Traffic and QoS Analysis of an Infrastructure CSMA/CA Network With Multiple Service Classes[*]

Ritabrata Roy

December 13, 2002

Abstract

This treatise aims at analysing a CSMS/CA network operating in an access point mode, in the presence of two classes of traffic— a non–real-time Poisson traffic and a real–time constant bit rate (CBR) session. The analysis further obtains a qualitative assessment of the network performance in terms of bit rate, traffic delay and quality of service parameters like jitter and bit error rate and proposes certain changes that can improve efficiency.

1 Introduction

This exercise considers a CSMA/CA network operating in an access point mode, which means that all data transfers are routed through the access point, which is responsible for maintaining fairness and Quality of Service. It is further assumed that the Request to Send (RTS) and Clear to Send (CTS) transmissions, originated by the access point, are received correctly by all network users, and that the RTS and CTS transmissions contain a field specifying the duration of the associated frame to be transmitted. It is necessary to design a system where this information can be used in a decentralised manner (*i.e.*, a protocol that runs independently at each node) for traffic management and QoS control.

The traffic pattern in the model is assumed to be Poisson and non–real-time in nature, while one user wishes to initiate a constant bit rate (CBR) session. The quality of the CBR session is to be analysed in terms of the expected delay and jitter associated with the admitted bit rate.

[*] Submitted as partial fulfillment of course requirement in Communication Networks I (ECE 330:543) to Dr. Ivan Marsic, Assistant Professor, Rutgers, The State University of New Jersey.

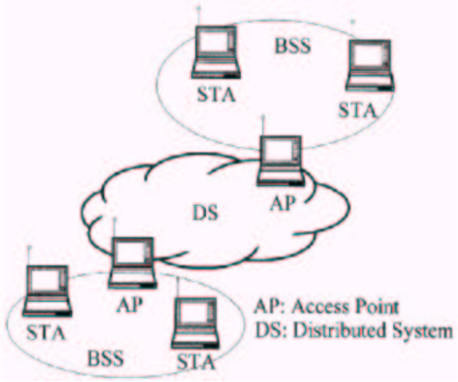

Figure 1: An infrastructure network, for instance, the IEEE 802.11 Wireless Local Area Network (WLAN) operating in access point (AP) mode [8].

1.1 Network Architecture

The IEEE 802.11 standard for Wireless Local Area Networks [10] is the most popular protocol to employ the CSMA/CA strategy. The *basic service set* (BSS) is the basic building block of IEEE 802.11 WLANs. The coverage area of a BSS is referred to as the *basic service area* (BSA). The IEEE 802.11 defines two types of network architecture, the *ad hoc* network and the *infrastructure*, or access point (AP), network. For the purposes of this exercise, only the infrastructure network will be studied, since the problem specification requires the CSMA/CA network to operate in the access point mode. A simple architecture of the infrastructure network ahs been illustrated in Fig. 1. The architectural component used to interconnect BSSs is the distribution system (DS), which enables support for mobile devices by providing logical services necessary to handle address-to-destination mapping and seamless integration of multiple BSSs. A BSS using an access point provides access to the DS and forms an infrastructure network. The AP operates in a manner analogous to operation of the base station in a cellular telephony system. [8]

In contrast to an ad-hoc network, infrastructure networks are established to provide wireless users with specific services and range extension. Infrastructure networks in the context of IEEE 802.11 are established using access points, which support range extension by providing the integration points necessary for network connectivity between mulitple BSSs, thus forming an *extended service set* (ESS). The ESS consists of multiple BSSs that are integrated together using a common *distribution system* (DS). The DS, as specified by IEEE 802.11, is implementation–independent, *i.e.* the DS could be a wired IEEE 802.3 Ethernet LAN, IEEE 802.4 token bus LAN, IEEE 802.5 token ring LAN or another IEEE 802.11 wireless medium [4].

1.2 IEEE 802.11

As has already been discussed, a basic service set (BSS) is formed of a group of wireless nodes, which are under the control of either a *Distributed Coordination Function* (DCF) or a *Point Coordination Function* (PCF). Access points link the nodes to a Distribution System (DS), thereby extending their range to other BSSs via other APs.

The IEEE 802.11 standard supports two services [2]:

- *Distributed Coordination Function* (DCF), which supports delay insensitive data transmissions (e.g. email, ftp).

