Mobile Ad Hoc Networks

Asis Nasipuri
Department of Electrical & Computer Engineering
The University of North Carolina at Charlotte
Charlotte, NC 28223-0001

I. INTRODUCTION

A mobile ad hoc network is a collection of digital data terminals equipped with wireless transceivers that can communicate with one another without using any fixed networking infrastructure. Communication is maintained by the transmission of data packets over a common wireless channel. The absence of any fixed infrastructure, such as an array of base stations, make ad hoc networks radically different from other wireless LANs. Whereas communication from a mobile terminal in an "infrastructured" network, such as a cellular network, is always maintained with a fixed base-station, a mobile terminal (node) in an ad hoc network can communicate directly with another node that is located within its radio transmission range. In order to transmit to a node that is located outside its radio range, data packets are relayed over a sequence of intermediate nodes using a store-and-forward "multihop" transmission principle. All nodes in an ad hoc network are required to relay packets on behalf of other nodes. Hence, a mobile ad hoc network is sometimes also called a multihop wireless network.

Since no base stations are required, ad hoc networks can be deployed quickly, without having to perform any advance planning or construction of expensive network infrastructure. Hence, such networks are ideally suited for applications where such infrastructure is either unavailable or unreliable. Typical applications include military communication networks in battlefields, emergency rescue operations, undersea operations, environmental monitoring, and space exploration. Because of its "on-the"fly" deployment quality and relatively low cost of implementation, ad hoc networks are also used in places where it is cheaper than its infrastructured counterparts. Examples of these applications consist of a network of laptop computers in conference rooms, network of digital electronic equipment and appliances (e.g. VCR, television, computer, printer, remote control, etc.) to form a home area network, networks of mobile robots, and wireless toys [43], [14], [49]. Recently, there is a growing interest of using ad hoc networks of wireless sensors to perform unmanned distributed surveillance and tracking operations [47].

The design of ad hoc networks faces many unique challenges. Most of these arise due to two principle reasons. The first is that all nodes in an ad hoc network, including the source node(s), the corresponding destination(s), as well as the routing nodes forwarding traffic between them, may be mobile. As the wireless transmission range is limited, the wireless link between a pair of neighboring nodes break as soon as they move out of range. Hence, the network topology, that is defined by the set of physical communication links in the network (wireless links between all pairs of nodes that can directly communicate with each other) can change frequently and unpredictably. This implies that the multihop path for any given pair of source and destination nodes also changes with time. Mobility also causes unpredictability

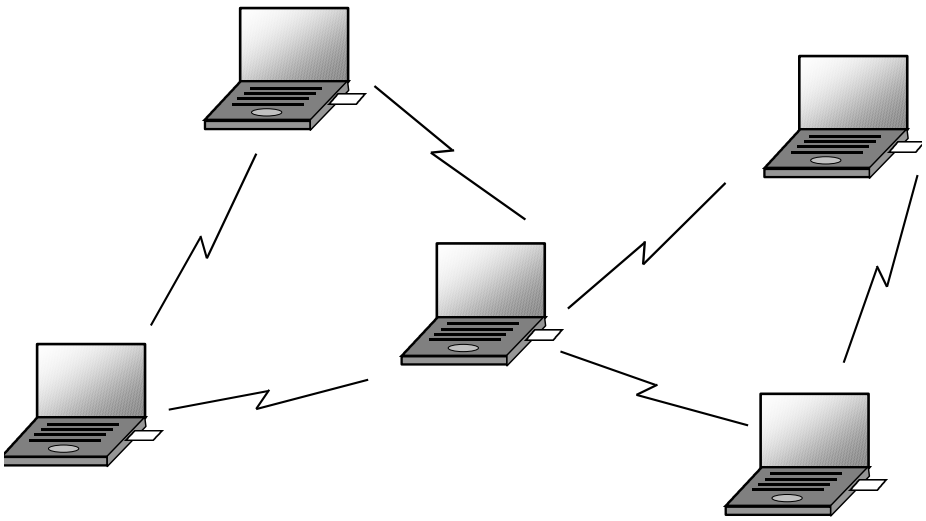

Fig. 1. A mobile ad hoc network.

in the *quality* of an existing wireless link between neighbors. A second reason that makes design of ad hoc networks complicated is the absence of centralized control. All networking functions, such as determining the network topology, multiple access, and routing of data over the most appropriate multihop paths, must be performed in a distributed way. These tasks are particularly challenging due to the limited communication bandwidth available in the wireless channel.

