In the Fifth IFIP International Conference on Mobile and Wireless Communications Networks (MWCN), Singapore, Oct. 27-29, 2003

Study of IEEE 802.11 Fairness and its Interaction with Routing Mechanism

Zhifei Li, Sukumar Nandi, and Anil K. Gupta
School of Computer Engineering, Nanyang Technological University, Singapore-639798

***Abstract*— The MAC protocol IEEE 802.11, used in the mobile ad hoc networks (MANETs), suffers from the well-known fairness problem. In this paper, using static routing, we first study the MAC fairness by considering three factors including the hidden-terminal, capture, and imprecise EIFS, which are all related to the signal attenuation property of the wireless networks. Since the MAC protocol and the routing protocol interact with each other, we then study the interaction between MAC fairness and routing mechanism by replacing static routing with AODV and DSDV. The main lessons learnt from our studies include: *(i)*we should distinguish between the MAC fairness and the overall system fairness, *(ii)*in studying the MAC fairness, we should isolate the factors (e.g., routing protocol) that are not directly related to the MAC layer, and *(iii)*to study the overall system fairness, in addition to the study of MAC fairness, we should also consider the effects of many other factors such as the routing mechanism, traffic pattern, mobility pattern etc.**

I. Introduction

In the Mobile Ad-hoc Networks (MANETs), MAC and routing are two major layers that need careful consideration. IEEE 802.11 [1] has been the most popular MAC protocol adopted in the research-work on MANETs, since it is the de facto industry standard for Wireless LANs. Meanwhile, many routing protocols have been proposed for the MANETs recently. Basically, they can be classified into two categories: proactive and reactive. Destination-Sequenced Distance-Vector (DSDV) [2] is an example of pro-active protocol, whereas Ad-hoc On-demand Distance Vector (AODV) [3] and Dynamic Source Routing (DSR) [4] belong to the reactive category.

Though IEEE 802.11 defines two MAC protocols, i.e., Point Coordination Function (PCF) and Distributed Coordination Function (DCF), only DCF is used in MANETs since PCF requires base stations. DCF is a CSMA/CA based protocol, which has two main components: Carrier Sensing (CS) and Collision Avoidance (CA). Since CS and CA work in a distributed manner without having precise information, they cause unfairness as pointed out in [5], [6]. While [5] mainly focuses on the unfairness caused by the CA algorithms, [6] points out that the imprecise EIFS value, which is related to the CS, is also a major reason for unfairnes. In [6], an enhanced CS (ECS) is proposed, which solves the imprecise EIFS problem to a great extent and thus improves the fairness substantially. In addition to CA algorithms and imprecise EIFS problem, the hidden-terminal problem and signal capture, which are the two most important issues in wireless networks, also affect the MAC fairness. Therefore, it is very important to simultaneously consider the effects of hidden-terminal, capture, and imprecise EIFS in studying the MAC fairness. Moreover, since the MAC protocol and the routing protocol interact with each other, which in turn affect the overall performance of an MANET, it is very important to characterize the interaction between the MAC fairness and the routing mechanism. However, to the best of our knowledge, no such work has been done. Bearing in mind these observations, we carry out our research work with the following two objectives: *(i)*using static routing, we illustrate how hidden-terminal problem, capture, and imprecise EIFS problem affect the fairness of IEEE 802.11, and *(ii)*replacing static routing with AODV and DSDV, we study the interaction between MAC fairness and routing mechanism.

The rest of the paper is organized as follows. In Section II, we present the factors affecting MAC fairness. In Section III, we study the MAC fairness. The interaction between MAC fairness and routing mechanism is studied in Section IV. The paper is concluded in Section V.

II. Factors Affecting MAC Fairness

In this section, we discuss the three factors described above, which affect MAC fairness.

A. Capture Effect

A special characteristic of wireless propagation is the attenuation of the transmission power over the distance traversed by the signal. Based on the attenuation, two ranges are defined: the transmission range (TR) and the sensing range (SR). Normally, the SR is greater than the TR. Correspondingly, we call a frame detected by a node within TR as the TR frame, and a frame detected by a node out of TR but within SR as the SR frame. If there is no collision or any other error, a node can receive a TR frame correctly. In contrast, a SR frame can be detected by the carrier sensing but it cannot be received correctly, and therefore it will be marked as an erroneous frame by the physical layer before being passed to the MAC layer. If capture is not allowed, a collision occurs whenever there are multiple frames (which may be a mixture of TR and SR frames) around a receiving node and all the frames will be destroyed. On the contrary, if capture is allowed, one of the multiple frames may be captured (i.e., received correctly) if both the following conditions are satisfied: *(i)*the Signal to Noise Ratio (SNR) of the captured frame is greater than a threshold, and *(ii)*the frame satisfying the first condition arrives at the receiving node earlier than any other frames. Intuitively, capture can improve the system throughput, but it results in unfair sharing of bandwidth with preference given to nodes, whose packets have higher SNR.

