WLAN Manager (WM): a Device for Performance Management of a WLAN

Malati Hegde[†], S.V.R Anand[†], Anurag Kumar[†], and Joy Kuri[‡]

Abstract

We propose a *Split MAC* (see RFC [1]) architecture based solution to enhance the Quality of Service (QoS) experienced by the wireless nodes in an IEEE 802.11b/e based WLAN. Our proposed solution achieves the objectives of a) providing a fair sharing of the wireless channel time among stations (STAs) while maximizing channel utilization in a situation where STAs are associated at different data rates with the Access Point (AP), b) exercising Connection Admission Control (CAC) for two-way voice calls, c) supporting policy based service differentiation among the flows within one IEEE 802.11e Access Category (AC), and d) maintaining performance levels during inter-AP roaming of STAs. The proposed solution is implemented on an extraneous device, called WM (WLAN Manager), which in the CAPWAP (see the RFC [1]) parlance can be termed as an Access Controller (AC). We report simulation and testbed results.

I. INTRODUCTION

In recent years there has been a proliferation of wireless devices such as laptops and PDAs accessing WLANs that are based on IEEE 802.11b/e standards. These devices run applications such as VoIP, video streaming, web browsing, etc. These applications have different QoS requirements in terms of delay and throughput. The recent IEEE 802.11e standard adds QoS support in WLANs in order to provide service differentiation in the multiple access WLAN framework. In WLANs, devices are associated with the AP at fixed rates that adapt slowly over time. Hence the task of the MAC protocol, or, equivalently, the distributed packet scheduling, is to share the wireless channel time between the various devices so that each traffic flow obtains its required QoS, the time sharing is fair in some sense, and the channel is utilized efficiently.

Consider, for example, a WLAN that is used to carry VoIP phone calls, and TCP controlled data transfers. The VoIP calls require hard real-time guarantees, and an intrinsically low packet loss rate. On the other hand the TCP controlled transfers require high throughputs while achieving some sort of fairness. The IEEE 802.11b MAC provides no service differentiation. In the default implementations, all packet transfers at each device are handled in a single queue in an FCFS mode. In the IEEE 802.11e EDCA MAC there are 4 access categories (ACs). All arriving packets are classified into one of four queues and service differentiation is provided by certain mechanisms in the Enhanced DCF protocol (EDCF). These mechanisms provide service differentiation but do not by themselves suffice to ensure that all aspects of the QoS strategy are achieved.

For example, with the default resource-sharing mechanism implemented in IEEE 802.11b (see [2]), the long-run average TCP controlled data download throughputs obtained by a group of STAs associated

---

[†] Dept. of Electrical Communication Engg., Indian Institute of Science, Bangalore, 560 012; malati, anand, anurag@ece.iisc.ernet.in

[‡] Centre for Electronics Design and Technology, IISc, Bangalore, 560 012; kuri@cedt.iisc.ernet.in

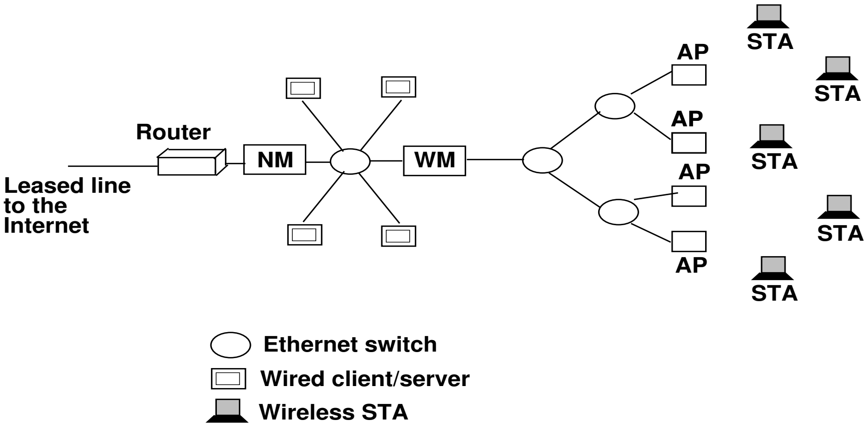

Fig. 1. An enterprise network, showing WLAN components and the proposed WLAN manager. Also shown is NM a WAN access link manager that is often utilized in many enterprises and campuses.

with a single AP will be close to each other, and the STA with the weakest link (least signal strength) strongly influences the bit rate seen by all, even those with better links. That is, the average bit rate experienced by a group of STAs is close to the average bit rate seen by the weakest link (see [3]). One QoS objective would be that the STAs get average throughputs that are *fair* in some well-defined sense. One commonly used criterion is proportional fairness, in which STAs see average throughputs that are proportional to their physical wireless link speeds. One should note here that the system can be proportionally fair while being very inefficient i.e., the channel time can be poorly utilized. Hence, it is important to obtain the maximum possible efficiency in conjunction with the fairness goal. We note here that another related observation has been that even with all STAs associated at the same rate, the upload and download aggregate throughputs are unequal, with the upload throughputs being higher (see [4]). This is because packet loss at the AP buffer affects download TCP transfer windows, thus reducing their share of the throughput.

In addition to fairness between data transfers, Connection Admission Control (CAC) is required for VoIP calls. If a minimum guaranteed throughput is needed for TCP controlled data transfers then CAC is required for them as well. Further, within a service category, say AC1 (normally used for TCP traffic), if some kind of fairness is required then additional mechanisms need to be developed. In addition, it is necessary to ensure that mobile STAs experience uninterrupted service even though they may dissociate with one AP and associate with another.

The additional management capabilities that we suggest above, cannot be achieved within the existing WLAN MAC standards. Also, with the possible move towards inexpensive and light-weight APs it can be expected that such QoS management capabilities will be implemented externally to the AP, perhaps in a centralized manner. We propose a network device called the Wireless Manager (WM) (see Figure 1), which is a host computer with two network interfaces, located in the path of all traffic to and from the STAs, and running appropriate software.

In this paper we describe the QoS mechanisms that will operate in WM, and show how several of the QoS management objectives outlined above can be solved by WM. We propose novel QoS mechanisms that follow Split MAC architecture as described in architecture taxonomy for CAPWAP (see RFC [1]). By virtue of using the Split MAC architecture that uses a central device controlling light weight APs across a WLAN, WM inherently has all the benefits like scalability, cost effectiveness, easy creation and enforcement of unified policies, multi-vendor environment support, and so forth. We provide simulation and testbed results to corroborate our proposals.

The rest of the paper is organized as follows. In Section II we discuss related literature and other approaches to QoS management in WLANs. Section III contains the discussion on WM architecture. Section IV deals with the proposed method of supporting time fairness in 802.11b WLAN environment, the method for estimating the channel rate using ping measurements and the testbed experiment results.

Section V discusses some simulation results to show that WM can be used to give policy based service differentiation to different clients. In Section VI we deal with the QoS for real-time traffic in the presence of TCP traffic. In Section VII we propose an algorithm for managing inter-AP handovers. In the following section we will discuss some of the related research and implementation work.