These challenges must be addressed in all levels of the network design. The physical layer must tackle the path loss, fading, and multi-user interference to maintain stable communication links between peers. The data link layer (DLL) must make the physical link reliable and resolve contention amongst unsynchronized users transmitting packets on a shared channel. The latter task is performed by the medium access control (MAC) sublayer in the DLL. The network layer must track changes in the network topology and appropriately determine the best route to any desired destination. The transport layer must match the delay and packet loss characteristics specific to such a dynamic wireless network. Even the application layer needs to handle frequent disconnections.

Although this area has received a lot of attention in the past few years, the idea of ad hoc networking started in the seventies, when the U.S. Defense Research Agency, DARPA, sponsored the PRNET (Packet Radio Network) [26] project in 1972. This was followed by the SURAN (Survivable Adaptive Radio Network) project [52] in the 1980s. These projects supported research on the development of automatic call set up and maintenance in packet radio networks with moderate mobility. However, interest in this area grew rapidly in the nineties due to the popularity of a large number of portable digital devices such as laptop and palmtop computers, and the common availability of wireless communication devices. The rising popularity of the Internet added to the interest to develop internetworking protocols for mobile ad hoc networks operating in license-free radio frequency bands (such as the Industrial-Scientific-Military or ISM bands in the U.S.). In an interest to develop IP based protocols for ad hoc

networking, a working group for Mobile Ad Hoc Networking (MANET) was formed within the Internet Engineering Task Force (IETF) [20]. The DoD also renewed their support on similar research objectives by starting the GloMo (Global Mobile Information Systems) and the NTDR (Near-term Digital Radio) projects. Spurred by the growing interest in ad hoc networking, a number of commercial standards were developed in the late nineties. This includes the IEEE 802.11 physical layer and MAC protocol in 1995 [10], which has since then evolved into the more updated versions. Today, one can build an ad hoc network by simply plugging in 802.11 PCMCIA cards into laptop computers. Bluetooth [13] and Hiperlan [53] are some other examples of related existing products. In this chapter we discuss some the key challenges, protocols, and future directions of mobile ad hoc networks.

II. PHYSICAL LAYER AND MAC

The main aspects of designing the physical transmission system are dependent on the characteristics of the radio propagation channel such as path loss, interference (co-channel), and fading. In addition, since mobile terminals usually have limited power resources, the transceiver must be power efficient. These aspects are taken into account while designing the modulation, coding, and power control features in the radio equipment. In principle, the radio equipment in the nodes forming a mobile ad hoc network can use any technology as long as it provides reliable links between neighboring mobile terminals on a common channel. Candidate physical layers that have gained prominence are infrared and spread spectrum radio.

The MAC plays the key role in determining the channel usage efficiency by resolving contention amongst a number of unsupervised terminals sharing the common channel. An efficient MAC protocol would allow the transmissions from independent nodes to be separated in time and space, thereby maximizing the probability of successful transmissions and maintaining fairness amongst all users. Though research on medium access schemes for wired local area networks (LANs) have been done for many years, the same concepts cannot be directly applied to wireless LANs. In a wired medium, a transmitted signal is received with the same signal strength at all terminals connected to the same shared medium. Hence a terminal in a LAN can avoid contention by sensing the presence of a carrier to determine if any other terminal is using the channel before it starts a transmission. This "listen before transmit" principle has led to a class of efficient random access protocols for wired LANs that are generally known as carrier sense multiple access (CSMA) schemes [28]. A popular example is CSMA/CD (CSMA with collision detection), which is the standard for Ethernet (IEEE 802.3) LANs [46].