B. Hidden Terminal Problem

In addition to the capture issue, there exists the well-known hidden-terminal problem in wireless networks. Specifically, when two nodes are out of the SR range of each other and they simultaneously try to transmit to nodes, which are in the overlapping area of the two transmitting nodes'SR, the two transmitting nodes are hidden from each other. Note that our definition of hidden-terminal problem is different from the traditional one because we assume that the SR is greater than TR. The example shown in Figure 1 illustrates this problem. In this example, there are two single-hop flows: flow from A to B and flow from D to C. The distance between any two neighboring nodes is about one hop. In other words, two neighboring nodes are within the TR of each other. The SR is assumed to be about two times of TR. Therefore, nodes A and D are hidden from each other and they may initiate transmission simultaneously, resulting in collision or capture. It will be clear in the later discussion that the hidden-terminal problem results in unfairness.

Fig. 1. 4-nodes with Two Single-hop Flows

C. Imprecise EIFS Problem

To combat with the hidden-terminal problem, IEEE 802.11 defines a four-way handshaking technique, where a sequence of Request To Send (RTS), Clear To Send (CTS), data, and Acknowledgement (ACK) frames, is transmitted whenever the length of the data frame is more than a threshold known as the RTSThreshold. For the convenience, we call the exchange of RTS/CTS/data/ACK frames as a frame exchange sequence (FES). FES (A, B) represents a FES between nodes A and B, initiated by node A.

In the IEEE 802.11, whenever a node detects an SR frame, which is treated as an error, this node defers its transmission by a duration indicated by the Extended Inter Frame Space (EIFS) constant. The idea, that a node defers its transmission with EIFS when the node gets an SR frame, is to allow the next frame in a FES between two other nodes to be complete. Since the next frame can be of any type (i.e., CTS, data, or ACK), the transmission time of this frame may differ substantially. Therefore, corresponding to different SR frames received, the EIFS value should be different. In particular, the EIFS value corresponding to a SR frame should be equal to the time needed to complete the transmission of the next frame in the sequence. However, the IEEE 802.11 does not distinguish among different SR frames and uses the same constant EIFS value in all the cases, which results in large- and small-EIFS problems and thus exhibits unfairness. For convenience, we refer to these two problems as imprecise EIFS problem.

To illustrate the large-EIFS problem, let us consider the scenario in Figure 1 and a situation where nodes C and D are deferring their transmission, while FES (A, B) is in progress (see Figure 2). After node B sends a CTS to A, since node D is in the sensing range of node B, node D gets a SR frame and defer its transmission by EIFS. Then, node A begins to send its data frame to B. However, since node D is out of the SR of A, node D cannot hear this frame. Moreover, since the EIFS is equal to SIFS + TxTime (ACK) + DIFS [1] which is much smaller than the transmission time of node A's data frame, after deferring for the EIFS duration, node D starts to contend for the medium, and may send out a frame even though node A is still transmitting the data frame to B. The frame sent by node D is definitely discarded by node C. We call this small-EIFS problem because the EIFS value is smaller than it should be to reflect the state of the medium. In fact, in this situation, the EIFS value should be large enough to allow for the complete transmission of the A's data frame.

Now let us consider that node B sends an ACK back to node A. Corresponding to this ACK, node D gets an SR frame and defers its transmission with EIFS. We can see that, after the completion of the ACK, before node D backs-off by a random period, it has to defer by EIFS duration rather than by DIFS as done by other nodes (e.g., nodes A and B). Since EIFS is greater than DIFS, the deferment at node D is certainly of much longer duration than that at the other nodes. We call this large-EIFS problem because the EIFS value is larger than it should be to reflect the state of the medium. In fact, the EIFS value should be equal to DIFS in this case.

