual function is larger than that for $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*(\bar{\boldsymbol{\mu}})$, i.e.

$$\begin{aligned} g(\boldsymbol{\mu}) &\geq U'(\boldsymbol{\alpha}^*(\bar{\boldsymbol{\mu}})) + \sum_{i=0}^{N-1} \mu_i (m_i d_i - D_i(\boldsymbol{\alpha}^*(\bar{\boldsymbol{\mu}}))) \\ &= g(\bar{\boldsymbol{\mu}}) + \sum_{i=0}^{N-1} (\mu_i - \bar{\mu}_i) (m_i d_i - D_i(\boldsymbol{\alpha}^*(\bar{\boldsymbol{\mu}}))) \end{aligned}$$

A sub-gradient of $g(\cdot)$ at any $\bar{\boldsymbol{\mu}}$ is thus given by the vector

$$\left[m_i d_i - D_i(\boldsymbol{\alpha}^*(\bar{\boldsymbol{\mu}})) \right]_{i \in \{0, 1, \dots, N-1\}}$$

and the projected sub-gradient descent update is

$$\mu_i^{(t+1)} = \left[\mu_i^{(t)} - \gamma \cdot (m_i d_i - D_i(\boldsymbol{\alpha}^*(\boldsymbol{\mu}^{(t)}))) \right]^+$$

where $\gamma > 0$ is a sufficiently small stepsize, and $[f(\cdot)]^+ := \max\{f(\cdot), 0\}$ ensures that the Lagrange multiplier never goes negative [31].

The subgradient updates for $\boldsymbol{\mu}$ can be carried out centrally by the AP. For the i th AC, the AP requires the knowledge of the number of stations n_i , the PHY data rate r and the packet size L . The algorithm to calculate optimal $\boldsymbol{\alpha}$ is detailed in Algorithm 1.

Algorithm 1 Calculate optimal $\boldsymbol{\alpha}$

- 1: Initialise $\boldsymbol{\mu}^{(1)} = [\mu_0^{(1)}, \mu_1^{(1)}, \dots, \mu_{N-1}^{(1)}]$, $t = 0$.
- 2: **repeat**
- 3: $t = t + 1$; $\gamma^{(t)} = 1/t^2$
- 4: solve for $\boldsymbol{\alpha}^*(\boldsymbol{\mu}^{(t)})$ by combining equations $f_i(\boldsymbol{\alpha}, \boldsymbol{\mu}^{(t)}) = 0 \quad \forall i \in \{0, 1, \dots, N-1\}$
- 5: $\forall i \in \{0, 1, \dots, N-1\}$, calculate

$$\mu_i^{(t+1)} = \left[\mu_i^{(t)} - \gamma^{(t)} \cdot (m_i d_i - D_i(\boldsymbol{\alpha}^*(\boldsymbol{\mu}^{(t)}))) \right]^+$$

- 6: **until**

- (i) $\exists i \in \{0, 1, \dots, N-1\}$, $|m_i d_i - D_i(\boldsymbol{\alpha}^*(\boldsymbol{\mu}^{(t)}))| \leq \epsilon$, where $\epsilon > 0$ and is sufficiently small;
 - (ii) $\forall i \in \{0, 1, \dots, N-1\}$, $m_i d_i - D_i(\boldsymbol{\alpha}^*(\boldsymbol{\mu}^{(t)})) \geq 0$.
-

IV. CENTRALISED CLOSED-LOOP CONTROL APPROACH

In this section, we design a centralized adaptive control approach to implement the desirable proportional fairness in real networks. Based upon the analysis in Section III, the proportional fairness is achieved when the station attempt

probability parameter α reaches its optimum value α^* . The variable α is only determined by the minimum contention window W_i with *AIFS* and *TXOP* taking the recommended values and $CW_{max} = CW_{min}$. Our approach uses a multivariable closed-loop control system to tune W to drive the station attempt probability to its optimum. As the station attempt probability is hard to measure in real networks, we measure the conditional collision probability $p^{col}(\tau)$ instead of $\alpha(\tau)$ in the proposed control approach.

red

A. Measuring p^{col}

The measuring of p^{col} is performed periodically at the AP every beacon interval. It requires messages passing from ordinary stations. A transmission from a station collides with a sign of a missing CTS after sending a RTS. Each station starts to count the numbers of transmitted RTS packets and missing CTS packets after receiving a beacon packet. The counted numbers continue being piggybacked to the AP along with following RTS packets. After a beacon interval, based on the piggybacked information, the AP gets to know the number of transmitted RTS packets from stations of AC i , denoted as N_i^{RTS} , and the number of missing CTS packets of AC i , denoted as N_i^{M-CTS} , within a beacon interval. If we ignore the transmission duration of the beacon packet, and assume that packets collide with a constant collision probability within each interval, the average number of collided packets of AC i is $N_i^{RTS}p_i^{col}$. These packets will be observed as missing CTSs, which yields $E(N_i^{M-CTS}) = (N_i^{RTS})p_i^{col}$. Therefore, we have

$$E\left(\frac{N_i^{M-CTS}}{N_i^{RTS}}\right) = \frac{(N_i^{RTS})p_i^{col}}{N_i^{RTS}} = p_i^{obs}$$

The observed conditional collision probability p_i^{obs} can be estimated as $\frac{N_i^{M-CTS}}{N_i^{RTS}}$. To simplify notations, we hereafter refer to p^{col} with p .