However, designing MAC protocols for wireless networks faces a different set of challenges. Propagation path losses in the wireless channel cause the signal power to decline with distance. This introduces the following problems, which are the main factors that affect the efficiency of the MAC in a mobile ad hoc network:

- **Carrier sensing is location-dependent:** Since the strength of the received signal depends on the distance from the transmitter, the same signal is not heard equally well by all terminals. Hence carrier sensing is not very effective in wireless. Typical problems of using carrier sensing to determine the availability of the wireless channel are:
  - **The hidden terminal problem:** a node may be hidden or out of range from a sender but

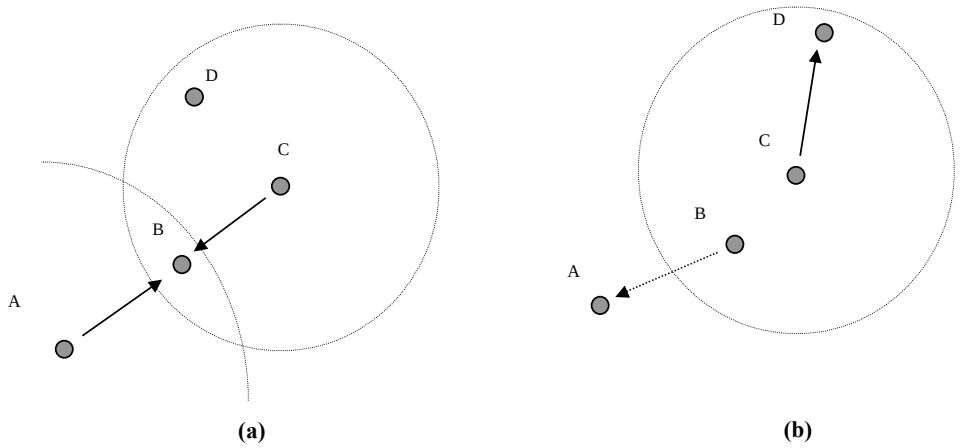

Fig. 2. Illistration of the hidden terminal problem (a) and the exposed terminal problem (b). The dotted lines represent the radio ranges.

within range of its intended receiver. For instance, in Figure 2(a), node **C** is out of range from **A**, and hence any transmission from **C** cannot be heard by **A**. So while **C** is transmitting to **B**, node **A** thinks that the channel is idle and simultaneously transmits a data packet to node **B**. This causes both packets to be lost at **B** because of interference, and the packets are considered to have suffered a "collision". A transmission from **A** to **B** will face the same consequence even if **C** is transmitting to some other node, such as **D**.

- **The exposed terminal problem:** this is the reverse problem, where a transmitting or "exposed" node is within range of a sender, but is out of range of the intended destination. The problem is illustrated in Figure 2(b), where node **B**, which wants to transmit a data packet to **A**, finds the channel to be busy due to the transmission from **C** to **D**. Hence, **B** might wait for the transmission from **C** to be over before transmitting to **A**, which is not necessary as the transmission from **C** would not interfere at **A**.

Both the hidden terminal and the exposed terminal problems arise due to the fact that carrier sensing is only performed at the transmitter, whereas its effect is determined by the interference power at the receiver, which are usually different due to propagation path loss characteristics.

- **Collision detection is not possible:** A wireless transceiver cannot transmit and receive at the same time as the transmitted signal will always be far stronger than any received signal. Hence a wireless terminal cannot detect if its transmission has been successful. To inform the transmitting node about a successful packet transmission, the receiver sends an ACKNOWLEDGEMENT (ACK) packet back to the transmitter after it receives a data packet. If the transmitter does not receive an ACK within a fixed period of time, it assumes that the transmitted packet has been lost. However, this is learnt only after completing transmission of the data packet and waiting for a further no-ACK timeout period.

Many different schemes have been designed for reducing these problems in wireless channel access. We

first present the IEEE 802.11 standard that is the most popular scheme for wireless LANs. Following that, a discussion on additional issues on the design of MAC protocols and current research directions are described.

A. IEEE 802.11

The IEEE 802.11 [10] is an international standard of physical and MAC layer specifications for WLANs. It provides mandatory support for 1 Mb/s data rate with optional support for 2 Mb/s. These original specifications have been upgraded to higher data rates in succeeding versions, with the projected goal of going up to 54 Mb/s for future systems. The standard can be applied to both infrastructure-based WLANs (that use fixed access points for wireless communication with mobile terminals) as well as infrastructureless ad hoc networks. In the following, we discuss the main features of this standard with relation to ad hoc networks.

