Achieving MAC Layer Fairness in Wireless Packet Networks

Thyagarajan Nandagopal Tae-Eun Kim Xia Gao Vaduvur Bharghavan
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
1308 W. Main Street, Urbana, IL 61801
Email:{ thyagu, tkim, xiagao, bharghav }@timely.crhc.uiuc.edu

Abstract

Link-layer fairness models that have been proposed for wireline and packet cellular networks cannot be generalized for shared channel wireless networks because of the unique characteristics of the wireless channel, such as location-dependent contention, inherent conflict between optimizing channel utilization and achieving fairness, and the absence of any centralized control.

In this paper, we propose a *general analytical framework that captures the unique characteristics of shared wireless channels* and allows the modeling of a large class of system-wide fairness models via the specification of per-flow utility functions. We show that system-wide fairness can be achieved without explicit global coordination so long as each node executes a contention resolution algorithm that is designed to optimize its local utility function.

We present a *general mechanism for translating a given fairness model in our framework into a corresponding contention resolution algorithm.* Using this translation, we derive the backoff algorithm for achieving *proportional fairness* in wireless shared channels, and compare the fairness properties of this algorithm with both the ideal proportional fairness objective, and state-of-the-art backoff-based contention resolution algorithms.

We believe that the two aspects of the proposed framework, i.e. the ability to specify arbitrary fairness models via local utility functions, and the ability to automatically generate local contention resolution mechanisms in response to a given utility function, together provide the path for achieving flexible service differentiation in future shared channel wireless networks.

1 Introduction

In recent years, wireless ad hoc networks have increasingly received critical attention in the networking research community. While most current non-military ad hoc network test-beds are experimental in nature, possible future deployment scenarios include deeply networked conglomerations of embedded devices, emergency rescue operations, "zero conf" meeting setups, and rapidly reconfigurable metropolitan wireless networks. Migrating from experimental environments to commercial environments, ad hoc network designers will need to address critical new challenges, such as "service differentiation" among contending users for the dynamic and scarce channel resources. In a pay-for-use model, the network must provide minimum performance requirements for paying users, at least in relative terms. Since link-layer fairness mechanisms serve as the basis for achieving network-layer quality of service (e.g. Weighted Fair Queueing [1] for the IntServ Guaranteed QoS service model), wireless MAC protocols in commercial ad hoc networks must support some notion of "weighted fairness", wherein flows with larger weights receive correspondingly better service in accordance with a system-wide fairness model.

To this end, *the goal of this work is to formally investigate the fairness properties that can be achieved by the class of multiple access wireless MAC protocols in shared channel wireless networks in general, and ad hoc networks in particular.* We define a "shared wireless channel" as a communication regime wherein all nodes communicate over the same logical channel using decentralized control, and there is no concept of a base station in the MAC layer. Shared wireless channels thus underlie both ad hoc networks and packet cellular networks, and most wireless multiple access protocols [2, 3], including the basic IEEE 802.11 MAC standard [4], are designed with these channel assumptions.

Naturally, the first question to ask is: "Is there something fundamental about the nature of shared wireless channels that prevents us from reusing the wealth of link-layer fairness techniques that have been developed for wireline and packet cellular environments?" It turns out that shared channel wireless networks have three unique characteristics that make it very difficult to achieve, or even consistently define, the notion of fairness:

1. *Spatial (location-dependent) contention for the wireless channel:* Consider a simple channel model where a transmitter has a fixed transmission range, and multiple transmissions in the neighborhood of a receiver will cause a collision at the receiver. Following typical collision avoidance protocols [2, 3, 4], a successful transmission precludes any station in the neighborhood of either the transmitter or the receiver from engaging in another simultaneous packet transmission/reception. In other words, *transmission of a packet involves contention over the joint neighborhoods of the sender and the receiver*, and the level of contention for the shared

wireless channel in a geographical region is *spatially dependent* on the number of contending nodes in the region. This is fundamentally different from wireline and cellular channel models, wherein all flows perceive the same contention.

2. *Trade-off between channel utilization and fairness:* In a shared channel wireless network, *spatial reuse* of the channel bandwidth may be obtained by simultaneously scheduling transmissions whose regions of contention are not in conflict. While spatial reuse is very useful for increasing the utilization of the wireless channel, it introduces a fundamental conflict between optimizing aggregate allocated bandwidth and achieving fairness, because allocating the channel to a flow with a large contention correspondingly reduces the channel reuse. In contrast, wireline and cellular networks do not face this problem because all flows perceive the same contention.

3. *Inaccurate state and decentralized control:* Even if we define the notion of fairness by addressing the two issues above, designing mechanisms for achieving such fairness is a major challenge since there is no centralized control and no station is guaranteed to have accurate knowledge of the contention even in its own neighborhood. Further, contention is really "per-flow" (i.e. a sender-receiver pair) rather than per-node (see point 1 above), which makes its estimation harder at the transmitter before it decides to contend for the channel. Finally, contention resolution must be achieved without assuming any explicit coordination or handshakes among the contenders in order to preserve the robustness of multiple access protocols.

To summarize, wireline and cellular fairness models are inappropriate for our target environment because the first two channel characteristics are fundamentally unique to shared wireless channels, and the third characteristic, though also present in part in cellular networks, is much more pronounced in our communication model.

In the past decade, there has been a wealth of wireless MAC research focusing on the design and evaluation of multiple access wireless MAC protocols. However, to the best of our knowledge, there are no structured studies devoted to the *formal investigation of fairness models in shared wireless channels*, evaluation of competing fairness models, or formal fairness characterizations of existing contention resolution mechanisms in wireless MAC protocols. To address these limitations in the state-of-the-art, this paper proposes what we believe to be the first structured study of fairness models in shared channel multiple access wireless networks.

Our first contribution is a *general analytical framework that captures the unique characteristics of shared wireless channels* identified above, and allows us to reason about a large class of fairness models. Our starting point is the recognition that link-layer fairness in a shared wireless channel has some commonalities with network-layer fairness in multi-hop wireline networks in the sense that different "flows" experience different "contention"[1]. The genesis of our framework lies in the fairness framework for network flows proposed by Gibbens and Kelly [5], which we enhance to address the constraints of link-layer contention resolution in wireless channels. Using this framework, we show that the definition of fairness is equivalent to specifying a "utility function" for the channel allocation for each flow, and that different fairness models can be achieved by enforcing correspondingly different utility functions locally at each contending station without explicit global coordination.

Our second contribution is a *general mechanism for translating a given fairness model into a corresponding backoff-based collision resolution algorithm that probabilistically achieves the fairness objective.* This is a powerful result because it shows that once we model fairness in our framework, the backoff algorithm for achieving this fairness model is automatically derived from the framework! Using this translation, we derive the backoff algorithm for achieving *proportional fairness* [5] in wireless shared channels, as a representative example. We compare the fairness properties of this backoff algorithm to the ideal proportional fairness objective, and fairness properties of IEEE 802.11 MAC, MACAW, and a more recent work called Connection-based Balanced MAC (CB-Fair) [6].

We believe that the two aspects of the proposed framework – (a) the ability to specify arbitrary fairness models (by specifying the corresponding utility function), and (b) the ability to automatically generate local collision resolution mechanisms in response to a given utility function – jointly provide the path for achieving flexible service differentiation in future shared channel wireless networks.

