Performance Analysis of IEEE 802.11 Multirate WLANs: Time Based Fairness Vs Throughput Based Fairness

A.V. Babu and Lillykutty Jacob

Department of Electronics and Communication Engineering
National Institute of Technology, Calicut-673601, India
Email: babu,lilly@nitc.ac.in

Abstract

This paper investigates the issue of fairness in a heterogeneous environment in which stations in IEEE 802.11 based wireless LANs use multiple data rates for transmission. CSMA/CA provides throughput based fairness and hence the saturation throughput of any station is limited by the saturation throughput of the station with the lowest bit rate in the network. We consider a new fairness notion called time based fairness in which each competing node receives an equal share of the wireless channel occupancy time. We demonstrate that the time based fairness can achieve larger aggregate throughput while still guaranteeing that no node receives worse channel access than it would in a single rate wireless LAN. We propose and analyze two mechanisms to achieve time based fairness; and develop an analytical model for computing saturation throughput and saturation delay for competing nodes in a 802.11 based multirate wireless LAN. Using Jain's fairness index, optimal MAC parameters required to achieve maximum fairness between slow and fast stations in terms of channel occupancy time are obtained. The impact of these mechanisms on throughput and frame delay of slow and fast stations is also explored.

Key words: IEEE 802.11 Wireless LAN, Throughput based fairness, Time based fairness, Performance Analysis

I. Introduction

The IEEE 802.11 protocol is the dominating standard for Wireless LANs (WLANs) and employs Distributed Co-ordination Function (DCF) as the essential medium access control method. CSMA/CA provides equal transmission opportunity for all participating stations when all nodes experience similar channel conditions. When frame size used by all stations are also same, they achieve equal throughputs (throughput based fairness). Because of varying channel conditions, 802.11 standards support multiple data rates with dynamic rate switching capability to improve the performance. In this case time to send a frame depends on the used data rate and a longer time will be necessary if a lower rate is used. As a result, medium underlies a completely unfair time allocation for stations with different transmission rates. This unfairness is reflected in the throughput of the station with the highest bit rate. Further the aggregate throughput is reduced to a level much closer to what one gets when all competing nodes are slow.

In this paper we propose and analyze two mechanisms to achieve time based fairness in which each competing node receives an equal share of the wireless channel occupancy time, on long term. We demonstrate that this new notion of fairness can achieve larger aggregate throughput than throughput based fairness, while still guaranteeing that no node receives worse channel access than it would in a single rate wireless LAN. The performance anomaly of IEEE 802.11b has been analyzed in [4] using a simplified model with saturated sources, but no solutions were proposed to solve the unfairness issue. In [5] authors compensate the unfairness in channel time allocation in IEEE 802.11 WLANs when stations are transmitting at different rates, by assigning shorter back off time to high rate stations. However, the authors focused on throughput as their metric and no analysis was done to verify the fairness of their proposed scheme. An algorithm to be implemented at the Access Point (AP) to achieve time based fairness in an infrastructure-mode WLAN was proposed in [6]. In [7], authors investigate the issue of fairness and use the new feature of IEEE 802.11e draft standard, transmission opportunity (TXOP), to provide temporal fairness. However authors adopt a simpler and approximate calculation of WLAN capacity. The throughput and delay analysis for IEEE 802.11 protocol was done in [8 & 9] assuming saturated sources and single transmission rate. The capacity of WLAN is a function of number of contending stations, minimum and maximum CWs and PHY rates used by the stations. This motivates us to develop a model for multi-rate WLANs with saturated sources, which can be used to determine the actual throughput and delay experienced by competing nodes. The rest of this paper is organized as follows: Our analytical model is presented in Section 2 to evaluate saturation throughput and saturation delay in multirate WLAN. In Section 3, we propose methods to solve unfairness of CSMA/CA.

0-7803-9305-8/05/$20.00 ©2005 IEEE

We define a new fairness index and propose mechanisms to achieve the fairness objective. Analytical solutions for optimal minimum contention window size and optimal frame size for slow station to meet the desired fairness objective are derived. Section 4 gives analytical and simulation results. The paper is concluded in Section 5.

