Medium Access Control Mechanisms in Mobile Ad Hoc Networks

Chansu Yu
Dept. of Electrical and Computer Engineering
Cleveland State Univ.
Stilwell Hall 340
Cleveland, OH 44115
c.yu91@csuohio.edu

Ben Lee
School of Electrical Engineering and Computer Science
Oregon State Univ.
Owen Hall 302
Corvallis, OR 97331
benl@ece.orst.edu

Sridhar Kalubandi
Dept. of Electrical and Computer Engineering
Cleveland State Univ.
Stilwell Hall 332
Cleveland, OH 44115
s.kalubandi@csuohio.edu

Myungchul Kim
School of Engineering
Information and Communications Univ.
58-4 Hwaamdong
Daejon, Korea
mckim@icu.ac.kr

Abstract

Medium access control protocol plays an important role in providing fair and efficient allocation of limited bandwidth in wireless LANs. The basic medium access model in the IEEE 802.11 standard, known as DCF (Distributed Coordination Function), is widely used in wireless LANs. Research efforts in wireless multihop networks, where wireless nodes need to forward packets on other's behalf, try to measure up to or improve upon this standard. This chapter presents an in-depth discussion on the problems with IEEE 802.11, especially those relevant in a multihop network, and discusses various techniques that have been proposed to enhance the channel utilization of multihop wireless networks.

Keywords: *Mobile ad hoc networks, medium access control, backoff algorithm, RTS/CTS mechanism, transmission power control, directional antenna.*

1. Introduction

Mobile devices coupled with wireless network interfaces will become an essential part of future computing environment consisting of *infra-structured* and *infrastructure-less* wireless LAN networks [1]. Wireless LAN suffers from collisions and interference due to the broadcast nature of radio communication and thus requires special *medium access control* (*MAC*) protocols. These protocols employ control packets to avoid such collisions but the control packets themselves and packet retransmissions due to collisions reduce the available channel bandwidth for successful packet transmissions. At one extreme, aggressive collision control schemes can eliminate the retransmission overhead but at the cost of large control overhead. At the other extreme, the lack of control over collisions offers zero control overhead but it may need to expense large amount of channel bandwidth for retransmissions.

Distributed coordination function (*DCF*) is the basic medium access method in IEEE 802.11 [4], which is the most popular wireless LAN standard, and it makes prudent tradeoffs between the two overheads. DCF supports best effort delivery of packets at the link layer and is best described as the *Carrier Sense Multiple Access with Collision Avoidance* (*CSMA/CA*) protocol. While DCF works reasonably well in infra-structured wireless LAN environment, this is not necessarily true in a *mobile ad hoc network* (*MANET*) environment. A MANET is an infrastructure-less multihop network that consists of autonomous, self-organizing and self-operating nodes, each of which communicates directly with the nodes within its wireless range or indirectly with other nodes via a dynamically computed, multi-hop route.

While the multi-hopping technique can potentially maximize the channel utilization by allowing multiple simultaneous transmissions occurring separated in space [2, 3], all participating nodes must undertake the role of routers engaging in some routing protocol required for deciding and maintaining the routes. In comparison to one-hop wireless networks with base stations, multihop networks suffer from more collisions because nodes are not partitioned into a number of disjoint cells but overlapped successively in space. Therefore, congestion at one particular area in a MANET may affect the neighboring areas and can propagate to the rest of the network. In addition, multi-hopping effectively increases the total data traffic over the network by a factor of the number of hops. Moreover, it potentially causes self-generating collisions in addition to those from other data streams since each node acts as a router and uses a single network interface to receive a packet as well as to forward the previous packet of the same data stream to the next hop node.

This chapter overviews key elements of DCF, discusses problems of DCF when used in a multihop MANET environment, and surveys various mechanisms that balance the abovementioned two overheads to enhance the channel utilization in the presence of increased chance of collisions. These mechanisms can be broadly classified as *temporal* and *spatial* approaches depending on their focus of optimization on the channel bandwidth. The temporal approaches attempt to better utilize the channel along the time dimension by optimizing the parameters or improving the *backoff algorithm* of the DCF protocol [5-8]. On the other hand, the spatial approaches try to find more chances of spatial reuse without significantly increasing the chance of collisions. These mechanisms include *busy tone channel* [9], *transmission power control* [10-12], and *directional antenna* [13-17].

The organization of the article is as follows. In Section 2, general description of MAC algorithms and DCF of IEEE 802.11 are discussed. Section 3 and 4 discuss the temporal and spatial MAC techniques, respectively, to enhance the channel utilization based on DCF. Finally, Section 5 presents concluding remarks.

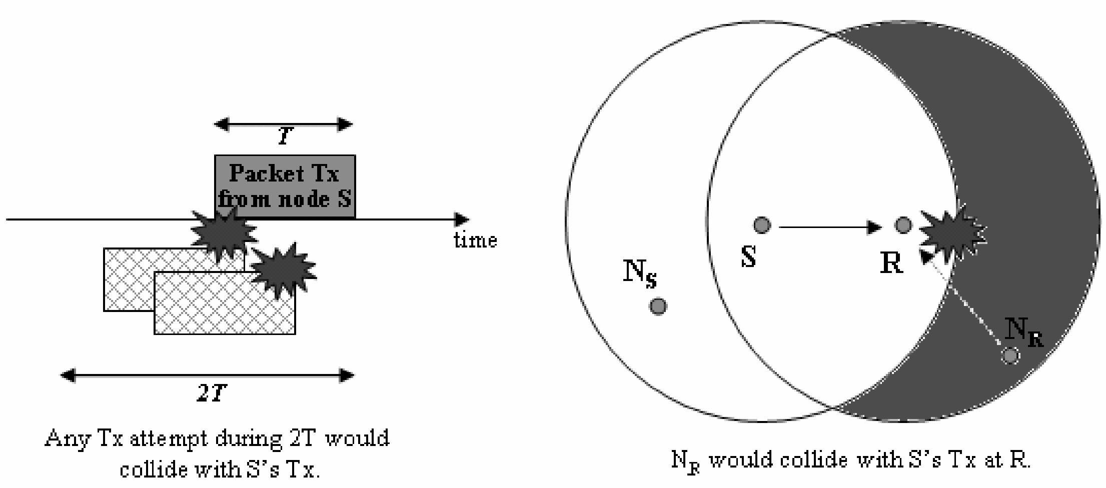
2. Medium Access Control (MAC) Protocols

A MAC protocol in a multi-access medium is essentially a distributed scheduling algorithm that allocates the channel to requesting nodes. Two commonly used access principles in wireless networks are *fixed-assignment channel access* and *random access* methods [18]. In the former method, a pair of nodes is statically allocated a certain time slot (frequency band or spread spectrum code), as is the case for most of voice-oriented wireless networks. On the other hand, in random access MAC protocols, the sender dynamically competes for a time slot with other nodes. This is a more flexible and efficient method of managing the channel in a fully distributed way, but suffers from collisions and interference. This section provides a general discussion on the random access MAC and then offers an in-depth discussion on DCF of IEEE 802.11.

