The following publication Hou, Y., Zheng, Y. & Li, M. Fair QoS multi-resource allocation for uplink traffic in WLAN. Wireless Netw 23, 467–486 (2017) is available at https://doi.org/10.1007/s11276-015-1170-2.

Fair QoS Multi-Resource Allocation for WLAN

Yuxiao Hou, *Member, IEEE,* Yuanqing Zheng, *Member, IEEE,* and Mo Li, *Member, IEEE*

Abstract

We consider the problem of allocating multiple types of network resources to achieve QoS fairness in WLAN. QoS fairness is a multi-resource fairness concept which aims to balance QoS and fairness in WLAN. To this end, we first transform user QoS requirements to multi-resource demands and apply the Dominant Resource Fairness scheme (DRF) to allocate network resources for each user. We prove several salient QoS-based fairness properties based on a model mapping between QoS and resources. We further discuss about more general conditions for diverse mapping models where QoS fairness properties can be satisfied. We find that the QoS fairness properties can be guaranteed as long as the mapping model meets a few practical requirements, indicating the wide applicability of our scheme. To consolidate our multi-resource allocation scheme, we design a practical protocol for WLAN. The simulation results validate that the QoS fairness can be guaranteed in practical WLAN scenario.

Keywords

WLAN, Multi-Resource Allocation, Fairness, QoS

I. Introduction

Current Wireless Local Area Network (WLAN) serves the last mile connection to the Internet for various wireless devices with diverse QoS requirements. A recent study [5] reveals that the

A preliminary version of this work was published in Proceedings of IWQoS 2014 [1].

Y. Hou and M. Li are with the School of Computer Engineering, Nanyang Technological University, Singapore 639798. E-mail: s120003@e.ntu.edu.sg, limo@ntu.edu.sg. Y. Zheng is with Department of Computing, The Hong Kong Polytechnic University, Hong Kong. E-mail: csyqzheng@comp.polyu.edu.hk.

main portion of end-to-end delay occurs at the WLAN rather than the Internet core. Traffic aggregation at the backhaul router can potentially cause packet loss due to buffer overflow. With increasing popularity of wireless devices (e.g., smartphones, tablets, etc.), it becomes challenging to meet all QoS requirements (e.g., delay and loss) with limited network resources.

We find that different QoS requirements are associated with demands of different network resources in WLAN. For instance, the delay is mainly due to the allocated wireless bandwidth, while the packet loss is mainly due to the queuing buffer overflow at the backhaul router. It is critical to fairly allocate those network resources among wireless devices so as to meet their diverse QoS requirements. Existing resource allocation schemes [9, 10, 18, 19, 22], however, fail to fairly accomodate the QoS diversity. Most of them do fair resource allocation based on a major bottleneck resource regardless of diverse QoS requirements.

In this paper, we first propose the QoS fairness in WLAN. To ensure QoS fairness, we investigate how to allocate multiple network resources. Dominant Resource Fairness (DRF) scheme [2] is a multi-resource allocation scheme which aims to fairly allocate multiple types of resources to users with diverse demands in these resources. Yet, the DRF scheme merely ensures fairness in terms of the resource quantity rather than the service quality that can be achieved with the allocated resources. Mapping different QoS requirements to multiple resource demands is non-trivial in WLAN networks. Moreover, it is unclear whether fairness properties (which can be trivially ensured in terms of resource quantity) would still hold after the mapping in terms of service quality.

We address the above challenges as follows. To translate user QoS requirements to network resource demands, we model the queuing buffer at the backhaul router and derive the mapping model based on it. We evaluate the service quality through a utility metric, which characterizes the perceived service quality of users with diverse QoS requirements. Based on the mapping model and the utility metric, we prove mathematically QoS-based fairness properties of our multi-resource allocation scheme. We further prove that these QoS-based fairness properties are not limited to specific mapping models and can be generalized in practice. We design a practical protocol to implement our allocation scheme in WLAN, with considerations of traffic monitoring, access control, etc.

We run extensive simulations to validate the QoS fairness properties of our multi-resource allocation scheme. The proposed scheme outperforms many existing allocation schemes in terms of both network utility and QoS fairness. Moreover, the proposed scheme can guarantee the strategy-proofness and robustness. We conclude with these results that our scheme achieves a good balance between fairness and QoS.

The contribution of this paper can be summarized as follows. We are the first to suggest QoS fairness in WLAN and design a multi-resource allocation scheme to achieve fairness in terms of QoS. We prove the generalities of QoS-based fairness properties for practical WLAN, regardless of the diversity in establishing the mapping between QoS requirements and network resource demands. Finally, we design a practical protocol to implement the proposed multi-resource scheme in WLAN.

The rest of this paper is organized as follows. Section II summarizes the previous resource allocation schemes and describe our design considerations in WLAN. Section III defines and proves QoS fairness properties. Section IV describes the practical protocol in WLAN that implements our proposed scheme. Section V evaluates our scheme with simulation results. Section VI presents related works. Section VII concludes this paper.

II. Background and Motivation

In this section, we first describe a QoS utility function, which can be used to measure user satisfaction in terms of the QoS performance. Next, we discuss why existing resource allocation schemes cannot satisfy QoS utility and strike fairness at the same time. After that, we introduce Dominant Resource Fairness, based on which a good tradeoff between QoS utility and fairness can be achieved. Finally, we demonstrate by an example in WLAN scenario to compare DRF with existing schemes in providing both fairness and utility.

A. Definition of QoS Utility

Many methods have been proposed to evaluate the quality of experience perceived by users. In general, these methods can be classified into two categories: subjective methods and objective methods [23]. The subjective methods ask users to vote for the quality perceived and thus directly reflect the users' experiences [28–30]. However, subjective methods are time-consuming and expensive as many users or testers have to evaluate the quality to get a meaningful and accurate result. In contrast, many automated objective metrics are suggested to evaluate perceived qualities of audio [24], speech [25], image [26], video [27], etc.

Different applications involve different types of data services and thus require different forms of QoE metrics. In this paper, we mainly consider end-to-end delay and packet loss rate as two main metrics of QoS and define the QoE metric as a QoS utility function U as follows:

$$U(d_0, r_0, d, r) = \min\left(\frac{d_0}{d}, \frac{1-r}{1-r_0}\right), \tag{1}$$

where d_0 and d are requested and perceived end-to-end delay experienced by the user, r_0 and r are requested and perceived packet loss rate, respectively.

We call d_0/d in Eq. (1) as the delay ratio which is the ratio of requested delay to actual delay. If the WLAN fully meets the delay requirement for the user, then we have $d_0/d = 1$; otherwise if the delay requirement is not fully satisfied, then we have $d_0/d < 1$. Hence, the value of the delay ratio reflects the extent to which the delay requirement is satisfied; a large delay ratio indicates the actual end-to-end delay is short, and vice versa.

Similarly, we call $(1 - r)/(1 - r_0)$ in Eq. (1) as the loss ratio which is the ratio of actual delivery rate to requested delivery rate. Similar to the delay ratio, the loss ratio reflects the extent to which the loss requirement is met.

The QoS utility function U is defined as the minimum of the delay ratio and the loss ratio, indicating that the overall utility is bottlenecked by one of the two QoS aspects (delay and loss) which is less satisfied than another by the WLAN.

The QoS utility function in Eq. (1) is an objective QoE definition which is independent of the application type. It also considers the subjective factor of the user, which is represented by its requirements (i.e., d_0 and r_0).

B. Existing Multi-Resource Fairness Schemes

A straightforward multi-resource allocation scheme is *per-resource fairness* (PF), which equalizes users' share in each resource regardless of users' diverse demands. In constrast, none fairness (NF) only considers user demands and ignore fairness. Specifically, for each user, NF allocates each type of resource with an amount proportional to his demand in the corresponding resource. Although it strikes equivalent QoS utility values for every user, it ignores fairness because a user with higher resource demands will get more shares, encouraging users to bluff about their true demands.

Another multi-resource allocation scheme is *bottleneck fairness* [17] (BF). It decides which resource is most bottlenecked (i.e., the resource type with the highest demand to supply ratio), and then equally shares this resource among users. The other types of resources are allocated to each user according to the resource demand proportionally.

BF is fair in that it equalizes user shares in the bottleneck resource, and meanwhile considers diverse user demands. However, it is vulnerable to a type of attack from a certain malicious user, who may lie about his true resource demands in order to shift the bottleneck resource from one to another, and consequently get more resource shares. In the last subsection of Section II we will explain this attack in detail.

C. Dominant Resource Fairness Scheme

Different from PF, NF and BF, a recent fairness scheme, named Dominant Resource Fairness [2], can achieve a good tradeoff between fairness and diversity of user demands. It can also overcome the attack by malicious users who lie about their true resource demands. In DRF, the *resource share* of a certain resource type is defined as the fraction of amount of resource obtained by the user over the total amount of this resource. The *dominant share* of each user is defined as the maximum resource share among all types of resources. DRF equalizes the dominant share of all users.