B. Control algorithm

The closed-loop control system consists of two modules as depicted in Fig. 2.

The controller module is installed at the AP and periodically carries out the adaptive algorithm every beacon interval, which is typically 100ms in 802.11 WLANs. redThe adaptive control algorithm is described in Algorithm 2:

The plant is the WLAN itself. The input of the plant is the contention window $W = [W_0, \dots, W_{N-1}]$, and the output is the observed conditional collision probability $p^{obs} = [p_0^{obs}, \dots, p_{N-1}^{obs}]$. The design objective is to obtain a stable system in closed-loop with desired performances and shape the output of the system to the given reference value. The reference value is the optimal conditional collision probability p^* given by Eqn. (2) for $\tau = \tau^*$.

C. Linearisation of the non-linear plant

As the proposed adaptive control algorithm is executed every beacon interval, the period is long enough to assume that

Algorithm 2 redAdaptive control algorithm

- 1: AP calculates the optimal α^* and the resulting p^* .
- 2: AP broadcasts the current W to all subscribed stations along with a beacon packet.
- 3: Each station starts to count the number of transmitted RTS packets and the number of missing CTS packets after receiving a beacon. The numbers are piggybacked to the AP along with following RTS packets.
- 4: At the end of a beacon interval (100ms), AP counts the total number of transmitted RTS packets of AC i , N_i^{RTS} , and the total number of missing CTS packets, N_i^{M-CTS} . The observed conditional collision probability p_i^{obs} in this interval is then calculated based on the counted numbers.
- 5: The reference p^* and the observed p^{obs} are input into the controller to calculate a new set of W .
- 6: If the output of the controller W is not an integer, it is rounded to the closest integer value, and it has to be at least larger than 1.
- 7: Go back to step 2.

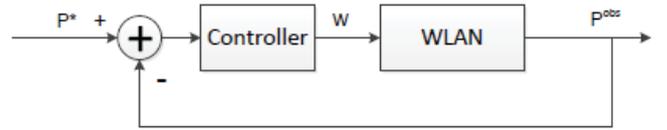


Fig. 2. Closed-loop control system

the measurement corresponds to stationary conditions. This implies that p^{obs} depends only on the current W , i.e. the system has no memory. Following this,

$$p_i^{obs} = 1 - (1 - \tau_i)^{n_i - 1} \prod_{\substack{j=0 \\ j \neq i}}^{N-1} (1 - \tau_j)^{n_j} \quad (7)$$

in which τ_i is a function of W_i , given by Eqn. (3).

Eqn. (7) and Eqn. (3) give a non-linear relationship between p^{obs} and W . In order to simplify the controller design, we proceed by working with a linear approximation to this non-linear relationship around the stable point of operation.

The perturbations of input around the stable point of operation is

$$W = W^* + \delta W$$

in which W^* is the W which yields the optimal value τ^* from Eqn. (3).

The perturbations suffered by p^{obs} can be approximated by

$$\delta p^{obs} = \delta W \cdot H$$

in which

$$H = \begin{pmatrix} \frac{\partial p_0^{obs}}{\partial W_0} & \frac{\partial p_1^{obs}}{\partial W_0} & \dots & \frac{\partial p_{N-1}^{obs}}{\partial W_0} \\ \frac{\partial p_0^{obs}}{\partial W_1} & \frac{\partial p_1^{obs}}{\partial W_1} & \dots & \frac{\partial p_{N-1}^{obs}}{\partial W_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_0^{obs}}{\partial W_{N-1}} & \frac{\partial p_1^{obs}}{\partial W_{N-1}} & \dots & \frac{\partial p_{N-1}^{obs}}{\partial W_{N-1}} \end{pmatrix}$$

The partial derivatives can be respectively calculated as

$$\frac{\partial p_i^{obs}}{\partial W_i} = \sum_{k=0}^{N-1} \frac{\partial p_i^{obs}}{\partial \tau_k} \cdot \frac{\partial \tau_k}{\partial W_i}$$

and

$$\frac{\partial p_i^{obs}}{\partial W_j} = \sum_{k=0}^{N-1} \frac{\partial p_i^{obs}}{\partial \tau_k} \cdot \frac{\partial \tau_k}{\partial W_j}$$

in which

$$\frac{\partial p_i^{obs}}{\partial \tau_i} = \prod_{k=0}^{N-1} (1 - \tau_k)^{n_k} \frac{n_i - 1}{(1 - \tau_i)^2}$$

$$\frac{\partial p_i^{obs}}{\partial \tau_j} = \prod_{k=0}^{N-1} (1 - \tau_k)^{n_k} \frac{n_j}{(1 - \tau_i)(1 - \tau_j)}$$

$$\frac{\partial \tau_i}{\partial W_i} = \frac{\tau_i^2}{-2(1 - P_i^{blk})(1 + (n_i - 1)(t_i - t_{min} + 1)\tau_i)}$$

$$\frac{\partial \tau_i}{\partial W_j} = \frac{\tau_j(1 - \tau_i)}{-2(1 - P_j^{blk})n_i(1 - \tau_j)(t_j - t_{min} + 1)}$$