2.1 Random Access MAC

Random access MAC protocol in radio networks has long been an active research area. The throughput of *ALOHA* and *carrier sensing* protocols in the presence of collisions has been analyzed with a wide range of system parameters, such as propagation delay and offered load. A key factor here is the "*vulnerable period*," during which for a node to transmit a packet successfully without collisions, other interfering nodes should not attempt to transmit during the node's transmission time [19]. In the *pure ALOHA* scheme, the vulnerable period is twice the packet transmission time as shown in Fig. 1(a). This is fairly large and cannot be ignored unless communication traffic is sufficiently light. It has been reported that the maximum achievable channel utilization is only 18% for pure Aloha and 36% for *slotted Aloha* even including retransmissions [19]. The *carrier-sensing mechanism* reduces this period substantially by sensing the medium before attempting to transmit a packet. The chance of collisions is reduced to the case where a node does not sense the medium correctly due to the propagation delay, which is fairly small compared to the packet transmission time.

Unfortunately, collisions are not completely avoidable in carrier sensing MAC protocols due to interfering "*hidden terminals*" [21]. When a mobile node is located near the receiver, but far from the sender, this node maybe unaware of the on-going communication and causes collisions at the receiver by initiating its own data transfers. In Fig. 1(b), N_R is an example of a hidden terminal when nodes *S* and *R* are the sender and the receiver, respectively. Here, the sender *S* cannot sense N_R's transmission, even though it is strong enough to corrupt the transmission from *S* to *R*. The shaded area shown in Fig. 1(b), where the hidden terminals can hide, is called the "*vulnerable region*".



(a) Vulnerable period (2T) in pure ALOHA (b) Vulnerable region in carrier sensing MAC

Fig. 1: Vulnerable period and vulnerable region in random access MAC protocols.

A *busy tone* is one approach used to avoid the hidden terminal problem in a carrier sensing radio network [20]. Whenever any node detects a packet being transmitted, it starts to send a signal, called a busy tone, in a separate frequency channel. For example, when node *S* starts to send a packet to node *R*, node *R* as well as node N_S will start to send a busy tone. All the nodes that can hear the busy tone will not initiate their own transmission and thus node *R* will experience no collision. A critical problem with the use of busy tones is that too many nodes (all 2-hop neighbors of node *S*) will be inhibited from transmitting. The number of nodes affected will typically be about four times the number of nodes within the transmission range of the receiver, which is the only set of nodes that should be inhibited. Therefore, while this approach almost completely eliminates collisions, it is not a very promising approach from a throughput standpoint [20].

2.2 DCF of IEEE 802.11 MAC

The IEEE 802.11 wireless LAN standard adopts a dynamic channel allocation scheme based on carrier sensing technique, called *DCF* (*Distributed Coordination Function*), as its basic MAC layer algorithm. Four key elements of DCF are *ACK*, *RTS/CTS* with *NAV*, *IFS* and *Backoff algorithm* with *CW*. This subsection introduces these four key elements, which is essential for understanding the utilization enhancing techniques in the following sections.

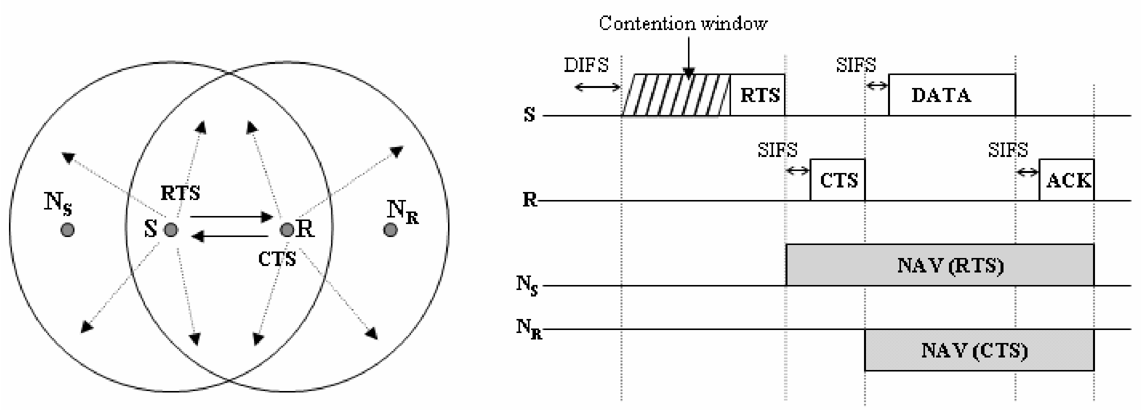
ACK for Collision Detection

ACKnowledgement (*ACK*) packets enable a mobile node to determine whether its transmission was successful or not since it cannot otherwise detect a collision. The sender is made aware of the collision after it times out waiting for the corresponding ACK for the packet transmitted. If no ACK packet is received or an ACK is received in error, the sender will contend again for the medium to retransmit the data packet until the maximum allowed number of retransmissions has been tried. If all fails, the sender drops the packet consequently leaving it to a higher level reliability protocol. Note that this sort of link level ACKs are not usually used in wired networks because wired links are quite reliable and collisions are easily detected.

RTS/CTS and NAV for Solving Hidden Terminal Problem

In DCF, collisions from the nodes hidden in the vulnerable region can be effectively avoided by *four-way handshake* based on *Request-To-Send* (*RTS*) and *Clear-To-Send* (*CTS*) packets. By exchanging the two short control packets between a sender and a receiver, all neighboring nodes recognize the transmission and back off during the transmission time advertised along with the RTS and CTS packets. Using this information, each node maintains a *Network Allocation Vector* (*NAV*), which indicates the remaining time of the on-going communication. Fig. 2(a) shows the transmission range of RTS and CTS control packets. Nodes N_S and N_R would receive RTS and CTS, respectively, and set their NAVs accordingly to refrain themselves from accessing the medium during the transmission of node *S*. Fig. 2(b) shows the four-way handshake between *S* and *R* as well as IFS and contention window, which will be described below.

However, as discussed in Section 1, the reduction in the chance of collisions occurs at the expense of increased control overhead involved with the exchange of RTS and CTS packets, which can be significant for short frames. For this reason, DCF allows the use of RTS/CTS mechanism but does not require it and suggests the use of the "*RTSTheshold*" parameter to determine the payload size for which RTS/CTS should be used [7]. This parameter is not fixed and has to be set separately by each mobile node.



(a) RTS/CTS mechanism (b) IFS, RTS/CTS and NAV

Fig. 2: Collision avoidance mechanism of DCF.

IFS for Prioritized Access to the Channel

Inter-Frame Spacing (*IFS*) is the time interval during which each node has to wait before transmitting any packet and is used to provide a prioritized access to the channel. For example, *Short IFS* (*SIFS*) is the shortest and is used after receiving a DATA packet to give the highest priority to an ACK packet. *DCF IFS* (*DIFS*) is larger than SIFS and is used when initiating a data transfer. When RTS/CTS is used, the RTS packet can be transmitted after waiting for DIFS duration of time. All other frames (CTS, DATA, and ACK) use SIFS before attempting to transmit. Fig. 2(b) shows the usage of DIFS and SIFS. Two other IFSs are *Point Coordination Function IFS* (*PIFS*) and *Extended IFS* (*EIFS*), which will be discussed shortly in this section.