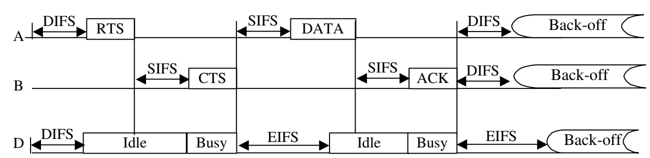

Fig. 2. Time diagram when FES (A, B) is in progress while D defers

In [6], we have shown that the large- and small-EIFS problems result in substantial unfairness. To solve these problems, Enhanced Carrier Sensing (ECS) has been proposed [6]. In ECS, whenever a node gets a SR frame, the node should identify the type of this SR frame based on its length, which is made different for different frames. Then the node adopts different EIFS values correspondingly, rather than a fixed EIFS value. As a result, the ECS reflects the medium state precisely and solves the imprecise EIFS problem greatly.

III. IEEE 802.11 Fairness Under Static Routing

Since the example of Figure 1 incorporates all the factors under consideration, we conduct an extensive study of this topology. Specifically, under a given routing protocol, we conduct simulations under four different scenarios, called S1 to S4, which are described in Table I. In this section, the simulation results are presented under the static routing. For convenience, we refer to the scenarios as static-S1 to static-S4.

The simulation environment is as follows. NS-2 with CMU wireless extensions [7] is used. For each single-hop flow, a Constant Bit Rate (CBR) traffic generates 200 packets per second. Each packet is 1000-bytes long, resulting in a traffic

TABLE I

FOUR DIFFERENT SCENARIOS

Scenario	IEEE 802.11/ECS	Allow capture
S1	IEEE 802.11	no
S2	IEEE 802.11	yes
S3	ECS	no
S4	ECS	yes

source rate of 1.6 Mbps, and therefore, nodes A and D are always backlogged. Due to the MAC overhead, the maximum throughput using a raw bandwidth of 2 Mbps is about 1.4 Mbps. Mobility and transmission errors are not explicitly considered in the simulation.

A. Throughput Comparison

Table II presents the average throughputs under the four scenarios. From the table, we can see that the throughput of each flow under static-S1 is much smaller than that under the remaining three scenarios, while the throughputs under the other three scenarios are very close to each other.

TABLE II

AVERAGE THROUGHPUTS UNDER STATIC ROUTING

Scenario	Flow	Average Throughput(M bps)
static-S1	A to B	0.314
	D to C	0.307
static-S2	A to B	0.683
	D to C	0.641
static-S3	A to B	0.662
	D to C	0.672
static-S4	A to B	0.665
	D to C	0.670

Let us first analyze the average throughputs under static-S1. When a node (say node A) is transmitting a data frame, as node D suffers from the small-EIFS problem if the IEEE 802.11 is used, node D may transmit its RTS several times until its contention window (CW) becomes very large. Since the capture is not allowed in this scenario, node B will drop the data frame, resulting in wastage of bandwidth. This explains why the throughputs under static-S1 degrade a lot. On the other hand, node A will be able to retransmit the collided packet successfully if the gap between node D's RTS, which collides with the tail of the data frame sent by node A in the previous try, and node D's another RTS, is long enough. This is very likely once the CW at node D becomes very large. This explains why the throughputs do not degrade to zero.

The dropping of the data frame at node B can be avoided in two ways. One is to prevent node D from sending out the RTS when node A is transmitting its data frame. The ECS achieves this by preventing the small-EIFS problem and thus the throughputs under S3 and S4 do not degrade. Another way to avoid frame dropping is to allow the capture of the data frames as being done in the static-S2. In this scenario, since node A's data is transmitted before node D's RTSs and the SNR of the data frame received at node B is greater than the threshold, the data frame can be captured by node B. Therefore, the throughputs in S2 also do not degrade.

When ECS is used without capture (scenario S3), we achieve throughput about the same as that in the case of S2, which highlights the benefit of the ECS. In static-S4, the capture is used with ECS, but the throughput does not increase compared to that under static-S3. The reason is that in the S4, as explained later, the capture occurs only of the RTS frames and this kind of capture does not enhance the performance.