At the stable point of operation, $\tau_i = \tau_i^*$, the non-linear plant is thus linearised as

$$\mathbf{p}^{obs} = \mathbf{W} \cdot \mathbf{H}(\boldsymbol{\tau}^*) - \mathbf{W}^* \cdot \mathbf{H}(\boldsymbol{\tau}^*) + \mathbf{p}^* \quad (8)$$

D. State feedback control

With the lumbarisation, the WLAN can be represented as a discrete MIMO LTI state-space model. According to the proposed adaptive algorithm, the conditional collision probability at instant $k + 1$ is determined by the contention window input to the WLAN at instant k , the state and measurement equations are therefore given by

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{B}\mathbf{u}(k) \\ \mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) \end{cases}$$

in which the system state is the conditional collision probability,

$$\mathbf{x}(k) = [\mathbf{p}^{obs}(k)]^T$$

the system input is the minimum contention window,

$$\mathbf{u}(k) = [\mathbf{W}(k)]^T$$

and the system model matrices are

$$\mathbf{B} = -\mathbf{H}^T$$

and

$$\mathbf{C} = \mathbf{I}_{N \times N}$$

The system output is thus

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) = [\mathbf{p}^{obs}(k)]^T$$

The control task can be accomplished by using the LQI control method [30] to design our controller. Fig. 3 shows the control block diagram for the system, in which $\mathbf{x}(k) \in \mathbb{R}^N$ is the system state, $\mathbf{y}(k) \in \mathbb{R}^N$ is the system output, $\mathbf{u}(k) \in \mathbb{R}^N$ is the controller output and $\mathbf{r}(k) \in \mathbb{R}^N$ is the controller input, which is the optimal collision probability $\mathbf{p}^*(k)$. $\mathbf{K} \in \mathbb{R}^{N \times 2N}$ is the control gain matrix, and $\mathbf{B} \in \mathbb{R}^{N \times N}$ and $\mathbf{C} \in \mathbb{R}^{N \times N}$

are the state-space system matrices. T_s is the sampling period of the system, i.e. the beacon interval 100ms.

The LQI controller computes an optimal state-feedback control law by minimising the quadratic cost function

$$J(\mathbf{u}(k)) = \sum_{k=0}^{\infty} (\mathbf{z}^T(k)\mathbf{Q}\mathbf{z}(k) + \mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k))$$

for any initial state $\mathbf{x}(0)$, in which $\mathbf{z}(k) = [\mathbf{x}(k); \mathbf{s}(k)]$ and $\mathbf{s}(k) = \mathbf{s}(k-1) + T_s \cdot (\mathbf{r}(k-1) - \mathbf{y}(k-1))$ is the output of a discrete integrator.

The matrices \mathbf{Q} and \mathbf{R} are the weighting matrices respectively indicating the state and control cost penalties. \mathbf{Q} and \mathbf{R} are required to be real symmetric and positive definite.

The state feedback control law is defined as

$$\mathbf{u}(k) = -\mathbf{K}\mathbf{z}(k)$$

The optimal state feedback gain matrix \mathbf{K} is computed by solving the associated discrete algebraic Riccati equation

$$\mathbf{P} = \hat{\mathbf{A}}^T(\mathbf{P} - \mathbf{P}\hat{\mathbf{B}}(\mathbf{R} + \hat{\mathbf{B}}^T\mathbf{P}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}^T\mathbf{P})\hat{\mathbf{A}} + \mathbf{Q}$$

in which

$$\hat{\mathbf{A}} = [\mathbf{0}_{N \times 2N}; \quad -\mathbf{C} * T_s \quad \mathbf{I}_{N \times N}]$$

and

$$\hat{\mathbf{B}} = [\mathbf{B}; \quad \mathbf{0}_{N \times N}]$$

\mathbf{K} is constructed from the solution of the above algebraic Riccati equation \mathbf{P}^* and weighting matrices \mathbf{Q} and \mathbf{R} , which is given by

$$\mathbf{K} = (\mathbf{R} + \hat{\mathbf{B}}^T\mathbf{P}^*\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}^T\mathbf{P}^*\hat{\mathbf{A}}$$

red

E. Selection of \mathbf{Q} and \mathbf{R}

The selection of weighting matrices \mathbf{Q} and \mathbf{R} affect the performance of the LQI controller. A simplified form using only 3 degrees of freedom is chosen in this work. The matrices are of the form

$$\mathbf{Q} = \begin{bmatrix} q_1 \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & q_2 \mathbf{I}_N \end{bmatrix}_{2N \times 2N} \quad (9)$$

and

$$\mathbf{R} = \rho \cdot \mathbf{I}_{N \times N} \quad (10)$$

in which $q_1 \in \mathbb{R}^+$ is the weight for the state feedback cost; $q_2 \in \mathbb{R}^+$ is the weight for the integral feedback cost and $\rho \in \mathbb{R}^+$ is the weight for the input cost.