- *Point Coordination Function* (PCF), which is an optional service that supports delay sensitive transmissions (for instance, real–time audio and video) and is used in combination with DCF.

In a BSS, the wireless nodes and the AP can either work in *contention mode* exclusively, using the DCF, or in *contention-free mode* using the PCF. In the first mode, wireless nodes have to contend for use of the channel at each data packet transmission. In the second mode the medium usage is controlled by the AP, polling the nodes to access the medium, thus eliminating the need for contentions. This last mode is not exclusive, and the medium can be alternated between contention mode and contention-free mode for CP (Contention Period) and CFP (Contention Free Period) respectively. This paper will focus on the contention mode using the distributed coordination function.

1.3 Distributed Coordination Function

As mentioned earlier, the distributed coordination function (DCF) is an asynchronous data transmission function, which best suits delay insensitive data. When used in an infrastructure network, DCF can be either exclusive or combined with PCF. The basic scheme for DCF is *Carrier Sense Multiple Access* (CSMA) [3]. This protocol has two variants:

1. Collision Detection(CSMA/CD), and

2. Collision Avoidance (CSMA/CA).

A collision can be caused by two or more stations using the same channel at the same time after waiting for the channel to become idle, or by two or more hidden terminals in a wireless network transmitting simultaneously. Since a wirel;ess node cannot listen to the channel, while it is transmitting, to avoid packet collision, the sender waits for an acknowledgment (ACK) from the receiver after each frame transmission, as shown in Fig. 2. The *source* axis shows

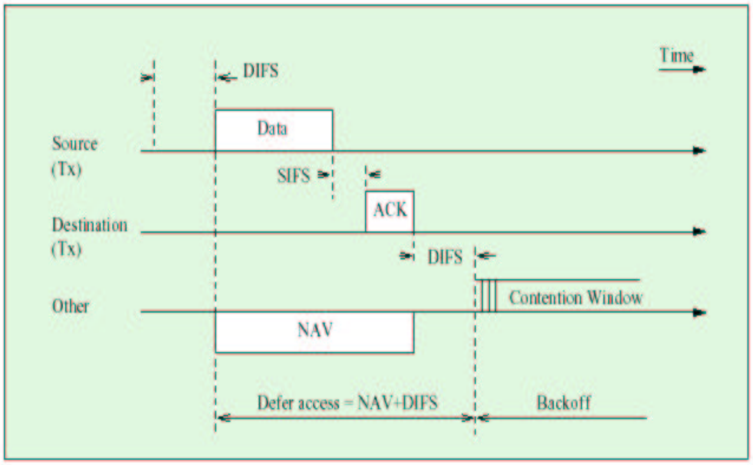

Figure 2: Basic access scheme for a CSMA/CA network [4].

the data transmitted by the source, and the destination, which replies with an ACK, is shown on the *destination* axis. The third axis shows the network status, as seen by the *other* wireless nodes. While this model does not consider transmission delays, it does include *interframe spacings* SIFS and DISF, which are explained later in this section. If no ACK is returned, a collisionis assumed to have occurred and the frame is retransmitted. This technique may waste a lot of time in case of long frames, keeping the transmission going on while collision is taking place (caused by a hidden terminal for example).

To solve the hidden terminal problem, an optional RTS/CTS (*Request To Send* and *Clear To Send* respectively) scheme is used in addition to the previous basic scheme, as shown in Fig. 3, whereby a station sends an RTS before each frame transmission to reserve the channel. Note that a collision of RTS frames (20 octets) is less severe and less probable than a collision of data frames (up to 2346 octets). The destination replies with a CTS if it is ready to receive and the channel is reserved for the packet duration. When the source receives the CTS, it begins transmitting its frame, making sure that the channel is reserved for itself for the entire frame duration. All the other wireless nodes in the BSS update their *Network Allocation Vector* (NAV) whenever they hear an RTS, a CTS or a data frame. Thus the NAV information is used for virtual carrier sensing.