II. RELATED LITERATURE AND COMMERCIAL PRODUCTS

Some of the recent works that address the problem of wireless bandwidth management are [5], [6], [7], [8], [9]. In [10] the authors discuss methods to overcome the performance anomaly seen in 802.11b WLAN (see [3]). In [11] the authors analytically prove that the phenomenon can be cleanly resolved through configuring the initial contention window size inversely proportional to the bit-rate. This approach needs considerable amount of modifications to the AP to adapt the initial contention window dynamically. Pilosof et al. [5] propose a solution that ensures that upstream and downstream throughputs are almost the same. The solution relies on manipulating the *receiver TCP window* in TCP ACK packets by software residing on the AP. This evidently requires modification to the AP firmware. Detti et al. [6] suggest rate control techniques that essentially reserve half the available wireless bandwidth for downstream connections by placing a rate-limiter in the AP. Obviously, this also requires modifications of the AP software, which is difficult or impossible to do with commercial APs. Malik et al. [7] also address the issue of unfairness between uplink and downlink TCP transfers in an 802.11 WLAN. They propose a scheme in which every STA opens a control TCP connection with the AP. For every control packet on this connection, a specific amount of data (called *virtual maximum segment size*) can be sent in the uplink direction; in addition, the AP exercises control by pacing ACKs. Again, it is evident that this proposed solution requires modifications to both the AP and the STAs, because the control TCP connection needs entities at both ends. The authors of [12] propose a mechanism called Weighted Fair - EDCA (WF-EDCA) to provide proportional fairness for 802.11 WLANs. With WF-EDCA, weighted fair service among different access categories (ACs) is provided, and strict priority service can also be implemented. This again requires changes to AP. Authors in [8] analyze fairness in the presence of impairments in location and interference, for the distributed contention based access mechanisms of the IEEE 802.11 and 802.11e standards. The notion of fairness is explored in terms of throughput and the amount of time permitted to transmit. They show how fairness can be achieved by judicious choice of parameters and study the impact of these choices on capacity.

In [9] the authors propose a scheduling algorithm called Multirate Wireless Fair Scheduling (MWFS) to ensure packet-level QoS in terms of minimum throughput, fair channel share, and maximum packet delay. Instead of providing throughput-based fairness, MWFS improves fairness in terms of time share, which allows flows in good channel condition to receive more service proportional to their higher rates. The simulation results confirm the effectiveness of this scheduling algorithm in multi-rate wireless LANs. In [13], the authors have given a brief tutorial on various MAC layer QoS mechanisms provided by 802.11e and they have shown that 802.11e standard provides a powerful platform for QoS support in WLANs. An extensive survey of admission control algorithms/protocols in 802.11e is found in this paper. This supports the fact that 802.11e provides platform to support QoS in WLAN. Our idea of using 802.11e's priority class to provide QoS for voice proves the point.

The work reported in paper [14] provides a method to enhance the VoIP performance over the 802.11 WLAN by implementing two queues along with a strict priority queuing on the top of the 802.11 MAC implemented as part of the firmware. The Real-time (RT) and Non Real-time (NRT) packets are classified and enqueued into one of the two queues and priority is given to the RT packets. This approach is somewhat similar to our idea but needs changes to the firmware running on the AP. Also their method does not consider service differentiation within a given class unlike ours.

Paper [15] suggests an admission control algorithm for 802.11 DCF access mechanism, taking throughput difference between saturated and unsaturated states as residual bandwidth, to utilize network resources effectively. The static mapping of application layer priority to MAC priority may lead to problems in non-cooperative network environment. Based on the analysis of per-flow throughput and delay, in [16] the authors have proposed an algorithm to enforce per service class admission control and priority re-allocation to provide QoS guarantee with service differentiation. This algorithm needs information on throughput, and delay. This information is estimated/computed per-flow on the AP. In

[17] the authors have addressed the issue of intra-class QoS differentiation in IEEE 802.11e networks. They suggest a reservation scheme of transmission channel using transmission opportunity (TXOPlimit) parameter. Each time a wireless station wins the contention; it is allowed to send a burst of packets. The number of packets present in this burst is computed dynamically and according to the flow's data rate and flow's priority. In [18] the author discusses a distributed scheduling algorithm which allocates bandwidth to flows proportional to its weights by modifying the DCF parameters in 802.11 standard. The paper [19] presents an Opportunistic Auto Rate (OAR) media access protocol, for multi-rate Ad Hoc networks, to better exploit durations of high-quality channel conditions. Nodes with good channel conditions are granted access to the channel for a duration that allows multiple packet transmissions vs. a single packet when nodes access the channel at the base rate. Consequently, by exploiting inherent variations in channel conditions, nodes will transmit more data during epochs of high-quality channels thereby increasing the total throughput of all users. An analytical model that characterizes the impact of channel conditions on the throughput is obtained. In [20] the authors suggest a Receiver Based AutoRate MAC protocol for optimizing the performance in WLAN. The channel quality (signal-to-noise ratio, signal strength, symbol error and so on) prediction method based on RTS/CTS is used to get the PHY rate. In [25] the authors propose a method for CAC for the 802.11 contention access mechanism. The proposed algorithm assumes intimate knowledge of the wireless channel conditions such as collision rate per flow, and tight control over the connection window and the TXOPs for flows. All this can be done only within an AP.

As one can see, all the above mentioned literature requires either changes to MAC parameters and/or modifications to the firmware running on the AP. We believe that implementing such mechanisms across a large-scale deployment consisting of heterogeneous WLAN infrastructure is impractical in terms of maintenance, management and cost-effectiveness. Our works does not require any modification to the MAC protocols, and works with any existing IEEE 802.11b/e based WLAN infrastructure. We propose an external device which will provide fairness among stations while maximizing channel utilization, policy based service differentiation, CAC for voice calls, and efficiently handles inter-AP handovers.

Some of the prominent vendors that manufacture WLAN controllers/switches for managing WLAN bandwidth are Meru Networks [21], Cisco/Airespace [22], Trapeze [23] and Extricom [24]. Meru Networks and Extricom use a single channel across all the APs in order to eliminate the problem of hand-off delays and achieve uninterrupted voice over WLANs. Their smart AP solution uses nonstandard and proprietary methods to realize the goal, and does not interwork in a multi-vendor WLAN scenario. On the other hand Cisco/Airespace, Trapeze, Aruba use Split-MAC technology wherein management, queuing/scheduling are handled at a WLAN controller, and only the core MAC functionality is performed at the AP. Their solution agrees with the upcoming CAPWAP open standard [1], and hence is expected to work in multi-vendor environment.

While the existing implementations as described above require changes to AP in a vendor specific manner, WM achieves the QoS objectives without requiring any modifications to APs. WM addresses certain QoS issues that lack in the above mentioned products. These are temporal fairness with maximal channel efficiency in a multi-rate WLAN and Call Admission Control (CAC) for voice calls that make use of 802.11e operating in EDCA mode.