As an extension from max-min fairness for single resource allocation to the multi-resource max-min fairness, DRF has many desirable properties: *pareto-optimality*, *share-guarantee*, *envy-*

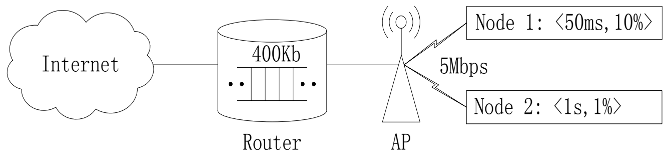


Fig. 1: The topology of wireless LAN.

freeness and *strategy-proofness*. *Pareto-optimality* means for any user, his overall resource share cannot be increased without either increasing the resource capacity or decreasing the resource share of other user. *Share-guarantee* means if there are n users in the system, the dominant share of each user should be larger than or equal to $1/n$, which encourages the participation of users in the system as a guarantee of resource share is promised. *Envy-freeness* means no user is willing to exchange his resource share with that of other users. *Strategy-proofness* means no user can concurrently increase his resource shares in all resource types by lying about his true resource demand, which encourages honest participation.

Note that DRF does not always equalize the dominant shares. If some users have their resource demands fully satisfied, the extra amount of resources left over by these users can be used to further meet the unsatisfactory demands of other users. As such, resources in the system can be utilized in a more efficient way.

D. Comparison of Different Fairness Schemes

In this subsection, we use an example scenario in WLAN to compare different fairness schemes, where two nodes have different QoS requirements.

Consider the illustrative example depicted in Fig 1, where node 1 and node 2, connect to a wireless access point (AP), which connects to the backhaul router. Suppose the wireless channel capacity is 5Mbps, the total amount of router queuing buffer is 0.4Mb and the backhaul link capacity is 5.13Mbps. The packet size for both nodes is 1500 bytes. Node 1 runs a real-time (delay-sensitive) application with an end-to-end delay requirement smaller than 50ms, and the loss rate smaller than 10%. We denote the QoS requirement for node 1 as $\langle$50ms, 10%$\rangle$. In the following we use the same notion (i.e., $\langle$d, r$\rangle$) to represent a QoS requirement of delay (d) and loss rate (r). Node 2 runs a loss-sensitive application with the QoS requirement of

$\langle$1s, 1%$\rangle$. Such QoS requirements can be transformed to resource demands of $\langle$5Mbps, 0.2Mb$\rangle$ and $\langle$0.25Mbps, 0.24Mb$\rangle$, respectively. We detail the mapping between a QoS requirement to demands in different types of resources in Section III-A.

Suppose a bandwidth allocation scheme allocates bandwidth proportional to the bandwidth demands of two nodes, and then each node gets 4.76Mbps and 0.24Mbps bandwidth respectively. Due to the different transmission rates of two nodes, the two nodes occupy 0.38Mb and 0.02Mb queuing buffer of the router. This scenario may happen for two uplink flows. With our mapping model, we compute that packets of node 1 has an actual end-to-end delay of 53ms and zero loss rate (denoted as $\langle$53ms, 0%$\rangle$), but node 2 experiences QoS performance of $\langle$1.05s, 92%$\rangle$. While node 1's QoS requirement is almost fully satisfied, node 2 suffers from as high as 92% packet loss, which is far worse than the required 1% loss. Based on the utility definition in Eq. (1), the two nodes achieve extremely unfair QoS utilities of 0.95 and 0.08.

On the other hand, a buffer allocation scheme allocates queuing buffer proportional to the buffer demands of two nodes, and then each node obtains 0.18Mb and 0.22Mb buffer respectively. This scenario may happen for two downlink flows. As such, the two nodes will obtain bandwidth shares of 2.27Mbps and 2.73Mbps. With our mapping model, the performances of the two nodes are computed as $\langle$111ms, 25%$\rangle$ and $\langle$92ms, 8%$\rangle$. Consequently the two nodes have utilities of 0.45 and 0.92, indicating an inefficient QoS utility result.

In the above settings, both the bandwidth allocation scheme and the buffer allocation scheme are single-resource schemes and cannot achieve fairness and utility at the same time. This is because they allocate the two resources separately and thus cannot achieve fair and efficient resource allocation. For the same scenario, however, there does exist much better resource allocation. If we allocated $\langle$4.75Mbps, 0.16Mb$\rangle$ to the first node, and $\langle$0.25Mbps, 0.24Mb$\rangle$ to the second node, the performances of two nodes are $\langle$53ms, 33.3%$\rangle$ and $\langle$1s, 1%$\rangle$. Consequently two nodes have utility of 0.8 and 1. Next we compare several multi-resource allocation schemes including PF, BF, NF and DRF, in providing fairness and utility.

In the above scenario, if per-resource fairness scheme (PF) is applied, both nodes will obtain $\langle$2.5Mbps, 0.2Mb$\rangle$, which maps to a QoS performance of $\langle$100ms, 16%$\rangle$. Such performance shows that node 1 gets none of its QoS aspect met, and node 2 only gets its delay requirement met. The two nodes have utilities of 0.5 and 0.83 respectively, indicating the inefficiency of PF in providing QoS utilities for both nodes, even though this scheme is absolutely fair in allocated resource shares.

Next we apply bottleneck fairness (BF). We first determine which resource type is most bottlenecked. Because $(0.2\text{Mb}+0.24\text{Mb})/0.4\text{Mb} > (5\text{Mbps}+0.25\text{Mbps})/5\text{Mbps}$, the bottleneck resource is queuing buffer. This scheme will allocate resource shares of ⟨4.8Mbps, 0.2Mb⟩ and ⟨0.2Mbps, 0.2Mb⟩ respectively to the two nodes. The QoS performances of two nodes are ⟨52ms, 20%⟩ and ⟨1.3s, 20%⟩. Hence, the two nodes have utilities of 0.96 and 0.8 respectively, which performs close to optimum.

However, the scheme of bottleneck fairness is vulnerable to strategic attackers. Suppose node 2 plays strategy by claiming a higher QoS requirement, e.g., ⟨100ms, 1%⟩, the corresponding resource demand becomes ⟨0.51Mbps, 0.24Mb⟩. The bottleneck resource transits from buffer to bandwidth. As a result, the two nodes get resource shares of ⟨2.5Mbps, 0.03Mb⟩ and ⟨2.5Mbps, 0.37Mb⟩ respectively, and their corresponding QoS performances are ⟨101ms, 88%⟩ and ⟨101ms, 0%⟩. Consequently, node 2 gets a significant increase in his QoS utility and outplays node 1 by strategy, which is not fair.

If we apply the none-fairness scheme (NF) in the above scenario, two nodes will obtain ⟨4.76Mbps, 0.18Mb⟩ and ⟨0.24Mbps, 0.22Mb⟩, respectively. The actual QoS performances of two nodes are ⟨52.9ms, 25%⟩ and ⟨1.05s, 8%⟩. Their corresponding utilities are 0.9 and 0.92. Although both users achieve high and comparable utility performances, this scheme suffers from strategic players. If node 1 proposes an extremely strict QoS requirement and thus has infinitely high resource demand, while node 2 fixes its QoS request, then node 1 will obtain almost all the network resources while node 2 will obtain negligible resource share.

Finally, if DRF is applied for allocation, ⟨3.33Mbps, 0.13Mb⟩ and ⟨0.28Mbps, 0.27Mb⟩ are initially allocated to node 1 and node 2 respectively. We note, however, node 2 is over-satisfied. Based on the initial allocation result, DRF adjusts the allocation as follows: node 1 gets a share of ⟨4Mbps, 0.16Mb⟩ and node 2 gets a share of ⟨0.25Mbps, 0.24Mb⟩. Their QoS performances are ⟨63ms, 33%⟩ and ⟨1s, 1%⟩, with utilities of 0.8 and 1 respectively. In terms of utility, DRF performs better than PF and comparable to BF and NF.

In contrast to BF and NF, DRF is strategy-proof. If node 2 lies about its QoS requirement and claims a resource demand of ⟨0.504Mbps, 0.24Mb⟩ after QoS mapping while node 1 keeps its QoS requirement unchanged, DRF will initially allocate ⟨3.33Mbps, 0.13Mb⟩ and ⟨0.56Mbps, 0.27Mb⟩ to the two nodes. The initial allocation result is only slightly different from the previous one: for node 2, only its bandwidth share increases from 0.25Mbps to 0.56Mbps, while its buffer share remains the same. After adjustment, the two nodes get the same resource share and thus same QoS performances as in the previous allocation. Hence we confirm that node 2 cannot benefit from this strategy by lying about his QoS requirements.

III. QoS-Based DRF

Although DRF has many good properties to ensure fairness, these properties are proved from the perspective of *resource quantity*, but not from *service quality*. In other words, the DRF fairness properties are proved to ensure that all users get fair amount of resource shares, but are not proved to guarantee the fairness in terms of QoS utilities.

To prove these properties, we need a model which maps resources (bandwidth and buffer) to QoS (delay and loss). Nevertheless, such a model cannot consider all related factors and some assumptions have to be made to simplify the model and facilitate the analysis. We list these assumptions in Section III-A. Mathematical expressions for the model are presented in Section III-B.

In Section III-C, we define QoS fairness formally through several fairness properties in terms of QoS and prove them based on the mapping model in Section III-B. Finally in Section III-D, we have more discussion on QoS fairness and the general assumptions for kinds of mapping models under which QoS fairness can be ensured.