Backoff Algorithm with CW to Provide Fair Access with Congestion Control

The abovementioned IFS is followed by an additional waiting time defined by the backoff algorithm used in DCF. The main purpose of the backoff algorithm is to reduce the probability of collisions when contention is severe. After waiting for the IFS duration, each competing node waits for a backoff time, which is randomly chosen in the interval (0, *CW*), defined as *contention window*. During the first transmission of a packet, CW is set to its minimum preset value, CWmin. If the channel continues to be idle during the backoff time, it transmits (winner). Other waiting nodes (losers) become aware of the transmission, freeze their backoff time, and contend again in the next competition cycle after the current transmission completes. Now, the frozen backoff time plays an important role in ensuring fairness. Definition of fairness may differ, but in general all nodes entering the competition for the first time should have on an average equal chance of transmitting, and nodes that have lost in the previous competition cycle should have higher priority than newly arrived nodes during the current competition cycle. The losers are given a higher priority by using the remaining frozen backoff time thereby preserving the first-come, first-serve policy.

The aforementioned access scheme has problems under heavy or light loads. If CW is too small compared to the number of competing nodes, it causes many collisions. On the other hand, if CW is too large, it causes unnecessary delay [21]. DCF adopts the *binary exponential backoff scheme* to allow an adaptive solution to this problem. When a node fails to receive an ACK in response to transmission of a DATA packet, it needs to contend in the next competition cycle. However, CW is doubled after the collision and this continues until CW reaches a preset limit, *CWmax*. It is noted that CW is restored to its minimum, *CWmin*, when a node successfully completes a data transmission. Fig. 3 shows the flow chart of the backoff algorithm used in DCF.

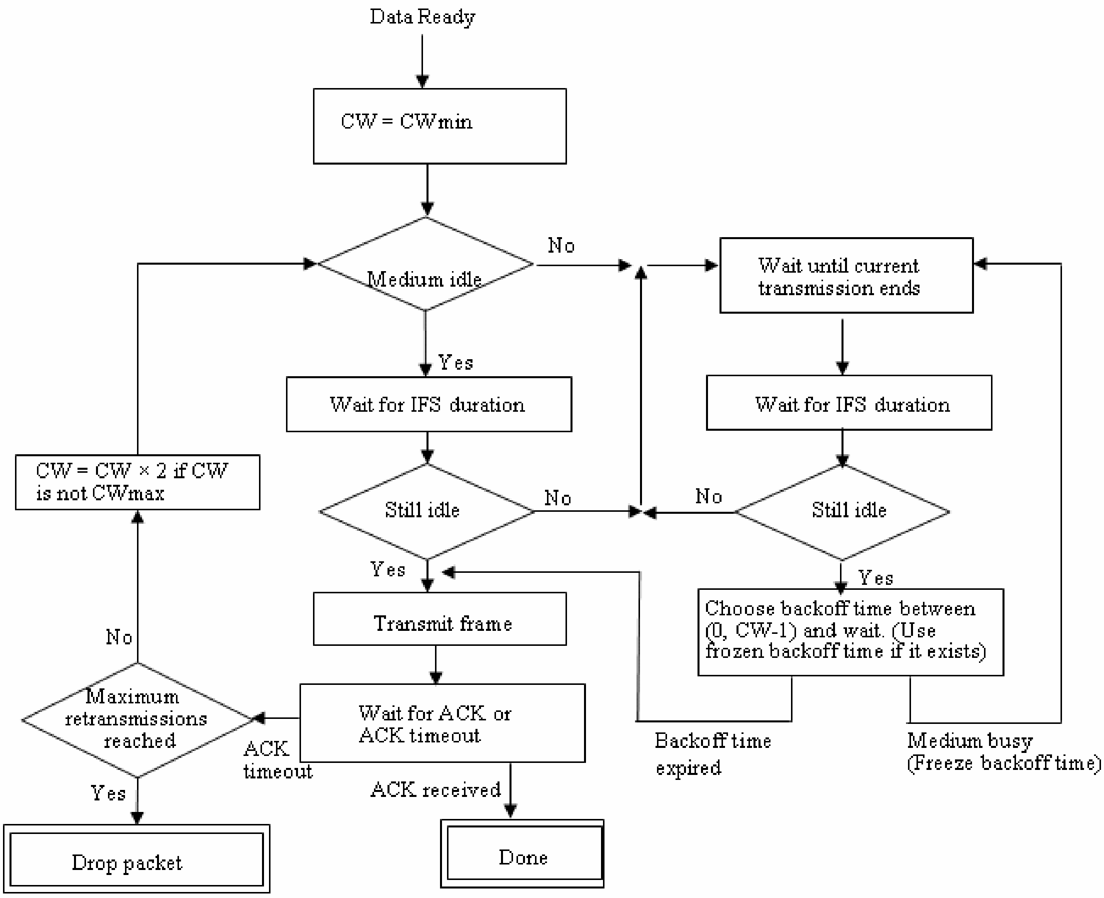


Fig. 3: Backoff algorithm used in DCF of IEEE 802.11 MAC.

EIFS to Protect ACK from Collisions

The RTS/CTS mechanism together with NAV effectively eliminates the vulnerable region introduced in Fig. 1(b). However, some packets are still vulnerable to collisions. For example, consider the coverage area of a radio transmitter, which depends on the power of the transmitted signal and the *path loss*. Each radio receiver has particular power sensitivity; e.g., it can only detect and decode signals with strength larger than this sensitivity [22]. There are two threshold values when receiving radio signals: *receive threshold* (*RXThresh*) and *carrier sense threshold* (*CSThresh*). If the power of the received signal is higher than RXThresh, it is regarded as a valid packet and passed up to the MAC layer. The corresponding distance for two nodes to communicate successfully is called the *transmission range*.

On the other hand, if the received signal power is lower than CSThresh, it is discarded as noise and thus the node can start its own transmission or reception. If the signal power is in between RXThresh and CSThresh, the node cannot receive the packet intelligibly but acknowledges that some active transmission is going on. The corresponding distance is referred to as *interference range*. Thus, when node *S* transmits a data packet to node *R*, there are four different groups of nodes in the network as shown in Fig. 4(a):

- A node is within the transmission range of *S* or *R* (Group I). Thus, it can receive RTS or CTS and sets its NAV accordingly.
- A node is outside of transmission range of *S* and *R* but is within the interference range of *S* and *R* (Group II). Thus, it cannot receive packets intelligently but recognizes the on-going communication.
- A node is outside of interference range of *R* but is within the interference range of *S* (Group III). Thus, it cannot sense CTS and ACK transmission from *R*.
- A node is outside of interference range of *S* but is within the interference range of *R* (Group IV). Thus, it cannot sense data packet transmission from *S*.

Nodes in Group I correctly set their NAVs when receiving RTS or CTS, and defer their transmission until the *S*-*R* communication is finished. Nodes in Group II cannot decode the packets and do not know the duration of the packet transmission, but they do sense on-going communications and thus do not cause collisions.