III. THE WM APPROACH

WM follows a centralized WLAN management approach similar to the Split MAC architecture as detailed in the architecture taxonomy for CAPWAP [1]. As defined in [1], the main idea behind the Split MAC architecture is to implement a part of the MAC functionality on a centralized access controller (AC) instead of the APs, in addition to providing the required services for managing and monitoring the APs. Usually, the decision of which functions of the 802.11 MAC need to be provided by the AC is based on the time-criticality of the services considered. Some of the motivating factors for adopting this architectures are in order. The first is to retain functionality that is specific and relevant only to the locality of each BSS in the AP, in order to allow the AC to scale to a large number of "light weight" APs. To elaborate on this, the functionality that goes into a "light weight" AP comprise of real-time services like beacon generation, probe response/transmission, processing of control frames such as RTS, CTS, ACK, and so on, in addition to retransmission, transmission, and rate adaptation. The locality

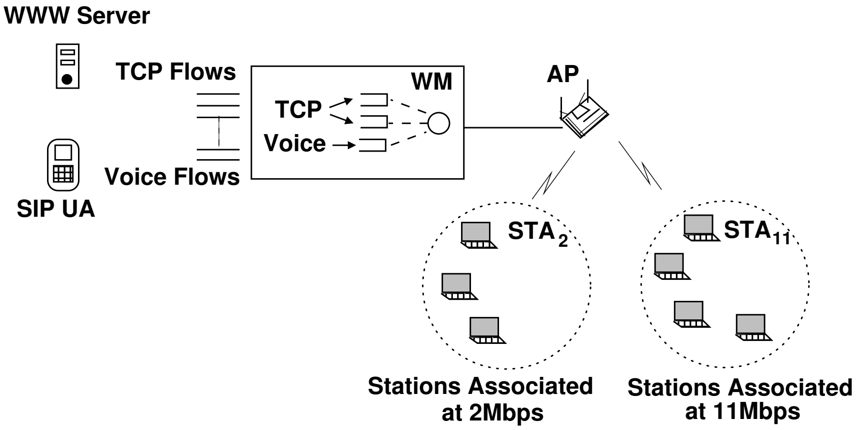

Fig. 2. A multi-rate WLAN environment in which laptops in clusters STA_2 and STA_{11} are associate with the AP at 2 Mbps and 11 Mbps respectively

of management functionality like authentication, association/re-association/ disassociation/distribution, privacy can be optionally implemented either in the light weight AP, or at the AC. The services like classification, scheduling, and queuing are handled externally at the AC.

Another consideration is cost reduction of the APs to make them as cheap and simple as possible. The centralization makes it easy to create and enforce unified policies that are consistent across a large WLAN. The Split MAC architecture facilitates building an AC that works in a multi-vendor environment.

WM comprises a general purpose computer that runs requisite software containing various traffic management techniques to realize the QoS objectives of a) providing fair sharing of the wireless channel time among the STAs while maximizing channel utilization in a multi-rate scenario, wherein STAs are associated with different data rates to the AP, b) exercising CAC for two-way voice calls, c) supporting policy based service differentiation among the flows within one service category, and d) maintaining performance levels during inter-AP roaming so that the ongoing voice calls do not experience glitches, and the throughputs of ongoing TCP connections do not drop because of packet losses.

The device is located (see Figure 1) so that all packets that pass through the AP(s) also pass through the machine; i.e., the software running on the machine is in the path of all traffic to and from the AP(s). In WM, for each AP managed by it, we create a queuing and scheduling environment that approximates the desired sharing of the wireless medium. The aim is to reduce the random access overheads substantially. This is achieved by allowing very little or no queuing at the wireless medium. Once this is achieved, then by time sharing of the medium among the various traffic flows the desired QoS is aimed to be achieved. For interactive voice, strict priority is given, and various throughput ratios are provided to TCP controlled traffic. We note that since the Internet access rates at the enterprise are in 10s of Mbps, WM cannot be a performance bottleneck.

IV. ENFORCING FAIRNESS AND EFFICIENCY

The following discussion refers to Figure 2 that depicts the test-bed experimental setup that has been used. In the figure, STA_{11} and STA_2 are two STAs that associate with the AP at 11 Mbps and 2 Mbps respectively. The box titled WWW is the web server from which the devices STA_{11} and STA_2 download files. The box titled SIP UA is one of the peers for the voice calls.

Let us first consider downlink file transfers to the clients. It can be seen (experimentally and also analytically) that the two sessions will obtain similar throughputs (if the maximum advertised TCP windows of both TCP receivers are the same)(see [26] and [27] to understand this phenomena). Since the throughput of the TCP transfer to the STA associated at 2 Mbps cannot be more than 2 Mbps, we find that the throughput of the transfers to the STA_{11} will also be substantially reduced. This phenomenon can be understood as follows. Suppose that the AP buffer size is large enough to accommodate the

To Client	Individually (without WM)	Simultaneous (without WM)	Simultaneous (with WM)
STA_2	1.4	0.96	0.55
STA_{11}	4.6	1.30	2.90
Total		2.26	3.45

TABLE I

DOWNLOAD TCP THROUGHPUTS (MBPS) TO STA_2 AND STA_{11}, WITH AND WITHOUT WM. THE FIRST COLUMN SHOWS THE DOWNLOAD THROUGHPUT TO EACH CLIENT WHEN THE OTHER CLIENT IS NOT DOWNLOADING.

maximum window worth of data of both the connections. For a long file transfer, the windows of both transfer will grow to their maximum values.

When contending for the channel the AP does not “care” whether its head-of-the-line (HOL) packet is for the STA associated at 2 Mbps or for the one associated at 11 Mbps. A certain aggregate AP packet throughput will be obtained. Once the AP wins a contention the HOL packet will be transmitted. Let W be the maximum window attained by the connections, Θ_2 and Θ_{11} represent throughputs seen by 2 Mbps STA and 11 Mbps STA respectively, then

$$\Theta_2 = \frac{W}{\Delta_{STA2} + \Delta_{AP}}$$

and

$$\Theta_{11} = \frac{W}{\Delta_{STA11} + \Delta_{AP}}$$

where Δ_{AP} is the mean delay in the AP until reaching the HOL position and Δ_{STAr} is the mean delay of a packet (of a connection at rate r) from the time it wins the contention at the AP until its ACK returns to the WWW server. Since the AP is the bottleneck, Δ_{AP} dominates, but Δ_{STA11} is a little less than Δ_{STA2}, leading to $\Theta_{11} > \Theta_2$, but the two values being not very different. Therefore, in the 11 Mbps case the packet will experience a slightly smaller round trip time until its ACK returns to the WWW server. Thus the amount of the window of the 11 Mbps connection that will be in the AP will be a little larger than that of the 2 Mbps connection. Hence the probability that the HOL packet at the AP is of an 11 Mbps connection is a little larger. Thus the throughputs of the two kinds of connections will be close to each other with the 11 Mbps throughput being a little larger.

It is evident that the problem can be solved if the HOL packet at the AP is more often a packet meant for the STA associated at 11 Mbps. This can be achieved by not letting the downlink packet queue build up in the AP, but forcing the queue to build up in WM. The packets are released from WM in order to achieve the correct ratio of throughputs.