V. PERFORMANCE EVALUATION

A. Throughput and delay performance

The main objective of this work is to achieve proportional fair allocation of station throughputs while satisfying specific delay constraints of different ACs. To verify if the proposed fairness algorithm meets this objective, we first evaluate the throughput allocation and delay performance. The results are obtained using Matlab based on the throughput and delay analysis in Section III-B, III-C and the proportional fairness

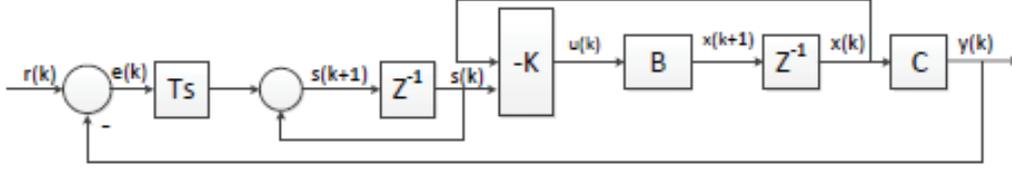


Fig. 3. LQI Controller

TABLE II
802.11 PROTOCOL PARAMETERS USED IN THE SIMULATIONS

σ	9 μ s	T_{PHYhdr}	20 μ s
<i>SIFS</i>	16 μ s	T_{RTS}	46.67 μ s
<i>DIFS</i>	34 μ s	T_{CTS}	38.67 μ s
<i>EIFS</i>	88.67 μ s	T_{ACK}	38.67 μ s

TABLE III
EDCA CHANNEL CONTENTION PARAMETERS FOR 802.11 OFDM PHY

Access categories	AIFSN	Max TXOP
Background (AC_BK)	7	0
Best Effort (AC_BE)	3	0
Video (AC_VI)	2	3.008ms
Voice (AC_VO)	2	1.504ms

algorithm described in Section III-D. As the throughput and delay analysis is based on the 802.11e performance model presented in [28] under the assumptions that stations have saturated traffic and $CW_{max} = CW_{min}$. The accuracy of this network model has been fully verified in [28] for different network scenarios. It is therefore fair enough to use numerical results to verify the correctness of the fairness optimisation algorithm proposed in this paper.

An example of an 802.11e WLAN with traffic of two ACs is considered, one of which is data traffic belonging to AC_BE (best effort), and the other is video traffic belonging to AC_VI (video). redFlows of both ACs are saturated. The 802.11 OFDM PHY layer is assumed to be used. The PHY data rates for two ACs are the same, i.e. $r = 54$ Mbps. The packet size is $L = 8000$ bits. The 802.11 protocol parameters used in the evaluation are listed in Table II. The EDCA contention parameters recommended for 802.11 OFDM PHY layer are listed in Table III. Note that the IEEE 802.11e standard also provides recommended values for CW_{min} and CW_{max} . As our control approach searches for the optimal contention window to achieve proportional fairness, the default CW_{min} and CW_{max} are not used in the proposed approach.

Fig. 4(a) and Fig. 4(b) shows the throughput and delay performance for two ACs versus the number of stations in AC_VI while keeping the number of stations in AC_BE fixed as 1. We let the TXOP burst reach the maximum limit as listed in Table III. The average delay deadline for a single video packet is $d_2 = 250\mu$ s, which is the successful transmission duration of a video packet. The delay deadline for a single data packet is four times that of a video packet, i.e. $d_1 = 1000\mu$ s. It can be seen that under these delay deadline constraints, the

resource allocation can be divided into four phases. redPhase I: When there are one video station and one data station, the network load is quite light, and both delay deadlines are not reached yet. Phase II: When the number of video stations increases to 2, the increased collision possibility leads to longer delays, and so the delay deadlines of both ACs are reached. As the number of video stations increases, in order to achieve a fair throughput allocation, data stations attempt more to access to the channel with an increased attempt probability, while video stations attempt less with a slightly decreasing attempt probability. The fairness algorithm makes the throughput of two ACs get closer to each other. Phase III: When the number of video stations increases up to 5, it comes to the turning point when video traffic is so aggressive that the proportional fairness algorithm allocates higher throughput to video traffic by assigning increased attempt probabilities to both ACs, but the increase for video traffic is larger than that for data traffic. The delay of video traffic is then reduced to be less than the deadline limit, while the delay constraint of data traffic remains tight. Phase IV: As the number of video stations continues increasing, the throughput ratio between video and data traffic remains around 1.5 in this phase. Even with 10 video stations and only one data station, the data station can still deliver a reasonable amount of throughput. Fig. 4(c) shows the corresponding station attempt probabilities. It can be seen that contrary to the 802.11e EDCA standard, our algorithm assigns data traffic with a higher attempt probability although it has lower priority. However not only does the delay performance satisfy the QoS requirement, the throughput is also fairly allocated between two ACs.

B. Air-time

The presence of collision losses and the coupling of station transmissions via carrier sense make the flow air-time in a WLAN not simply be the successful transmission duration but also include air-time expended in collisions. We define the *flow total air-time* as the fraction of time used for transmitting a flow, including both successful transmissions and collisions. For a flow of AC i , the flow total air-time is

$$\begin{aligned}
 T_i^{air} &= \frac{P_i^{succ} T_i^{succ} + \tau_i P_i^{col} T^{col}}{P_{idle} \sigma + \sum_{i=0}^{N-1} n_i P_i^{succ} T_i^{succ} + (1 - P_{idle} - P^{succ}) T^{col}} \\
 &= \frac{1}{X} \cdot \left(\alpha_i \left(\frac{T_i^{succ}}{T^{col}} - 1 \right) + \frac{\tau_i}{P_{idle}} \right)
 \end{aligned}$$