The rest of the paper is organized as follows. Section 2 describes representative fairness mechanisms from related literature, and Section 3 shows inconsistencies in the fairness of IEEE 802.11. Section 4 presents a general analytical framework for specifying fairness in shared channel networks. Section 5 presents the translation procedure from the fairness model to the specific backoff-based contention resolution algorithm, using proportional fairness as an example. Section 6 evaluates the backoff algorithm with reference to the ideal objective as well as related work. Section 7 summarizes the paper.

2 Contention Resolution in Wireless Multiple Access Protocols

In this paper, our investigations focus on the class of shared channel multiple access collision avoidance protocols in the CSMA/CA family. While we will investigate the fairness properties of multiple access protocols for general shared channel configurations, this work is more applicable to ad hoc networks than cellular networks, wherein other techniques for ensuring fairness, such as centralized scheduling at the base station, may be overlaid on top of the multiple access MAC protocol [7]. Fairness in multi-hop wireless networks has also been addressed in a recent related work [8]. This work focuses on maximizing aggregate channel reuse subject to a minimum fairness guarantee, which is very different from the issues we seek to address in this paper.

[1] A link layer flow is between a pair of neighboring nodes, and has a location-dependent contention for channel allocation. A network layer flow is between a pair of end hosts, and has a path-dependent "contention" for network bandwidth.

Wireless multiple access protocols typically have two components that work in concert: (a) collision avoidance and (b) contention resolution. A majority of the wireless multiple access research in the past decade has focused on achieving perfect collision avoidance [2, 3, 9], and we will not address this component further in this paper. For the rest of the work, we will assume some form of efficient collision avoidance that almost completely eliminates non-contention induced collisions [3, 9].

The second component, contention resolution, is the focus of this work. In multiple access protocols, contention resolution has typically been achieved through two mechanisms: (a) *backoff* and (b) *persistence*. In the backoff mechanism, each contending station maintains a "backoff counter" and defers for a random amount of time bounded by the backoff counter prior to a transmission. In the persistence mechanism, each contending station maintains a "persistence probability" and contends for the channel with this probability when it perceives a clear channel. In both cases, multiple simultaneous transmissions in a shared region cause collisions, and the goal is to adjust the backoff counter/ persistence parameters appropriately so that collisions are reduced, and ideally, eliminated. Thus, both the backoff counter and the persistence probability are functions of the estimated contention, and different contention resolution algorithms differ in terms of how they adjust these parameters in response to collisions and successful transmissions. In fact, the key to achieving the desired fairness properties is to map the fairness objective to a corresponding persistence/backoff adjustment algorithm. Deriving this mapping is a challenging exercise, which we will address in Section 5.

Let us now briefly consider three representative contention resolution mechanisms. Most of the multiple access protocols in related literature employ mechanisms that are variants of one or more of these three mechanisms.

Binary Exponential Backoff (BEB): Following the contention resolution scheme of Ethernet [10], early CSMA/CA protocols used a mechanism wherein the backoff counter of a station is doubled upon every contention loss, and reset to 1 upon a successful completion of the packet handshake. Unfortunately, it has been shown that BEB is highly short-term unfair under high offered loads [10]. The IEEE 802.11 MAC protocol adopts a variant of BEB by starting from a "base backoff counter" value for each new packet, and then using binary exponential backoff for subsequent retransmissions of the packet subject to a maximum upper bound (31 and 1023 respectively for DS PHY) [4]. While this alleviates the short-term unfairness of BEB to some extent, it induces more collisions and unpredictable fairness properties in the presence of heavy contention (Section 3 investigates these issues further).

Multiplicative Increase/Linear Decrease (MILD) with Backoff Copy: MACAW uses a more sophisticated contention resolution algorithm, wherein it tries to "assign" contention losses to each station and then computes a backoff "per-flow" rather than per-node. The backoff adjustment algorithm at a station maintains a backoff counter per-flow, doubles the backoff counter for the flow on each loss assigned to the station, and reduces the backoff counter by 1 for each successful transmission. MACAW also has a "backoff copy" mechanism wherein transmitters advertise their backoff counters in their packets, and idle stations snoop this backoff to keep track of the current contention estimates in their region [2]. The stated goal of MACAW is to set the backoff counter of a flow proportional to the contention experienced by the flow [2]. Unfortunately, this ends up being highly unfair to flows experiencing relatively high contention in asymmetric network configurations because such flows experience contend less aggressively while also experiencing contention from a larger number of flows (examples 2 and 3 in Section 6 illustrate this phenomenon).

Combining Persistence and Backoff: In Connection-based Balanced Medium Access (CB-Fair) [6], the authors design a contention resolution algorithm that combines persistence and backoff. A transmitter contends with a persistence probability that is proportional to the ratio of the degree of the receiver to the maximum degree in the transmitter's neighborhood (more generally, the persistence probability is a function of the transmitter's second neighborhood as well as the receiver's neighborhood). The backoff adjustment algorithm doubles the backoff counter on each loss and halves the backoff counter on each successful transmission. CB-Fair also uses backoff copying similar to MACAW. It turns out that the multiplicative nature of both the increase and decrease of backoff causes short-term unfairness similar to BEB. Further, the persistence algorithm tries to increase the persistence of highly contending flows, which leads to artificially higher backoffs and highly inconsistent short-term behavior over different time windows (examples 2 and 3 in Section 6 illustrate this phenomenon).

We have only presented brief high level descriptions of these three algorithms (though we have simulated the algorithms in detail for the purposes of comparison in Section 6) because what we really want to show is that, regardless of the details, these mechanisms are based on somewhat ad hoc reasonings about fairness rather than a structured realization of a formally defined fairness model. In the next section, we explore the fairness properties of the IEEE 802.11 MAC protocol standard in more detail, using it as a representative study. We show that because it does not have a clearly defined fairness model, the protocol has inconsistent fairness behavior over different time windows and is, in fact, both short-term and long-term unfair. Both MACAW and CB-Fair also exhibit similar characteristics (see Section 6 for examples). This motivates our approach, which starts with an arbitrary fairness model and ends with a contention resolution algorithm design that achieves the fairness model probabilistically.

3 Fairness Characteristics of the IEEE 802.11 MAC Protocol

We now illustrate the fairness properties of IEEE 802.11 through three tests in the *ns-2* simulator [11]. The simulation scenarios consist of both simple hand-configured topolo-

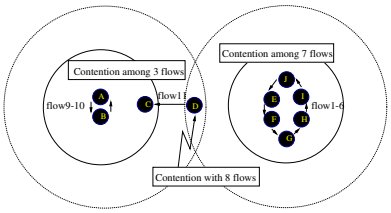

Figure 1: Topology - Asymmetric Contention

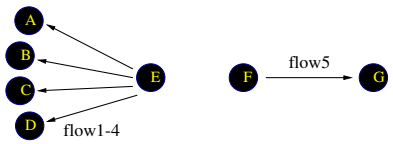

Figure 2: Topology: Per-flow vs. per-node fairness

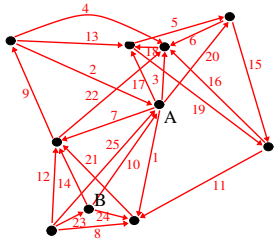

Figure 3: Network topology where all flows share the channel

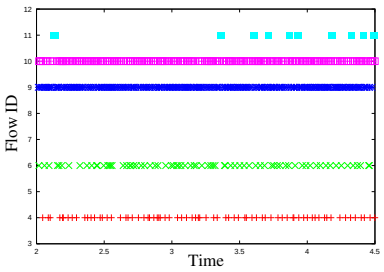

Figure 4: Packet sending timing: Flow11 exhibits short-term unfairness

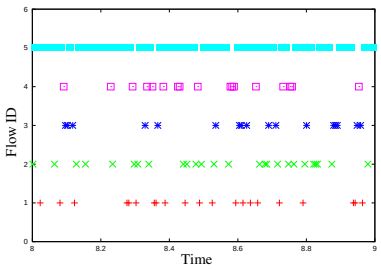

Figure 5: Packet sending timing: Flow 5 receives more bandwidth than other flows. Flows 1 - 4 exhibit short-term unfairness