2. Performance Analysis Under Multi-rate Scenario

To overcome unfairness of conventional CSMA/CA in a multi rate scenario, we use priority control according to the data rate of the stations. By selecting appropriate priority depending on the data rate and giving more transmission opportunity for high data rate station, the system capacity can be improved. Consider a heterogeneous WLAN of nodes with two different rates, lower rate corresponding to slow station and higher rate corresponding to fast station, and assume that each station is in saturation, i.e., has always a frame ready for transmission. Let there be n_S slow stations and n_f fast stations and let $n = n_s + n_f$. The maximum backoff stage, m', and the retry limit m, are assumed to be equal for the two priority classes. Let $W_{i,j}$ represents the contention window (CW) size in the j-th retry/retransmission for rate-i station, $j \in \{0,1,\ldots m\}$. For rate-i station, contention window size in the j-th retry/retransmission, $W_{i,j}$, and the retry limit, m, are related as follows:

$$W_{i,j} = \begin{cases} (2)^j W_{i,0} \; ; j = 0,1,\ldots m' - 1; m \geq m' \\ (2)^{m'} W_{i,0} \; ; j = m',\ldots\ldots m \; ; \; m \geq m' \end{cases} \quad (1)$$

Let $s_i(t)$ and $b_i(t)$ be the stochastic processes representing the back off stage and the back off time counter value, respectively, for rate-i traffic flow. We assume that the frame collision process is Bernoulli which allows us to describe the state of each station, slow or fast, with two dimensional DTMC $\{s_i(t), b_i(t)\}$ shown in Figure 1. The state transition occurs at the beginning of the next slot time, where a transition can happen after a transmission or an empty slot time. Let $p_{c,i}$ and τ_i, respectively, represent the frame collision probability and frame transmission probability for rate-i station. From the DTMC the frame transmission probability can be derived as follows. Let $q_i(j,k) = \lim_{t\to\infty} \{P(s_i(t) = j, b_i(t) = k)\}$, $j \in (0,m)$, $k \in (0, W_{i,j} - 1)$ be the stationary distribution for the chain. In steady state the following relations can be obtained:

$$q_i(j,0) = (p_{c,i})^j q_i(0,0); 0 \prec j \leq m$$

$$q_i(j,k) = \frac{W_{i,j} - k}{W_{i,j}} q_i(j,0); 0 \leq j \leq m; 1 \leq k \leq W_{i,j} - 1$$

$$\sum_{j=0}^{m} \sum_{k=0}^{W_{i,j}-1} q_i(j,k) = 1$$

$$\tau_i = \sum_{j=0}^{m} q_i(j,0)$$

$$\tau_i = \begin{cases} \dfrac{2(1-2p_{c,i})(1-(p_{c,i})^{m+1})}{[W_{i,0}\,(1-(2p_{c,i})^{m'+1})(1-p_{c,i}) + (1-2p_{c,i})(1-(p_{c,i})^{m+1})} ; \; m \geq m' \\ + W_{i,0}(1-2p_{c,i})(2p_{c,i})^{m'} p_{c,i}(1-(p_{c,i})^{m-m'})] \end{cases} \quad (2)$$

The collision probability for the slow and fast stations are given by:

$$p_{c,s} = 1 - (1-\tau_s)^{n_s - 1}(1-\tau_f)^{n_f}$$
$$p_{c,f} = 1 - (1-\tau_s)^{n_s}(1-\tau_f)^{n_f - 1} \quad (3)$$

Equations (2) and (3) form a system of non linear equations which can be solved using numerical techniques. Let p_{tr} be the probability that there is at least one transmission in a given time slot and $p_{tr,i}$ be the corresponding probability for rate-i station.

Clearly, $p_{tr} = 1 - (1-\tau_s)^{n_s}(1-\tau_f)^{n_f}$;

$$p_{tr,i} = 1 - (1-\tau_i)^{n_i} \; ; i = s, f. \quad (4)$$

Let $p_{s,s}$ denotes the probability of having a successful transmission by one of the slow stations assuming that at least one slow station transmits a frame and let $p_{s,f}$ be the corresponding probability for the fast station. These are calculated as follows:

$$p_{s,s} = \frac{n_s \tau_s (1-\tau_s)^{n_s - 1}(1-\tau_f)^{n_f}}{p_{tr,s}}$$

$$p_{s,f} = \frac{n_f \tau_f (1-\tau_f)^{n_f - 1}(1-\tau_s)^{n_s}}{p_{tr,f}} \quad (5)$$