The mechanism for achieving the desired proportional throughput fairness is depicted in Figure 2. There is a per flow or per class (“downlink”) queue in WM. In practice there can be a per-STA queue. This queue is served by a weighted fair queuing (WFQ) packet scheduler, that is dynamically configured for the desired ratio of throughputs. For example, this ratio would be nominally 11:2 if proportional fairness is required between STA_2 and STA_{11} in the above example. An important issues remains. What should be the aggregate service rate of the WFQ scheduler? If the service rate is set to a small value then the WLAN will unnecessarily idle, and we will have an inefficient design. On the other hand, if it is set to a high value, then all the queuing will again move to the AP, defeating the basic purpose. We need to obtain the *optimal* WFQ service rate that provides the desired throughput ratio, while keeping the WLAN maximally utilized.

In Table I and Figure 3 we report experimental results from a test-bed. The first column of the table shows that when individual file transfers are performed separately then an STA_2 obtains a throughput of 1.4 Mbps, and an STA_{11} obtains 4.6 Mbps. On the other hand when the file transfers are performed simultaneously (without WM) then the rates are, respectively, 0.96 Mbps and 1.3 Mbps. As discussed earlier, when operating together the throughput to STA_{11} also drops below 2 Mbps. If the channel time was devoted entirely to STA_{11} then it would have obtained a throughput of 4.6 Mbps and the aggregate rate would be also 4.6 Mbps. However, owing to the way the IEEE 802.11 MAC works the two clients share the channel inefficiently, with an aggregate throughput of just 2.26 Mbps, and the 11 Mbps STA

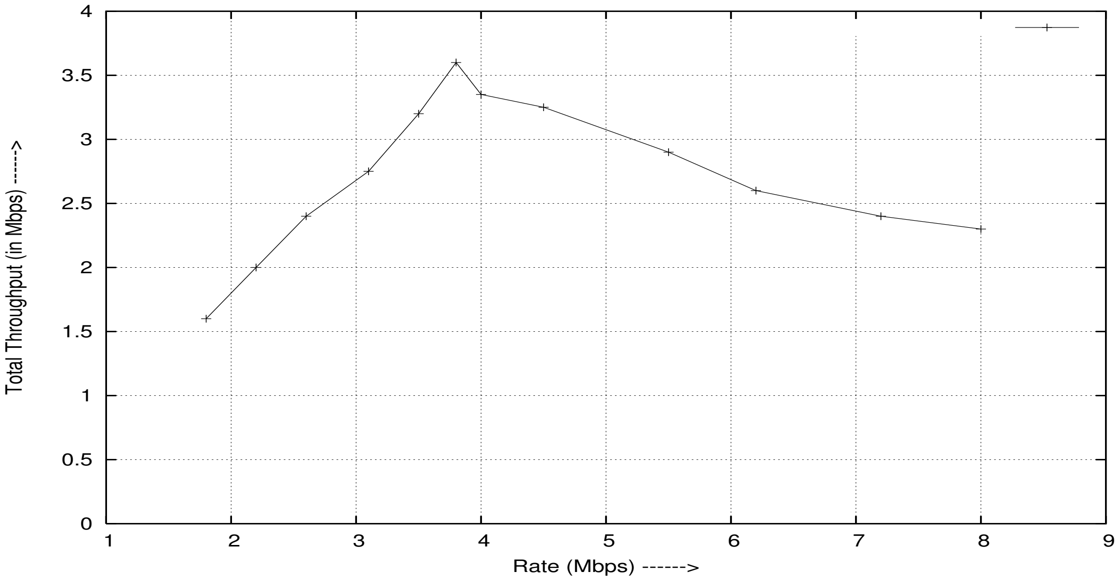

Fig. 3. Measured results on a testbed showing the aggregate download rate to STA_2 and STA_{11} as we vary the service rate of the WFQ server in WM.

getting a very low throughput. Next we impose a 5.5:1 service ratio at WM, and vary the WFQ service rate from a small value up to 8 Mbps. Figure 3 shows the aggregate download throughput when one file transfer is performed to each STA. We find that the aggregate rate increases until the service rate is about 3.8 Mbps, and then the throughput drops. The last column of Table I shows that if the service rate is set at 3.8 Mbps then the two STAs obtain 0.55 Mbps and 2.90 Mbps, respectively. The ratio is about 5.27, not 5.5. This is because, in setting the value 5.5 Mbps we have not taken care of the time taken by TCP ACKs; this will be considered later in this section.

The observations from Figure 3 can be explained as follows. When the service rate of the WFQ is small, the AP's queue is frequently empty and the throughput ratio is governed entirely by the WFQ service ratio. As we increase the WFQ service rate in order to increase the WLAN utilization, at a certain point the AP's queue begins to fill up. Eventually the queuing shifts from WM to the AP, and the ratio is no longer governed by the WFQ ratio, but instead by the functioning of the IEEE 802.11 DCF MAC. We observe that at 8 Mbps WFQ service rate the aggregate throughput is close to that without WM.

In order to obtain a desired throughput ratio, the actual service ratio in the WFQ would depend on the actual channel time that each connection utilizes and this would in turn depend on the overheads of each connection. There is only one channel in WLANs, unlike in the cellular case; i.e., the spectrum sharing is time division duplex. So it is appropriate to combine the data as well as the ACKs traveling in the uplink direction, and consider the aggregate amount of channel time used by clients. Considering this factor, we now provide an example of how the service ratio can be determined.

An Example of Service Ratio Computation: The total transmission time of a packet (data or ACK) comprises two parts, namely, the transmission time of the payload packet, and a constant protocol specific overhead, which can be seen to be 892 μs. This constant protocol overhead T_{OVH} is obtained by summing up all the timing components that accompany the transmission of a TCP packet. These time components include the transmission times for RTS, CTS, MAC, PHY, MAC ACK and the associated inter-frame delays SIFS, and DIFS. Assuming TCP controlled data transfer, the channel time utilized by an STA should take into account the time consumed by TCP ACK in the reverse direction as well.

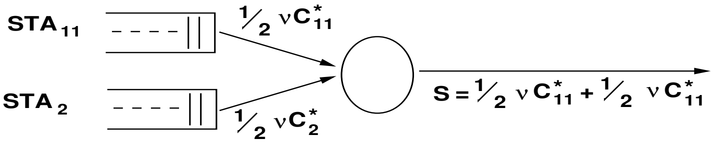

Fig. 4. WFQ scheduler in WM

The effective physical rate C_r^* seen by the STA with data rate C_r can be written as

$$C_r^* = \frac{(d \times L_{DATA} + L_{ACK})}{(d \times (L_{DATA}/C_r) + T_{OVH}) + ((L_{ACK}/C_r) + T_{OVH})}$$

where d is the delayed ACK parameter, usually 2. We can obtain C_{11}^* and C_2^* for STAs connected at 11 Mbps and 2 Mbps respectively by substituting the following values. $L_{DATA} = 1534$ Bytes (1500 MSS + 34 IEEE 802.11 MAC header), $L_{ACK} = 86$ Bytes (52 Bytes packet size + 34 IEEE 802.11 MAC header). This yields $C_{11}^* = 4.559$ Mbps and $C_2^* = 1.591$ Mbps, and hence the ratio of C_{11}^* and C_2^* is 2.865:1. By sharing the channel in this ratio one can achieve time fairness between the STAs that are operating at 11 Mbps and 2 Mbps. The same method can be extended to the case of other data rates.