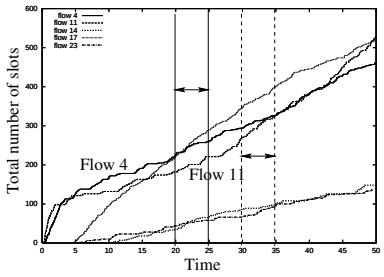

Figure 6: Fairness in shared channel: Flows 4 and 11 experience varying throughput at different intervals

gies and more complex randomly generated topologies. The carrier sense/data reception threshold ratio is set appropriately to eliminate all hidden/exposed sender/receiver problems [9], and no random channel loss is induced – thus all losses are due to contention. Each flow has enough packets to individually saturate the channel for the duration of the simulation to simulate the highly loaded case. The wireless channel bandwidth is 2 Mbps (approximately 180 packets per second for 1 KB packets). The duration of each simulation is 10 seconds. We will show the channel allocation over small 1-2 second time windows as well as over the entire simulation in order to observe both the short-term and long-term fairness characteristics.

We perform three tests. In the first test, we observe the inconsistency of short-term fairness in IEEE 802.11 in a simple scenario where contention in a neighborhood is asymmetric, i.e. nodes experience location-dependent contention. In the second test, we observe the unfairness in per-node versus per-flow fairness in IEEE 802.11 in a simple scenario where one node is communicating with many other nodes. In the third test, we illustrate both the short-term and long-term unfairness in IEEE 802.11 in a more complex randomly generated topology. The conclusion of this study is that the 802.11 MAC protocol is unable to achieve any consistent notion of fairness, and can be shown to be both short-term and long-term unfair.

3.1 Asymmetric Contention Neighborhoods

Consider the topology shown in Figure 1. The receiver threshold range is denoted by a solid circle, and the dotted circle denotes the carrier-sensing range. All flows in 1 – 6 contend with each other and with flow 11. Flows 9 and 10 contend with flows 9 - 11. Flow 11 contends with all the flows in the topology. The contention resolution algorithm is the modified binary exponential backoff described in Section 2. Figure 4 shows the channel allocation for the time window[2,4.5] for a few flows. The key point to note is the variation in the channel allocation for flow 11. Specifically, the throughput ratio $R_{1-6} : R_{11} = 6.5$ in [2,3.5], and $R_{1-6} : R_{11} = 2.5$ in [3.5,4.5]. In more complex simulation scenarios, asymmetric contention can cause larger variations in short-term and long-term fairness, as see from the third example as well as the examples in Section 6.

3.2 Per-flow versus Per-node fairness

While protocols such as MACAW and CB-Fair use per-flow queues with per-flow backoffs (thereby effectively treating a node with multiple flows as multiple co-located nodes with one flow each), IEEE 802.11 uses a per-node queue with a per-node backoff. This has two effects: (a) in scenarios with asymmetric contention, a head-of-line packet to a receiver in a high contention neighborhood can block other flow transmissions to lightly loaded neighbors, thereby distorting the fairness characteristics of the network as a whole, and (b) a node with many flows penalizes its flows unfairly. Figure 2 shows a simple scenario wherein all flows contend with each other, and Figure 5 shows the channel allocation for this scenario. It is important to note that we are not arguing about whether this specific fairness model is desirable or not. It may well be that the system fairness policy mandates per-node rather than per-flow fairness. The issue here is that IEEE 802.11 cannot achieve weighted fairness in a reason-

able way, even if we try to weight the backoff increase. In other words, the protocol is implicitly able to provide only per-node fairness, and that too falls apart in the presence of asymmetric contention neighborhoods. This behavior is illustrated in the next example.

3.3 Randomly Generated Topology

Figure 3 shows a randomly generated topology consisting of 10 nodes spread over a 400m by 400m grid in a way that the network is not a clique but only one flow can transmit at a time. There are 25 flows for which the source and receiver are within the receiving threshold range, and all flows start between 0 to 4 seconds and run for 50 seconds. Figure 6 shows the channel allocation for a few selected flows in this simulation. This example illustrates both short-term and long-term unfairness, both per-node fairness and per-flow fairness. Specifically, simply looking at the throughput evolution of flows 4 and 11 in two time intervals [20,25] and [30,35] illustrates the different fairness characteristics observed in different time intervals. Per-flow unfairness in both the short-term and the long-term is obvious from the figure. The allocation also happens to be per-node unfair, because node A receives significantly higher channel allocation than node B, even though every node has the same contention region.

This example illustrates that even for a relatively simple scenario of a few nodes dispersed in a geographical region with relatively small asymmetry in contention, the fairness of the IEEE 802.11 MAC protocol starts to fall apart. We show in Section 6, and describe in detail in [12] , that both MACAW and CB-Fair suffer from the same problems. The fundamental issue here is not whether any of these protocols achieves a certain definition of fairness in a given configuration over a given time window. The real problem is that none of the wireless multiple access protocols that we have come across in recent literature have a well defined formal definition of the fairness model that they seek to achieve. Thus, it is difficult to quantify their fairness properties for arbitrary topologies. In the next section, we seek to move the design of fairness mechanisms from arcane art to well defined science.

4 A General Analytical Framework for MAC Fairness

In wireline and cellular environments, all flows experience the same contention. Specifying a fairness model is thus fairly straight-forward, e.g. the fairness model of WFQ [1] states that over any arbitrarily small time window, each backlogged flow receives channel allocation in proportion to its weight. Unfortunately, simply using flow weights to determine the channel allocation in shared wireless channels does not adequately capture the unique characteristics of the shared channel. Recall from Section 1 that because of spatial location-dependent contention, different flows have different resource utilization for transmitting the same amount of data. Specifically, flows that experience more contention will "shut up" more contending flows while they are transmitting. Consequently, a general framework for modeling fairness for shared wireless channels must provide the ability to capture the location-dependent nature of contention. Within this framework, each fairness model can then trade-off aggregate channel utilization and fairness according to its own optimization criteria.

We now present the general analytical framework for modeling MAC layer fairness in shared wireless channels. Modeling fairness in this framework is a 4-step process.

1. Step 1: From the network topology, we generate an undirected graph that captures the neighborhood property of nodes (i.e. nodes that are within range of each other are neighbors).

2. Step 2: From the graph and the set of active flows (wherein an active flow is a transmitter-receiver pair that has packets to send), we generate a *flow contention graph* that captures the contention among flows (i.e. each flow is a node in this graph and two flows that contend for the same channel region are neighbors).