A. Concerns in Establishing the Mapping Model

QoS performances consist of two important aspects: end-to-end delay and packet loss rate. We first discuss on factors that have impacts on the end-to-end delay. According to [5], a single-hop transmission delay of a packet in the wired link is less than 20 microseconds, which is ignorable in the end-to-end delay compared with that in the wireless network, which takes a magnitude of several tens of milliseconds. Moreover, it is rare for packets to suffer from a long queuing delay in the routers of the Internet core, as these routers usually have powerful packet processing abilities and large storage capacities for packet accommodations. In contrast, packets may suffer from notable queuing delay at the backhaul router in WLAN. This is because small internet service providers (ISPs) tend to configure the backhaul router with small memory storage to reduce economical cost, even if the backhaul router serves as the only gate to the Internet core and aggregates all traffics in the WLAN. Consequently, we approximate the end-to-end delay as the sum of wireless transmission delay and queuing delay at the backhaul router.

Packet loss due to channel fading in the physical layer or 802.11 random access collisions in MAC layer can be compensated through the 802.11 retransmission mechanism in the MAC layer. Since router is a network-layer device, we do not consider packet loss that happens in physical or MAC layer. Rather, we only consider packet loss due to queuing buffer overflow at the backhaul router, which often has limited buffer storage capacity compared to routers with large buffer storage at the Internet core.

To estimate the queuing delay and packet loss rate, we use a M/D/1 queuing model for the queuing buffer at the backhaul router. We further prove that QoS-based fairness properties are held in a general way and not just limited to some specific queuing models. In the following, we build a mapping model that maps between user QoS requirement (end-to-end delay and packet loss rate) and network resource demands (wireless bandwidth and queuing buffer at the backhaul router).

B. A Mapping Model between QoS and WLAN Resource

We assume all users have a fixed packet size P. Each user has a QoS performance in delay and loss rate, denoted as $\langle d, r\rangle$. The corresponding WLAN resources (i.e., wireless bandwidth and queuing buffer) allocated to a certain user are expressed with $\langle BW, L\rangle$.

The transmission delay d_t in the wireless link of a packet from the node is:

$$d_t = \frac{P}{BW}$$

Suppose the packet arrival follows the Poisson Distribution with an arrival rate $\lambda = BW/P$ and the service rate of the queue α. We define $\rho = \lambda/\alpha$ as the traffic load. Generally, $\rho < 1$.

According to [16], the average queuing delay d_q is

$$d_q = \frac{1}{2\alpha}\frac{2-\rho}{1-\rho}.$$

Consequently, the end-to-end delay d is

$$d = d_t + d_q = \frac{1}{\lambda} + \frac{1}{2\alpha}\frac{2-\rho}{1-\rho}. \quad (2)$$

For the packet loss due to queuing buffer overflow, we provide the average queuing length as follows:

$$E[Q] = \rho + \frac{1}{2}\frac{\rho^2}{1-\rho} \quad (3)$$

However, the M/D/1 model assumes an infinite buffer size. There will be packet loss if $E[Q] > L/P$. Here L should be normalized by packet size P when it compares with $E[Q]$, as L has the unit of bit while $E[Q]$ has the unit of packets. We approximate the loss rate r as follows:

$$r = \frac{E[Q] - L/P}{E[Q]} \quad (4)$$

We assume the traffic load ρ is a constant value. When the traffic load is heavy, we can approximate ρ as C/C_b, where C is the wireless link capacity and C_b is the capacity of the wired link connected to the backhaul router. With this assumption we have

$$\lambda = \frac{S}{d},$$

where

$$S = 1 + \frac{\rho(2-\rho)}{2(1-\rho)}.$$

Similarly, BW is

$$BW = \lambda P = \frac{SP}{d}. \quad (5)$$

Since $E[Q]$ is a function of ρ according to Eq. (3), $E[Q]$ is also a constant for the user. Based on Eq. (4), we derive the expression for queuing buffer length:

$$L = PE[Q](1 - r) \quad (6)$$

In Section IV, we discuss on how to choose the α value and allocate the queuing buffer for each node.

C. Proof of QoS-Based Fairness Properties

In this subsection, we define and prove QoS fairness properties based on DRF, the predefined mapping model, and the utility definition.

Suppose the wireless link capacity is C, the total queuing buffer is L_Q, and M users contend for network resources. For each user i, we denote its QoS requirement as $\langle d_{i0}, r_{i0} \rangle$ and its corresponding resource demand vector as $\langle BW_{i0}, L_{i0} \rangle$. $\mu_i = max\left\{BW_{i0}/C, L_{i0}/L_Q\right\}$ denotes the required dominant share of user i. If $\mu_i = BW_{i0}/C$, user i is bandwidth-dominant; otherwise user i is buffer-dominant.

After the WLAN performs the allocation for all users, user i obtains a resource share of $\langle BW_i, L_i \rangle$. Its corresponding QoS performance is denoted as $\langle d_i, r_i \rangle$. We denote $x_i = BW_i/BW_{i0} = L_i/L_{i0}$ as the ratio of provision over demand. For DRF allocation scheme, such ratios for both bandwidth and buffer resources are the same (i.e., x_i).

We first formulate the allocation problem. Without loss of generality, we assume that the first K of M users are bandwidth-dominant, and the remaining $M - K$ users are buffer-dominant. The resource demands of these users are computed based on their QoS requirements and our mapping model.

Based on the definition of μ_i, we have

$$\mu_i = \begin{cases} \dfrac{BW_{i0}}{C} & \text{if } i = 1, 2, ..., K \\ \dfrac{L_{i0}}{L_Q} & \text{if } i = K+1, K+2, ..., M \end{cases}.$$

The DRF allocation problem is formulated as:

$$\begin{aligned} \text{maximize} \quad & (x_1, x_2, ..., x_M) \\ \text{subject to} \quad & \sum_{i=1}^{M} BW_i = \sum_{i=1}^{M} x_i BW_{i0} \leq C \\ & \sum_{i=1}^{M} L_i = \sum_{i=1}^{M} x_i L_{i0} \leq L_Q \\ & \mu_1 x_1 = \mu_1 x_2 = ... = \mu_M x_M \end{aligned} \tag{7}$$

where x_i $(i = 1, 2, ..., M)$ are optimization variables.

In the above formulation, we maximize the share of all users subject to resource constraints in bandwidth and queuing buffer. The DRF contraint equalizes users' dominant shares.

The analytical solution is given as follows:

$$\begin{aligned} BW_i = x_i BW_{i0} &= \begin{cases} qC & \text{if } i = 1, 2, ..., K \\ \dfrac{qL_Q}{m_i} & \text{if } i = K+1, K+2, ..., M \end{cases} \\ L_i = x_i L_{i0} &= \begin{cases} m_i qC & \text{if } i = 1, 2, ..., K \\ qL_Q & \text{if } i = K+1, K+2, ..., M \end{cases} \end{aligned} \tag{8}$$

where $q = \mu_i x_i$, $(i = 1, 2, ..., M)$ represents the actual dominant share of user i. Solving the DRF optimization problem, we have q as follows:

$$q = min(\frac{1}{K + a\sum_{i=K+1}^{M} 1/m_i}, \frac{a}{\sum_{i=1}^{K} m_i + a(M-K)}), \tag{9}$$

where $a = L_Q/C$ denotes the ratio of total queuing buffer size over wireless channel capacity, and $m_i = L_{i0}/BW_{i0}$ is defined as the resource demand ratio for user i.

By applying the definition of dominant resource for a user i, we can get the following inequalities with respect to m_i:

$$m_i \begin{cases} \leq a & \text{if } i = 1, 2, ..., K \\ \geq a & \text{if } i = K+1, K+2, ..., M \end{cases} \tag{10}$$

According to Eq. (9), there are 2 cases for the value of q. We only show the proofs of the two properties for the first case where $q = 1/(K + a\sum_{i=K+1}^{M} 1/m_i)$. The proof for the second case where $q = a/(\sum_{i=1}^{K} m_i + a(M-K))$ is similar to that for the first case.

All QoS fairness properties are related to the utility, while all DRF fairness properties are described from the view of resources. Consequently, to prove QoS fairness properties, we first need to formulate the utility as a function of resources.

Based on Eq. (5) and Eq. (6), we get

$$\frac{d_{i0}}{d_i} = \frac{BW_i}{BW_{i0}},$$

and

$$\frac{1-r_i}{1-r_{i0}} = \frac{L_i/PE[Q_i]}{L_{i0}/PE[Q_i]} = \frac{L_i}{L_{i0}}$$

Thus from Eq. (1), we have

$$U_i = \min\left(\frac{BW_i}{BW_{i0}}, \frac{L_i}{L_{i0}}\right) \tag{11}$$

Based on Eq. (8), we can further derive that

$$\frac{BW_i}{BW_{i0}} = \frac{L_i}{L_{i0}} = x_i \Rightarrow U_i = x_i \tag{12}$$

In the following we define and prove QoS fairness properties based on the above preliminary results.

QoS-based Pareto-Optimality: With fixed amounts of all types of network resources (i.e., wireless link capacity and queuing buffer at the backhaul router) in WLAN, a user can only get his utility increased by decreasing the utility of any other user.

Proof: Suppose among M users, user i increases his utility. For any other node, say node j, does not have his utility decreased.