1) 802.11 Physical Layer: IEEE 802.11 supports three different physical layers in order to allow designers to match price and performance to applications: one layer based on infrared and two layers based on radio transmission in the 2.4 GHz ISM band, an unlicensed band of radio frequencies available worldwide. The infrared specification is designed for indoor use only using line-of-sight and reflected transmissions of lightwaves in wavelengths from 850 to 950 nm. Both of the two RF specifications are based on spread spectrum, but employ different principles. While one uses frequency hopping (FH), the other is based on direct sequence (DS) spread spectrum. Either one can be used for the physical transmission system in ad hoc networks.

Frequency hopping spread spectrum: As the name implies, a frequency hopping spread spectrum radio hops from one carrier frequency to another during transmission. The transmission at any carrier frequency is narrow band, however frequency spreading is achieved by hopping from one carrier to another over a wide frequency band. The transmitter and receiver use the same sequence of carrier frequencies, which is pseudorandom (i.e. a long random sequence that repeats itself). The time for which the FH radio dwells in each frequency depends on the application requirements, government regulations, and adherence to standards. A *slow* FH system has a dwelling time that is longer than a bit period, whereas a *fast* FH system hops over many carrier frequencies during a single bit period. Since the hopping pattern is random, a FH system may experience interference during a few of the hops but achieve error-free transmission on other hops. One of the advantages of this property is that there is no hard limit on the total number of users that can be accommodated in a particular FH system. Rather, the limitation is decided by the amount of errors caused by multi-user interference that the users are willing to tolerate (known as the soft capacity). Such systems are especially beneficial in interference limited communication systems, where the transmission capability is constrained by a large number of contending users who are not all active at the same time.

The 2.4 GHz ISM band in the U.S. (i.e. 2.4000 - 2.4835 GHz) has 79 channel frequencies in the hopping set, with a channel spacing of 1 MHz. The specified channel spacing allows 1 Mb/s transmission rate using two-level Gaussian frequency shift keying (GFSK), which is the modulation scheme specified by the 802.11 standard. To achieve 2 Mb/s transmission rate, four-level GFSK modulation may be used, where 2 bits are encoded at a time using four frequencies. There are three different hopping sequence

sets in the U.S., with 26 hopping sequences in each set. All the terminals in any given ad hoc network must use the same hopping sequence. However, the availability of multiple sets allow multiple systems or networks to coexist in the same location.

Direct sequence spread spectrum: The DS system achieves frequency spreading by multiplying each data bit by a sequence of chips ($+1/-1$ symbols that are shorter than a bit) before modulation. This has the effect of artificially increasing the transmission bandwidth. The receiver uses the same chip-sequence to correlate the received signal. This technique achieves excellent interference rejection due to auto and cross correlation properties of the random chip sequences. Usually the chip sequences are pseudorandom sequences having a long period. Multiple pairs of transmitters and receivers using different chip sequences can co-exist in the same region. A DS system also has a soft capacity and can coexist with other narrow band radio systems without causing significant interference.

The IEEE 802.11 standard specifies an 11-chip Barker sequence for spreading each data bit. The modulation scheme is differential binary phase shift keying (DBPSK) for 1 Mb/s data rate, and differential quadrature phase shift keying (DQPSK) for 2 Mb/s. This effectively spreads the data stream over a 11 MHz band. Multiple systems can use different bands of frequencies whose center frequencies are separated by at least 30 MHz. As usual, all terminals of the same ad hoc network must use the same chip sequence (spreading code) for transmission as well as reception.

2) 802.11 MAC: The 802.11 MAC is designed to provide mandatory asynchronous data service along with an optional time-bounded service that is only usable in an infrastructured wireless networks with access points. The asynchronous data service is usable by both ad hoc networks and infrastructured wireless networks and supports "best effort" packet exchange without delay bounds.