3. Step 3: From the flow contention graph, we generate a *resource constraint graph* that represents each "distinct contention region" as a resource server and each flow as a client. While the precise generation of the graph will become clear later, it is sufficient to know at this point that at most one flow can transmit at any time in a distinct contention region, and that a flow can transmit only if it is the sole transmitter in all the distinct contention regions to which it belongs.

4. Step 4: Given a resource constraint graph, we will show that achieving fairness in the system is equivalent to solving a utility maximization problem subject to the transmission constraints described above in the resource constraint graph. The fairness model of the system is determined by the utility function that must be maximized.

The high level overview of our approach is that Step 3 captures the location-dependent contention constraints of the shared wireless channel, and Step 4 optimizes the fairness model-specific utility function subject to these constraints. It will turn out that the formulation of the optimization problem will naturally lead to the ideal fairness objective, as well as translate automatically to a backoff-based contention resolution algorithm. With this overview, we will now describe the steps leading from modeling of fairness to the design of the backoff algorithm.

4.1 Steps 1, 2, and 3

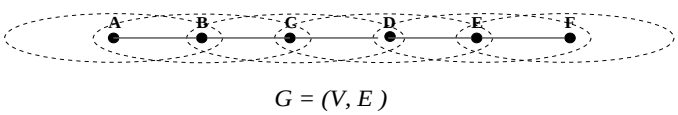

Figure 7: Network graph G: No two neighbors can simultaneously transmit

In Step 1, the network is represented as an undirected graph $G = (V, E)$. V is the set of nodes in the network. $e = (u, v)$ is an edge in E iff nodes u and v are within range

of each other[2]. Figure 7 shows an example, wherein nodes $A - F$ are hosts, and the range of each flow is denoted by the dotted ellipses around the host.

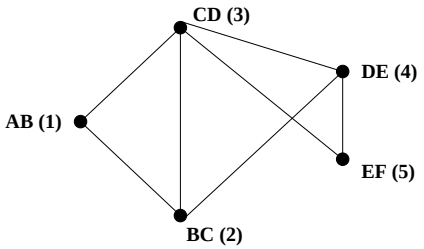

Figure 8: Flow-contention graph G': In every clique, only one node can transmit

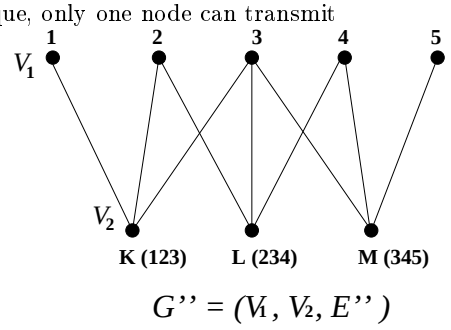

Figure 9: Resource contention graph G'': Each maximal clique is a server and each flow is a client.

In Step 2, we consider all the "active flows" in the network, i.e. transmitter-receiver pairs that have packets to send. Let us assume in Figure 7 that all links are active flows. Note that when BC is active, AB, CD, and DE are constrained not to transmit or receive simultaneously, because at least one of the two end-points for each of these flows is a neighbor of either B or C. Thus any two active flows that are within a distance 2 in G contend with each other[3].

We generate the flow-contention graph $G' = (V', E')$. $V' \subset E$, i.e. each node in G' is a link with packets to transmit in G. $e' = (u', v')$ is an edge in E' iff $d_G(u', v') \leq 1 + \delta$, where $d_G(e, e')$ is the shortest distance between the two links e and e' in graph G. Figure 8 represents the flow contention graph for the graph in Figure 7.

Let us now consider the maximal cliques in G'. A *maximal clique* C in G' is a maximal complete subgraph of G' (e.g. in Figure 8, nodes AB, BC and CD form a maximal clique). A maximal clique in the flow-contention graph represents a "distinct contention region" because at most one node in the clique can transmit at any time, and adding any other node to the maximal clique will enable two non-colliding simultaneous transmissions among the nodes under consideration.

In Step 3, we generate the resource contention graph G''. $G'' = (V1, V2, E'')$ is a bi-partite graph such that $V1 = V'$, and each node in $V2$ represents a maximal clique in G'. $e'' = (u'', v'')$ is an edge in E'' iff $u'' \in V1$, $v'' \in V2$, and u'' belongs to the maximal clique in G' represented by v''. Figure 9 represents the resource contention graph for the flow-contention graph shown in Figure 8.

Each maximal clique in G' represents a "channel resource" with the nodes in the clique contending for exclusive access to the resource. Let us consider a simple slotted model of channel allocation. We represent each node in $V2$ as a server which grants a *token* in each slot to at most one of the edges that is incident on it. Then, the allocation of a token in a clique represents the successful transmission by a node within the clique, and a node in $V1$ (i.e. a flow in G) accesses a channel slot successfully if and only if it simultaneously obtains a token from all the edges incident on it.

Let us now consider a time window $[0, T]$. Let $I_{i,j}(t)$ be an indicator function such that $I_{i,j}(t) = 1$ if the node $j \in V2$ allocates its token in slot t to node $i \in V1$, and $I_{i,j}(t) = 0$ otherwise. Let $x_i(t)$ be the channel allocation for flow i in time $[0, t]$. Then the channel allocation problem can be represented as a set of the following linear constraints.

$$\forall j, \quad \sum_i I_{i,j}(t) \leq 1, \quad \forall t$$

$$\begin{aligned} \forall t, \forall i, \; x_i(t) &= x_i(t-1) + 1, \quad if \; I_{i,j}(t) = 1, \forall (i,j) \in E'' \\ &= x_i(t-1), \qquad otherwise. \end{aligned}$$

Note that this set of constraints captures the location- dependent contention characteristics of shared wireless channels.

4.2 Step 4: Modeling Fairness

In the rest of this section, we will only deal with unweighted fairness for simplicity of explanation. We will see at the end of the section that we can easily extend our discussions to achieve weighted fairness.

Consider a *utility function* $U(r)$ for a channel allocation rate r that is continuous, differentiable, increasing, and strictly concave over the range $r \geq 0$. If $U(r)$ is concave, then flows with a lower channel allocation rates will have a higher marginal utility than flows with higher channel allocation rates. If the flow contention graph is a complete graph, $\sum_i U(r_i)$ is maximized when $\forall i, j, \; r_i = r_j$, i.e. aggregate utility is maximized when the channel allocation rate for every flow is the same. For this simple case, it is easy to see that a maximizing the sum of concave utility functions achieves fairness.