Saturation throughput of rate-i station, Z_i, defined as the ratio of successfully transmitted payload bits corresponding to slow station traffic to the average length of a slot, is calculated as follows [8]:

$$Z_i = \frac{p_{tr,i} p_{s,i} E[L_i]}{E[\sigma]} \quad (6)$$

Where $E[L_i]$ denotes the average payload corresponding to rate-i station. $E[\sigma]$ represents the average length of a slot time which is determined as follows: Let σ_0, $T_{s,i}$, and $T_{c,i}$ respectively, represent the duration of an empty time slot, the average time channel is sensed busy because of successful transmission of rate-i station traffic, average time the channel is sensed busy because of a transmission failure due to collision of rate-i station.

$\frac{1}{W_{i,o}}$ if it comes from stage (m, 0); otherwise $\frac{1-p_{c,i}}{W_{i,0}}$

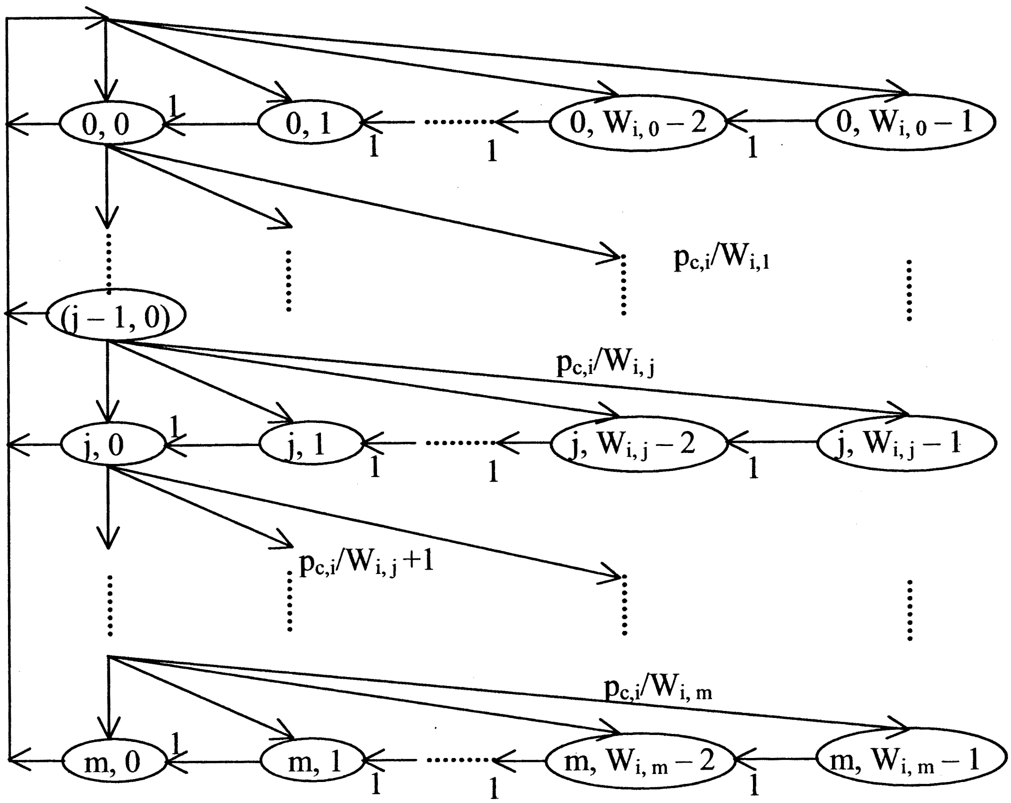

Figure 1: The state transition diagram for rate-i station

Then

$$E[\sigma]=(1-p_{tr})\sigma_0+p_{tr,s}p_{s,s}T_{s,s}+p_{tr,f}p_{s,f}T_{s,f}$$
$$+p_{tr,s}(1-p_{s,s})(1-p_{tr,f})T_{c,s}+p_{tr,f}(1-p_{s,f})(1-p_{tr,s})T_{c,f}$$
$$+p_{tr,s}p_{tr,f}\max(T_{c,s};T_{c,f}) \quad (7)$$