B. Jain's Index Comparison

Since the average throughput is computed over the entire simulation time, it only tells us whether the bandwidth is distributed fairly between the two flows on a long-term basis. In order to study fairness on a short-term basis, we need to make use of the popular Jain's fairness index [8], which is defined as follows:

$$F_J = (\textstyle\sum_{i=1}^{N} \gamma_i)^2 / (N \sum_{i=1}^{N} \gamma_i^2) \tag{1}$$

where N is the total number of flows sharing the common medium, and γ_i is the fraction of the bandwidth utilized by flow i over a certain number of packets transmitted, say w. As the computation of γ_i depends on w, the ultimate value of the Jain's index also depends on w, though w does not appear in the formula directly. Generally, the F_J value increases with w. According to [8], complete fairness is achieved when F_J =1 and complete unfairness is achieved when F_J =1/N. In our case, as N is equal to two, 0.5 will indicate complete unfairness.

Figure 3 presents the Jain's index with reference to w. The index has been averaged over all sliding windows of w packets, which occur in the simulation run. We notice that all the scenarios exhibit short-term unfairness to a certain extent. Since the common problem existing in all the four scenarios is the hidden-terminal problem, we can conclude that the hidden-terminal problem results in short-term unfairness as observed by experiments in [9]. Interestingly, though the static-S1 attains the worst bandwidth utilization, it achieves the best fairness. This shows the fundamental conflict between optimizing throughput and achieving fairness, which has also been pointed out in [5]. Compared to the static-S2, which shows the worst fairness, the ECS (S3 and 4) improves the fairness while the throughput does not degrade. Moreover, it seems that capture dose not play much role in the fairness under ECS as S3 and S4 almost show the same fairness.

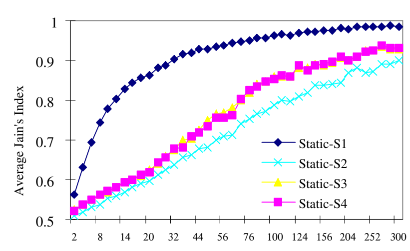

Fig. 3. Fairness Comparison under Static Routing

Before analyzing the results of Jain's index, let us first consider the situations under which the frames sent by nodes A and D overlap. One such situation is, node D (or A) starts transmitting its RTS frame even though node A (or D) is transmitting a data frame. This situation occurs whenever the imprecise EIFS problem exists. As analyzed in subection A, when capture is allowed in such situations, the data frame can be received correctly. Otherwise, both the frames are dropped. Another overlapping situation, which arises due to the hidden-terminal problem, is that both the nodes transmit their RTSs. In particular, since nodes A and D are hidden from each other, both of them may transmit their RTSs during a certain interval before any one of them gets a CTS. Consider the situation when RTS of A arrives at node B earlier than the RTS sent by node D, and therefore, node B captures the RTS of A if the capture is allowed. However, after completely receiving the captured RTS, carrier sensing at node B indicates busy medium since node D is still transmitting its RTS. This forbids node B from sending a CTS back to node A, and thus the four-way handshake initiated by node A fails even though its RTS is captured. Therefore, both the RTS frames are dropped irrespective of whether the capture is allowed. Note that, if a data frame is captured, the four-way handshake can be successful because the node capturing the data frame always sends back an ACK without considering whether the medium is busy [1]. In fact according to [1], in addition to the data frame, the capture of an ACK frame or a broadcast frame is also effective since these frames are the last ones in the sequence. The capture of other frames, such as RTS, or CTS, is not effective in the handshake because ultimately they will be discarded. This implies that a capture is effective only if the receiver processes the captured frame.

Now we analyze the results of the Jain's index shown in Figure 3. In static-S1 and S2, the imprecise-EIFS problem leads to the overlapping between the data and RTS frames. When the capture is not allowed (static-S1), both the frames are dropped and then the CWs at nodes A and D are increased. On the other hand, if capture is allowed (static-S2), the data frame will be captured. Then, the node that sends the data frame will get an ACK and reset its CW while the node that sends the RTS frame will retransmit its RTS in vain, leading to a larger CW. This obviously results in unfairness and explains why the static-S2 shows worse fairness than that under static-S1. In scenarios static-S3 and S4, the overlapping of frames occurs only when both nodes A and D transmit their RTSs, and the difference between these two scenarios is that the capture is not allowed in S3. However, a capture of a RTS does not lead to any advantage as explained above, and as a result, the two scenarios show similar fairness. Moreover, they show better fairness than that under static-S2 as the ECS solves the imprecise EIFS problem, which is a major reason of unfairness [6]. The presence of hidden-terminal, imprecise EIFS, and absence of capture each plays an important role in determining the fairness of static-S1, and the combined effect of these makes S1 the fairest among the four scenarios.