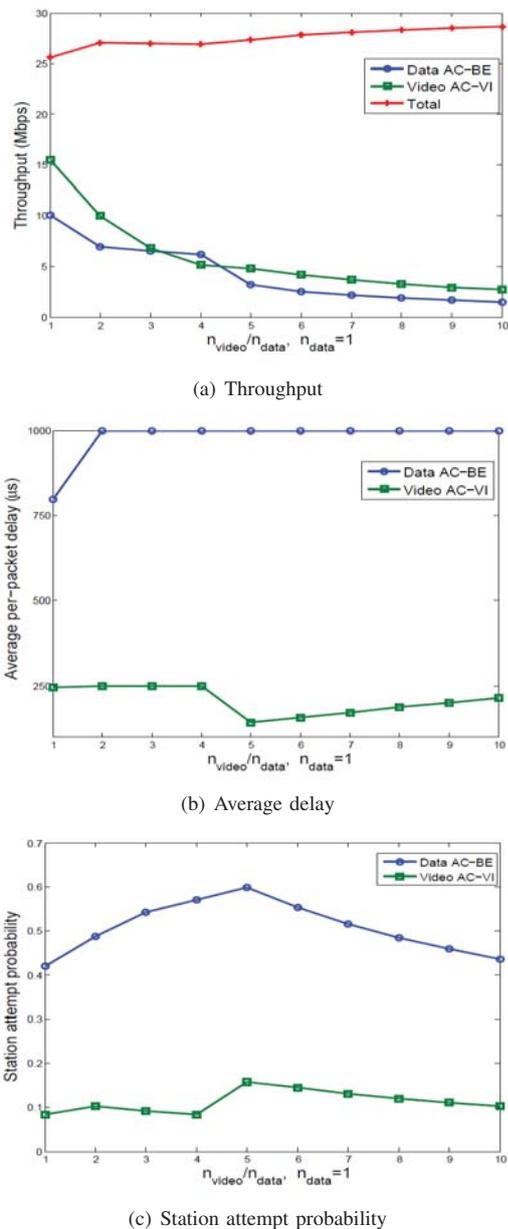


Fig. 4. Throughput and delay performances in an 802.11e WLAN with two ACs, the number of stations in AC_BE $n_1 = 1$, $r = 54\text{Mbps}$, $L = 8000\text{bits}$, $d_1 = 1000\mu\text{s}$ and $d_2 = 250\mu\text{s}$.

The work in [17], [18] finds that the proportional fair allocation assigns equal total air-time to each flow in a WLAN, and the air-times sum to unity. We investigate the air-time allocation by considering an 802.11e WLAN with four ACs, in which AC_VI and AC_VO have two stations, i.e. $n_2 = n_3 = 2$, and AC_BE and AC_BK have one station, i.e. $n_1 = n_4 = 1$. The PHY rates for four ACs are the same, i.e. $r = 54\text{Mbps}$. The packet size is $L = 8000$ bits. Table IV compares the flow total air-time allocations under different delay deadline constraints. In Case I, the average per-packet delay deadline for the four ACs are respectively $d_1 = 900\mu\text{s}$, $d_2 = 300\mu\text{s}$, $d_3 = 250\mu\text{s}$ and $d_4 = 1800\mu\text{s}$, while in Case II, the average delay deadlines are relaxed as

TABLE IV
COMPARISON OF FLOW TOTAL AIR-TIME ALLOCATIONS UNDER DIFFERENT DELAY DEADLINE CONSTRAINTS.

Flow	1	red2 and 3	red4 and 5	6	Sum
AC	BE	VI	VO	BK	
Case I	0.1565	0.1530	0.1550	0.1562	0.9287
Case II	0.1667	0.1667	0.1667	0.1667	1

$$d_1 = d_2 = d_3 = d_4 = 5000\mu\text{s}.$$

It can be seen that different from observations in [17], [18], the flow total air-time is not equalised in Case I, and the sum of air-times is less than 1. We also notice that in this case the delay constraints for voice traffic Flow 2 and Flow 3 are tight. Nevertheless when the delay constraints are relaxed in Case II, none of the delay deadline constraints is tight. Flows are now allocated with equal total air-times, and the air-times sum to 1. It is thus found that the air-time allocation in our algorithm is affected by the imposed delay constraints. With tight delay constraints, the proportional fair allocation assigns flows with the exact amount of air-time needed by each of them. Air-time resource in the network is not completely occupied in such a case. However with loose delay constraints, flows can occupy all the available air-time resource, and air-time is evenly distributed amongst flows in a network as discovered in previous work. It is worth pointing out that since the flow air-time usage overlaps due to collisions, the flow total air-times summing to unity does not imply that the channel idle probability $P^{idle} = 0$.

red

C. Q and R tuning

The tuning of q_1 , q_2 and ρ can be performed using trial and error method. The influences of the three parameters are illustrated using an example with three ACs, AC_BE, AC_VI and AC_VO. The number of stations in AC_BE, AC_VI and AC_VO is respectively $n_1 = 1, n_2 = 2$ and $n_3 = 1$. Three ACs use the same PHY rate, i.e. $r_1 = r_2 = r_3 = 54\text{Mbps}$. The packet size is $l = 8000$ bits. The average packet delay limit for AC_BE, AC_VI and AC_VO is respectively $d_1 = 900\mu\text{s}$, $d_2 = 300\mu\text{s}$ and $d_3 = 250\mu\text{s}$. The TXOP burst reaches the maximum limit. Fig. 5 plots the system output response for AC_BE with different sets of q_1 , q_2 and ρ values. The reason we do not put the output responses of AC_VI and AC_VO in Fig. 5 is that we notice that three outputs have the same convergence speed with a fixed set of q_1 , q_2 and ρ values. The effects of the three parameters are outlined as follows:

- q_1 imposes the constraints to the state dynamics. It is directly related to the overshoot. A higher q_1 corresponds to a lower overshoot. As shown in Fig: 5(a), $q_1 = 700$ results in an overshoot, while $q_1 = 800$ corresponds to an undershoot.
- q_2 impacts on integral action dynamics and so on the system dynamics. As shown in Fig: 5(b), the higher it is, the smaller rising time will be, and the higher overshoot will be.
- ρ affects the dynamics of the controller input, and so on the system dynamics. It is related to the overshoot.

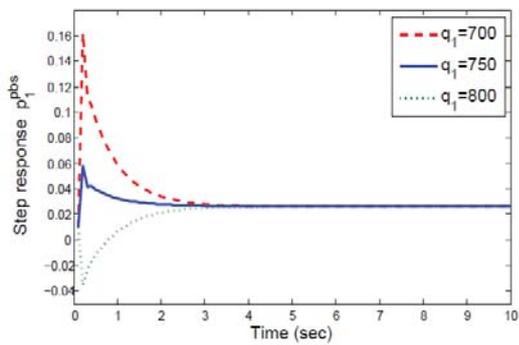
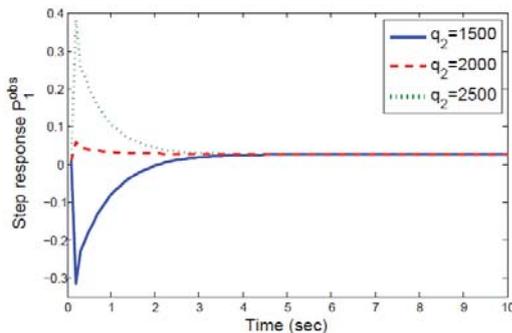
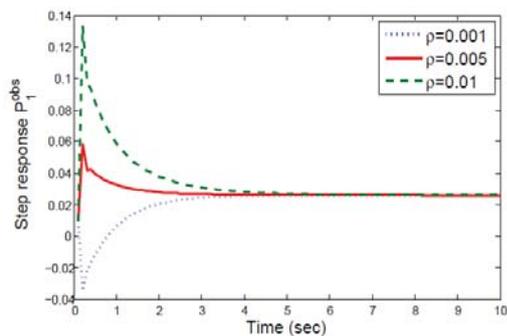
(a) $q_2 = 2000, \rho = 0.005$ (b) $q_1 = 750, \rho = 0.005$ (c) $q_1 = 750, q_2 = 2000$

Fig. 5. System output step response, $n_1 = 1, n_2 = 2, n_3 = 1, r = 54\text{Mbps}$, $L = 8000$ bits, $d_1 = 900\mu\text{s}, d_2 = 300\mu\text{s}, d_3 = 250\mu\text{s}$.

A higher ρ results in a higher overshoot, as shown in Fig: 5(c).

D. Adaptivity to changes in the WLAN

We will next evaluate the adaptivity of the proposed method to changes in the network size. The scenario being considered is depicted in Fig. 6. The algorithm starts at $t = 0\text{s}$ with four saturated stations, one in AC_BE, two in AC_VI and one in AC_VO. One more station in AC_VO joins the network at $t = 100\text{s}$ and leaves at $t = 200\text{s}$. At $t = 300\text{s}$ one AC_BK station joins the network, and after 100s one AC_VI station leaves. The PHY data rates for the three ACs are the same, i.e. $r = 54\text{Mbps}$. The packet size is $l = 8000$ bits. The average

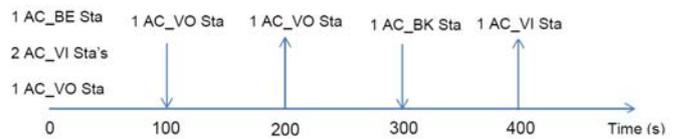


Fig. 6. Injection and/or removal of stations in the WLAN.

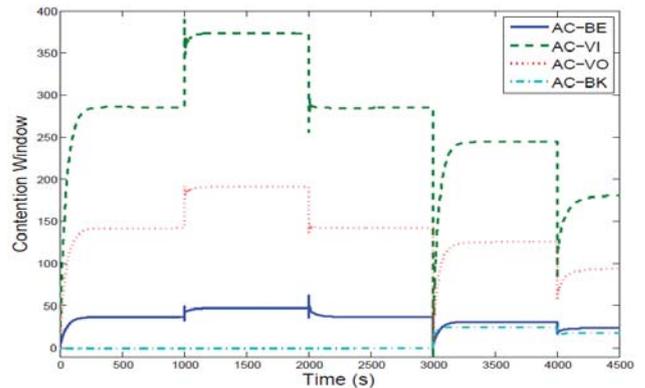


Fig. 7. Contention window

packet delay limit for data, video, voice and background traffic are respectively $d_1 = 900\mu\text{s}$, $d_2 = 300\mu\text{s}$, $d_3 = 250\mu\text{s}$ and $d_4 = 1800\mu\text{s}$. Fig. 7 plots the variation of contention windows over time. Fig. 8 plots the corresponding station throughput for each AC. red \mathbf{Q} and \mathbf{R} take the form as displayed in Eqn. (9) and (10). We choose $q_1 = 750$, $q_2 = 2000$ and $\rho = 0.005$ to make a fast convergence speed. It can be seen that when the network condition changes the contention window converges to the desirable value very quickly as long as proper \mathbf{Q} and \mathbf{R} are chosen. Moreover, the steady-state errors can be neglected, which means the control system has high accuracy performance.