The values of $T_{s,s}$, $T_{s,f}$, $T_{c,s}$, and $T_{c,f}$ for basic access method are calculated as follows [8]:

$$T_{s,i}=T_{H,i}+T_{E(L_i)}+SIFS+\beta+T_{ACK,i}+DIFS+\beta$$
$$T_{c,i}=T_{H,i}+T_{E(L_i)}+DIFS+\beta \quad (8)$$

Where β is the propagation delay, $T_{H,i}$ is the transmission time for physical and MAC layer headers of rate i station, and $T_{E[Li]}$ is the transmission time for average payload corresponding to rate-i station. Upon successful receipt of a packet, an ACK is transmitted at the physical rate of the received packet.

Saturation Delay which is the expected delay from when a frame becomes head-of-line (HOL) until it is successfully delivered, $E[D_i]$ is calculated as follows:

$$E[D_i]=\sum_{j=0}^{m}\frac{(p_{c,i})^j(1-p_{c,i})}{1-(p_{c,i})^{m+1}}\sum_{k=0}^{j}\frac{W_{i,k}}{2}E[\sigma]$$
$$+\sum_{j=0}^{m}\frac{j(p_{c,i})^j(1-p_{c,i})}{1-(p_{c,i})^{m+1}}E[T_{c,i}]+T_{s,i} \quad (9)$$

Where the first term represents the total waiting time in a typical backoff interval, which is the sum of waiting times in as many back off stages as the number of transmission attempts before the successful transmission of the frame; the second term represents the total time wasted due to its own collision in all the successive transmission attempts and the third term represents the time required to actually accomplish the successful transmission and receipt of ACK. $E[T_{c,s}]$ and $E[T_{c,f}]$, respectively, represent the average time a slow and fast station spend in a collision. These quantities are given as follows:

$$E[T_{c,s}]=p_{tr,s}(1-p_{s,s})(1-p_{tr,f})T_{c,s}+p_{tr,s}p_{tr,f}\max(T_{c,s};T_{c,f})$$
$$E[T_{c,f}]=p_{tr,f}(1-p_{s,f})(1-p_{tr,s})T_{c,f}+p_{tr,s}p_{tr,f}\max(T_{c,s};T_{c,f}) \quad (10)$$

3. Solutions for Performance Anomaly

It has been proven that if there are two bit rates under the same channel conditions, the saturation throughput of any station will be equal to the saturation throughput of station with the lower data rate [4]. To solve the throughput degradation occurring in a multi rate environment, we propose two mechanisms: In the first method CW_{min} of slow host is increased to decrease its transmission opportunity. In the second method the frame size of slow station traffic is reduced to decrease its transmission time for each transmission opportunity. Our objective is to make the long term channel occupancy time of slow and fast stations equal. As mentioned above, this is done in one of the two ways: (i) by increasing the CW_{min} of the slow stations, which causes the mean contention delay ($E[D_s]$) of slow station to increase; (ii) by reducing the frame size of slow stations, which causes successful transmission time ($T_{s,s}$) at each transmission opportunity to decrease. Therefore, for station i, we define $y_i = \dfrac{T_{s,i}}{E[D_i]}$. We use Jain's fairness index to evaluate how fair a particular allocation is [10]: $F = \dfrac{(\sum_{i=1}^{n} y_i)^2}{n\sum_{i=1}^{n} y_i^2}$ where n is the total number of stations. It has been proved that $F \le 1$ and equality holds iff $y_i = y$ for all i. We find the optimal CW_{min} and optimal frame size for the slow station, which maximize the fairness index. Analytical expressions for optimal CW_{min} and optimal frame size for the slow station are obtained next.

3.1. An approximate expression for optimal CW_{min} for slow station, $W_{s,0}^*$

F=1 when
$$\frac{T_{s,s}}{E[D_s]} = \frac{T_{s,f}}{E[D_f]} \tag{11}$$

Assume minimum contention window sizes $W_{s,0}$ & $W_{f,0}$ >>1 and the transmission probabilities τ_s & τ_f <<1. Then we have $p_{c,s} = p_{c,f}$. From Eqn (2) for τ_i and, assuming the retry limit to be infinite, $\dfrac{\tau_s}{\tau_f} \approx \dfrac{W_{f,0}}{W_{s,0}}$. Further from Eqn (6), it is possible to find ratio of the saturation throughputs of slow and fast stations. Then by substituting Eqn (5), we have

$$\frac{Z_s}{Z_f} = \frac{n_s\tau_s(1-\tau_s)^{n_s-1}(1-\tau_f)^{n_f}E[L_s]}{n_f\tau_f(1-\tau_f)^{n_f-1}(1-\tau_s)^{n_s}E[L_f]} \approx \frac{n_s\tau_sE[L_s]}{n_f\tau_fE[L_f]} \tag{12}$$