Based on Eq. (12), we can infer that x_i increases and x_j remains the same. However, according to the DRF formulation, at least one type of resource is completely utilized by users. Without lossing generality we assume bandwidth is completely consumed, i.e., $\sum_{i=1}^{M} x_i BW_{i0} = C$. After x_i increases, the individual component $x_i BW_{i0}$ also increases, resulting in $\sum_{i=1}^{M} x_i BW_{i0} > C$ if x_j $(j \neq i)$ remains the same. The reason why the contradiction happens is the wrong assumption that user i can get U_i increased without decreasing U_j $(j \neq i)$. Hence, we prove the QoS-based pareto-optimality property. ■

QoS-based Envy-Freeness: A user will not envy the QoS performances of any other users. This property indicates that the utility of a user will not increase if this user exchanges his QoS performances with another user's.

Proof: As the mapping between QoS and resource is one-to-one mapping, exchanging QoS performances is equivalent to exchanging actual resource shares.

Suppose two users, i and j, have obtained resource shares of $\langle BW_i, L_i \rangle$ and $\langle BW_j, L_j \rangle$, respectively. The corresponding utility values for the two users are U_i and U_j. We assume user i is bandwidth-dominant after we map his QoS requirements to resource demands. According to DRF allocation results in Eq. (8), we have $BW_i = qC$.

If user j is also bandwidth-dominant, we have $BW_j = BW_i = qC$. Otherwise if user j is buffer-dominant, we have $BW_j = qL_Q/m_j$. We can derive that

$$\frac{BW_j}{BW_i} = \frac{L_Q}{m_j C} = \frac{a}{m_j}.$$

Since user j is buffer-dominant, according to Eq. (10), we have $m_j \geq a$. Hence,

$$\frac{BW_j}{BW_i} = \frac{a}{m_j} \leq 1 \;\Rightarrow BW_j < BW_i$$

Combining two cases, we prove that if user i is bandwidth dominant, his actual bandwidth share will be no less than that of any other users, i.e.,

$$BW_i \geq BW_j, \ \forall j \neq i.$$

Now we prove QoS-based envy-freeness. After user i exchanges his resource shares with user j, user i's new utility U_i^{new} is expressed as follows:

$$U_i^{new} = min(\frac{BW_j}{BW_{i0}}, \frac{L_j}{L_{i0}}).$$

We know that $BW_j < BW_i$. If $L_j < L_i$, then both resource shares of user j are smaller than that of user i, and obviously, $U_i^{new} < U_i$. If $L_j > L_i$, we derive based on Eq. (12) that

$$U_i^{new} = min(\frac{BW_j}{BW_{i0}}, \frac{L_j}{L_{i0}}) = \frac{BW_j}{BW_{i0}} \leq \frac{BW_i}{BW_{i0}} = x_i = U_i.$$

■

QoS-based Strategy-Proofness: A user cannot improve his QoS utility by lying about his true QoS requirement. In other words, the QoS utility is maximized if the user tells the true QoS requirement, which encourages cooperation among nodes.

Proof: Suppose the true QoS requirement of user i is $\langle d_{i0}, r_{i0} \rangle$. The actual QoS performance, decided by DRF scheme, is denoted as $\langle d_i, r_i \rangle$ and the corresponding utility is U_i. If user i provides another requirement $\langle d_{i0}^*, r_{i0}^* \rangle$, his actual QoS performance is $\langle d_i^*, r_i^* \rangle$ and the new utility is U_i^*. Then, the QoS-based strategy-proofness property can be expressed with the following inequility:

$$U_i(d_{i0}, r_{i0}, d_i, r_i) \geq U_i^*(d_{i0}^*, r_{i0}^*, d_i^*, r_i^*) \tag{13}$$

We denote the true resource demands as $\langle BW_{i0}, L_{i0} \rangle$. The actual resource shares for the true and fake QoS requirements are $\langle BW_i, L_i \rangle$ and $\langle BW_i^*, L_i^* \rangle$, respectively.

According to the property of resource-based strategy-proofness of DRF, the resource shares of a user cannot increase concurrently. Without lossing generality, if $BW_i^* > BW_i$, we can infer based on this property that $L_i^* \leq L_i$. Similar to the proof for QoS-based envy-freeness, we can derive based on Eq. (12) that

$$U_i^* = min(\frac{BW_i^*}{BW_{i0}}, \frac{L_i^*}{L_{i0}}) = \frac{L_i^*}{L_{i0}} \leq \frac{L_i}{L_{i0}} = x_i = U_i.$$

Note that in the above U_i^* is computed based on $\langle BW_{i0}, L_{i0} \rangle$, which is the true resource demand of user i. ■

D. More Discussion on DRF and QoS Fairness

DRF has another resource-based fairness property named share-guarantee, which means if M users are in the network, then the dominant share of each user would be larger than or equal to $1/M$. Note that we cannot ensure "QoS-based Share-Guarantee", because utility is a subjective metric while a QoS performance is an objective one. Since QoS utility is also a function of QoS requirement, an initial high requirement leads to low utility while an initial low requirement leads to high utility, for the same QoS performance.

DRF focuses more on fairness than on utility. In DRF, user QoS utilities are only considered to some limited extent by allowing users to specify their dominant resources. Detailed user demands cannot affect the allocation results, which prevents malicious users from lying about their demands.

In contrast to DRF which focuses more on fairness, we can formulate the multi-resource allocation problem so as to maximize the systematic utility and improve utilization efficiency of resources. However, the new formulation suffers from malicious users, as QoS utility is a subjective metric and they can easily get much better QoS performances by lying about their true QoS requirements.

In the previous section, we prove several QoS fairness properties based on our mapping model. However, from the process of proof, we only utilize some basic features of the model: first, end-to-end delay is a monotonic decreasing function of actual bandwidth share; second, packet loss rate is a monotonic decreasing function of actual queuing buffer share. These two features are general and well-recognized. Consequently, the QoS fairness properties are held for a wide range of mapping models which satisfy the above two features.

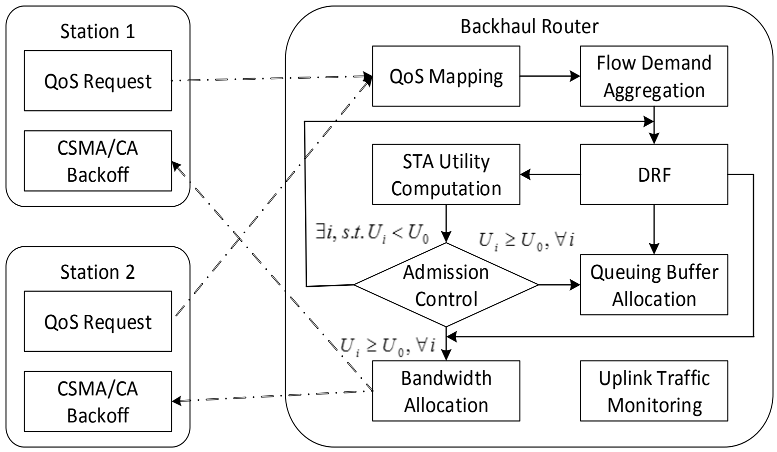


Fig. 2: Router functionalities and interactions between wireless node and the router.

IV. Protocol design

In this section, we illustrate our design protocol for providing QoS fairness in WLAN. As shown in Fig. 2, the protocol mainly consists of many allocation-related components at the backhaul router, the interaction process between end users and the backhaul router, etc. Extra components are further implemented to enhance the protocol such as admission control, flow demand aggregation, etc. In the following, we first summarize how the WLAN system works under our protocol and then highlight several key components in more details.

When a user has packet for transmission in the WLAN, it first sends its QoS request message to the backhaul router. After the router collects all QoS request messages from all users, it maps QoS requirements of each flow of each user to network resource demands. To ensure per-node fairness, it performs the flow demand aggregation functionality to keep that one user only has one demand presented. Next, it computes the DRF allocation results with the information of wireless link capacity C and queuing memory size L_Q.

With the allocation results, it calculates the expected utilities for all users according to Eq. (11), which are used for access control. If for a certain user i, its expected utility U_i is smaller than U_0, a predefined threshold, user i will be denied of access. If at least one user is denied of access, the DRF component is performed again without considering these users. The above process is repeated until no user is denied of access.

To allocate queuing buffer, the router can perform it directly on its own; while to allocation wireless bandwidth, the router needs to compute each user's persistency probability, a parameter in 802.11 CSMA/CA backoff scheme, according to the bandwidth allocation results. Users can acquire their bandwidth allocations by adjusting this parameter on their owns. In case some malicious users do not follow the parameter setting from the router, we design traffic monitoring component to detect such behavior and punish these users. Finally, the router periodically broadcasts a message to detect whether some nodes join or leave, or whether any flows are added or deleted for certain users, based on user replies. If any events happen, the router repeat the allocation process in the above to update allocation.

A. Adaptive Flow Demand Aggregation

Before applying DRF scheme, the router has to check if any node creates multiple flows. If so, the flow demand aggregation has to be performed for these nodes so that per-node fairness is achieved. This is because a node may increase resource share by transmitting multiple flows and each flow has a corresponding resource share from DRF.