It has been formally shown in [13], that there is a general equivalence between maximizing concave utility functions and achieving some system-wide notion of fairness. Specifically, each concave utility function achieves a corresponding fairness model. Thus, any fairness objective in shared wireless channels can be modeled by the following system of equations:

$$Maximize \sum_i U(r_i), \quad where \;\; r_i = \Delta x_i(t)/\Delta t,$$

[2] As in collision avoidance based CSMA/CA protocols, we assume that links are bidirectional

[3] More generally, if the *carrier sense threshold/packet reception threshold* $= \delta$, then any two flows that are within a distance of $1 + \delta$ potentially contend with each other.

subject to

$$\begin{aligned} \forall j,\ \sum_i I_{i,j}(t) &\leq 1, \quad \forall t \\ \forall t,\ \forall i,\ x_i(t) &= x_i(t-1)+1,\ \ if\ I_{i,j}(t)=1, \forall (i,j)\in E'' \\ &= x_i(t-1), \qquad otherwise. \end{aligned}$$

The constraints in the above system are always satisfied for a central channel allocation scheme. However, our goal is to achieve fully decentralized channel allocation. In this case, each flow controls its own rate allocation, which it adjusts in response to success or failure feedback. Now consider that flow i has a locally adjusted channel allocation rate of r_i. In every distinct contention region $j \in V2$, the "loss probability" $P_j \leq (\frac{\sum_{\{i|(i,j)\in E''\}} r_i - 1}{\sum_{\{i|(i,j)\in E''\}} r_i})^+$. In other words, if the sum of the channel allocation rates in a maximal clique exceeds 1 slot, then some of the contenders will experience contention loss. Further, the contention loss probability that is experienced by any flow i is $p_i = 1 - \Pi_{\{j|(i,j)\in E''\}}(1-P_j)$. In other words, the probability of a flow successfully accessing a channel slot is the product of its success probabilities over all the distinct contention regions to which it belongs. If the loss probability in each distinct contention region is small, i.e. $P_j \to 0$, then the flow loss probability becomes $p_i = \sum_{\{j|(i,j)\in E''\}} P_j$. Note that we do not need to explicitly compute the loss probability in each distinct contention region. Each flow simply monitors its own loss probability that is a function of the sum of the loss probabilities over all the contention regions that it belongs to, and the flow's channel allocation rate as a function of its own success/failure feedback, as we will see later.

For a flow with a channel allocation rate r_i, utility function $U(r_i)$, and perceived contention loss probability p_i, we represent its objective function as

$$maximize\ \ J(r_i) = \alpha U(r_i) - \beta p_i r_i \tag{1}$$

where α and β are system parameters that respectively represent the "utility constant" and the "penalty constant" and can be tuned to achieve the desired trade-off between maximizing utility and minimizing loss rate. It has been shown in [14] that for a system of equations of this form, the aggregate utility $\sum_{i\in V1} J(r_i)$ is maximized when each individual flow i maximizes its own objective function $J(r_i)$. Further, as the penalty constant β becomes large, the optimal value of $\sum_{i\in V1} J(r_i)$ converges to $\alpha \sum_{i\in V1} U(r_i)$, i.e. the fully decentralized solution also converges to a channel allocation scheme that maximizes the aggregate utility over all the flows.

Having shown that the distributed scheme converges to the optimal aggregate utility, what remains is to derive the general framework for the adaptation of the channel allocation rate in each flow. Note that the flow objective function $J_i(r_i)$ is maximized when

$$\begin{aligned} \frac{dJ(r_i)}{dr_i} = 0 \quad &\Longleftrightarrow \quad \alpha U'(r_i^*) - \beta p_i = 0 \\ &\Longleftrightarrow \quad \alpha - \frac{\beta p_i}{U'(r_i^*)} = 0 \end{aligned} \tag{2}$$

where the optimal channel allocation rate is denoted by r_i^*. Since $U(x)$ is concave and the penalty function is linear, $J(r_i)$ is a unimodal function. This suggests that the following scheme can be used for oscillating the channel allocation rate about the optimal point:

$$\dot{r}_i = \alpha - \frac{\beta p_i}{U'(r_i^*)} \tag{3}$$

where $\dot{r}_i$ is the rate of change of channel allocation rate. It is easy to see that this scheme has an equilibrium point that is equal to the optimal value r_i^* that satisfies $\frac{dJ(r_i)}{dr_i} = 0$. It can be proved that the unique maximizing point is also a stable point of the system as follows. Note that

$$\begin{aligned} \frac{dJ(r_i)}{dt} &= \sum_{i\in V_1} \frac{dJ(r_i)}{dr_i}\,\frac{dr_i}{dt} \\ &= \sum_{i\in V_1} U'(r_i)(\alpha - \frac{\beta p_i}{U'(r_i)})^2 \\ &\geq 0 \end{aligned} \tag{4}$$

since the derivative of the strictly concave, increasing function $U(r_i)$ is always positive. Thus, $J(r_i)$ is also a Lyapunov function for the channel allocation scheme given in Equation (3). Therefore the unique maximizing value of $J(r_i)$ to which the flow converges is a stable point of the system. This results in the following proposition:

Proposition 1 *Let all the flows in the system have the utility function $U(r_i)$. Then, if the channel allocation rate r_i for each flow changes according to the following algorithm*

$$\dot{r}_i = \alpha - \frac{\beta p_i}{U'(r_i)} \tag{5}$$

the network utility and the flow utility functions converge to their optimal value, and the system converges to the optimal point. Further, for weighted fairness, we can replace α by $w_i.\alpha$ which results in an optimal rate allocation that is a function of both the weight and the utility function.

The choice of the utility function for flows determines the fairness model for the system. The general form of flow utility functions is $U(r_i) = -1/r_i^\nu$ for $\nu > 0$. Among these, three fairness models have been of particular interest to the research community. When $\nu = 1$, the utility function $U_i(r_i) = -1/r_i$ characterizes the minimum potential delay fairness model [15]. For the special case when $\nu = 0$, the utility function is given by $U_i(r_i) = \log r_i$ and this is associated with proportional fairness [5]. The third is the max-min fairness model, and it has a very different utility function. For a normalized allocation x_i, it is characterized by [15] $U_i(x_i) = -(-\log x_i)^\alpha \quad 0 < x_i < 1, \ \alpha \to \infty$.

It should be clear at this point that from Proposition 1, we can derive the contention resolution algorithm as a function of the utility function. In the next section, we will present the realization of a simple, robust, and local contention resolution algorithm that achieves proportional fairness, as a representative example of the use of our fairness framework for translating an arbitrary fairness model to a corresponding contention resolution mechanism.

5 Proportionally Fair Contention Resolution

The general analytical framework for fairness presents a very powerful result – for any arbitrary fairness model that is represented by a concave differentiable utility function, we derive the corresponding channel allocation rate adaptation algorithm from Proposition 1. In this section, our goal is to start from this point and end up with a fully decentralized, purely local contention resolution algorithm that requires no explicit coordination among flows. Specifically, we seek to derive the contention resolution algorithm for achieving *proportional fairness*, as an example.

5.1 Proportional Fairness and Rate Control

A vector of channel allocation $r = (r_i, i \in V_1)$ is *proportionally fair* if it is feasible (i.e. $r \geq 0$ and $\sum_{i \in j} r_i < 1 + \epsilon$ for every $j \in V2$)) and if for any other feasible vector r', the aggregate of the proportional changes is not positive [5]:

$$\sum_{i \in V1} \frac{r_i' - r_i}{r_i} \leq 0.$$

Specifically, for the proportionally fair rate allocation vector r,

$$\sum_{i \in V_1} \frac{dr_i}{r_i} = 0$$

which translates to the maximization of the utility function $U(r) = log(r)$. *Thus proportional fairness is represented by the utility function* $log(r)$.

Now, substituting for $U'(r_i) = log'(r_i) = 1/r_i$ in Proposition 1 from the previous section, the channel allocation rate adaptation algorithm is given by

$$\dot{r}_i = \alpha - \beta p_i r_i. \tag{6}$$

Combining the results in the previous section and this section, we have now shown how to start with a fairness model (e.g. proportional fairness), convert it to a corresponding utility function, then use the framework to generate a channel allocation rate adaptation function. What remains is to use this rate adaptation algorithm to derive a contention resolution mechanism.