Hence the ratio of the throughputs per station is given by:

$$\frac{z_s}{z_f} = \frac{Z_s/n_s}{Z_f/n_f} \approx \frac{E[L_s]/W_{s,0}}{E[L_f]/W_{f,0}} \tag{13}$$

According to Little's law, for any queueing system, the average number of customers in the system is equal to the average experienced delay multiplied by the average customer departure rate. Because each of the n_s (resp. n_f) stations is contending with an HOL frame and $Z_s/E[L_s]$ (resp. $Z_f/E[L_f]$) represents throughput in frames/seconds,

$$E[D_s] = \frac{n_s}{Z_s/E[L_s]}$$
$$E[D_f] = \frac{n_f}{Z_f/E[L_f]} \tag{14}$$

Hence using Eqns (13) and (14), the following relation is obtained:
$$\frac{E[D_s]}{E[D_f]} = \frac{E[L_s]}{E[L_f]}\frac{z_f}{z_s} = \frac{W_{s,0}}{W_{f,0}} \tag{15}$$

On combining this with Eqn (11), we obtain

$$W_{s,0}^* = \frac{T_{s,s}}{T_{s,f}}W_{f,0}. \tag{16}$$

3.2. Optimal frame size, $E[L_s^*]$

Because we have assumed that all nodes are in saturation, if the MAC parameters of all nodes are same, they all experience equal back off delay. Assuming one slow node and one fast node, the desired fairness objective can be met if $T_{s,s} = T_{s,f}$. We use Eqn (8) to find the optimal frame size for slow station:

$$E[L_s^*] = (T_{H,f} - T_{H,s} + T_{ACK,f} - T_{ACK,s} + T_{E[L_f]})R_s \tag{17}$$

Where R_s represents the slow station bit rate.

4. Numerical Results

In this section we present the numerical results of our analysis. To validate our analytical model, we also present simulation results. Our simulation model is developed using network simulator-2 based on IEEE 802.11b standard [11]. We select 1 Mbps for the slow station and 11 Mbps for the fast station. Initially MAC parameters of slow and fast stations are kept same: m = 7; $m' = 5$; $W_{s,0} = W_{f,0}=16$; $n_s = n_f = 10$; $E[L_s]=E[L_f]=1450$ bytes. We find that the saturation throughput achieved by slow and fast stations are equal (throughput based fairness).

Next we keep $W_{f,0}=16$; $n_s=n_f=10$; $E[L_s]=E[L_f]=1450$ bytes, and increase $W_{s,0}$, the minimum CW of slow station. As $W_{s,0}$ increases, the saturation throughput of slow host decreases and that of fast host increases. We evaluate the fairness index defined earlier for each $W_{s,0}$ and find that there is an optimal $W_{s,0}$ which makes F = 1 and its value 134. The numerical solution of the equations from our analytical model gives the optimal value to be 132 while our approximate model based on equation (16) gives the optimal value to be 130.2. We change the number of slow and fast stations and find that the optimal $W_{s,0}$ which achieves F = 1 is the same and is independent of the number of slow and fast hosts. The fairness index is plotted against minimum CW of slow station in Figure 2.

Next we keep $W_{s,0}=W_{f,0}=16$; $n_s=n_f=10$; $E[L_f]=1450$ bytes and vary $E[L_s]$, the frame size of slow station traffic. We find that there is an optimal frame size that achieves F=1 and is equal to 64 bytes while through mathematical analysis we find that the optimal value is 60 bytes. We also find that the optimal frame size is independent of the number of competing slow and fast hosts. The fairness index is plotted against frame size of slow station in Figure 3. We find that simulation and analytical results are approximately equal.