Upload-Download Fairness: It has been shown in [28] and [4] that, because of buffer overflow at the AP, uplink transfers obtain a higher throughput than downlink transfer even if the PHY rate are the same. This problem is easily solved with WM, since all the queuing is forced to take place in WM, and hence buffer tail drop losses at the AP are eliminated. Since there is no queuing and packet drops at the AP the TCP flow direction should not matter.

Two practical issues remain to be addressed:

1) How does WM determine the optimal WFQ service rate?
2) How does WM determine the PHY rates at which the STAs are associated?

A. *An Adaptive Algorithm for the WFQ Service Rate*

In the above experiment we determined the optimal service rate experimentally. In this section, we describe an algorithm to dynamically arrive at the optimal WFQ service rate S that maximizes the channel efficiency in conjunction with time fairness. Figure 4 depicts the approach. After having determined the rates C_2^* and C_{11}^*, for proportional fairness we would like the effective throughput of the transfers to be $\frac{1}{2}C_2^*$ and $\frac{1}{2}C_{11}^*$, respectively. However, owing to collision overheads this will not be achievable. We denote the fraction of the these rates that is achievable as ν. Then the aggregate throughput will be $S = \nu\left(\frac{1}{2}C_2^* + \frac{1}{2}C_{11}^*\right)$, and this will be the appropriate choice of the WFQ service rate. This and the resulting WFQ weights are shown in Figure 4. For $\nu << 1$ the channel capacity will be underutilized. When $\nu >> 1$ then the queuing will shift to the AP and the fairness criterion cannot be achieved. It remains to “tune” ν on-line so that the aggregate throughput is maximized. This can be done via a simple stochastic gradient type algorithm. The value of ν is updated periodically after a certain measurement interval. During each measurement interval the derivative of the aggregate throughput is estimated by a perturbation technique.

Consider the k^{th} iteration, if ν_k is the iterate in a measurement interval, during half the interval the system is operated with $\nu = \nu_k + \epsilon$ and in the remaining half with $\nu = \nu_k - \epsilon$. This yields the measurements $R_{+\epsilon,k}$ (measured rate at k^{th} interval with $\nu = \nu_k + \epsilon$) and $R_{-\epsilon,k}$ ((measured rate at k^{th} interval with $\nu = \nu_k - \epsilon$). The derivative is then estimated as $\frac{R_{+\epsilon,k} - R_{-\epsilon,k}}{2\epsilon}$, and ν_k is updated to ν_{k+1}.

Algorithm for Self-Tuning of ν

1) Start with an initial value ν_0
2) At each measurement interval k do
    - Compute ν_k using

Station Rate (Mbps)	Packet size (Bytes) 64	 512	 1200	 1500
11 Mbps	2.0ms	4.0ms	6ms	12ms
5.5 Mbps	2.2ms	4.8ms	8ms	16ms
2 Mbps	2.7ms	8.0ms	14ms	22ms

TABLE II

PING MEASUREMENTS FOR TRACKING THE STATION RATES.

$$\nu_k = \nu_{k-1} + a\left(\left(\frac{(R_{+\epsilon,(k-1)} - R_{-\epsilon,(k-1)})}{2\epsilon}\right)\right)$$

where a is constant algorithm parameter. let S_k denote the service rate applied at the WFQ.

- Apply $S_k = \nu_k + \epsilon\left(\frac{1}{2}C_2^* + \frac{1}{2}C_{11}^*\right)$, as WFQ service rate for half of the interval and measure $R_{+\epsilon(k)}$
- Apply $S_k = \nu_k - \epsilon\left(\frac{1}{2}C_2^* + \frac{1}{2}C_{11}^*\right)$ as WFQ service rate for the remaining half of the interval and measure $R_{-\epsilon(k)}$

The test-bed experiments have been done using the setup as shown in Figure 2. Two experiments were done, one with the initial service rate $S = 2.6$ Mbps, which corresponds to $\nu = 0.85$, and another with the initial service rate $S = 4.6$ Mbps which corresponds to $\nu = 1.25$. Since all the queuing is moved to WM, the WLAN contention is minimized and the experimental results show that ν value close to but less than 1 can be used to get the optimal utilization. The same experiment shows that the approximate convergence time for ν is 15 to 20 seconds with the measurement interval of 15 sec. The convergence time can be improved by reducing the measurement interval and by tuning the constant parameter a.

This experiment demonstrates that dynamically one can adapt the WFQ serving rate, so that the link utilization is optimal and also time fairness is achieved. This experiment also demonstrates that differentiated QoS can be given to different TCP traffic for different stations.

The adaptive algorithm can be easily extended to the multi-rate scenario where the STAs are simultaneously connected at the other possible rates.

B. How WM Can Determine PHY Rates

We propose to use pings sent by WM to estimate the channel rate at which an STA is associated to the AP. This method is consistent with the Split MAC architecture followed by WM. Since 802.11 MAC devices adapt their rates depending on the signal strength and interference, estimating rates dynamically becomes necessary. One important point here is that the channel rates are very distinct. Initially the rate chosen is 11 Mbps, which drops to 5.5 Mbps and then further to 2 Mbps, depending on the number of observed packet losses. Since the rates are distinct we can use ping round trip times to estimate the rate at which a client is connected to the AP.

To distinguish the station rates reliably, we have to choose ping packet sizes so that the RTTs are well separated at different station operating rates. From Table II, it is clear that a normal ping packet of 64 bytes does not serve the purpose, as the min RTT is 2 msec at all the rates. This is due to fixed 802.11 protocol overhead that is dominating over the transmission time of the small ping packet. In the last row of the same table we notice that 1500 byte ping packets give distinct RTTs of 12, 16, and 22 msec for 11, 5.5 and 2 Mbps, respectively. So, we have chosen 1500 byte ping packets for rate tracking. The minimum RTT gives the delay due to the transmission rate alone without queuing and channel access delays, thus making the PHY rate estimate more accurate.

In [20] the authors have suggested a method to estimate the PHY rate based on RTS/CTS information. The information about RTS/CTS frames is available on the AP. So this mechanism needs to be implemented on AP and need considerable modification to the AP. Another alternative that can be potentially considered for WM is to use SNMP to get signal-to-noise ratio information from the AP, and use this information to arrive at PHY rate of an STA attached to this AP. This method therefore does not require any changes to the AP software. But it however requires that SNMP is supported on all the APs which is not guaranteed in an heterogeneous WLAN environment. So a ping based technique which can work across a wide variety of AP implementations appears to be more suitable.

V. POLICY BASED SERVICE DIFFERENTIATION

WM can be used to provide policy based service differentiation among different clients as well as traffic flows that are associated with the same AP.

Let us consider the situation where we want to differentiate between two clients by providing differential share of the air link between these two clients in the ratio of $9:1$. This can be achieved by queuing the packets into different queues corresponding to these clients at the input of WM, and adjusting the weights of the WFQ scheduler in WM to a $9:1$ ratio. In general, WM can be used in the scenarios wherein one can define various policy definitions based on multi-field packet/flow classification, DiffServ Code Points, and so on. Further, the same approach can be extended to provide service differentiation to the flows belonging to a given Access Category of 802.11e.