The mandatory basic asynchronous service is provided by a method known as *carrier sense multiple access with collision avoidance (CSMA/CA)* and an optional channel reservation scheme based on a four-way handshake between the sender and receiver nodes. These two methods provides the mechanism for achieving distributed coordination amongst uncoordinated wireless terminals that do not use a fixed access point, and are known as the **distributed coordination function (DCF)**. A third method, known as the **point coordination function (PCF)**, offers both asynchronous and time-bounded service, but needs an access point to control medium access to avoid contention.

Basic DCF using CSMA/CA: The basic channel access scheme uses two fundamental ideas to avoid collisions amongst contending transmitting stations:

- *Carrier sensing:* to determine that the medium is not being used by a neighboring transmitter (channel idle) before accessing the channel
- *Random backoff:* a terminal that senses the channel to be busy, waits for a random period of time for which it has to see the channel in the idle state before initiating transmission.

A terminal that intends to transmit and senses the presence of a carrier (channel busy) waits till the end of the current transmission and considers the channel to be idle only when it detects the absence of the carrier for certain duration of time, known as the DCF INTER-FRAME SPACE (DIFS). At the end of the DIFS period, in order to avoid collision with other terminals that might also be waiting for the current transmission to end before transmitting their packets, the terminal does not access the channel

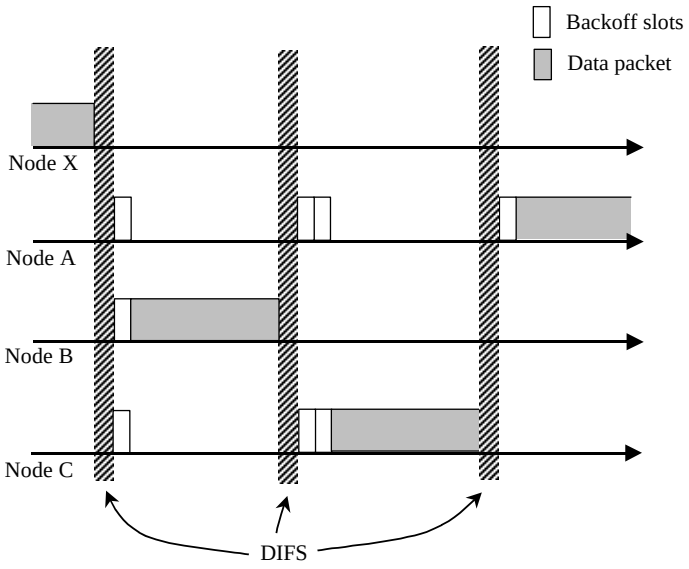

Fig. 3. Illistration of the CSMA/CA protocol.

immediately. Instead, each terminal starts a backoff timer, which is initiated at a random value and counts down as long as the channel is sensed idle. The backoff timer is frozen whenever the channel is sensed busy, resuming the countdown again after it goes idle (i.e. senses absence of the carrier for at least DIFS period). The terminal initiates transmission only when its backoff timer reaches zero. The backoff interval is slotted, and may be expressed as $CW_{rand} \times$ SLOT TIME, where CW_{rand} is a random integer chosen uniformly between 0 and CW and SLOT TIME is a predetermined slot duration. CW is the contention window, which can take one of the following set of integer values: 7, 15, 31, 63, 127, 255. Initially a node uses the smallest value of CW and uses the next higher value in the set after each unsuccessful transmission.

In order to indicate that a transmission has been successful, a receiver transmits an ACK packet after a SHORT INTER-FRAME SPACE (SIFS) period (which is shorter than DIFS) immediately following the reception of the data packet. In case an ACK is not received, the transmitter assumes that the transmitted data packet is lost and it schedules a retransmission of the same. This will be continued for a maximum number of allowable retries at the MAC before the data packet is discarded.

An illustration of the access control scheme is shown in Figure 3. Here, nodes **A**, **B**, and **C**, all have data packets to transmit when they find the channel busy (due to a transmission from some node **X**). After the channel is idle for a DIFS period, each node selects a random backoff period. In this illustration, the backoff timers of **A**, **B**, and **C** are chosen as 4, 1, and 3, respectively. So **B**'s backoff timer reaches zero first, when it initiates transmission and the timers of both **A** and **C** are frozen. The transmissions of the data frames from **A** and **C** take place subsequently, as shown in Figure 3.