In particular, suppose a user creates J flows. The QoS requirements of each flow j are denoted as $\langle d_{j0}, r_{j0} \rangle$. With our mapping model the corresponding resource demand for each flow is $\langle BW_{j0}, L_{j0} \rangle$. The aggregated resource demand for this node, denoted as $\langle BW_0, L_0 \rangle$, is as follows

$$BW_0 = \sum_{j=1}^{J} BW_{j0} \ , \ L_0 = \sum_{j=1}^{J} L_{j0}. \tag{14}$$

Hence, this user contends for WLAN resources with the above aggregated demand.

Note that one-time flow demand aggregation cannot adapt well to the varying user demands over time. Instead, an adaptive aggregation scheme, which requires the router periodically collects QoS information of flows of all users, is desired. If a user creates a new flow or ends an existing flow, or has any changes in any of his flow QoS requirements, the router will update the corresponding user QoS and adjusts the resource allocation results.

B. DRF Algorithm

After flow demand aggregation, the router can compute the DRF allocation result based on the formulation in Eq. (7). We have to note that DRF does not always equalize dominant shares of users. If a user gets more resources than he requires, the extra resource can be re-allocated to other users.

We describe the adjustment process as follows. First, we perform regular DRF scheme according to Eq. (7). Then we compute the utility of each user based on Eq. (11). Users whose utilities are larger than 1 will obtain their resource shares as they require. For other users, regular DRF will be performed again with the remaining network resources. This process is repeated until no users can get a utility larger than 1 or all users have their resource demands satisfied.

C. Admission Control

When the network is in heavy load, admission control scheme is used to deny access of new users and ensure network performances. In traditional admission control schemes, only if the user request can be completely satisfied by the network, the user is admitted access. In contrast, we suggest an admission control scheme that admits user access even if the user request cannot be totally met, which allows more users to join the network while ensuring certain level of quality of services.

We describe our admission control scheme as follows. First, we compute the expected utility for each user according to DRF allocation results (after adjustment). Next, we check if the expected utility of a user is above a certain threshold, e.g., 0.9. If so, this user is admitted access and otherwise not. In our simulation, we set the minimum utility threshold to 0.7, which works well across various experiment settings.

Under our admission control scheme, users with strict QoS requirements may be denied access. This is because the fairness properties of DRF will ensure users get fair resource shares. A user with larger demands will get lower QoS utility based on Eq. (11) and hence may have his utility below the minimum utility threshold. The admission control scheme can discourage potential malicious users from bluffing about their QoS requirements.

D. Bandwidth Allocation in WLAN

Now we come to the issue of bandwidth allocation. We differentiate the bandwidth allocation into two cases: uplink and downlink. For downlink flows, the router can perform fair queuing algorithms [31–34] to allocation wireless bandwidth, as all downlink packets have to go through the backhaul router to the end user. To proportionally allocate bandwidth based on the DRF allocation result, we can apply weighted fair queuing algorithms, e.g., [35], where the weight w_i for node i is proportional to the actual bandwidth share BW_i and normalized to the wireless link capacity C, i.e., BW_i/C.

For uplink flows, we adopt the distributed allocation algorithm in [36]. It assumes a low contention loss probability as flows (nodes) can adaptively adjust their channel rate when loss occurs. As such, it infers that the channel rate of flows (or nodes) can be approximated by the persistence probability in the 802.11 backoff mechanism. In this suggestion, all contending nodes have the same backoff bounds (i.e., minimum and maximum contention window size).

To achieves proportional fairness, [36] allocates transmission rates to all nodes proportionally to their persistence probability. In our scenario, similar to the weight setting for downlink weighted fair queuing, for uplink case the persistence probability x_i of node i is set to BW_i/C.

After determining x_i, we briefly illustrate how [36] achieves the wireless bandwidth allocation through the 802.11 backoff mechanism. Node i has three states: ***IDLE***, ***CONTEND***, ***ACQUIRE***. Initially node i is in idle state. When it has a packet to send and senses the idle channel, it changes its state to ***CONTEND*** with probability x_i. If it fails to change the state, it increases x_i to $x_i + \alpha$ and contends for the channel in the next slot with the new x_i. In the ***CONTEND*** state, node i first waits for a random time interval B_i and next senses the channel. B_i is a random interger chosen in $[0, B]$ with uniform distribution. B is a constant for all nodes. If node i senses that the channel is free, it changes to the ***ACQUIRE*** state, adjusts x_i to $x_i+\alpha$ and starts transmission; otherwise if the channel is busy or it experiences a collision, it changes to the ***IDLE*** state, adjusts x_i to $(1-\beta)x_i+\alpha$, and waits for the contending process in the next slot. If the node is in ***ACQUIRE*** state, after transmission it goes back to the ***IDLE*** as well. According to [36], α, β and B are system constants with values of 0.1, 0.5 and 32 respectively.

Since the uplink bandwidth is allocated distributedly at the side of end nodes, it is possible that some malicious nodes will not follow the backoff scheme mentioned above. Hence it is necessary to monitor uplink traffics and detect such malicious behaviors.

Intuitively, if a malicious node wants to get more bandwidth share, it should increase its persistence probability to a high value. As such, it has more chances to win the channel than other nodes with lower persistence probability. Consequently, the malicious node will have a better bandwidth share while the bandwidth shares of other normal nodes will be degraded.

With this observation, we design a lightweight uplink bandwidth monitoring scheme at the backhaul router. The router periodically computes the average transmission rate of each node over the period based on the records of packets that pass through the router. Next the router computes the ratio of the actual transmission rate to the allocated bandwidth for each node. If this ratio of a certain node exceeds a certain threshold (e.g., 1.3), the router identifies this node as malicious and punishes this node by dropping all its packets in the next period. In the later section we run simulations to explore the optimal values for the ratio threshold and the period interval.

E. Buffer Allocation at the Backhaul Router

To allocate queuing buffers to M users, we first need to divide the whole memory block at the backhaul router into M logical queues, each of which accomodates the packets from the corresponding user. Next the router can execute the buffer allocation by assigning the amount of memories for each user (i.e., buffer size of the corresponding queue) based on DRF allocation results. If the amount of packets of a certain user exceeds the logical queue size, then the extra packets that cannot be accomodated in this queue will be dropped. The router dequeues a packet from all these logical queues through a round-robin manner.

In our mapping model we assume that the traffic load ρ is fixed. In practical WLAN scenario we design a simple method to achieve this. When the traffic load is heavy in WLAN, wireless bandwidth is nearly used up by end users, while the wired line rate C_b is a constant value. We thus approximate the traffic load of any user i, denoted as ρ_i, as C/C_b, where C is the wireless

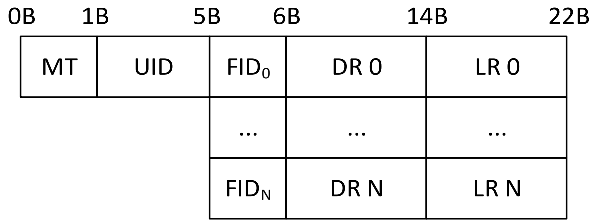


Fig. 3: Message format for end user.

link capacity. Since different users may have different input rates determined by the wireless bandwidth allocation results, it is necesssary for the backhaul router to adjust their corresponding queuing service rates to keep the constant traffic load ρ_i. With the notions in Section III, this implies that

$$\alpha_i = \frac{\lambda_i}{\rho_i} = \frac{BW_i/P}{C/C_b} = \frac{BW_iC_b}{PC}, \tag{15}$$

where BW_i is the bandwidth allocation result based on DRF scheme and P is the packet size. Note that here the unit of service rate α_i and input rate λ_i is packet/sec.

In our mapping model, we further assume that $\lambda_i < \alpha_i$. This is ensured at the backhaul router since the wired link rate C_b is larger than the wireless link capacity C and $\rho_i = C/C_b < 1$.

Finally, for uplink traffics, we can implement the differentiated service rates through a weighted queuing algorithm. Note for downlink traffic there is no need to adjust the service rate as the input traffic is from the internet core and thus does not follow the input traffic rate BW_i.

F. Message Format & Communication Overhead

In Fig. 3, we regulate the message formats for the end user. In Fig. 3, the MT field, which indicates message type, takes value '1' if it is the user QoS request message and '0' if it is the user QoS update message. The UID field indicates the user identity. Note that we assume the router ID is unique and known to all users. The FID field represents the flow ID. The DR field is the QoS delay requirement and the LR field is the QoS loss requirement. A user can have multiple flows and each flow has its corresponding DR and LR values. The size of each field is clearly labeled in Fig. 3.

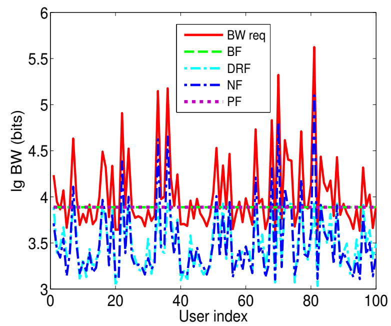
The message format for the router includes three fields: MT, UID/BID and related information field. The MT field is the message types and occupies 1 byte. If it is a bandwidth allocation message, MT takes '0'; if it is a periodical and broadcasted QoS update request message, MT takes '1'; if it is a notification message of denied access to a certain user, MT takes '2'. The next field, UID/BID, represents either the user ID or the broadcast ID, depending on the message type. It occupies 4 bytes. Finally, the related information field also depends on the message type: if MT is 0, this field is the bandwidth allocation result for certain user; if MT is 1, this field is empty; if MT is 2, this field indicates the reason why the user is denied of access. There can be two reasons for denied access: first, the user provides too strict Qos requirements and the network cannot provide good enough service for him; second, the user does not follow the bandwidth allocation result from the router. This field occupies 8 bytes.