A. A Simulation Experiment

We have used ns2 [29] for our simulation work. WM is realized on ns2 by inter-connecting the available ns2 components such as node, link, and the WFQ scheduler. Figure 5 depicts the functional elements and their connectivity that form the scenario for simulation tests. In the figure, node $W(1)$, simplex link $W(1) \rightarrow W(2)$, and the WFQ scheduler, operating at rate C on simplex link $W(1) \rightarrow W(2)$, together constitute WM. The multiple queues served by WFQ is shown within the node $W(1)$ in the Figure 5. Client 1 and Client 2 are connected to the same AP $W(3)$ and downloading files from the server $W(0)$. The nodes are connected using simplex links in such a way that both downlink traffic towards WLAN and uplink traffic from WLAN towards servers are sent through the WFQ scheduler of WM. The simplex link $W(3) \rightarrow W(1)$ in the Figure 5 models the clients uplink traffic passing through the WFQ. In Figure 5, the path taken by the downlink traffic from the server to STAs is $W(0) \rightarrow W(1) \rightarrow W(2) \rightarrow W(3) \rightarrow STA$, and the path for the uplink traffic from the clients to server is $STA \rightarrow W(3) \rightarrow W(1) \rightarrow W(2) \rightarrow W(0)$ where STA refers to Client 1, or Client 2.

In WM, the weights of Client 1 and Client 2 are $9:1$. In Figure 6 we have plotted throughputs of Client 1 and Client 2 downloading files from web server for various WM serving rates. The top most curve is the total throughput that Client 1 and 2 together gets. The Second from the top line is for Client 1 and bottom most curve is for Client 2. It can seen from the figure that Client 1 gets 9 times the bandwidth of 2 up to WFQ serving rate of 3.8-4.0Mbps. This is the optimum WFQ operating point at which the total throughput is maximum as well as the service differentiation criterion is met. Beyond this serving rate, the service differentiation is not observed because the queues served by WFQ at WM will shift to AP resulting in FCFS service at AP. We therefore have to operate WM at the optimal operating point.

In Table III we tabulate the throughputs seen by Client 1 and Client 2 for various weight ratios configured for WFQ in WM. The serving rate for WM is chosen to be 3.8 Mbps which is the desired optimum rate. We can clearly see from the table that WM is able to give bandwidth to two Clients 1 and 2 in the assigned ratio.

VI. SUPPORTING COMBINED VOICE AND DATA OPERATIONS IN 802.11E BASED WLANS

In this section we describe our algorithm for achieving the QoS objectives in a mixed traffic scenario comprising both voice and data. The algorithm provides i) Call Admission Control (CAC) for two-way voice calls based on end-to-end delay and packet loss constraints, ii) Time fairness with maximal channel efficiency, iii) Service differentiation among competing flows within a 802.11e AC based on administratively configured policies.

The real-time voice bypasses WM and goes through to the AP without any queuing and scheduling within WM. This is because the QoS required for voice traffic is handled well at 802.11e enabled AP. Hence, so far as voice is concerned, WM handles only Connection Admission Control. The service

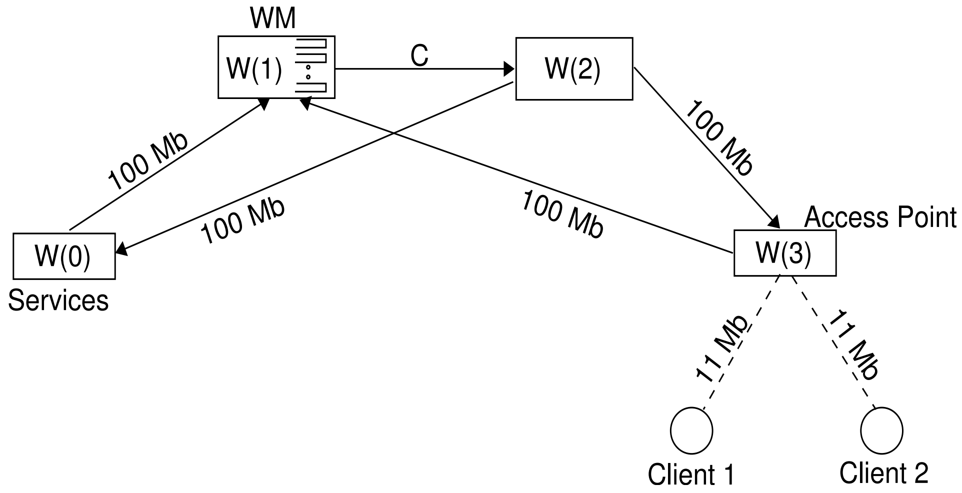

Fig. 5. Realizing WM using ns2 simulation

Ratio Bandwidth share	Client 1 Kbits	Client 2 Kbits
1:1	224.0	224.0
2:1	299.3	149.6
3:1	336.8	112.2
4:1	359.2	89.9
5:1	374.0	75.0
6:1	384.8	64.2
7:1	392.7	56.3
8:1	399.1	49.8
9:1	404.0	45.0

TABLE III

THROUGHPUTS SEEN BY CLIENT 1 AND 2 FOR VARIOUS CHANNEL SHARING RATIOS

differentiation within a given AC of 802.11e is done by classifying and queuing IP packets based on DSCP, 802.1q, or any other user defined policy, and subsequently scheduling the packets using WFQ scheduler within WM.

The Resource Management Algorithm

1) WM maintains the total wireless channel time used up by all the existing voice traffic flows per unit interval of time.
2) From the computation done in Step 1, WM arrives at the remaining available wireless channel time.
3) When a new voice call arrives at WM, the decision as to whether it should be admitted or not is based on a) the channel time requirement, b) available channel time obtained from Step 2), and c) administratively configured policies if any. The channel time requirement for a new voice call can be obtained during codec negotiation of a SIP call.
4) If Step 3 results in call admission, the wireless channel time will be updated. The expression used to estimate the channel time takes as its input the RTP payload profile of the codec that would be used for the call, and the rate at which the STA gets associated with the AP. The RTP payload profile is described in the SDP portion of a SIP message. When the call is released, the wireless channel time is updated again with the corresponding channel time used by that call.
5) The channel time remaining after Step 4 is available for data. The algorithm then adaptively adjusts the optimum WFQ operating rate of the scheduler in WM in order that the channel time fairness condition is met and the throughput on the wireless medium is maximized.