From the study, we deduce that: *(i)*hidden-terminal problem results in substantial short-term unfairness, *(ii)*capture, if effective, increases the capacity at the cost of fairness, showing that there is a conflict between achieving fairness and maximizing throughput. On the other hand, if the capture is not effective in the handshake, it does not have any effects on the fairness as well as the throughput, and *(iii)*imprecise EIFS results in unfairness, and the ECS improves the fairness without degrading the throughput performance.

IV. Interaction between MAC Fairness and Routing Mechanisms

To study the effects of the interaction, we replace static routing with AODV and DSDV, one at a time. The simulation environment is the same as discussed in Section III.

A. Throughput Comparison

Table III presents the average throughputs under different routing protocols. We notice that the average throughputs under DSDV are almost the same as those under the static routing. On the other hand, under the AODV-S1 to S3, the flow from node D to C completely starves while the other flow gets the entire bandwidth. This is because the AODV at node D cannot even find a route to node C as explained later. Note that the results presented here are obtained under the situation where the flow from node A to B starts first. However, if we start the flow from D to C first, exactly opposite results are obtained, i.e., the flow from node A to B starves.

TABLE III

Throughputs under different routing protocols

Scenario	Flow	Average Throughput(M bps)		
		Static	DSDV	AODV
S1	A to B	0.314	0.314	1.401
	D to C	0.307	0.308	0
S2	A to B	0.683	0.684	1.402
	D to C	0.641	0.669	0
S3	A to B	0.662	0.675	1.402
	D to C	0.672	0.659	0
S4	A to B	0.665	0.650	0.660
	D to C	0.670	0.685	0.678

In order to explain the flow starving, we first describe the route discovery process of AODV [3]. The AODV uses two kind of messages for the route discovery: Route Request (RREQ) and Route Reply (RREP). When a node, say node S, needs a route to some destination node, say node X, node S first broadcasts a RREQ and then waits for a RREP. When a node receives the RREQ and has a fresh enough route to node X, the node generates a RREP message and then unicasts it to the next hop toward the originator of the RREQ (i.e., node S). If a RREP is not received by the time RREQ_TIMEOUT expires, to find the route, node S will broadcast another RREQ, up to a maximum of RREQ_RETRIES times. To reduce congestion in the network, the value of RREQ_TIMEOUT follows a binary exponential back-off mechanism. In our simulation, the initial value of RREQ_TIMEOUT is 1.8 second while the RREQ_RETRIES is set to 3.

Now we explain why the AODV's route discovery process at node D always fails in the S1 scenario. Whenever an RREQ is handed to the MAC layer at node D, this packet will be transmitted immediately by the MAC protocol if the medium is determined to be idle. On the other hand, if the medium is determined busy, the MAC protocol backs off and defers the transmission of the RREQ. Figure 4 demonstrates when node D will determine the medium to be idle or busy while FES (A, B) is in progress. When the medium is determined to be idle, D's MAC layer decrements its back-off timer and transmits the RREQ when the timer becomes zero. We notice from the diagram that the RREQ can be successfully transmitted if it is sent in a very small crucial interval of duration ν. At other instances, however, the carrier sensing at node D may indicate the medium to be idle though really it is not idle, and thus the transmission of the RREQ will result in a collision. This has been shown as collision in the figure. If we denote the sum by λ of all idle durations determined by node D within a single FES, then the ratio ν/λ reflects the probability that node D can successfully transmit the RREQ. In the AODV-S1, the ν and λ are as follows:

$$\nu = y - (EIFS - DIFS) - Txtime(RREQ) \quad (2)$$

$$\begin{aligned}\lambda &= Txtime(RTS) + SIFS + Txtime(Data) + SIFS \\ &\quad -(EIFS - SIFS) + y - (EIFS - DIFS) \quad (3)\end{aligned}$$

where y is the back-off time at node A. As ν dependents on the y, which in turn is determined by the CW value, we first consider the CW value at node A. One should note that the RREQ_TIMEOUT between two consecutive RREQs sent by node D is normally long enough for nodes A and B to exchange several FESs successfully, leading to the reset of the CW at node A. Therefore, whenever the RREQ is transmitted by node D, the CW at node A is always equal to CW_{min}. Since the maximum value of y is equal to CW_{min} (i.e., 31), which is smaller than the transmission time (i.e., 40 slots in the case) of RREQ, the ν in equation (2) is equal to zero as ν cannot be negative. Therefore, ν/λ is zero. In other words, whenever node D transmits the RREQ, it will always collide with a frame (either RTS or data) sent by A. This explains why the AODV at node D cannot start in the S1 scenario. Similarly, we can show why the AODV cannot start in S2 and S3 scenarios but due to lack of space, we exclude the explanation. However, in scenario S4, ν/λ is significant enough to allow the AODV to start at node D.