Suppose a network has 1000 active users and a user creates at most 10 flows currently. From the above we know each user message occupies at most $5 + 17 \times 10 = 175$B and each router message occupies at most $1+4+8 = 13$B. The total amount of messages sent in the protocol is $(175 + 13) \times 1000 \approx 0.19$MB, or around 1.5Mb. If the wireless link capacity is 5Mbps, it takes around 0.3 second to run through the protocol and such delay is acceptable in practical WLAN.

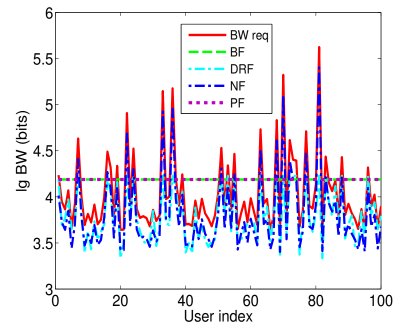
V. Evaluation of QoS-based Fairness Properties

In this section we run extensive simulations to verify QoS-based fairness properties of DRF and compare DRF with other fairness allocation schemes including bottleneck fairness (BF), none fairness (NF) and per-resource fairness (PF) in terms of fairness and network utility. The QoS mapping component in our protocol is used to map QoS requirements to resource demands for all schemes.

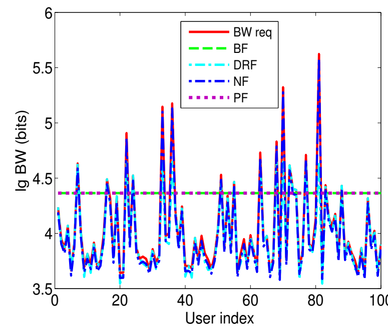
The basic simulation settings are as follows. There are 1000 users in the network. The wireless link capacity is 5Mbps. The backhaul link rate is 10Mbps. The available queuing buffer at the router is 20Mb. Packet size is 12Kb. QoS requirements of each user are randomly generated. Specifically, the delay requirement is chosen in $[0.1\text{sec}, 5\text{sec}]$ and the loss requirement is chosen in $[1\%, 50\%]$. The DRF allocation result is the one after adjustment, which is described in Section IV-B. Such settings are assumed for the following simulations unless otherwise noted.



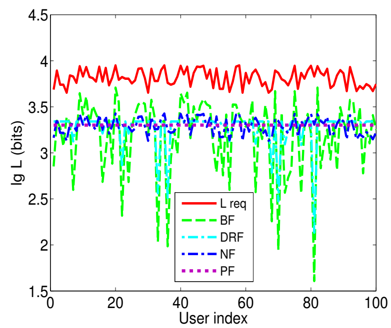
(a) Bandwidth distribution under heavy load



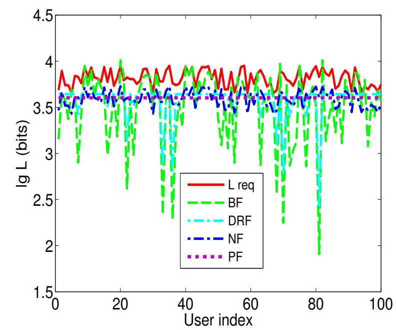
(b) Bandwidth distribution under median load



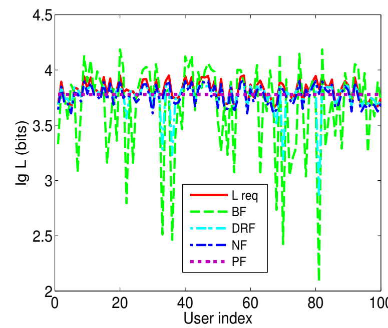
(c) Bandwidth distribution under light load



(d) Queuing buffer distribution under heavy load



(e) Queuing buffer distribution under median load



(f) Queuing buffer distribution under light load

Fig. 4: Distribution of allocated resource shares of 100 users under different traffic loads.

A. Comparison of Fair Allocation Schemes

In this subsection we compare the fairness and efficiency properties of multiple fair allocation schemes including BF, DRF, NF and PF under different traffic load settings. We define traffic load factor θ as follows:

$$\theta = min(\frac{C}{\sum_{i=1}^{M} BW_{i0}}, \frac{L_Q}{\sum_{i=1}^{M} L_{i0}}). \tag{16}$$

In Eq. (16), C and L_Q are total capacities of wireless bandwidth and router queuing buffer. $\sum_{i=1}^{M} BW_{i0}$ and $\sum_{i=1}^{M} L_{i0}$ are total amount of requested bandwidth and buffer from all participating users. Obviously, the larger θ is, the lighter the traffic load would be. In this subsection we differentiate three types of traffic load: $\theta = 0.3$ (heavy load), $\theta = 0.6$ (median load) and $\theta = 0.9$ (light load). The user resource demands are fixed and the total resource amounts are varied so as to evaluate different traffic load.

As mentioned in the beginning of this section, we simulate 1000 users with random QoS requirements. After mapping from QoS requirements to network resource demands, we perform each of these schemes separately. Based on the requested and actual QoS performances, we compute a utility value for each user. The above simulation is performed under all three types of traffic loads.

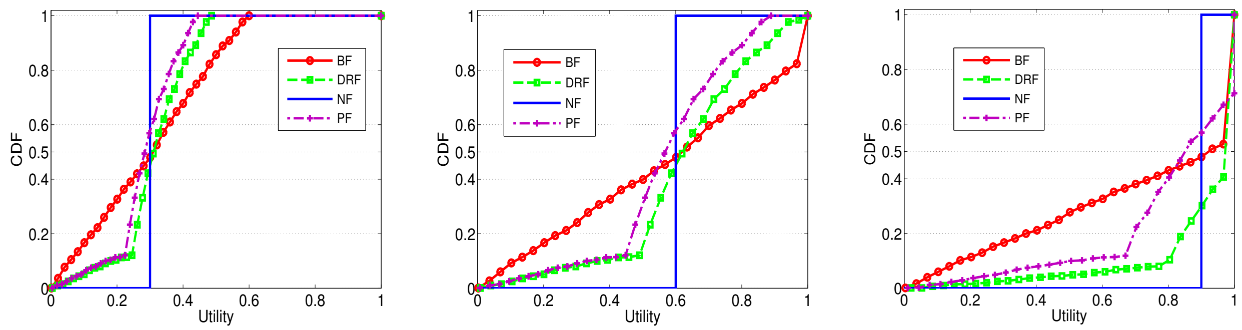
We first compare the fairness feature of these fairness schemes. Fig. 4 plots the resource shares in both bandwidth and buffer of first 100 users of the 1000 users under three traffic load settings. Fig. 4(a) - Fig. 4(c) plot the bandwidth allocation and Fig. 4(d) - Fig. 4(f) plot the queuing buffer allocation (in logarithm scale).

In Fig. 4(a), the red line shows the requested bandwidth of users. Both BF and PF distribute the total bandwidth equally to users. NF exhibits similar shapes with the requested bandwidth, but when users have a large gap in their bandwidth requests, the amount of bandwidth they obtain would also be largely different. As a result, NF would unfairly favor the users who tend to bluff about their requests. DRF also shares similar trend as the requested bandwidth, but in contrast with NF, it does not allow users with high demands to get unlimitedly high shares and thus reduce the allocation gap between users.

Fig. 4(d) plots the distribution of queuing buffer shares of users. The red line shows the requested queuing buffer of users. NF shares similar shape as the request buffer. PF equally distributes the total queuing buffers. BF and DRF have similar shapes: the valleys in buffer distribution under the two schemes are coupled with the peaks in the corresponding bandwidth distribution in Fig. 4(a). These valleys can be regarded as penalties to users with very high bandwidth requests. The difference between DRF and BF is that DRF has higher valley values than BF for each valley position, indicating DRF controls the buffer gap between users better than BF and thus strikes higher fairness than BF.

The distribution of user resource shares with median (Fig. 4(b) and Fig. 4(e)) and high traffic loads (Fig. 4(c) and Fig. 4(f)) have similar observations as that with light traffic load (Fig. 4(a) and Fig. 4(d)), except that when the traffic load becomes lighter, users can get more resource shares in each scheme.

To summarize Fig. 4, PF achieves absolute fairness but does not consider diverse user demands.



(a) CDF of user utilities under heavy load

(b) CDF of user utilities under median load

(c) CDF of user utilities under light load

Fig. 5: CDF of user utilities with varying traffic loads.

NF only considers varying user demands and ignores fairness. Both DRF and BF consider fairness and user demands at the same time, but their effects are different. BF equally shares one bottleneck resource and allocates another based on user demands, leading to absolute fairness in one resource and a large share gap in another. DRF allocates two resources in a coherent way: it considers diversities in user demands and controls the diversities of allocated shares to some extent.