Not all packet types have the same priority. For example, ACK packets should have priority over RTS or data frames. This is done by assigning to each packet type a different *Inter Frame Spacing* (IFS), after the channel turns idle, during which a packet cannot be transmitted. In DCF two IFSs are used:

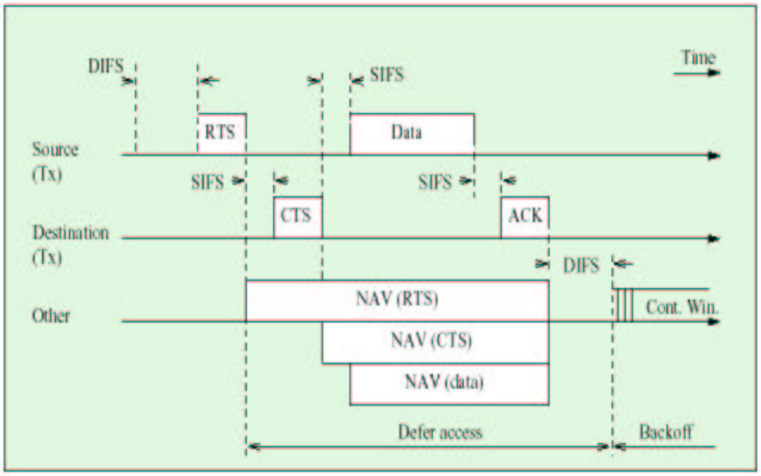

Figure 3: RTS/CTS access scheme for a CSMA/CA network [4].

Short IFS (SIFS) and DCF IFS (DIFS), where SIFS is shorter than DIFS, as shown in Fig. 2 and 3. As a result, if an ACK (assigned with SIFS) and a new data packet (assigned with DIFS) are waiting simultaneously for the channel to become idle, the ACK will be transmitted before the new data packet (the first has to wait SIFS whereas the data has to wait DIFS).

The *collision avoidance* part of CSMA/CA consists of avoiding packet transmission right after the channel is sensed idle for DIFS time, so it does not collide with other "waiting" packets. Instead, a node with a packet ready to be transmitted waits for the channel to become idle for DIFS time, then waits for an additional random *backoff time* after which the packet is transmitted, as shown in Fig. 2 and 3. The backoff time of each node is decreased as long as the channel is idle (during the so-called *contention window*). When the channel is busy, the backoff time is freezed, and when the backoff time reaches zero, the wireless node transmits the frame. If, however, a collision occurs, then the node computes a new random backoff time with a higher range to retransmit the packet with lower collision probability. This range increases exponentially as 2^{2+i}, where i (initially equal to 1) is the transmission attempt number.

Thus the backoff time equation is [1]:

$$\text{backoff\_time} = \lfloor 2^{2+i} \times \text{ rand(0,1) } \rfloor \times \text{ slot\_time}$$

where, slot_time is a function of physical layer parameters,
rand(0,1) is a random function with a uniform distribution in [0,1].

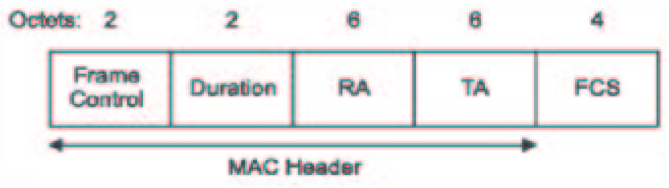

Figure 4: Frame format of Request to Send (RTS) signal [10].

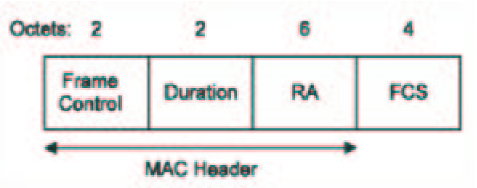

Figure 5: Frame format of Clear to Send (CTS) signal [10].

2 Analysis

Before proceeding to the details of the proposed protocol, it might be instructive to revisit the order of packet exchanges, illustrated in Fig. 3, that precedes the actual exchange of data. According to the access point model, when a source A wishes to transmit data to destination B, it senses the medium and sends an RTS packet to the access point when the channel is idle. Since the source node may not be within the transmission range of all the wireless nodes associated with the access point, it is not until the access point makes a transmission, that *all* the nodes become aware of the traffic and desist from transmitting for the duration of A's transmission. The original IEEE 802.11 protocol requires another exchange of RTS and CTS signals between the access point and the destination B, and it is this RTS from the access point that is "heard" by all the wireless nodes within the cluster [9].

The frame format of the RTS signal is shown in Fig. 4, where RA is the address of the immediate recipient of the transmission, and TA is the address of the originating station. The duration value is the time, in microseconds, required to transmit the pending data or management frame, plus one CTS frame, one ACK frame and three SIFS intervals. The CTS frame, shown in Fig. 5, is similar to the RTS frame except that the RA of the CTS is copied to the TA field, and the duration field carries the RTS duration value less the time to transmit the CTS frame and its corresponding SIFS interval.