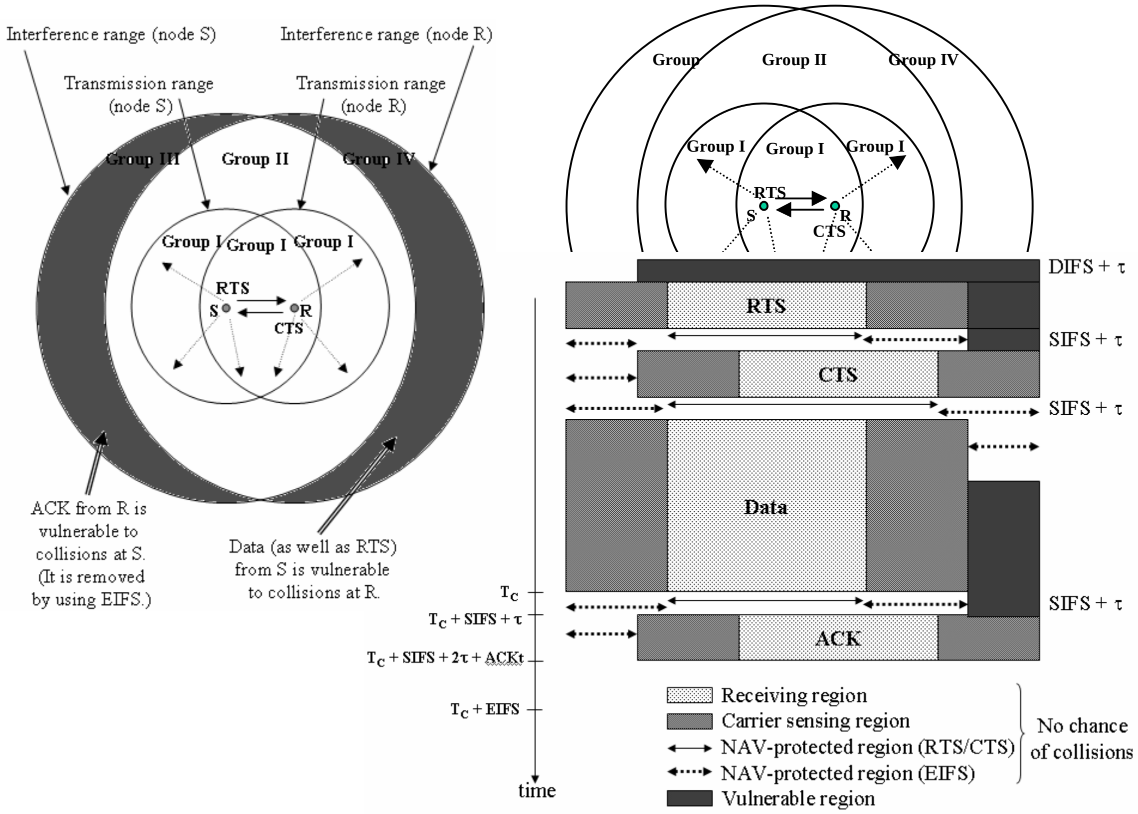


Fig.4: Vulnerable region with considering the interference range (τ: propagation delay).

However, ACKs (from *R* to *S*) and DATA (from *S* to *R*) are vulnerable to collisions due to nodes in Group III and IV, respectively. Collisions are critical for any type of packets but ACK collisions are a more serious problem because an ACK packet forms a vital piece of information as the last step of the four-way handshake. A loss of ACK results in retransmission of long DATA packet and thus significantly degrades the performance. *Extended IFS* (*EIFS*) is used in DCF to prevent collisions with ACK receptions at the sender. When nodes detect a transmission but cannot decode it, they set their NAVs for the EIFS duration. For example, in Fig. 4(b), when *S* completes its data transmission at T_C, nodes in Group II and III would set their NAV to T_C+EIFS. At $T_C+SIFS+\tau$, *R* replies back to *S* with an ACK and the transmission is completed at $T_C+SIFS+2\tau+ACK_t$, where τ is the propagation delay of the channel and ACK_t is the transmission time for the ACK packet. If EIFS is larger than $SIFS+2\tau+ACK_t$, nodes in Group II and III would not corrupt the ACK packet from *R* to *S*. These nodes have to wait an additional DIFS to start the competition, thus EIFS is set to $SIFS+ACK_t+DIFS$ in the IEEE 802.11 MAC standard.

Table 1 summarizes the characteristics of a typical radio transceiver and the four key elements of DCF with typical values for the related parameters.

Table 1: Radio transceiver characteristics and key elements of DCF.
(914 MHz, 1Mbps Lucent WaveLAN using Direct Sequence Spread Spectrum)

Key elements	Parameters	Typical values	Comment
Radio transceiver	Transmission power	0.2818 W	
	RxThresh	3.652×10^{-10} W	Transmission range 250m (with two-ray ground model)
	CSThresh	1.559×10^{-11} W	Interference range 550m (with two-ray ground model)
ACK	ACK frame size	376 µsec	184-bit ACK packet with 144 and 48 bits of physical layer preamble and header over 1Mbps link
RTS/CTS and NAV	RTS frame size	424 µsec	232-bit RTS packet with 144 and 48 bits of physical layer preamble and header over 1Mbps link
	CTS frame size	376 µsec	184-bit CTS packet with 144 and 48 bits of physical layer preamble and header over 1Mbps link
	RTSThreshold		Not specified
	Retry limit for a long packet	4	For DATA packet longer than RTSThreshold
	Retry limit for a short packet	7	For RTS and shorter DATA packet
IFS	SIFS (Short IFS)	10 µsec	For CTS, DATA and ACK packet
	DIFS (DCF IFS)	50 µsec	For RTS and short DATA packet
	EIFS (Extended IFS)	436 µsec	SIFS (10) + ACKt (376) + DIFS (50)
Backoff algorithm	Slot time	20 µsec	
	CWmin	32	Equivalent to 640 µsec
	CWmax	1024	Equivalent to 20.48 msec

Performance Limit of DCF

There has been active research on estimating the performance of IEEE 802.11 MAC analytically as well as via simulation [7, 8, 18, 23-27]. Among them, Cali *et al*. have provided a mathematical model for the maximum achievable throughput [8]. According to their results, the theoretical throughput is bounded by around 80% when the typical DCF parameters are used (with propagation delay of 1 µsec and packet size of 50µsec~5msec). In reality, DCF operates very far from the theoretical limits due to collisions and control overhead associated with RTS/CTS and the backoff algorithm.

In a multihop MANET, the situation becomes worse due to the reasons discussed in Section 1. Li *et al.* showed that the end-to-end throughput is at most 1/4 of the channel bandwidth even without any other interfering nodes [28]. In other words, when IEEE 802.11-based 2 Mbps wireless network interface is used, a source-destination pair in a MANET cannot support more than 500 kbps. This is mainly due to collisions among intermediate forwarding nodes of the same data stream. In addition, the control overhead of DCF aggravates the situation and the maximum throughput is reduced to about 1/7 of the channel bandwidth [28]. When other data traffic exists, the throughput is reduced even further. For example, *Xu* and *Saadawi* reported that multiple simultaneous TCP sessions in a MANET result in unreasonably low aggregate throughput and suffers from severe unfairness [23].

3. Enhancing Temporal Channel Utilization

As pointed out previously, the performance limitation is mainly due to the limited capability of MAC protocols in a multihop communication environment. A key idea for improving DCF for MANET is *adaptivity*. That is, each node should be able to behave adaptively according to traffic intensity in its vicinity. This section discusses the non-adaptive characteristics of DCF and the temporal approaches proposed in the literature [5-8]. They attempt to enhance the effective channel utilization by reconsidering the DCF parameters such as RTSThreshold (Section 3.1) and the backoff algorithm (Section 3.2) in order to better schedule the channel along the time dimension.