We then compare the efficiency of these schemes in utilizing network resources. Fig. 5 plots the CDF distribution of utilities under three different traffic loads. Since the three graphs share similar trends, we focus on Fig. 5(b) (i.e., median traffic load) for illustration.

In Fig. 5(c), we find under NF scheme all users have the same utility value (i.e., 0.6). PF performs worst among all schemes: 50% of users have utility value less than 0.6. This is because PF equally distributes resources and does not consider any diversity in user demands. BF and DRF performs better than PF and NF. However, DRF concentrates on the median utility better than BF, indicating that the number of users with too low or too high utilities under DRF scheme is smaller than that under BF scheme. To summarize the results in Fig. 5(b), BF, NF, PF and DRF achieve 50^{th} percentile utility of 0.618, 0.6, 0.568 and 0.622 respectively. Such comparison leads to the conclusion that DRF adapts to the diverse user requests well and maximizes network user utilities with limited resources.

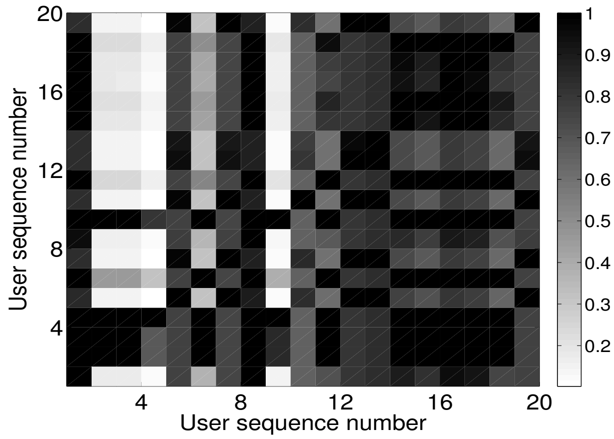


Fig. 6: Verification of QoS-based envy-freeness property of DRF.

B. Validation of QoS-based Envy-Freeness Property of DRF

Here we validate the QoS-based envy-freeness property of DRF. We randomly choose 20 users from 1000 users and record their true QoS requirements. The chosen users are labeled with index from 1 to 20. First, DRF scheme is performed on the whole 1000 users. Next we exchange the allocated resource shares by DRF of user i with that of user j ($i, j = 1, 2, ..., 20$). Note that the exchange in resource share also indicates the exchange in actual QoS performances. Finally we compute the utility of each user who exchanges resource shares with pairs and normalize it over the original utility of this user with unexchanged resource shares. Note that all other users keep their actual resource shares unchanged. The results are plotted in Fig. 6.

In Fig. 6, the x-axis and y-axis represent the user index. A grid at the position (i, j) indicates an exchange in resource share between user i and user j. The color of a grid reflects the utility value: white color reflects low utility and black color reflects high utility. The darker the color is, the higher the utility would be. From Fig. 6 we find that the grids on the diagonal, i.e., grids with position (i, i) ($i = 1, 2, ..., 20$) always have the highest utility value of 1, while other grids with position (i, j) ($i \neq j$, $i = 1, 2, ..., 20$) have lower or equal utility values. This demonstrates that DRF allocates the most optimal resource share to each user and any exchange in resource share between users will not lead to an increased QoS performances. Hence, users will not envy each other for their allocated shares.

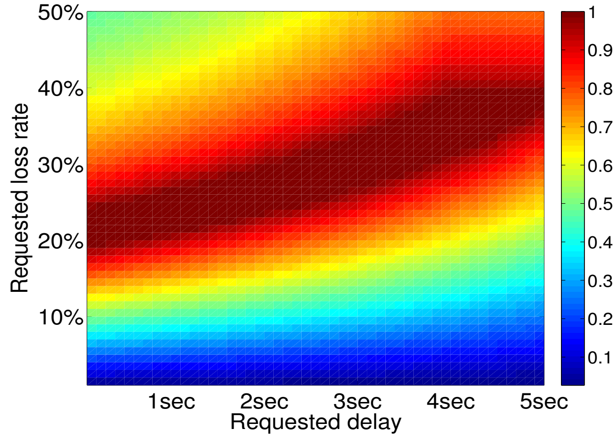


Fig. 7: Normalized utility over QoS requirements of a user that is randomly chosen from 1000 users.

C. Validation of QoS-based Strategy-Proofness Property of DRF

In the following, we validate the QoS-based strategy-proofness property of DRF. We randomly select a user and vary his QoS requirement in the range $[0.1\text{sec}, 5\text{sec}] \times [1\%, 50\%]$, with the granularity $[0.1\text{sec} \times 1\%]$. DRF scheme is performed among the whole 1000 users each time when a different QoS requirement of this user is proposed. All other users keep their QoS requirements unchanged. We compute the utilities of the user with different QoS requests and normalize them over the original utility when the true QoS request is proposed. The true QoS request of this user is $[2.5\text{sec}, 25\%]$. Fig. 7 plots the result.

In Fig. 7, the x-axis is the delay range and the y-axis is the loss rate range. The color indicates the normalized utility value, ranging from 0 to 1. We find in this figure that at the middle point $[2.5\text{sec}, 25\%]$, i.e., the true QoS request, the corresponding color grid has the highest utility value. This finding validates the QoS-based strategy-proofness property.

Note that there exists a linear ribbon region in Fig. 7 where all the normalized utility values are 1. This is because the DRF allocation result is dependent on the resource demand ratio of a user (defined in Section III-C), rather than the separate resource demand in each resource. Consequently, if a user changes his QoS requirements but keeps his resource demand ratio unchanged after the QoS mapping, he will get the same resource share and the same utility performance.

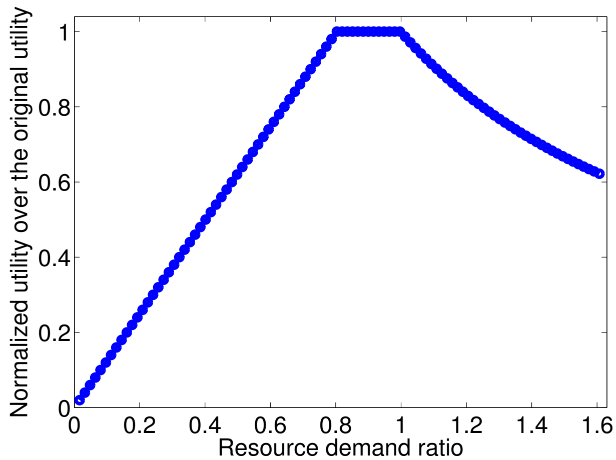


Fig. 8: Normalized utility over resource demand ratio of a user that is randomly chosen from 1000 users.

To view the strategy-proofness property in another perspective, we plot the normalized utility over resource demand ratio in Fig. 8. We fix the loss rate requirement of the user and vary his delay requirement to vary the resource demand ratio in the range $[0.1\text{sec}, 5\text{sec}]$. The true resource demand ratio for the request $[2.5\text{sec}, 25\%]$ is computed as 0.80. In Fig. 8 we observe that with the resource demand ratio increasing, the utility first increases linearly until the true demand ratio (i.e., 0.80) is reached; when the ratio is in the range $[0.80, 1]$, the utility keeps highest and unchanged; when the ratio further increases, the utility starts to drop. We hence conclude that a user cannot get better utility performance no matter how he varies the resource demand ratio, which also validate the QoS-based strategy-proofness property.

D. Performance of Our Admission Control Scheme

In this subsection we explore the impact of the admission control component in our protocol on the network utility performance. If the expected utility of a certain user is lower than a utility threshold U_0 after performing DRF scheme, it would be denied of access. We use the number of users that are admitted to the network to represent the network capacity. In Fig. 9 we observe this metric over varying U_0 which ranges from 0 to 1. When U_0 is 0, all users would be admitted; when U_0 is 1, only users that are completely satisfied with their QoS performances are admitted.

From Fig. 9 we find that the number of admitted users decreases elegantly if U_0 increases.

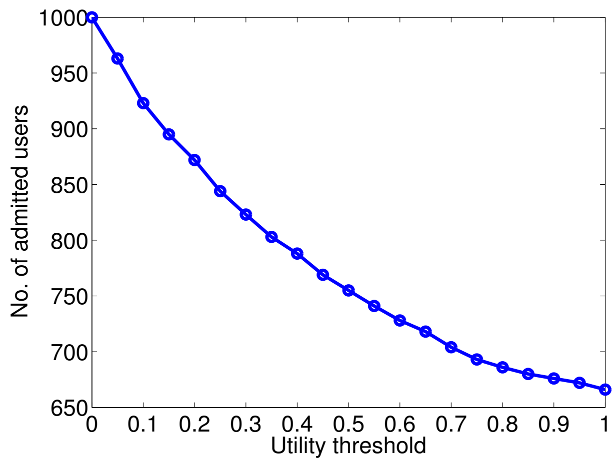


Fig. 9: Impact of the utility threshold on number of users that can be admitted to the network.

In our simulation settings, even if U_0 is 1, there are still more than 650 users that get access to the network. A good tradeoff between network capacity and user utility is achieve if U_0 is set to 0.7, in which case around 700 users are admitted.