It is possible to reduce the number of RTS/CTS exchanges between source–

AP and AP–destination by making a subtle change to the protocol, that would reduce redundancy and improve efficiency. Since the CTS signal sent by the access point to the source is heard by all the nodes in the wireless medium, if the signal had an additional field that encapsulated information about the source's intended recipient, then all the nodes would desist from transmission for the period indicated in the duration. This increase in complexity of the CTS field structure would partly offset the RTS/CTS overhead by cutting it almost by half.

Although this modification does improve the efficiency of the CSMA/CA model in access point mode, it plays no direct role in the design of the protocol that handles a constant bit rate (CBR) session in the presence of background Poisson traffic, and this forms the subject of the next subsection.

2.1 Traffic Management

The aim of this section is to propose a traffic management algorithm that can provide a user with constant bit rate service with a sufficient Quality of Service, while not underutilising the medium. As will be seen, the two important requirements of minimum jitter and maximum channel utilisation are contradictory, and the optimum design algorithm makes a tradeoff between the two. Since the CBR session is routed through the access point, the latter can provide no–delay, no–jitter service by dedicating itself entirely to the CBR source. However, this would put all other traffic on hold and will grossly underutilise the wireless medium. An alternative is to take advantage of the fact that the capacity available to the access point is usually much greater than that demanded by the CBR source.

When the throughput capacity of the access point is more than that demanded by the CBR data source, it can allocate part of its capacity to other users while maintaining the constant bit rate demanded by the data source. For instance, if the CBR session requires a data rate of n bits second^{-1}, and the AP can provide a bit rate of mn bits second^{-1} (where $m > 1$), then the AP can service the CBR source every $\frac{1}{r}$ second (where $r \leq m$) at a data rate of rn bits second^{-1} so that an average bit rate of n bits second^{-1} is maintained. In the remaining $\frac{r-1}{r}$ second, the AP can provide service to other nodes without affecting the throughput of the CBR session.

It is evident from the above discussion that the higher the value of r, the more will be the utilisation of the channel and hence the net average delay will be lower. However, the disadvantage to high values of r is that it might increase the jitter in the CBR session, since there is an increased likelihood of frames arriving with different delays. This is particularly true for frames requiring retransmission, that could cause other frames to be dropped. Thus, optimum service is obtained for some value of r that would not cause too much

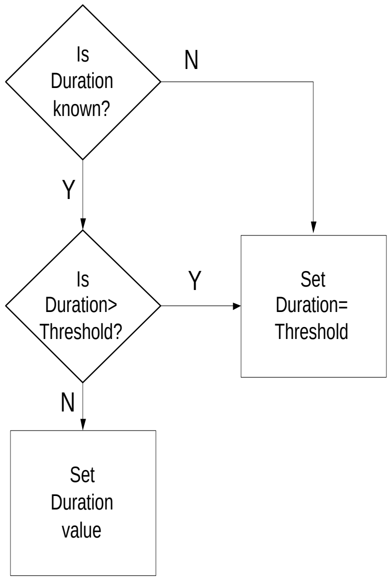

Figure 6: Algorithm used to assign transmit duration to CBR session to ensure fair utilisation of the wireless medium.

jitter while sharing the channel with other users. It may also be pointed out that scheme will require some further modification to the RTS/CTS packet exchange between the CBR source and the AP, so that they are synchronised and the source buffers the data and transmits it in the interval that it is allowed.

A practical implementation of this scheme would require the calculation of the optimum value of r, and might form the basis of an interesting research topic. For real–life applications, in order to ensure fair and equitable traffic distribution, there should also be a limit to the CBR duration that is allowed to a "greedy" source, *before it has to contend for the medium again.* To cite an extreme example, it would be unwise to grant the complete mn bits second^{-1} capacity to a source that demands such a CBR capacity for an infinitely long period of time, since it would backlog data on all the other nodes and eventually clog up the network. Thus, a *threshold* may be defined which would be the upper limit of the duration that can be assigned to a CBR session at a particular capacity. The flowchart in Fig. 6 illustrates the logic in deciding the duration of time for which the constant bit rate data transfer will be in session.