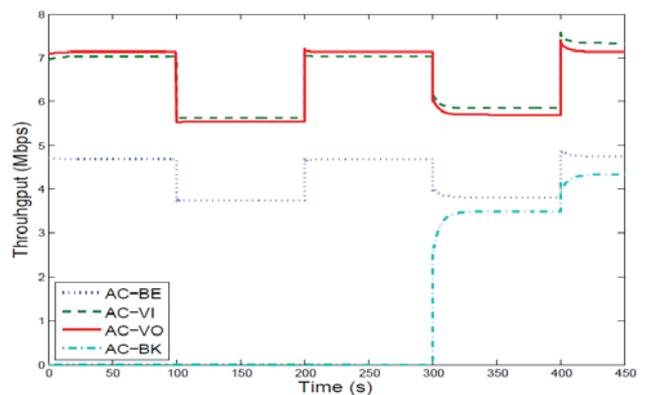


Fig. 8. Station throughput

VI. CONCLUSIONS

This paper considers using a closed-loop control approach to achieve proportional fair allocation of station throughputs in a multi-priority EDCA WLAN. The optimal station attempt probability that leads to proportional fairness is derived given the average delay deadline constraints of different ACs present in an WLAN. To achieve the desirable proportional fairness, a centralised adaptive control approach is proposed. The WLAN is represented as a discrete MIMO LTI state-space model. The LQI control is used to tune the CW_{min} value to the optimum. We have demonstrated using numerical results that the proposed control approach has high accuracy and fast convergence speed, and is adaptive to general network scenarios. To the best of our knowledge this might be the first detailed study of using a closed-loop control approach to achieve proportional fairness amongst ACs in EDCA WLANs. The optimisation of controller parameters is not considered in this paper. We leave that for future work.

VII. ACKNOWLEDGEMENTS

This work was conducted as part of the project ISS-EWATUS (issewatus.eu) and has been financially funded by European Commission through the FP7 program (Grant agreement no:619228).