VI. Related Work

Fairness is an important metric when performing resource allocation to demanding entities. In current 802.11 WLAN, resources are allocated separately and independently. Many prior works [6, 7, 20, 21] study the problem of allocating single resource in a fair manner. For instance, there are many works study the problem of allocating the resource of wireless bandwidth [8, 9]. Some fair queuing schemes manage to allocate link bandwidth by scheduling packet transmission order in order to achieve fairness in throughput [10–12]. As current mobile applications require multiple types of resources, single resource allocation designs cannot satisfy heterogeneous QoS requirements.

More recently, multiple resources are jointly optimized and scheduled in resource allocation [2–4, 13, 17]. Egi *et al.* applies bottleneck fairness for resource management and packet processing in multi-core software routers, where CPU, memory and NIC resources are considered. Ghodsi *et al.* [2] proposes DRF for multi-resource allocation and apply it in the datacenter environment to allocate CPU and memory. Ghodsi *et al.* [13] suggests a multi-resource fair queuing algorithm called DRFQ to achieve time-based DRF for packet processing in middleboxes.

Joe-Wong *et al.* [3] proposes a unifying framework for multi-resource allocation, balancing the allocation fairness and resource utilization efficiency. Wang *et al.* [4] generalizes the conventional processor sharing scheduling algorithm and propose DRGPS for multi-resource settings. Our work builds on the concept of DRF and allocates network resources in WLAN to support user QoS.

VII. Conclusion

In this paper, we are the first to propose QoS fairness in WLAN and implement a multi-resource allocation scheme to achieve QoS fairness based on Dominant Resource Fairness (DRF) scheme. We define a QoS utility metric which considers both subjective user QoS request and objective QoS performances perceived by users. To perform the multi-resource allocation, we build a mapping model between QoS requirements and resource demands. Based on the QoS utility and the mapping model, we define QoS fairness through several QoS-based fairness properties and prove them analytically. We further prove that these properties are not limited to specific mapping models and are held in general. We design a practical protocol to implement the allocation scheme in WLAN. Extensive simulation results validate that the proposed scheme strikes an optimal balance between QoS and fairness.

References

[1] Y. Hou, M. Li, Y. Zheng, "Fair QoS Multi-Resource Allocation for Wireless LAN", in *Proceedings of IWQoS*, 2014.

[2] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, I. Stoica, "Dominant Resource Fairness: Fair Allocation of Multiple Resource Types", in *Proceedings of USENIX NSDI*, 2011.

[3] C. Joe-Wong, S. Sen, T. Lan, M. Chiang, "Multi-Resource Allocation: Fairness-Efficiency Tradeoffs in a Unifying Framework", in *Proceedings of IEEE INFOCOM*, 2012.

[4] W. Wang, B. Liang, B. Li, "Multi-Resource Generalized Processor Sharing for Packet Processing", in *Proceedings of IEEE/ACM IWQoS*, 2013.

[5] K. Papagiannaki, S. Moon, C. Fraleigh, P. Thiran, C. Diot, "Measurement and Analysis of Single-Hop Delay on an IP Backbone Network", *IEEE Journal on Selected Areas in Communications*, vol. 21, issue 6, pp. 908–921, August 2003.

[6] T. Lan, D. Kao, M. Chiang, A. Sabharwal, "An Axiomatic Theory of Fairness in Network Resource Allocation", in *Proceedings of IEEE INFOCOM*, 2010.

[7] B. Briscoe, "Flow Rate Fairness: Dismantling a Religion", *ACM SIGCOMM Computer Communication Review*, vol. 37, issue 2, pp. 63–74, April 2007.

[8] D. Giustiniano, E.Goma, A. Toledo, L. Dangerfield, J. Morillo, P. Rodriguez, "Fair WLAN Backhaul Aggregation", in *Proceedings of ACM MobiCom*, 2010.

[9] X. Gong, S. Vorobyov, C. Tellambura, "Joint Bandwidth and Power Allocation with Admission Control in Wireless Multi-User Networks with and without Relaying", *IEEE Transactions on Signal Processing*, vol. 59, issue 4, pp. 1801–1813, April 2011.

[10] S. Lu, V. Bharghavan, R. Srikant, "Fair Scheduling in Wireless Packet Networks", *IEEE/ACM Transactions on Networking*, vol. 7, issue 4, pp. 473–489, August 1999.

[11] J. C. R. Bennett, H. Zhang, "WF^2Q: Worst-Case Fair Weighted Fair Queuing", in *Proceedings of IEEE INFOCOM*, 1996.

[12] A. Demers, S. Keshav, S. Shenker, "Analysis and Simulation of a Fair Queuing Algorithm", *ACM SIGCOMM Computer Communication Review*, vol. 19, issue 4, pp. 1–12, September 1989.

[13] A. Ghodsi, V. Sekar, M.Zaharia, I. Stoica, "Multi-Resource Fair Queuing for Packet Processing", in *Proceedings of ACM SIGCOMM*, 2012.

[14] M. A. Marson, M. Gerla, "Fairness in Local Computing Networks", in *Proceedings of IEEE ICC*, 1982.

[15] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function", *IEEE Journal on Selected Areas in Communications*, vol. 18, issue 3, pp. 535–547, March 2000.

[16] D. G. Kendall, "Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain", *The Annals of Mathematical Statistics*, 1953.

[17] N. Egi, A. Greenhalgh, M. Handley, G. Iannaccone, M. Manesh, L. Mathy, S. Ratnasamy, "Improved Forwarding Architecture and Resource Management for Multi-Core Software Routers", *Sixth IFIP International Conference on Network and Parallel Computing*, 2009.

[18] A. Trunganont, V. Visoottiviseth, "Adaptive Wireless Bandwidth Allocation for Per-Station Fairness", in *ISCIT*, 2009.

[19] T. Nandagopal, T. Kim, X. Gao, V. Bharghavan, "Achieving MAC Layer Fairness in Wireless Packet Networks", in *MOBICOM*, 2000.

[20] K. Xie, J. Cao, X. Wang, J. Wen, "Optimal Resource Allocation for Reliable and Energy Efficient Cooperative Communications", *IEEE Transactions on Wireless Communication*, vol. 12, issue 10, pp. 4994-5007, 2013.

[21] X. Wang, H. Schulzrinne, "Pricing Network Resources for Adaptive Applications in a Differentiated Services Network", *IEEE/ACM Transactions on Networking*, vol. 2, pp. 943-952, 2001.

[22] J. Dai, F. Liu, B. Li, B. Li, J. Liu, "Collaborative Caching in Wireless Video Streaming Through Resource Auctions", *IEEE Journal on Selected Areas in Communications*, vol. 30, issue 2, pp. 458-466, 2012.

[23] U. Engelke, H. Zepernick, "Perceptual-based Quality Metrics for Image and Video Services: A Survey", in *NGI*, 2007.

[24] ITU, "Method for objective measurements of perceived audio quality", ITU-R, Rec. BS.1387-1, 2001.

[25] ITU, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow band telephone networks and speech codecs", ITU-T, Rec. P.862, 2001.

[26] ITU, "Methodology for the subjective assessment of the quality of television pictures", ITU-R, Rec. BT.500-11, 2002.

[27] H. Koumaras, A. Kourtis, D. Martakos, "Evaluation of video quality based on objectively estimated metric", *IEEE Journal of Communications and Networks*, vol. 7, no. 3, pp. 235242, 2005.

[28] P. Calyam, E. Ekici, C. Lee, M. Haffner, N. Howes, "A GAP-Model based Framework for Online VVoIP QoE Measurement", *IEEE Journal of Communications and Networks*, vol. 9, issue 4, pp. 446-456, 2007.

[29] R. Schatz, S. Egger, A. Platzer, "Poor, Good Enough or Even Better? Bridging the Gap between Acceptability and QoE of Mobile Broadband Data Services", in *Proceedings of IEEE ICC*, 2011.

[30] T. Hobfeld, M. Fiedler, T. Zinner, "The QoE Provisioning-Delivery-Hysteresis and Its Importance for Service Provisioning in the Future Internet", in *NGI*, 2011.

[31] S. Lu, V. Bharghavan, R. Srikant, "Fair Scheduling in Wireless Packet Networks", *IEEE/ACM Trans. on Networking*, vol. 7, no. 4, pp. 473-489, 1999.

[32] I. Stoica, H. Wahab, J. Gehrke, K. Jeffay, C. Plaxton, S. Baruah, "A Proportional Share Resource Allocation Algorithm for Real-Time, Time-Shared Systems", in *IEEE Real-Time Systems Symposium*, 1996.

[33] V. Bharghavan, S. Lu, T. Nandagopal, "Fair Queuing in Wireless Networks: Issues and Approaches", *IEEE Personal Communications*, vol. 6, no. 1, pp. 44-53, 1999.

[34] T. Ng, I. Stoica, H. Zhang, "Packet Fair Queueing Algorithms for Wireless Networks with Location-Dependent Errors", in *IEEE INFOCOM*, 1998.

[35] A. Banchs, X. Perez, "Distributed Weighted Fair Queuing in 802.11 Wireless LAN", in *IEEE ICC*, 2002.

[36] T. Nandagopal, T. Kim, X. Gao, V. Bharghavan, "Achieving MAC Layer Fairness in Wireless Packet Networks", in *ACM MobiCom*, 2000.