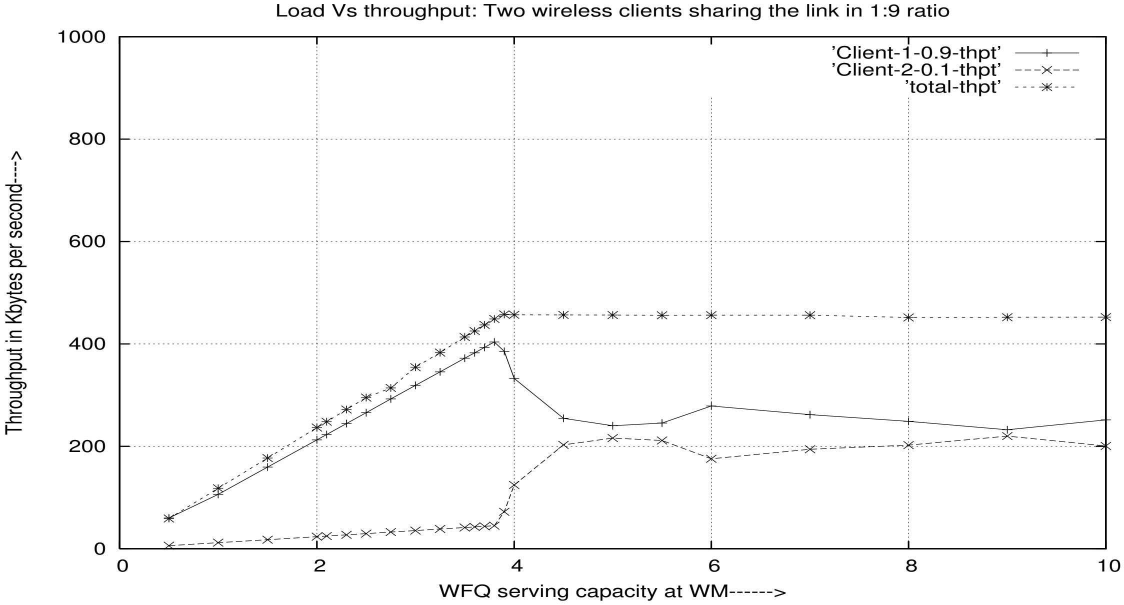

Fig. 6. Throughputs seen for Client 1 and Client 2 using WM

In the absence of this step, the STAs running at lower speeds will consume more channel time compared to the STAs operating at higher speeds thus bringing down the overall throughput on the wireless medium.

6) Service differentiation among flows within one Access Category (AC) is achieved by appropriately classifying the packets, and scheduling using WFQ scheduler within WM.

The above algorithm can be implemented by having a table lookup or by using an analytical expression that directly gives the optimal WFQ rate for a given number of admitted voice calls belonging to STAs operating at different data rates. Conversely, for a given operating rate we can obtain the connection admission region for the voice. The table lookup is simple to implement considering the number of possible rates for a typical IEEE 802.11b based WLAN (4 different rates) and the number of voice calls that can be supported per BSS. For instance, the maximum number of voice calls that can be supported on a 11 Mbps channel is approximately 10. The table lookup approach may be impractical in the real world due to, for instance, the wide variety in voice codecs, thereby increasing the combinations of entries we need to maintain in the table in a multi-rate WLAN environment. In such scenarios, an analytical expression that considers these factors to estimate the WFQ rate is better suited. In the following section, we build one such table using simulation under different operating scenarios for a simple case of two-way G.711 PCM voice calls.

A. A Simulation Experiment

Since at the time of writing this paper, we did not have IEEE 802.11e equipment, we have tested our proposal on the 802.11e patch for *ns2* available at [30]. Refer to Figure 5 for the simulation setup considered in this experiment.

The two-way voice calls use G.711 PCM with 20 msec packetization intervals. Some of the STAs operate at 11 Mbps and some at 2 Mbps. Each of the STAs generates only one flow of traffic. The flow is either a voice call or a down-link TCP transfer. The voice traffic and TCP traffic have been assigned to AC3, and AC1 respectively in accordance with 802.11e access classes. All the voice traffic going through WM enters one common strict priority queue effectively "bypassing" queuing in WM, whereas, all TCP flows pertaining to a given STA a enter separate queue assigned for them. During all the simulation runs, the number of TCP down-link transfers were fixed to 5 on 11 Mbps, and 3 on 2 Mbps. For each simulation run the number of voice calls from 2 Mbps and 11 Mbps was varied, and

Number of calls on 2 Mbps (N_2)	Number of Calls on 11 Mbps (N_{11}) (Rate in Mbps)								
	1	2	3	4	5	6	7	8	9
1	3.2	3.0	2.8	2.6	2.4	2.0	2.0	1.6	1.6
2	2.6	2.4	2.2	2.0	1.9	1.8	1.6	1.4	CAC
3	2.2	2.0	1.6	1.6	CAC	CAC	CAC	CAC	CAC
4	1.8	1.6	CAC	CAC	CAC	CAC	CAC	CAC	CAC
5									
6					CAC Region				
7									
8									

TABLE IV

WM OPTIMAL OPERATING RATE AND CONDITION FOR CAC FOR VOICE CALLS IN A MULTI-RATE WLAN

the optimal operating rate of WM has been obtained. This is the rate at which the proportional fairness is met with maximum throughput.

Our result are summarized in Table IV. The top row of the table is the number of voice calls (N_{11}) from 11 Mbps STAs and the first column indexes the number of voice calls (N_2) from 2 Mbps STAs. The entry in the i^{th} row and j^{th} column is the optimal operating rate for WM for i 2 Mbps and j 11 Mbps voice calls. The entries marked CAC in the i^{th} row and j^{th} column implies that i 2 Mbps and j 11 Mbps calls cannot be accepted simultaneously. Thus if $N_{11} = 4$, and $N_2 = 3$, then one more call cannot be accepted to an 11 Mbps associated STA. It can be seen from the same table the optimal operating point is 3.0Mbps for the case of 2 voice calls of 11 Mbps, and 1 voice call of 2 Mbps.

VII. MANAGING INTER-AP HAND-OVERS

Finally, we propose a buffering mechanism to minimize the interruptions of ongoing voice calls and reduce degradation in TCP performance due to inter-AP handover delays, and the resultant packet losses during roaming. The inter-AP hand-off delays are of the order of 100s of msec and hence needs attention especially for voice traffic. The buffering procedure is significant and challenging because of the issues involved in anticipating the time epoch and duration of the hand-off phase and accordingly buffer the packets during this phase. We believe that there is no other solution that addressees this issue in the SplitMAC framework. A policy based capacity provisioning scheme to accommodate roaming users can be used to maintain the same service rate when a station changes its AP.

Handover Management Algorithm

1) WM buffers a copy of the most recently forwarded packets to an STA. The number of packets that are buffered corresponds to the inter-AP reassociation delays.
2) When the STA reassociates itself from one AP to another AP during roaming, WM detects this event as an updation in the MAC address of the AP to which the STA is newly associated.
3) This updation process takes place during post reassociation procedure using IAPP and IEEE 802.11d protocols.
4) After this, WM releases the buffered packets to the STA at the new AP which would not have been delivered to the STA by the AP during the re-association phase.

It is important to note that IAPP context transfer does not mandate that the older AP forwards all the buffered packets meant for an STA to the newer AP after change in the STA location. WM on the other hand ensures that the receiver STA does not lose voice packets and makes the voice reception smooth. By minimizing the packet losses, WM ensures that TCP sender does not timeout and retransmit, thus preventing performance drop during the handovers.