5.2 PFCR: A local mechanism for Proportional Fair Contention Resolution

Recall that our goal is to design contention resolution mechanisms within the commonly followed guidelines of multiple access protocols. We thus have two instruments available to us: (a) persistence, and (b) backoff. Using these two instruments, we need to achieve a channel allocation rate adaptation algorithm of $\dot{r}_i = \alpha - \beta p_i r_i$.

There are two important observations that we make about this adaptation equation. First, the contention loss probability for each distinct contention region is very small because flows will adapt their channel allocation rate when they observe loss – the penalty constant β can be tuned to make the throttling-upon-loss aggressive. Second, considering a single clique in isolation, every contending flow must observe the same loss probability for the derivation in the previous section to hold. Based on these observations, we obtain two results:

- The first observation implies that the sum of the channel allocation rates in any maximal clique will be close to 1. In other words, *we can approximate the channel allocation rate by a persistence probability, using which each flow decides whether or not to contend for a slot.*
- The second observation implies that *all flows must have the same backoff bound for fair contention loss distribution in a clique.*

In concert, these two results naturally define a contention resolution algorithm wherein a flow contends for a channel with a persistence probability that adapts according to the equation in proposition 1, and a contending flow defers for a random time bounded by a system-wide backoff counter before commencing transmission. Of course, this contention resolution algorithm co-exists with standard collision avoidance algorithms which preclude contention if the carrier is busy or if some other flow in the contention region has already acquired the channel.

An interesting feature of this algorithm is that it moves the burden of contention adaptation away from the backoff mechanism and into the persistence mechanism. This feature enables weighted versions of the contention resolution mechanisms to better approximate the ideal model, because weighted persistence is easier to achieve than weighted backoff [12]. Of course, the fact that persistence is highly adaptive to loss implies that in any contention region, there are very few flows (expected value of 1) contending for the channel at any time, and hence a fixed backoff algorithm is both efficient and robust.

5.2.1 Protocol Details

There are three states in which a flow can exist:
(a) *NO_CONTEND*, (b) *CONTEND* and (c) *ACQUIRE*. At the beginning of each slot, all flows are in the *NO_CONTEND* state. There are two steps involved acquiring a channel for a particular slot: (a) a flow decides to contend for a slot (*NO_CONTEND* $\rightarrow$ *CONTEND*), and (b) if it decides to contend, it tries to acquire the channel (*CONTEND* $\rightarrow$ *ACQUIRE*). The pseudo-code for the PFCR algorithm is illustrated in Figure 10.

Each flow has a transmission probability x_i. When a flow has a packet to transmit and does not sense a carrier, it moves from the *NO_CONTEND* to the *CONTEND* state with a probability of x_i (Lines 2 - 4). Each contending host chooses a wait time of B_i, where B_i is uniformly distributed in the interval $[0, B]$. B is a system-wide backoff counter. After the waiting time, the flow senses the carrier. If the channel is free, then the flow tries to acquire the channel (Lines 6 - 8). If either the channel is busy or there is collision, the flow declares the slot as lost (Lines 9, 10, 12), and reduces its transmission probability by βx_i (Lines 10, 13). Note that the notion of contention loss subsumes collisions in that a contending flow also declares a loss if it did not win the contention round (e.g. if it perceives a busy carrier at the end of the waiting period). This is consistent with the derivation in the analytical framework in Section 4. If the flow successfully acquires the channel, it moves from

```
    contend_and_acquire_slot()
1   for each slot
2       State ← NO_CONTEND;
3       if uniform( 0, 1 ) ≤ x_i
4           State ← CONTEND;
5           B_i = uniform( 0, B );
6           wait ( B_i );
7           if carrier_sense( ) == FREE
8             acquire_channel( );
9             if acquire_status( ) == COLLISION
10              x_i ← x_i ( 1 − β );
11            else State ← ACQUIRE;
12          else
13            x_i ← x_i ( 1 − β );
14      x_i ← x_i + α;
```

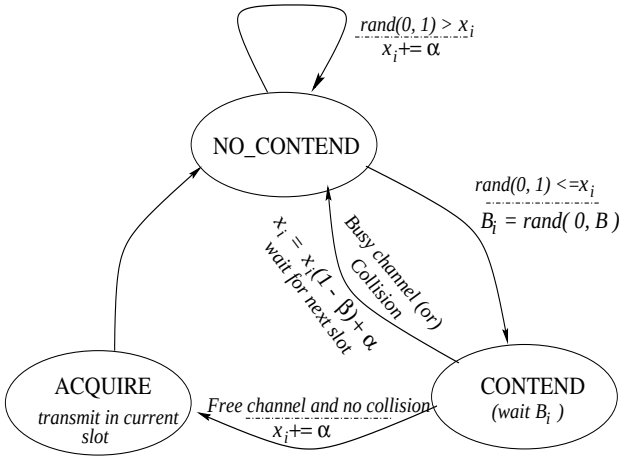

Figure 10: Pseudo-code and State Transition Diagram for the PFCR Algorithm

the *CONTEND* state to the *ACQUIRE* state (Line 11). At the end of the transmission slot, all flows that have packets to send increase their transmission probability by α (Line 14). Since a flow experiences a contention loss with probability p_i, the expected change in the channel allocation rate is $\alpha - \beta p_i x_i$. This exactly reflects the adaptation algorithm in Proposition 1 of Section 4.

The algorithm described above is executed independently for each flow and thus is fully decentralized. Flows do not need to have any knowledge of the topology of the network, and adaptation to the dynamics of the flows, channel conditions and topology occurs implicitly (via increased or decreased contention loss feedback). Most importantly, the contention resolution algorithm is derived naturally from the framework, and is very general. For any given concave and differentiable utility function, the contention resolution algorithm that achieves the corresponding fairness property is automatically generated.

5.2.2 Practical Considerations

While the analytical framework guarantees stability and fairness in a probabilistic sense, a number of practical issues need to be addressed in order to make the contention resolution algorithm robust, efficient, and closely approximate the expected fairness model. The most important of these is the selection of the α and β parameters.

In order to approximate the ideal fairness model perfectly, $\beta \to 1$ because a large penalty constant ensures that the sum of the transmission probabilities in a distinct contention region never exceeds 1. This has two effects. First, setting a large β value causes large oscillations in short-term fairness due to large fluctuations in the persistence. Second, since the transmission probabilities must never exceed 1, α must be small. However, this reduces efficiency under low loads. Fundamentally, there is a trade-off between achieving higher efficiency ($\alpha \uparrow$, $\beta \downarrow$) and achieving better approximations to the ideal fairness objective ($\alpha \downarrow$, $\beta \uparrow$). In our simulations, we set $\alpha = 0.1$, $\beta = 0.5$, and $B = 32$.

6 Performance of the PFCR algorithm

In this section, we present simulation results to compare the fairness characteristics of the Proportional Fair Contention Resolution (PFCR) algorithm, with some existing medium access protocols that were discussed earlier in Section 2. In particular, we have chosen three protocols with different contention resolution schemes: IEEE 802.11 [4], MACAW [2] and CB-Fair [6], for comparison. IEEE 802.11 uses binary exponential backoff with backoff reset for each new packet. MACAW uses contention measurement at the sender and the receiver with multiplicative increase and linear decrease as well as backoff copy. CB-Fair [6] uses multiplicative increase and multiplicative decrease for backoff with copying, and adjusts its persistence based on the characteristics of the second neighborhood of the sender. To show the contention among various flows better, we have chosen to illustrate the topology using the flow-contention graph (G') and the resource contention graph (G'') for each of the simulation scenarios. For each flow, we present the bandwidth obtained under different protocols, and evaluate them against the ideal proportionally fair rate allocation for that topology. At the outset, we recognize that proportional fairness is the appropriate benchmark objective only for PFCR. However, none of the other approaches has a well defined fairness objective and they do not match any of the well known fairness objective functions such as max-min, proportional, or min-potential delay fairness.