3.1 RTS/CTS Mechanism

Optimal Setting of RTSThreshold to Tradeoff between Control and Collision Overhead

As discussed in Section 2.2, the parameter RTSThreshold is used to determine whether RTS/CTS is used or not. However, this parameter is not fixed in the DCF standard as discussed previously. Khurana *et al*. studied the throughput of an IEEE 802.11-based ad hoc network to obtain the optimal parameters for DCF including the RTSThreshold [5]. Assuming that the physical layer uses *Direct Sequence Spread Spectrum* (*DSSS*) and DCF uses typical parameters as in Table 1, they recommend a value of 250 bytes for the RTSThreshold [5]. In other words, the RTS/CTS exchange is beneficial only when data packet size is larger than 250 bytes. Weinmiller *et al*. performed a similar study and concluded via simulation that the best throughput is obtained when 200-500 bytes is used for the RTSThreshold [7]. Note that this size should take into account the necessary physical layer preamble and header according to the MAC packet format called *MPDU* (*MAC Protocol Data Unit*) as noted in Table 1.

A better idea is to adjust the parameter depending on the traffic and the collision probability. Even if data packet size is large, the RTS/CTS exchange is a waste of bandwidth if the number of hidden terminals is small and collisions are unlikely. Therefore, the optimal value for RTSThreshold depends on the traffic intensity, which can be estimated indirectly by noting the number of collisions experienced [5, 7].

3.2 Exponential Backoff Algorithm

Conservative CW Restoration to Reduce Collisions

In DCF of IEEE 802.11, the contention window is reduced to the minimum value (CWmin) for every new packet whether the last packet was successfully delivered or not. Even if the network area is congested with many competing data streams, each packet transmission starts with the minimum window size and thus experiences a large number of collisions before its window size becomes appropriate [8, 24]. In addition, restoration of CW to CWmin makes the backoff algorithm unfair since it favors the mobile node that has most recently transmitted [23]. In Fig. 5(a), node *A* wins in the first competition cycle because it chooses the smaller backoff time ($BOFF_A$) than nodes *B* and *C* ($BOFF_B$ and $BOFF_C$). While node *A* restores its CW to CWmin in the next competition cycle, nodes *B* and *C*, being losers, keep the same CW as in Fig. 5(b). Even though nodes *B* and *C* reduce their backoff time by using the frozen values ($BOFF_B$-$BOFF_A$ and $BOFF_C$-$BOFF_A$, respectively), node *A* has a better chance of winning in the next competition cycle again due to the reduced CW size.

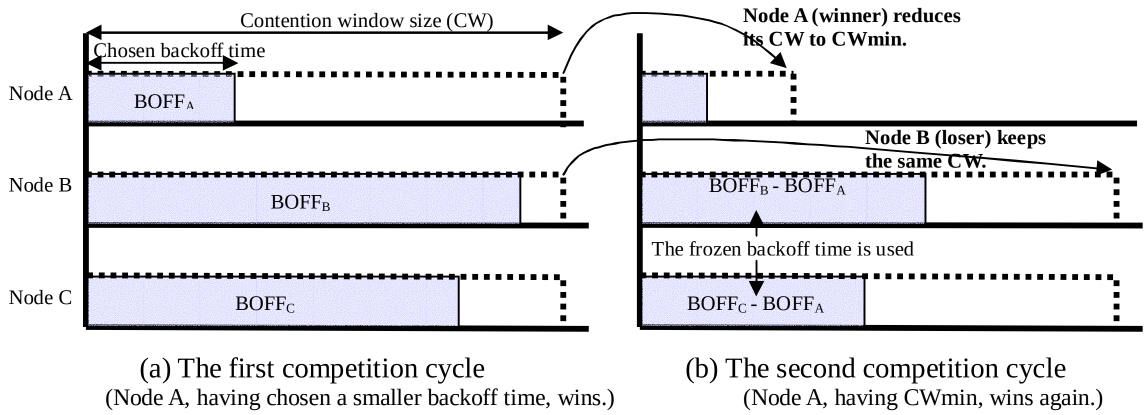


Fig. 5: Unfairness problem in DCF due to backoff algorithm.

In order to solve the collision and fairness problem, Bharghavan *et al*. proposed a *Multiplicative Increase and Linear Decrease* (*MILD*) algorithm where the contention window size increases multiplicatively on collisions but decreases linearly on successful transmission [6]. MILD algorithm works well when the network traffic is high, but under light traffic condition, it incurs additional delay to return the CW to CWmin, which is not required in the original backoff algorithm.

Different Treatment of New and Lost Nodes for Fairness

Weinmiller *et al*. investigated the effect of CW restoration to CWmin together with the frozen backoff time [7]. In the initial state, the backoff algorithm in DCF results in an equally distributed probability for each slot to be selected. However, in the following competition cycle, the probability is not equally distributed. Consider an example in Fig. 5(b). Since $BOFF_A$ is the winner's backoff time in the first competition cycle and the losers use the frozen backoff time in the next competition cycle, the contention window of these nodes is effectively reduced to (0, CW-$BOFF_A$). Still within this reduced contention window, all slots are selected with the same probability by these nodes. However, newly entered nodes will choose their slot with equally distributed probability within the whole range of the contention window (0, CW). Therefore, slots later than CW-$BOFF_A$ have a significantly lower probability to be chosen compared to the earlier slots. After several competition cycles, the slot selection probability becomes a decreasing staircase function.

As far as the collision probability is concerned, this leads to a high chance of collisions at earlier slots because these slots will most probably be selected twice or more times. An equally distributed probability for every slot to be chosen is the favored situation in terms of collision avoidance. Weinmiller *et al*. suggested two alternative solutions for this fairness problem, both of which attempt to offer the later slots in (CW-$BOFF_A$, CW) to the newly entering nodes and earlier slots in (0, CW-$BOFF_A$) to the nodes that have lost the previous competition [7]. These schemes assume that a newly arriving node knows the winning slot of previous competition, which may not be the case under certain conditions.

Dynamic Tuning of CW to Minimize the Collision Probability

Cali *et al*. observed that the collision probability increases as the number of active nodes increases, but it cannot be dynamically controlled due to the static backoff algorithm of DCF [8]. In other words, the optimal setting of CW, and thus the optimal backoff time, can be achieved by estimating the number of active nodes in its vicinity at run time. Since each node can estimate the number of empty slots in a virtual transmission time by observing the channel status, the number of active nodes can be computed and exploited to select the appropriate CW without paying the collision costs [8].

Table 2 summarizes the channel utilization enhancing techniques discussed in this section.

Table 2: Enhancing temporal channel utilization.