2.2 Quality of Service

There are several ways to characterize Quality of Service (QoS). Generally, QoS is the ability of a network element (for instance, an application, a host or a router) to provide some level of assurance for consistent network data delivery. Some applications are more stringent about their QoS requirements than others, and the types of QoS available may be broadly classified as [6]:

- **Resource reservation** (integrated services): network resources are apportioned according to an application's QoS request, and subject to bandwidth management policy.
- **Prioritization** (differentiated services): network traffic is classified and network resources apportioned according to bandwidth management policy criteria. To enable QoS, network elements give preferential treatment to classifications identified as having more demanding requirements.

Priority may be introduced into the IEEE 802.11 protocol using either the DCF (distribution coordination function) or the PCF (point coordinate function) techniques. In this paper, only DCF–based techniques have been studied.

Introducing different levels of priorities for two different classes of services (non–real-time and constant bit rate) is akin to analysing an M/G/1 for two different service classes. Thus, assuming that the first class is the CBR session and has a higher priority, the average waiting time W_k for each of the two classes can be written as follows [3]:

$$W_1 = \frac{R}{1-\rho_1}$$
$$W_2 = \frac{R+\rho_1 W_1}{1-\rho_1-\rho_2}$$

where, R= mean residual service time

$\rho_k = \frac{\lambda_k}{\mu_k}$ = system utilisation for priority k

Thus, it is seen that the waiting time to get access to the common wireless medium is less for the higher priority CBR session.

- **Backoff Increase Function or Distributed Fair Scheduling**
  An access scheme called Distributed Fair Scheduling (DFS) uses the backoff mechanism of IEEE 802.11 to determine which station should send first [5]. Before transmitting a frame, the backoff process is always initiated. The backoff interval calculated is proportional to the size of the packet to send and inversely proportional to the weight of the flow. This causes stations with low weights to generate longer backoff intervals than those with high weights, thus getting lower priority. Fairness is achieved by including the packet size in the calculation of the backoff interval, causing flows with smaller packets to get a chance to send more often. If a

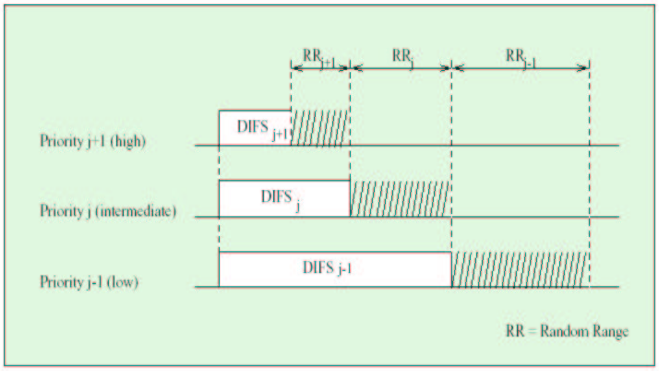

Figure 7: Introducing priority into DIFS [1].

collision occurs, a new backoff interval is calculated using the backoff algorithm of the IEEE 802.11 standard. This protocol could be modified to tackle the problem at hand by using longer backoff intervals for the non–CBR traffic, so that the CBR session is initiated quicker than the Poisson traffic. Once the CBR session is initiated, the algorithm proposed in Section 2.1 is executed to maintain it.

- **Varying DIFS**
  Priorities can also be assigned by varying the DIFS for differentiation [1]. It has already been discussed that in IEEE 802.11, ACK packets get higher priority than RTS packets, simply because the SIFS length is shorter than the DIFS, associated with the latter. This idea can be extended to introduce priority for data packets, like constant bit rate (CBR) sessions. In this approach, each priority level is given a different DIFS with $\text{DIFS}_{j+1} < \text{DIFS}_j$, as shown in Fig. 7.

- **Limiting the Maximum Frame Length**
  This mechanism limits the maximum frame length used by each wireless node, so that one of the following two scenarios occurs:

  1. Drop packets that exceed the maximum frame length assigned to a given node.
  2. Fragment packets that exceed the maximum frame length assigned to the node.