FlowID	802.11	MACAW	CB-Fair	PFCR	IDEAL
1	22253	22550	22550	22501	22460
4	23009	22491	22334	22525	22460

Table 1: Example 1: Total number of packet transmitted

Example 1 We first look at a very simple ring topology, as in Figure 11. All flows have the same number of contending flows. We present the throughput results for each protocol in Table 1 along with the ideal throughput expected under a proportional fair system. We also show the relative performance of the various protocols with respect to the proportionally fair throughput in Figure 12. For this scenario, binary exponential backoff performs worse than the other

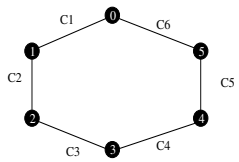

Figure 11: Flow contention graph(G') for example 1

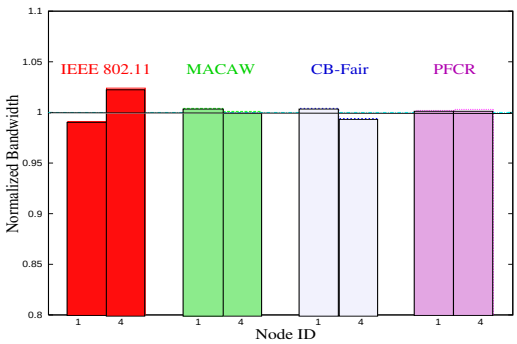

Figure 12: Example 1 - Relative throughput compared to Proportional Fair allocation

protocols, although all protocols obtain within 98% of the desired throughput.

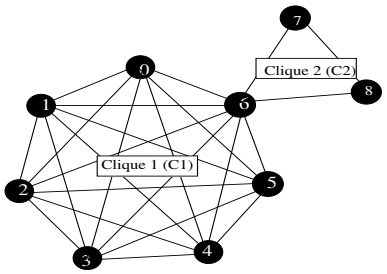

Figure 13: Flow contention graph(G') for example 2

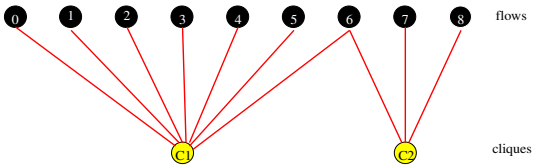

Figure 14: Contention clique graph(G") for example 2

Example 2 In the next example, we consider two flow cliques, shown in Figures 13 and 14, with one clique having a small number of flows and the second clique having a large number of flows. The throughput and relative performance are shown in Figures 15 and 16 respectively. The throughput results show us something very interesting. Flow 6 is part of two cliques, and as a result, it has to win in both the cliques to transmit. In such cases, binary exponential backoff and contention measurement mechanisms cause short-term unfairness as a result of asymmetric contention. Flow 6 receives a rate that is 16% less than the proportional fair share under IEEE 802.11; it receives 54% less than the ideal under MACAW. The bandwidth lost by flow 6 is used by the other flows in both the cliques, which obtain above-the-ideal bandwidth. In the case of CB-Fair, however, flow 6

FlowID	IEEE 802.11	MACAW	CB-Fair	PFCR	IDEAL
5	7645	7834	9076	7865	7633
6	4825	2683	10322	5357	5724
7	23543	23979	15433	22900	22898

Figure 15: Example 2: Total number of packet transmitted

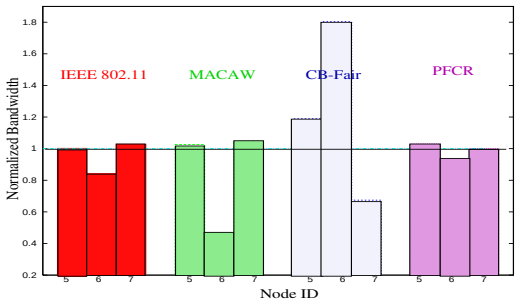

Figure 16: Example 2 - Relative throughput compared to Proportional Fair allocation

has the highest degree among all flows, and hence, it receives more than ideal throughput, while other flows receive lesser throughput. However, the PFCR algorithm is able to match the proportionally fair allocation for every node, in particular, the rate for flow 6 is within 6% of the ideal throughput.

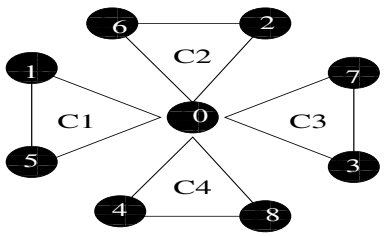

Figure 17: Flow contention graph(G') for example 3

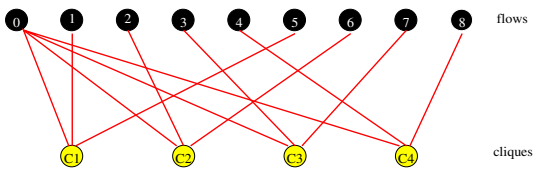

Figure 18: Contention clique graph(G") for example 3

Example 3 In this example, we illustrate the case when there are many flow-cliques in the network and one flow is part of all the cliques, as in Figure 17. Figure 18 shows that flow 0 belongs to all the cliques while all other flows are just part of a single clique. The results are presented in Figure 19 and Figure 20. As in the previous case, the PFCR algorithm does not exhibit any short-term unfairness and follows the ideal proportional fair allocation for all flows, in particular, for flow 0 which belongs to all the cliques. However, 802.11 gives flow 0, a rate that is 3/5 of the proportional fair share

FlowID	IEEE 802.11	MACAW	CB-Fair	PFCR	IDEAL
0	3571	1424	27830	5764	5931
4	24091	24271	20298	23613	23724

Figure 19: Example 3: Total number of packet transmitted

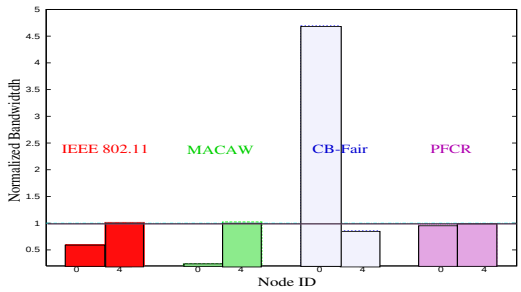

Figure 20: Example 3 - Relative throughput compared to Proportional Fair allocation

and MACAW gives a rate of 1/5 of the proportional fair share. CB-Fair, on the other hand, gives flow 0 five times the proportionally fair throughput, since its degree is far higher than any of the other flows. Compared to this, the PFCR algorithm gives a rate that is just 3% less than the proportional fair share.