In fact, what we have analyzed above applies not only to the AODV, but also to all the on-demand routing protocols. To verify this, we carried out simulation using DSR. The results show that the DSR at node D in S1, S2 and S3 does not start.

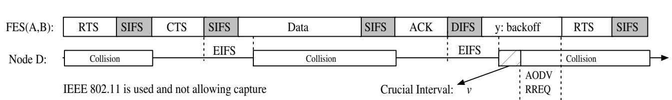

Fig. 4. Medium State Determined by Node D when FES (A, B) is in Progress

B. Jain's Index Comparison

Similar to the throughput, the Jain's index under DSDV is also almost the same as that under the static routing. It is easy to see that the Jain's indices under AODV-S1, -S2 and -S3 are always equal to 0.5 as complete unfairness occurs in these three scenarios. Performance for scenario S4 under all three routing protocols is compared in Figure 5.

Since the DSDV is very similar to the static routing except that the DSDV periodically exchanges route messages, they show similar fairness. We notice that the AODV-S4 shows worse fairness in comparison to other two. The reason is as follows. Whenever the MAC layer at a node drops a packet after several unsuccessful attempts, the AODV views this as an indication of the link breakage, and thus will initiate a fresh route discovery process that may take a long time during which the other node gets the entire bandwidth of the medium. This situation may occur very frequently due to the hidden-terminal problem. On the contrary, DSDV and static routing do not view the dropping of a packet as an indication of link breakage, and therefore they do not initiate the route discovery process. Therefore, AODV shows worse fairness than that under the static routing or DSDV.

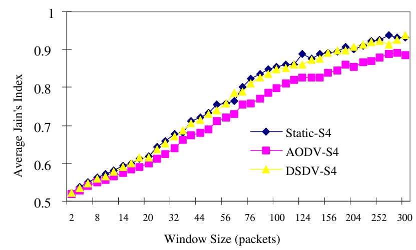

Fig. 5. Fairness Comparison under S4

From the above discussion, we can deduce that: *(i)*the short-term MAC unfairness may lead to the stalling of the on-demand routing protocols (e.g., AODV and DSR), and *(ii)*the routing mechanisms (e.g., route discovery and link breakage detection) also affect the system fairness.

V. Conclusions

From this paper, the main lessons learnt are: *(i)*we should distinguish between the *MAC fairness* and the *overall system fairness*, since under identical MAC conditions different overall fairness performance may be obtained if different routing protocols are used, *(ii)*in studying the *MAC fairness*, we should isolate the factors that are not directly related to the MAC layer, and therefore, the study is to be carried out under the static routing, CBR backlogged traffic, static topology etc, and *(iii)*to study the *overall system fairness*, in addition to the study of MAC fairness, we should also consider many other factors such as the routing mechanism, traffic pattern, mobility etc.

References

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

[2] C. E. Perkins and P. Bhagwat. Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers Proc. SIGCOMM'94, 1994.

[3] C. E. Perkins and E. M. Royer. Ad-hoc On-Demand Distance Vector Routing. Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, February 1999

[4] D.Johnson and D. Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. Mobile Computing, Tomasz Imielinski and Hank Korth, Eds. Chapter 5, pages 153-181, Kluwer Academic Publishers, 1996

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

[6] Li Zhifei, "Improving Performance of MAC Protocols in Multi-hop Wireless Ad hoc Networks", First year report for confirmation of Ph.D. Candinate, SCE, Nanyang Technological University, June 2003.

[7] CMU Monarch Group. CMU Monarch Extensions to NS, http://www.monarch.cs.cmu.edu/.

[8] Chiu, D.-M., and Jain, R. Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks. Computer Networks and ISDN Systems 17 (1989),1-14.

[9] C.Ware, J.Judge, J.Chicharo, E. Dutkiewicz "Unfairness and capture behavior in 802.11 ad-hoc networks," IEEE ICC, 2000