- **Varying the Contention Window**
  As has already been discussed, the MAC protocol requires each station to wait for a randomly–chosen time period before attempting a transmission in order to reduce the probability of collision. This time frame, or

contention window, may be used to give higher probability to some stations over others. Assigning a shorter contention window to stations with higher priority ensures that in most cases, the higher–priority stations will be able to transmit ahead of the lower–priority ones. For instance, in case of a collision between a low–priority Poisson traffic and the CBR session, the backoff time for the CBR session may be assigned a lower value by either of the following methods:

1. Choosing a random number from a smaller range, for instance [0,0.4] instead of [0,1].
2. The value of the exponential backoff timer may be prevented from increasing by fixing i at a small value.

Thus, the backoff time for the CBR session will be smaller than the other data frames involved in the collision and the expression is a modification of the equation discussed in Section 1.3. Backoff_time for the case when $i = 0$ and the random number generated is between [0,0.4] is shown below:

$$\text{backoff\_time} = \lfloor 2^2 \times \text{ rand(0,0.4) } \rfloor \times \text{ slot\_time}$$

where, slot_time is a function of physical layer parameters,
rand(0,1) is a random function with a uniform distribution in [0,1].

3 Conclusion

This paper has successfully proposed an algorithm for an infrastructure CSMA/CA network that can accomodate two classes of traffic— a non–real-time Poisson traffic and a higher priority constant bit rate session. Methods have been suggested to ensure an adequate QoS in terms of bit rate, delay and jitter, while allowing a fair and equitable medium access for all the wireless nodes.

It has also been demonstrated that the protocol, while designed for an access point mode of operation, does in fact allow the nodes to operate in a *decentralised* manner. For instance, when the AP sends the *modified* CTS to the initiating wireless node, all the nodes in the cluster become aware of the duration of the ensuing transmission and adjust their own transmissions *independently*, thus reducing the risk of collisions. As explained in Section 2, the AP–source clear to send frame is modified to include information about the final destination, so that the exchange of RTS/CTS signals between the AP and the destination becomes unnecessary. Since the CTS from the AP is heard by all nodes, *including the destination*, this simple modification will not affect the logical behaviour of the protocol, while improving the overall system efficiency.

It has also been discussed in detail about the effect of the algorithm on jitter, and how minimising the jitter will lead to underutilisation of the channel as well as increased average system delay. An interesting problem would be to find out the optimisation conditions that would minimise jitter for a given allowable value of delay, and could form the basis of future work in this area. An analytical result for the waiting time for medium contention was also obtained by treating the problem as an M/G/1 system with different priority classes. Other ways of improving the quality of service were also proposed, which involved minor adjustments to the backoff timer, the size of the contention window, the DIFS, and limiting the maximum frame length.

Finally, it may be pointed out that certain MAC modifications have been incorporated into IEEE 802.11e draft D0D and IEEE 802.15.3 draft D0D [7] that efficiently handle prioritised and QoS–based traffic. These modifications were kept out of the purview of this study.

References

[1] Imad Aad and Claude Castelluccia. Introducing Service Differentiation into IEEE 802.11. In *Proceedings of ISCC 2000*, Antibes, France, 2000.

[2] Imad Aad and Claude Castelluccia. Differentiation Mechanisms for IEEE 802.11. In *Proceedings of IEEE INFOCOM 2001*, Anchorage, AL, 2001.

[3] Dimitri Bertsekas and Robert Gallager. *Data Networks.* Prentice-Hall, Englewood Cliffs, NJ, 1992.

[4] Brian P. Crow, Indra Widjaja, Jeong Geun Kim, and Prescott T. Sakai. IEEE 802.11 Wireless Local Area Networks. *IEEE Communications Magazine*, pages 116–126, September 1997.

[5] D-J. Deng and R-S. Chang. A Priority Scheme for IEEE 802.11 DCF Access Method. *IEICE Transactions on Communications*, pages 96–102, January 1999.

[6] Yasir Drabu. A Survey of QoS Techniques in 802.11. Technical report, Kent State University., 2001.

[7] Pierre T. Gandolfo. Trade–Off Analysis: 802.11e versus 802.15.3 QoS Mechanism. Technical report, XtremeSpectrum Inc., 2002.

[8] Hung Huan Liu and Jean-Lien C. Wu. A Scheme for Supporting Voice Over IEEE 802.11 Wireless Local Area Network. In *Proceedings of the National Science Council*, volume 25, pages 259–268, 2001.

[9] Ivan Marsic. Wireless Networks. Technical report, Rutgers, The State University of New Jersey, 2002.

[10] IEEE Computer Society. IEEE 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Technical report, ANSI/IEEE Std 802.11, 1999 Edition, 1999.