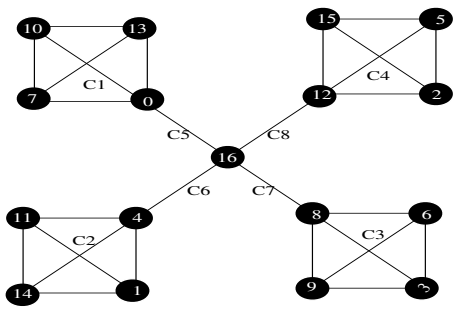

Figure 21: Flow contention graph(G') for example 4

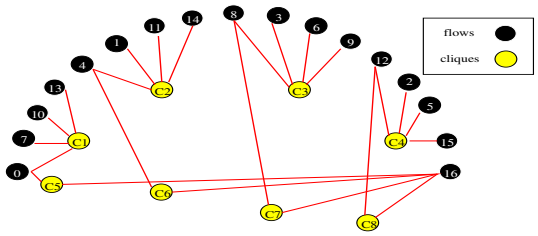

Figure 22: Contention clique graph(G") for example 4

Example 4 Consider the flow-contention graph shown in Figure 21 and its corresponding contention-clique graph in Figure 22. Flows 0, 4, 8 and 12 belong to two cliques, one of size 4 and the other of size 2. Flow 16 is part of four cliques with two flows in each. All the other flows are part of a clique of size 4. When we look at the throughput behavior of the various protocols, shown in Figure 23 and Figure 24 for all the protocols, IEEE 802.11 and MACAW allocate

FlowID	IEEE 802.11	MACAW	CB-Fair	PFCR	IDEAL
0	6208	8406	11784	10686	11973
1	14454	14171	13934	13317	12970
16	42449	37351	28763	39283	38910

Figure 23: Example 4: Total number of packet transmitted

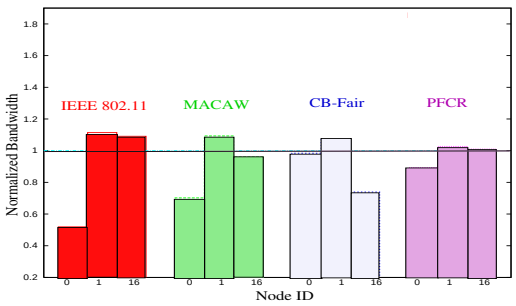

Figure 24: Example 4 - Relative throughput compared to Proportional Fair allocation

bandwidth in a very unfair manner. Flows that experience asymmetric contention suffer, while flows that experience uniform levels of contention achieve better throughput. In IEEE 802.11, flows 0, 4, 8 and 12 receive only 1/2 of the ideal proportional fair share, while flow 16 receives 9% more than its fair share. In MACAW, flow 16 gets 4% less than its fair share, while flows 0, 4, 8 and 12 receive just 70% of their fair share. CB-Fair gives a better bandwidth allocation to flows 0, 4, 8, 12 while penalizing flow 16 which gets only 50% of the ideal bandwidth, even though they have the same degree. This is because flows 0, 4, 8 and 12 are in a region of high contention, and in CB-Fair, higher contention flows win over flows experiencing lower contention. Note that the proportional fairness model does not penalize asymmetric contention flows severely, and the PFCR algorithm, which is modeled to achieve proportional fairness, closely follows the ideal allocation of rates for all flows. Using PFCR, flows 0, 4, 8 and 12 receive 91% of their fair share, and flow 16 gets just 1% more than its fair share.

Of course, as we pointed out in the previous section, the closeness to which PFCR approximates proportional fairness is dependent on the α and β values chosen for the simulations. While it is of course possible to tune these parameters according to each simulation configuration to achieve best performance, we chose to set them a priori and make them non-adaptive because in practice, nodes do not have non-local state and there is no reliable way to achieve coordination of parameter adaptation among the nodes. We found that the results typically did not change significantly for the range of values $\alpha \in [0.01, 0.15]$ and $\beta \in [0.3, 0.7]$. We thus used $\alpha = 0.1$ and $\beta = 0.5$ consistently for all our simulations. In all our experiments, it turns out that PFCR approximates the ideal proportional fairness objective quite well.

7 Summary and Future Work

State-of-the-art multiple access protocols for shared channel wireless networks often lack a well-defined notion of fairness, and provide somewhat ad hoc mechanisms for achieving fairness among contending flows. However, as shared channel wireless networks move from the academic domain to the commercial domain, these environments must support network-level quality of service and service differentiation, which translates to MAC layer weighted fairness. While there is a wealth of work on achieving fairness in wireline and cellular networks, these approaches are inappropriate for shared wireless channels because of the unique location-dependent nature of contention in such networks. Given the spatial distribution of contention, different systems may choose to enforce different fairness models depending on their requirements, goals, and deployment scenarios.

Based on these requirements, we perceived the need to develop a general analytical framework in which to reason about fairness in shared channel wireless networks. This paper makes two fundamental contributions: (a) we propose perhaps the first general framework in which a large class of fairness objectives (represented by concave differentiable utility functions) can be modeled, and (b) we propose the first translation mechanism for taking a fairness specification and automatically generating a corresponding contention resolution algorithm. Together, these two techniques are very powerful because they allow future wireless network designers to deploy different service differentiation models as well as alter fairness/service models on the fly.

We demonstrated the techniques in this paper by starting with the proportional fairness model, generating the corresponding utility function, and then plugging the derivative of this function into the generic channel allocation rate adaptation algorithm to obtain a fully decentralized purely local contention resolution algorithm. We showed via simple simulation experiments that the automatically generated contention resolution algorithm does in fact approximate the ideal fairness objective closely.

While this work is very promising, we are still investigating a number of issues. We will explore the fairness, sensitivity, and convergence properties of the contention resolution algorithm in the presence of user mobility and random channel error. Both of these are important considerations for the practical deployment of MAC protocols, and we will evaluate the behavior of our solution in the presence of mobility as well as random loss. We are also investigating the relationship between MAC layer fairness and network-layer service differentiation in shared channel wireless networks. There are subtle interactions between the two layers because both are affected by spatial contention, and these need to be studied more carefully in the future. Finally, we are investigating the possibility of instantiating our fairness mechanisms as a part of the IEEE 802.11 protocol standard.

References

[1] A. Demers, S. Keshav, and S. Shenker, "Analysis and simulation of a fair queueing algorithm," in *ACM SIGCOMM*, August 1989.

[2] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A Media Access Protocol for Wireless LANs," in *ACM SIGCOMM*, September 1994.

[3] J. Garcia-Luna-Aceves and C. Fullmer, "Floor Acquisition Multiple Access (FAMA) in Single-Channel Wireless Networks," *Mobile Networks and Applications*, vol. 4, pp. 157–174, 1999.

[4] IEEE, "Wireless LAN Medium Access Control(MAC) and Physical Layer(PHY) specifications." IEEE Standard 802.11, June 1999.

[5] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, March 1998.

[6] T. Ozugur, M. Naghshineh, P. Kermani, C. Olsen, B. Rezvani, and J. Copeland, "Balanced Media Access Methods for Wireless Networks," in *ACM MOBICOM*, October 1998.

[7] S. Lu, T. Nandagopal, and V. Bharghavan, "Fair scheduling in wireless packet networks," in *ACM MOBICOM*, October 1998.

[8] H. Luo, S. Lu, and V. Bharghavan, "A New Model for Packet Scheduling in Multihop Wireless Networks," in *ACM MOBICOM*, August 2000.

[9] V. Bharghavan, "Performance Analysis of a Medium Access Protocol for Wireless Packet Networks," in *IEEE Performanc*