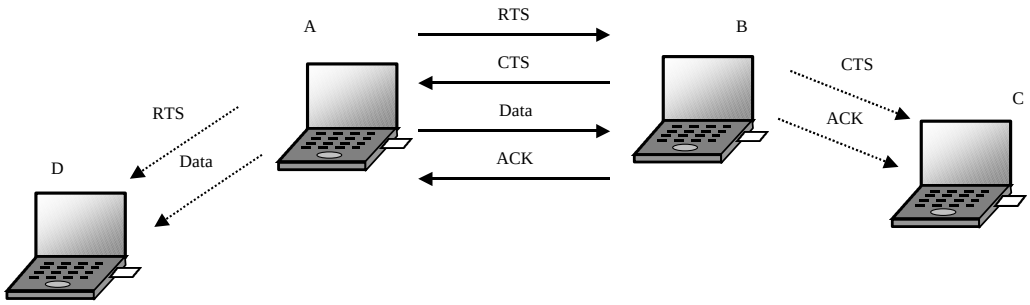

Fig. 4. CSMA/CA with RTS/CTS handshake.

According to this scheme, a collision may happen when multiple stations select the same backoff time. A large value CW will ensure a small probability of collision as it results in a smaller probability of two nodes selecting the same backoff time. However, a larger CW may cause a node to wait longer before transmission. When very few nodes are transmitting, a large value of CW causes inefficient usage of the channel. Hence, initially all nodes set the CW to the smallest value of 7. With heavier traffic, some of the transmissions will collide and eventually higher and higher values of CW may be chosen by the nodes to ensure collision free transmission.

CSMA/CA with RTS/CTS extension: Though the basic CSMA/CA scheme has excellent mechanisms to avoid collisions amongst a number of uncoordinated nodes that can hear one another, it does not solve problems due to hidden and exposed terminals. In order to address the hidden terminal problem, 802.11 has the option of adding the mechanism of an exchange of REQUEST TO SEND (RTS) and CLEAR TO SEND (CTS) control packets between a transmitting and receiving nodes before initiating the transmission of a data packet.

The principle behind the use of the RTS and CTS packets can be seen from Figure 4. Here, node **A**, which intends to send a data packet to **B**, first broadcasts an RTS packet using the basic CSMA/CA scheme. The RTS frame contains the identity of the destination **B**, and the time that would be required for the entire transmission to complete. If **B** receives the RTS packet, it replies with a CTS packet after waiting for SIFS. The CTS packet also contains the time required for completion of the intended data exchange and the identity of the transmitting node. Upon receiving the CTS packet from **B**, **A** waits for SIFS and then transmits the data packet. When the data packet is received, **B** sends an ACK packet after SIFS, thus completing one entire data packet transfer protocol. All neighbors of **A** and **B** that receive either the RTS and/or the CTS learn about the intended exchange process and cooperate by remaining silent for the period of time that is required for the data exchange to be over.

The exchange of RTS and CTS packets serves two purposes:

- First, if **A** receives the CTS, it is ascertained that **B** is ready to receive and there are no interfering transmissions near node **B**. This process thus serves as a "virtual carrier sensing" mechanism.
- Second, all neighboring nodes of the destination, including those hidden from **A** (such as **C**), are

expected to hear the CTS packet and remain silent for as long as it is required for the data transmission to be over. Neighbors of **A** (such as **D**) also remain silent for the period specified by the RTS so that their own transmissions do not experience any interference from the data packet to be transmitted from **A**. A silent period is implemented in a listening node by setting its NET ALLOCATION VECTOR (NAV) in accordance to the duration field in the RTS or CTS, which specifies the earliest possible time at which it can access the channel again. Hence, the RTS/CTS exchange effectively *reserves* the channel for the intended data transmission from **A** to **B**.