Key elements	Parameter	Problem	Solution technique
RTS/CTS and NAV	RTSThreshold	Undetermined or fixed RTSThreshold	Optimal preset value: - 250 bytes MPDU [5] - 200-500 bytes MPDU [7] Adaptive adjustment based on - traffic and collision probability [5] - experienced collisions [7]
Backoff algorithm	CW restoration to CWmin	Many collisions or large delay	Multiplicative Increase and Linear Decrease (MILD) [6]
	Frozen backoff time	Staircase-like slot selection probability and more collisions	Offer later slots to new nodes and earlier slots to old and lost nodes [7]
	Backoff algorithm	CW is not optimal	Dynamic tuning with the estimation of the number of active nodes in its vicinity at run time [8]

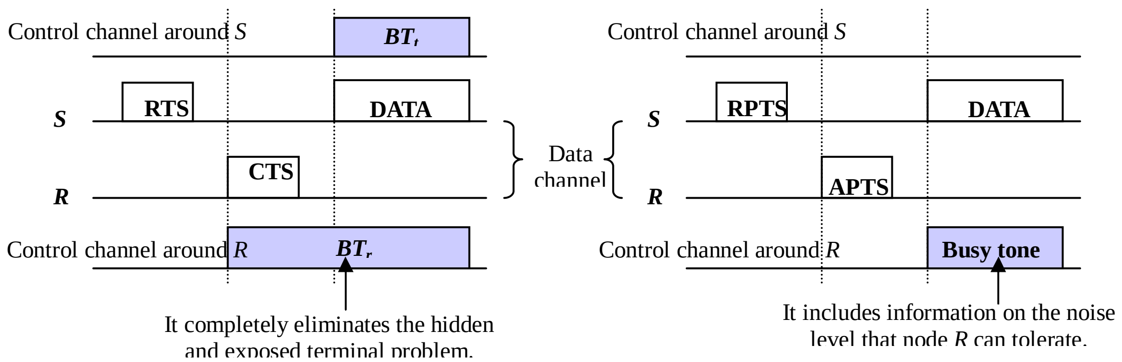
4. Enhancing Spatial Channel Utilization

In this section, we discuss MAC protocols that better utilize the channel along the spatial dimension. While the temporal approaches in Section 3 can be applied to single-hop wireless LANs as well as multihop MANETs, the spatial approaches discussed in this section focus on multihop MANETs and exploit the characteristics unique to the multihop communication environment. The *Dual Busy Tone Multiple Access* (*DBTMA*) protocol [9] employs a busy tone to reserve only the space around the receiver to encourage spatial reuse. Based on the same concept of busy tone, the *Power Controlled Multiple Access* (*PCMA*) scheme [10] further reduces the interference range by employing the *transmission power control*. An alternative to these two approaches is the use *directional antenna* to transmit or receive data only along a certain direction, and thus reserves only a fraction of space compared to that of omni-directional antenna [13-17]. The following three subsections discuss the three approaches, respectively.

4.1 Busy Tone to Solve the Exposed Terminal Problem

In order to avoid interference from other transmissions, a source-destination pair should reserve some spatial area, but the area should be as small as possible to encourage more spatial reuse. One example of excessive space reservation in DCF is the RTS/CTS mechanism: Since collisions occur only at the receiver side, it is not necessary to reserve space around the sender. This is known as the *exposed terminal problem* [21], which means that some nodes around the sender are overly exposed to the on-going communication and experience unnecessary delay until the sender completes its data transmission.

The *Dual Busy Tone Multiple Access* (*DBTMA*) protocol [9] uses busy tone with RTS/CTS to solve the exposed terminal problem. A separate control channel is used for both control packets (RTS and CTS) and two busy tones (transmit and receive busy tones, BT_t and BT_r). The main feature of DBTMA is the use of the control channel to completely eliminate the hidden as well as the exposed terminal problem. BT_t and BT_r on the control channel indicate that the node is transmitting and receiving on the data channel, respectively. All other nodes sensing the BT_r signal (hidden terminals) defer their transmissions, and nodes sensing the BT_t signal do not attempt to receive. Thus, exposed terminals can sense BT_t but not BT_r so that they can safely reuse the space by transmitting their packets. Fig. 6(a) shows the DBTMA protocol with two busy tones.



(a) DBTMA with two busy tones (b) PCMA with power control & busy tone

Fig. 6: DBTMA and PCMA protocols.

In addition, busy tone can help solve the collision problem due to mobility. The conventional RTS/CTS scheme may not work well in a network with highly mobile nodes. This is because nodes may come within the range of either the sender or receiver after the RTS/CTS exchange. With DBTMA, such hidden terminals do not exist because the receiver continuously sends the BT_r signal to its neighbors.

4.2 Transmission Power Control to Reduce Interference Range Radially

When a node's radio transmission power is controllable, its direct communication range as well as the number of its immediate neighbors is also adjustable. While higher transmission power increases the transmission range, lower transmission power reduces the collision probability by reducing the number of competing nodes. In the *Power Controlled Multiple Access* (*PCMA*) protocol [10], a source-destination pair uses *Request power to send* (*RPTS*) and *Acceptable power to send* (*APTS*) control packets to compute the optimal transmission power based on their received signal strength, which will be used when transmitting data packets. PCMA also uses the busy tone channel to advertise the noise level the receiver can tolerate. A potential transmitter first senses the busy tone to detect the upper bound of its transmission power for all control and data packets. Fig. 6(b) shows the PCMA protocol with busy tone.

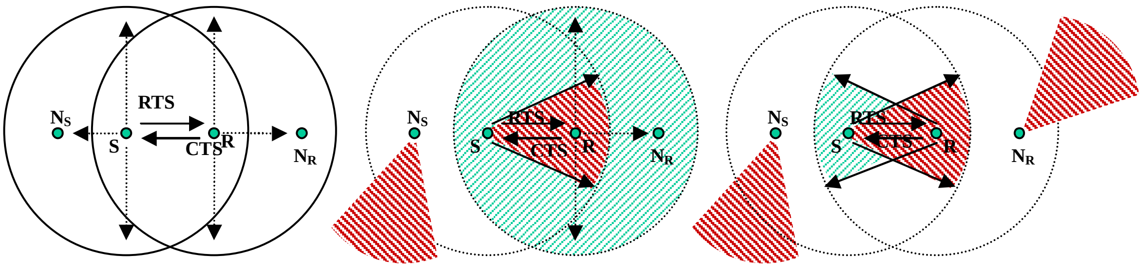
Transmission power control approach has been actively studied for other purposes, such as energy saving or topology control. For example, Gomez *et al.* proposed using the maximum power level for RTS and CTS packets and lower power levels for data packets [11]. This does not increase or decrease the collision probability but nodes can save substantial amount of energy by using a low power level for data packets. However, this approach has a problem with respect to ACK reception because EIFS (used to protect ACK) is only effective when data packets are transmitted at full power as discussed in Section 2.2. The *Power Control MAC* (*PCM*) protocol addresses this problem by transmitting data at a reduced power level most of the time, but periodically transmits at the maximum power level to inform to its neighboring nodes about the current transmission. Another related area of research is routing protocols based on transmission power control [29-31]. We do not discuss these protocols in detail in this chapter because they are designed to save energy rather than improve channel utilization. For a detailed discussion on this subject, please refer to [32].

4.3 Directional Antenna to Reduce Interference Range Angularly

Unlike an omni-directional antenna, a *directional antenna* has a directional radiation pattern making it possible to transmit to a subset of its neighbors [33]. When it is used for transmission, it can significantly reduce the unwanted interference to nodes outside its directional pattern. Similarly, when it is used for reception, the receiver can eliminate the interference signals from directions other than the signal source [13]. Thus, directional antennas improve spatial reuse and reduce multi-path propagation, which can result in better channel utilization.