VIII. CONCLUSION

In this paper we have proposed a centralized WLAN management solution, namely WM, using Split MAC architecture to enhance the Quality of Service (QoS experienced by end applications running on STAs in an IEEE 802b/e based WLAN. We have described various algorithms that go into WM to achieve the desired QoS objectives. The objective of proportional fairness with maximal channel utilization in a multi-rate WLAN is met by dynamically adapting the weights of the WFQ scheduler operating in WM. We have proposed a method to estimate the link speed between STA and AP using periodic ping measurements. The emulated testbed results of the prototype implementation of WM have been presented. We have also described an algorithm for exercising Connection Admission Control for two-way voice calls and to support policy based service differentiation among the flows within one AC of 802.11e service category. A buffering mechanism to minimize the voice interruptions of an ongoing voice calls and reduce degradation in TCP performance due to inter-AP handover delays, and the resultant packet losses during roaming is discussed.

Both the testbed and simulation results that have been presented in the paper establish proof-of-concept. Future work involves enhancing the algorithms for more complex WLAN scenarios and implementing a prototype WM.

REFERENCES

[1] L. Yang, P. Zerfos UCLA and E. Sadot,“Architecture Taxonomy for Control and Provisioning of Wireless Access Points (CAPWAP)”, June 2005,http://www.ietf.org/rfc/rfc4118.txt

[2] ANSI/IEEE Standard 802.11b Part 11: “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY)”, Specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz band, 1999

[3] Martin Heusse, Franck Rousseau, Gilles Berger-Sabbatel and Andree’s Duda, “Performance Anomaly in 802.11b”, in *Proc. of IEEE Infocom 2003, San Francisco* April 2003.

[4] Mingwei Gong, Qian Wu and Carey Williamson, “Queue management strategies to improve TCP fairness in IEEE 802.11 wireless LANs”, RAWNET 2006 workshop, Boston, April 3, 2006

[5] Saar Pilosof, Ramachandran Ramjee, Danny Raz, Yuval Shavitt, and Prasun Sinha, “Understanding TCP fairness over Wireless LAN”, *IEEE INFOCOM* 2003

[6] Andrea Detti, Emiliano Graziosi, Valerio Minichiello, Stefano Salsano, and Vincenzo Sangregorio, “TCP fairness issues in IEEE 802.11 based access networks”,*http://www.eln.uniroma2.it/Stefano Salsano/papers/salsano-tcp-fair-wlan.pdf*

[7] S. Malik, M. Jaseemuddin, G. Ravindran, and H. El-Sayed, “Bandwidth Management for Improving Peformance and Fairness in IEEE 802.11 based Wireless Networks”, in *ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-05)*, January 3-6, 2005, Cairo , Egypt.

[8] Dennis Pong and Tim Moors, “Fairness and Capacity Trade-off in IEEE 802.11 WLANs” , in *29th Annual IEEE International Conference on Local Computer Networks* 2004, pp. 310-317

[9] Yuan Yuan, Daqing Gu, Williams Arbaugh and Jinyun Zhang, “Achieving Packet Level Quality of Service through Scheduling in Multi-rate WLANs”,in *Vehicular Technology Conference* September 2004, Volume 4, pp: 2730- 2734.

[10] See-Hwan Yoo, Jin-Hee Choi, Jae-Hyun Hwang and Chuck Yoo, “Eliminating the Performance Anomaly of 802.11b”,in *Lecture Notes in Computer Science* Publisher: Springer Berlin/Heidelberg Volume 3421/2005, p.1055

[11] Hyogon Kim, Sangki Yun, Inhye Kang, and Saewoong Bahk, “Resolving 802.11 Performance Anomalies through QoS Differentiation”,*IEEE COMMUNICATIONS LETTERS*, VOL. 9, NO. 7, JULY 2005

[12] Jeng Farn Lee, Wanjiun Liao and Meng Chang Chen, “Proportional Fairness for QoS Enhancement in IEEE 802.11e WLANs”, in *The IEEE Conference on Local Computer Networks 30th Anniversary* November 2005, pp. 503-504.

[13] Deyun Gao, Jianfei Cai, King Ngi Ngan,“Admission Control in IEEE 802.11e Wireless LANs”, *IEEE Network* July/August 2005, pp 6-13.

[14] Jeonggyun Yu,Sunghyun Choi and Jaehwan Lee,“Enhancement of VoIP over IEEE 802.11 WLAN via dual Queue Strategy”, in *IEEE International conference on communication* ICC’04, Paris, France, June 2004.

[15] Lidong Lin, Haohuan Fu and Weija, “An Efficient Admission Control for IEEE 802.11 Networks Bases on Throughput analysis of (Un)saturated Channel”, IEEE Globecom 2005.

[16] “MAC Layer Admission Control and Priority Re-allocation for Handling QoS Guarantees in Non-cooperative wireless LANs”, *ACM/Springer Mobile Network and Applications* (MONET) 2005, Vol 10, pp 947-959.

[17] Adlen Ksentini, Abdelhak Gueroui and Mohamed Naimi,“ Adaptive Transmission Opportunity with Admission Control for IEEE 802.11e Networks”, *ACM/IEEE MSWIM 2005*, Montreal, quebec, Canada, October 2005.

[18] Nitin H. Vaidya, Paramvir Bahl and Seema Gupta, “Distributed Fair Scheduling in a Wireless LAN”, in *Proceedings of ACM MOBICOM*, Boston, 2000, pp167-178.

[19] B. Sadeghi, V. Kanodia, A.Sabharwal and E. Knightly, “Opportunistic Media Access for Multirate Ad HOC Networks”, in *Proceedings of ACM MOBICOm’02*, Atlanta, Gerogia, September 2002.

[20] Gavin Holland, Nitin Vaidya and Paramvir Bahl"A Rate-Adaptive MAC protocol for Multi-Hop wireless Network", in *ACM/IEEE Int. conf on Mobile Computing and Networking (MOBICOM'01)*, Rome,Italy, Jul 2001.

[21] Meru Networks, *http://www.merunetworks.com*

[22] Cisco/Airespace, *http://www.cisco.com*

[23] Trapeze networks, *http://www.trapeze.com*

[24] Extricom networks, *http://www.extricom.com*

[25] Dennis Pong and Tim Moors, "Call Admission Control for IEEE 802.11 Contention Access Mechanisum", in *Global Telecommunications Conference* December 2003, Volume 1, pp 174- 178

[26] Sri Harsha, Anurag Kumar and Vinod Sharma, "Analytical Models for Capacity Estimation of Combined VoIP and TCP file transfers over EDCA in an IEEE 802.11 WLANs", in *IEEE IWQoS 2006*

[27] Raffaele Bruno, Marco Conti and Enrico Gregori, "Throughput analysis of TCP clients in Wi-Fi hot spot Networks", in *Networking 2004*, LNCS 2042, pages 626-637, 2004

[28] D.J. Leith and P. Clifford, "Using the 802.11e EDCF to Achieve TCP Upload Fairness over WLAN Links",in *Third International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, WiOpt'05* 2005, pp. 109-118.

[29] Network Simulator - ns-2,*http://www.isi.edu/nsnam/ns/*

[30] IEEE802.11e for the ns-2 simulator,*http://sourceforge.net/projects/ieee80211e-ns2/*