REFERENCES

- [1] IEEE 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements, 2005.
- [2] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, MACAW: A media access protocol for wireless LANs, in Proc. ACM SIGCOMM, 1994 , pp. 212-225.
- [3] J. Dunn, M. Neufeld, A. Sheth, D. Grunwald and J. Bennett, "A practical cross-layer mechanism for fairness in 802.11 Networks", *Mobile Networks and Applications*, 11(1), pp. 37-45, 2005.
- [4] L. B. Jiang and S. C. Liew, Improving throughput and fairness by reducing exposed and hidden nodes in 802.11 networks, *IEEE Transactions on Mobile Computing*, 7(1), pp. 34-49, 2008.
- [5] A. Kochut, A. Vasani, A. U. Shankar, and A. Agrawala, Sniffing out the correct physical layer capture model in 802.11b, in Proc. IEEE ICNP, 2004, pp. 252-261.
- [6] D. J. Leith, P. Clifford, and A. Ng, TCP fairness in 802.11e WLANs, *IEEE Communications Letters*, 9(11), pp. 964-966, Nov. 2005.
- [7] D. Malone, P. Clifford, and D. J. Leith, MAC layer channel quality measurement in 802.11, *IEEE Communications Letters*, 11(2), pp. 143-145, 2007.
- [8] C. Wang, "Achieving per-flow and weighted fairness for uplink and downlink in IEEE 802.11 WLANs", *EURASIP Journal on Wireless Communications and Networking*, 2012(1), 2012.
- [9] W. Li, S. Wang, Y. Cui, X. Cheng, R. Xin, M. A. Al-Rodhaan and A. Al-Dhelaan, "AP association for proportional fairness in multirate WLANs", *IEEE/ACM Transactions on Networking*, 22(1), pp. 191-202, 2013.
- [10] F. Kelly, "Charging and rate control for elastic traffic", *European Transactions on Telecommunications*, 8(1), pp. 33-37, 1997.
- [11] A. Banchs, P. Serrano and H. Oliver, "Proportional fair throughput allocation in multirate IEEE 802.11e wireless LANs", *Wireless Networks*, 13(5), pp. 649-662, 2007.
- [12] J. F. Lee, W. Liao M. C. Chen, "Proportional fairness for QoS enhancement in IEEE 802.11e WLANs", Proceedings of The IEEE Conference on Local Computer Networks, pp. 503-504, 2005.
- [13] V. A. Siris and G. Stamatakis, "Optimal CWmin selection for achieving proportional fairness in multi-rate 802.11e WLANs: test-bed implementation and evaluation", Proceedings of the 1st international workshop on Wireless network testbeds, experimental evaluation and characterization, pp. 41-48, 2006.
- [14] P. Serrano, A. Banchs, P. Patras and A. Azcorra, "Optimal configuration of 802.11e EDCA for real-time and data traffic", *IEEE Transactions on Vehicular Technology*, 59(5), pp. 2511-2527, 2010.
- [15] B. A. H. S. Abeysekera, T. Matsuda and T. Takine, "Dynamic contention window control scheme in IEEE 802.11e wireless LANs", Proceeding of IEEE VTC, 2009, pp. 1-5.
- [16] R. Cheng, C. Chang, C. Shih and Y. Chen, "A new scheme to achieve weighted fairness for WLAN supporting multimedia services", *IEEE Transactions on Wireless Communications*, 5(5), pp. 1095 - 1102.
- [17] X. Chen and D. Leith, "Proportional Fair Coding for 802.11 WLANs", *IEEE Wireless Communications Letters*, 1(5), pp. 468-471, 2012.
- [18] A. Checco and D. J. Leith, "Proportional Fairness in 802.11 Wireless LANs", *IEEE Communications Letters*, 15(8), pp. 807-809, 2011.
- [19] D. Leith, Q. Cao and V. G. Subramanian, "Max-Min fairness in 802.11 mesh network", *IEEE/ACM Transactions on Networking*, 20(3), pp. 756-769, 2012.
- [20] D. J. Leith, V. G. Subramanian and K. R. Duffy, "Log-convexity of rate region in 802.11e WLANs", *IEEE Communications Letters*, 14(1), pp. 57-59, 2010.
- [21] K. Premkumar, X. Chen and D. J. Leith, "Utility optimal coding for packet transmission over wireless networks Part I: Networks of binary symmetric channels", Proceeding of 49th Annu. Allerton Conference Communication, Control and Computing, pp. 1592-1599, 2011.
- [22] V. Hollot, V. Misra, D. Towsley and W. Gong, "A control theoretic analysis of RED", Proceeding of IEEE INFOCOM, 2001, pp. 1511-1519.
- [23] D. Cavendish, M. Gerla and S. Mascolo, "A control theoretic approach to congestion control in packet networks", *IEEE/ACM Transactions on Networking*, 12(5), pp. 893-906, 2004.
- [24] E. Park, N. Kwak, S. K. Lee, J. Kim and H. Kim, "Provisioning QoS for WiFi-enabled portable devices in Home Networks", *KSI Transactions on Internet and Information Systems*, 5(4), pp. 720-740, 2011.
- [25] P. Patras, A. Banchs and P. Serrano, "A control theoretic approach for throughput optimization in IEEE 802.11e EDCA WLANs", *Mobile Networks and Applications*, 14(6), pp. 697-708, 2008.
- [26] P. Patras, A. Banchs and P. Serrano, "A control theoretic scheme for efficient video transmission over IEEE 802.11e EDCA WLANs", *ACM Transactions on Multimedia Computing, Communications, and Applications*, 8(3), 2012.
- [27] P. Patras, A. Banchs, P. Serrano and A. Azcorra, "A control theoretic approach to distributed optimal configuration of 802.11 WLANs", *IEEE Transactions on Mobile Computing*, 10(6), 897-910, 2010.
- [28] K. Kosek-Szot, M. Natkaniec and A. R. Pach, "A simple but accurate throughput model for IEEE 802.11 EDCA in saturation and non-saturation conditions", *Computer Networks*, 55(3), pp. 622-635, 2011.
- [29] X. Yao, W. Wang, S. Yang and Y. Cen, "Bio-inspired self-adaptive rate control for multi-priority data transmission over WLANs", *Computer Communications*, Volume 53, , pp. 73-83, 2014.
- [30] P. C. Young and J. C. Willems, "An approach to the linear multivariable servomechanism problem", *International Journal of Control*, 15(5), pp. 961-979, 1972.
- [31] D. P. Bertsekas, A. Nedic and A. E. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, Belmont, MA, 2003.

APPENDIX

Proof.

$$U_1(\boldsymbol{\eta}) = \sum_{i=0}^{N-1} n_i (\eta_i - \log X(\boldsymbol{\eta}))$$

in which

$$\begin{aligned} \log X(\boldsymbol{\eta}) &= \log \left(\frac{\sigma}{T_{col}} + \sum_{i=0}^{N-1} n_i \left(\frac{T_i^{succ}}{T_{col}} - 1 \right) e^{\eta_i} + \prod_{i=0}^{N-1} (1 + e^{\eta_i})^{n_i} - 1 \right) \\ &= \log \left(\frac{\sigma}{T_{col}} + \sum_{i=0}^{N-1} n_i \frac{T_i^{succ}}{T_{col}} e^{\eta_i} + \prod_{i=0}^{N-1} (1 + e^{\eta_i})^{n_i} - 1 - \sum_{i=0}^{N-1} n_i e^{\eta_i} \right) \\ &= \log \left(\frac{\sigma}{T_{col}} + \sum_{i=0}^{N-1} n_i \frac{T_i^{succ}}{T_{col}} e^{\eta_i} + \sum_{k=2}^n \sum_{A \subseteq \mathcal{N}, |A|=k} \prod_{j \in A} e^{\eta_j} \right) \end{aligned}$$

and $\mathcal{N} = \{1, 2, \dots, n\}$ denotes the set of stations in the WLAN.

As the logarithm of a sum of exponentials is a convex function, $\log X$ is convex in the transformed variable $\boldsymbol{\eta}$, and $U_1(\boldsymbol{\eta})$ is thus concave in $\boldsymbol{\eta}$. \square