With omni-directional antennas, one-hop neighbors within the range of the sender (S) or the receiver (R) defer their transmission based on RTS/CTS as shown in Fig. 7(a). While a hidden terminal N_R should defer its transmission in order to protect node R's reception, an exposed terminal N_S unnecessarily defers its transmission because it would not have interfered with the ongoing S-R communication. This wastes the spatial channel bandwidth around node S. Directional antennas can eliminate this problem by using *directional RTS* (*DRTS*) and *directional CTS* (*DCTS*) instead of *omni-directional RTS* (*oRTS*) and *omni-directional CTS* (*oCTS*) as shown in Fig. 7(b) and 7(c).

A key question then is how can collisions be avoided with DRTS and DCTS packets. For example, in Fig. 7(c), when N_R wishes to transmit directly to R, it simply transmits because N_R did not receive DCTS from node R and thus it is not aware of the S-R communication (*deafness problem* [16]). This may or may not cause collisions at node R depending on the underlying antenna model (*directional hidden terminal problem*). Another important question is how to find the desired direction for the transmission and reception when initiating DRTS or replying with DCTS. This section discusses three representative *directional MAC* (*DMAC*) algorithms based on oRTS/oCTS [13], DRTS/oCTS [14] and DRTS/DCTS [15], respectively, as shown in Fig. 7.



(a) oRTS/oCTS
(Very conservative. N_S and N_R defers.)

(b) DRTS/oCTS
(N_R defers. N_S can transmit but can corrupt CTS and ACK from R at S.)

(c) DRTS/DCTS
(Very aggressive. N_S and N_R can transmit.)

Fig. 7: Three MAC algorithms based on directional antenna.

oRTS/oCTS-based DMAC

Naspuri *et al.* proposed the oRTS/oCTS-based DMAC protocol [13], where all control packets are transmitted omni-directionally and only data packets are transmitted directionally. Collisions are avoided as in conventional omni-directional MAC algorithms, and the additional benefit is the significant reduction in interference by transmitting and receiving data packets over a small angle.

The key feature of this scheme is a mechanism to determine the direction of the other party of the communication. Here, the radio transceiver is assumed to have multiple directional antennas and each node is capable of switching any one or all antennas to active or passive modes, known as *directional reception capability*. An idle node listens to on-going transmission on every direction. When it receives an oRTS addressed to itself, it can determine the direction of the sender by noting the antenna that received the maximum power of the oRTS packet[1] [13]. Similarly, the sender estimates the direction of the receiver by receiving the oCTS packet. Thus, a receiver is not influenced by other transmissions from other directions. Fig. 7(a) shows the oRTS/oCTS-based DMAC scheme.

DRTS/oCTS-based DMAC

Ko *et al.* proposed two DMAC schemes based on DRTS [14]. The first scheme trades off between spatial reuse and collision avoidance by using DRTS and oCTS. While oCTS helps avoid the collisions from hidden terminals, such as N_R in Fig. 7(b), DRTS helps improve the spatial channel utilization by eliminating the exposed terminal problem. (N_S is free to attempt its transmission during the *S*-*R* communication.) The second scheme uses both DRTS and oRTS to reduce the probability of collisions of control packets in the sender's vicinity caused by the exposed terminal. The usage rule is if there is no on-going communication in every direction around a sender, then it transmits an oRTS. Otherwise, the sender transmits a DRTS. In both schemes, nodes require external location tracking support such as GPS to determine the direction of the nodes they would like to communicate with. Based on the location of the receiver, the sender may select an appropriate directional antenna to send packets (DRTS and data packets) to the receiver.

DRTS/DCTS-based DMAC

Wang and Garcia-Luna-Aceves observed that the benefit of spatial reuse achieved by a DMAC protocol can outweigh the benefit of a conservative collision avoidance mechanism that sends some omni-directional control packets to silence potential interfering nodes [15]. Their approach uses both DRTS and DCTS and aggressively reuses the channel along the spatial dimension at the cost of

[1] Several directional antenna models have been proposed. *Sectored antenna* is assumed for the oRTS/oCTS-based scheme. It consists of multiple (M) directional antennas, each of which has a conical radiation pattern spanning an angle of $2\pi/M$ radians. A mobile node can look out simultaneously with all of its M antennas and recognize the direction of arrival by noting the antenna on which the gain is the maximum. *Directional beam-forming antenna* is used for directional transmission or reception by beam-forming towards intended receiver or sender. Thus, it is usually used along with an omni-directional antenna for listening on all directions. *Multi-beam adaptive array model* is based on an antenna array, capable of forming multiple beams for several simultaneous receptions or transmissions.

increased chance of collisions. In Fig. 7(c), N_S and N_R can initiate their own transmissions during an *S*-*R* communication. It is noted that nodes have directional reception capability as discussed previously and thus the transmission from N_R does not cause collisions at node *R*. Location tracking support is required for implementing this scheme.

Other DMAC Protocols

Before concluding this section, we introduce two additional DMAC protocols: *Multihop RTS MAC* (*MMAC*) [16] and *Receiver-Oriented Multiple Access* (*ROMA*) [16]. Choudhury *et al*. made an important observation that the gain of directional antennas is higher than that of omni-directional antennas, and thus they have a greater transmission/reception range [16]. Even if the receiver is within the sender's transmission range, the receiver may not be able to communicate with the sender if its reception range does not include the sender. This is quite possible when the sender transmit directionally knowing the receiver's location (via GPS), but the receiver tries to receive omni-directionally since it does not know about the transmission attempt from the sender. Therefore, even though data packets can be transmitted over a single hop using directional antenna at both nodes, it is possible for control packets such as DRTS to take more than one hop. MMAC takes into account this fact and uses *multihop RTS* for delivering DRTS to the receiver over a number of hops.

Another recent DMAC protocol proposed by Bao and Garcia-Luna-Aceves is not based on RTS/CTS but uses a transmission schedule determined statically based on node identifier and time slot number [17]. While on-demand medium access schemes determine the communicating pair by exchanging short control signals such as RTS/CTS before each transmission session, scheduled medium access schemes prearrange or negotiate a set of timetables for individual nodes or links. ROMA is such a schedule-based MAC protocol where the communicating nodes are paired with the designated time slots based on the schedule, and thus the transmissions are collision-free [17].

Table 3 summarizes the channel utilization enhancing techniques discussed in this section.

Table 3: Enhancing spatial channel utilization.

Conventional facility	Problem	Additional facility	Solution technique
Single channel for data and control packet	Unnecessary space reservation around the sender by RTS (Exposed terminal problem)	Separate busy tone channel	Advertise the communication over the busy tone channel (DBTMA) [9]
Single power model	Unnecessary interference and space reservation when the communicating distance is short	Transmission power control of radio transceiver	Advertise the tolerable noise level over the busy tone channel (PCMA) [10] Use low power for data packets [11] Periodic power adjustment when delivering data packets (PCM) [12]
Omni-directional antenna model	Unnecessary interference and space reservation since communication is omni-directional	Directional antenna	Omni-directional control packet transfer but directional data packet transfer [13] Directional RTS [14] Directional RTS and CTS [15] Multihop RTS to take into account the difference in antenna gain [16] Schedule-based directional MAC [17]

6 Conclusions

Mobile ad hoc networks are composed of nodes that are self-organizing and communicate over wireless channels usually in a multi-hop fashion. They exhibit dynamic topology, share limited bandwidth, with most nodes having limited processing abilities, and energy constraints. In this chapter, we have considered some of the techniques in the design of medium access control protocols with DCF of IEEE 802.11 as a reference model. Each of these schemes tries to maximize network capacity, reduce congestion at the MAC layer, and ensure fairness by balancing the control overhead to avoid collisions. Key techniques used to enhance temporal utilization is to optimize the DCF parameters such as RTSThreshold and those associated with the backoff algorithm, which is used to avoid collisions in DCF. Spatial reuse assumes special importance in multi-hop networks. Busy tone method, transmission power control, and directional transmissions are the key techniques in this direction. Among these, the possibilities provided by directional transmissions are most promising since it can reduce interference and collisions considerably, and can be used in conjunction with the other two techniques. Transmission power control methods not only help in reducing interference but also in energy conservation.