Next we use the optimal values of minimum CW, $W_{s,0}$ and frame size, $E[L_s]$, obtained in the above analysis to investigate further results. We consider 1 Mbps/11 Mbps scenario and keep $n_s=1$. The total throughputs for the two mechanisms (optimal CW and optimal frame size) are determined for increasing number of fast stations. We also determine the total throughput for the case where all stations use the same MAC parameters and frame size (basic configuration). We find that the aggregate throughputs obtained for optimal $W_{s,0}$ and $E[L_s]$, are much greater than the aggregate throughput achieved for basic configuration. Figure 4 shows the variation of total throughput against n_f, the number of fast stations. Here we conclude that by using optimal value of $W_{s,0}$ or $E[L_s]$ for the slow station, it is possible to achieve larger aggregate throughput compared to throughput based fairness scheme.

We compare the station throughput, that are achievable under the notions of throughput based fairness and time based fairness. Total throughput of two nodes simultaneously exchanging data at the same rate is shown in Table 1. Table 2 shows the throughput of each node in throughput based fairness scheme and the proposed time based fairness scheme (optimal CW method). We observe the following: The total throughput in time based fairness scheme is much greater than that in throughput based fairness scheme (improved by 175%). The throughput achievable by a fast station is much greater in time based fairness while slow station achieves lesser throughput. In fact under time based fairness, the 1 Mbps station achieves the throughput it would have achieved if both the nodes were running at 1 Mbps.

4.1. Effect on frame delay

When we use optimal CW method to achieve time based fairness, the delay experienced by slow station increases. We find that slow station experiences comparatively lesser delay when optimal frame size method is used to achieve time based fairness. But we also find that the throughput achieved by slow station and the aggregate throughput achieved is greater in optimal CW method. Figures 5 shows the delay experienced by slow and fast stations as number of active fast stations increases.

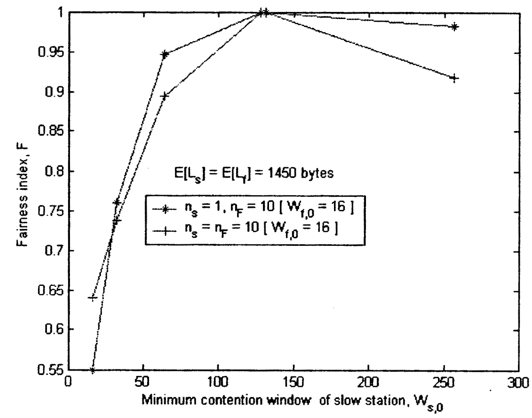

Figure 2: Fairness index versus minimum CW of slow station

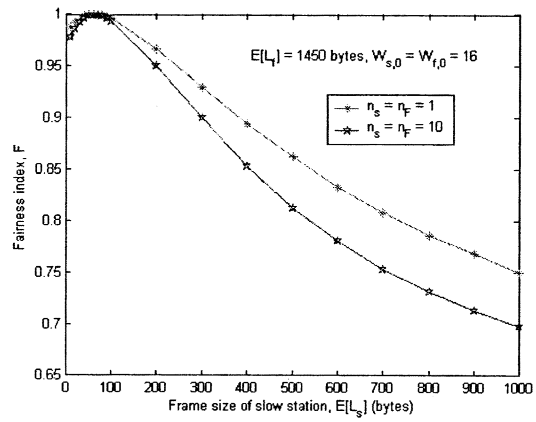

Figure 3: Fairness index versus frame size of slow station

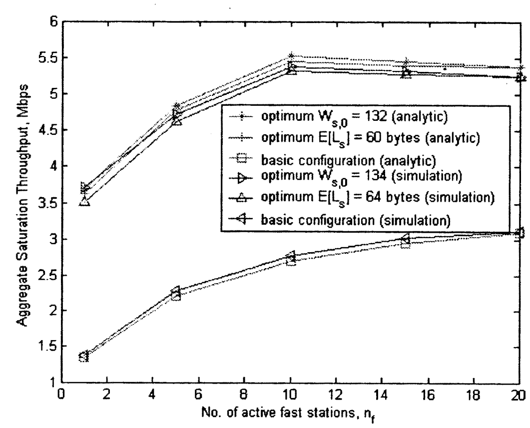

Figure 4: Aggregate saturation throughput versus no. of active fast stations

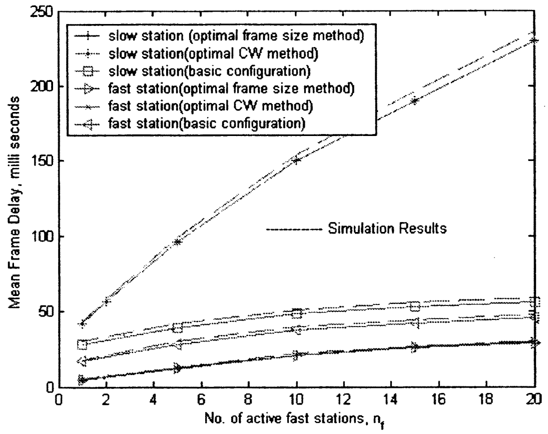

Figure 5: Mean frame delay versus no. of active fast stations.

Table 1: Throughput achieved by two nodes in single rate WLAN

Data Rate (Mbps)	Packet Size (Bytes)	No. of Nodes	Total Throughput (Mbps) Analytic/simulation
11	1450	2	6.7461/6.9124
1	1450	2	0.8618/0.8804

Table 2: Throughput achieved by two nodes in 1 Mbps/11 Mbps multirate WLAN

Fairness Criterion	Throughput of 1 Mbps node (Mbps) Analytic/ Simulation	Throughput of 11 Mbps node (Mbps) Analytic/ Simulation	Total (Mbps) Analytic/ Simulation
Through put based fairness	0.6723/ 0.6893	0.6723/ 0.6893	1.3446/ 1.3786
Time based fairness	0.3762/ 0.3752	3.3283/ 3.3322	3.7045/ 3.7074

5. Conclusions

In this paper we have investigated the issue of fairness in IEEE 802.11 based heterogeneous wireless LAN. The notion of time based fairness was proposed to achieve equal long term channel occupancy time for all the participating stations. We proposed and analyzed two mechanisms to achieve this new notion of fairness. We presented an analytical model to determine saturation throughput and saturation delay experienced by contending nodes transmitting at different rates. Using Jain's fairness index, we determined optimal MAC parameters required to achieve fairness objective in terms of channel occupancy time. The impact of these mechanisms on throughput and frame delay of slow and fast stations were determined. We established that this new notion of fairness can achieve larger aggregate throughput while still guaranteeing that no node receives worse channel access than it would in a single rate wireless LAN. Extensive simulations were conducted to validate the analytical results.

References

[1] IEEE 802.11 WG, Part II: Wireless LAN Medium Access Control and Physical Layer specifications, 1999.

[2] IEEE 802.11 WG, Draft Supplement to part II: Wireless LAN Medium Access Control and Physical Layer specifications: MAC enhancements for QOS, Nov.2001.

[3] IEEE 802.11a WG, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High speed Physical Layer in the 5 GHz band, September 1999.

[4] Martin Heusse et al, " Performance Anomaly of IEEE 802.11b," *IEEE Infocom* 2003

[5] K. Saitoh et al, " An Effective Data Transfer Method by Integrating Priority Control into Multirate Mechanisms For IEEE 802.11 Wireless LANs," *IEEE VTC* 2002, Vol. 1.

[6] Godfrey Tan and John Guttag, "Time based Fairness Improves Performance in Multirate WLANs," *The USENIX Annual Technical Conference*, Boston, MA, June 2004.

[7] Dennis Pong and Tim Moors, "Fairness and Capacity Trade off in IEEE 802.11 WLANs," *IEEE LCN* 2004.

[8] G.Bianchi, "Performance Analysis of The IEEE 802.11 Distributed Co-ordination Function," *IEEE JSAC*, Vol. 18, No.3, March 2000.

[9] P. Chatzimisios , A. C. Boucouvalas and V. Vitsas, " IEEE 802.11 Packet Delay - A Finite Retry Limit Analysis ", *Proceedings of the* IEEE *Global Telecommunications Conference (*Globecom 2003), vol. 2, pp. 950-954.

[10] R.K. Jain et al, " A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems, " *DEC report*, DEC-TR-301,September 1984.

[11] Y. Xiao and J. Rosdahl, "Throughput and Delay Limits of IEEE 802.11," *IEEE Comm. Letters*, Vol. 6, No. 8, Aug. 2002, pp. 355-357.