References

[1] Forman, G., Zahorjan, J., “The challenges of mobile computing,” IEEE Computer, Vol. 27, No. 4, pp.38-47, 1994.

[2] Frodigh, M., Johansson, P., and Larsson, P., “Wireless ad hoc networking – The art of networking without a network,” Ericsson Review, No. 4, pp. 248-263, 2000.

[3] Perkins, C., Ad Hoc Networking, Addison-Wesley Pub. Co., 2001.

[4] Stallings, W., “IEEE 802.11 Wireless LAN Standard,” Chapter 14, Wireless Communications and Networks, Prentice Hall, Inc., 2002.

[5] Khurana, S., Kahol, A., Gupta, S. K. S. and Srimani, P. K., “Performance Evaluation of Distributed Co-Ordination Function for IEEE 802.11 Wireless LAN Protocol in Presence of Mobile and Hidden Terminals,” MASCOT'99, pp. 40-47, 1999.

[6] Bharghavan, V., Demers, A., Shenker, S., and Zhang, L., “MACAW: A Media Access Protocol for Wireless LANs,” ACM SigComm, pp. 212-225, 1994.

[7] Weinmiller, J., Woesner, H., Ebert, J.-P., Wolisz, A., “Analyzing and Tuning the Distributed Coordination Function in the IEEE 802.11 DFWMAC Draft Standard,” MASCOT’96, 1996.

[8] Cali, F., Conti, M., and Gregori, E., “Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit,” IEEE/ACM Tr. Networking, Vol. 8, No. 6, pp. 785-799, Dec. 2000.

[9] Deng, J., and Haas, Z. J., “Dual Busy Tone Multiple Access (DBTMA): A New Medium Access Control for Packet Radio Networks,” IEEE ICUPC'98, Florence, Italy, October 5-9, 1998.

[10] Monks, J. P., Bharaghavan, V., and Hwu, W. W., “A Power Controlled Multiple Access Protocol for Wireless Packet Networks,” IEEE Infocom, 2001.

[11] Gomez, J., Campbell, A. T., Naghshineh, N., Bisdikian, C., “Conserving Transmission Power in Wireless Ad Hoc Networks,” ICNP’01, 2001.

[12] Jung, E.-S., and Vaidya, N. H., “A Power Control MAC Protocol for Ad Hoc Networks,” ACM/IEEE MobiCom’02, 2002.

[13] Nasipuri, A., Ye, S., You, J., Hiromoto, R. E., “A MAC Protocol for Mobile Ad Hoc Networks Using Directional Antennas,” IEEE WCNC, 2002.

[14] Ko, Y.-B., Shankarkumar, V., Vaidya, N. H., “Medium Access Control Protocols Using Directional Antennas in Ad Hoc Networks,” IEEE Infocom, 2000.

[15] Wang, Y., and Garcia-Luna-Aceves, J. J., “ Spatial Reuse and Collision Avoidance in Ad Hoc Networks with Directional Antennas ” IEEE Globecom, 2002.

[16] Choudhury, R. R., Yang, X., Ramanathan, R., and Vaidya, N. H., “Using Directional Antennas for Medium Access Control in Ad Hoc Networks,” ACM/IEEE MobiCom, 2002.

[17] Bao, L., and Garcia-Luna-Aceves, J. J., “Transmission Scheduling in Ad Hoc Networks with Directional Antennas,” ACM/IEEE Mobicom, 2002.

[18] Pahlavan, K., and Krishnamurthy, P., “Wireless Medium Access Alternatives,” Ch. 4, Principles of Wireless Networks, Prentice Hall PTR, Upper Saddle River, New Jersey, 2002.

[19] Kleinrock, L., Tabagi, F. A., “Packet Switching in Radio Channels: Part I – Carrier Sense Multiple-Access Models and Their Throughput-Delay Characteristics,” IEEE Tr. Communications, Vol. COM-23, No. 12, Dec. 1975.

[20] Bertsekas, D., and Gallager, R., “Multiaccess Communication,” Ch. 4, Data Networks, 2nd Edition, Prentice Hall PTR, Upper Saddle River, New Jersey, 1992.

[21] Schiller, J., “Wireless LAN,” Ch. 7, Mobile Communications, Addison-Wesley, Harlow, UK, 2000.

[22] Pahlavan, K., and Krishnamurthy, P., “Characteristics of Wireless Medium,” Ch. 2, Principles of Wireless Networks, Prentice Hall PTR, Upper Saddle River, New Jersey, 2002.

[23] Xu, S., Saadawi, T., “Does the IEEE 802.11 MAC Protocol Work Well in Multihop Wireless Ad Hoc Networks?,” IEEE Communications Magazine, pp. 130-137, Jun. 2001.

[24] Chhaya, H. S., and Gupta, S., “Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol,” Wireless Networks, Vol. 3, pp. 217-234, 1997.

[25] Bruno, R., Conti, M., Gregori, E., “Optimization of Efficiency and Energy Consumption in p-Persistent CSMA-Based Wireless LANs,” IEEE Tr. Mobile Computing, Vol. 1, No. 1, pp. 10-31, Jan-Mar, 2002.

[26] Foh, C., and Zukerman, M., “Performance Analysis of the IEEE 802.11 MAC Protocol,” European Wireless, pp. 184-190, 2002.

[27] Kwak, B.-J., Song, N.-O., and Miller, L. E., “Analysis of the Stability and Performance of Exponential Backoff,” IEEE WCNC, 2003.

[28] Li, J. *et al.*, “Capacity of Ad Hoc Wireless Networks,” ACM/IEEE MobiCom'01, pp. 61-69, 2001.

[29] Chang J.-H. and Tassiulas L., “Energy Conserving Routing in Wireless Ad-hoc Networks,” IEEE Infocom, pp. 22-31, 2000.

[30] Doshi, S. and Brown, T. X., “Minimum Energy Routing Schemes for a Wireless Ad Hoc Network,” IEEE Infocom, 2002.

[31] Narayanaswamy, S., Kawadia, V., Sreenivas, R. S., Kumar, P. R., “Power Control in Ad-Hoc Networks: Theory, Architecture, Algorithm and Implementation of the COMPOW Protocol,” European Wireless, pp. 156-162, 2002.

[32] Yu, C., Lee, B., and Youn, H. Y., “Energy Efficient Routing Protocols for Mobile Ad Hoc Networks,” Wireless Communications and Mobile Computing (WCMC) Journal, John Wiley and Sons, Ltd., 2003.

[33] Horneffer, M., and Plassmann, D., “Directed Antennas in mobile broadband system,” IEEE Infocom, 1996.