Even though the data packets have higher probability of success due to this channel reservation technique enacted by the RTS/CTS exchange, the RTS and CTS packets themselves are susceptible to the same rate of failure as that of the basic CSMA/CA scheme. Many of these control packets may suffer loss due to collisions and require retransmissions before the channel reservation is performed successfully. However, since the RTS and CTS control packets are shorter than the data packets, the scheme usually has a better throughput performance than the basic CSMA/CA in the presence of hidden terminals. Comprehensive analysis of the performance of the DCF under various conditions in mobile ad hoc networks have been reported in [7], [58], [5].

B. ADDITIONAL ISSUES ON MAC

Several concerns with the IEEE 802.11 MAC has motivated researchers to explore newer techniques to improve the channel utilization and throughput in mobile ad hoc networks. The basic access method of the 802.11 MAC protocol is susceptible to inefficiencies due to the hidden and exposed terminal problems. The RTS/CTS option reduces the hidden terminal problem but not the inefficiency caused by the exposed terminal problem. Some other concerns of 802.11 DCF using the RTS/CTS dialog are:

Additional overhead of control packets: The transmission RTS and CTS control packets consume an additional amount of bandwidth and may lead to significant delays in data transmission if they experience a high amount of collisions and retransmissions. Usually this is justified when the size of the data packets is large and the advantage gained from saving the loss of data packets far outweighs the additional overhead incurred by the transmission of RTS and CTS packets. However, it has been observed that especially in higher loads and under high mobility, most of the transmission bandwidth may be consumed for repeated transmissions of the RTS and CTS control packets with very little data traffic being possible [21].

Collisions of control packets: Since the RTS and CTS packets are susceptible to collisions, the channel reservation scheme may fail leading to loss of data packets as well. Figure 5 illustrates such an example scenario. Here **A** starts an RTS-CTS dialog with **B** before transmitting a data packet to it. The CTS reply from **B** is received by **A** correctly, but it is not received by **C**, which is hidden from **A**, due to a collision with an RTS packet sent from **D** to **E**. This happens because **D**, being far away from both **A** and **B**, does not hear either the RTS or the CTS packet and is unaware of the communication between **A** and **B**. Node **A** assumes that the channel is successfully reserved and proceeds with transmission of the data packet to **B**. This data transmission is vulnerable to interference from **C**, which has not been able to set its NAV accordingly, and may initiate a transmission to any of its neighbors before the data transmission is over.

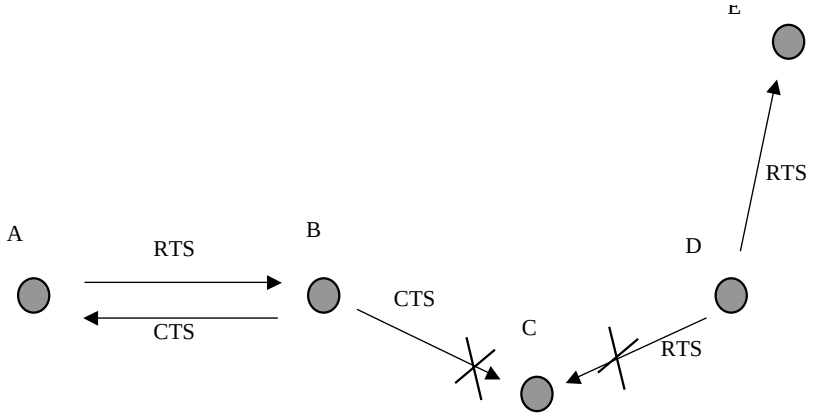

Fig. 5. Example where the node **C** that is "hidden" from **A** misses the CTS packet from **B** due to a collision with an RTS packet from **D**.

Problems such as these are common because the RTS and CTS packets themselves are sent using the basic CSMA/CA access method which is prone to the hidden and exposed terminal problems. A technique described in [17] tries to resolve this problem by making the duration of the CTS *longer* than the RTS packets. This ensures that in the event that an RTS packet collides with a CTS at a receiver (such as in **C** in Figure 5), it would still be able to detect a part of the CTS packet. This might allow it to set its NAV to avoid interfering with the data exchange.

Radio Interference: Since wireless transmission is mostly limited by interference rather than noise, it is important to study the nature of interference and its effect on packet success probability. Figure 6 depicts the strengths of signals that would be received at node **B** from nodes **A** and **C** located at distances d_1 and d_2, respectively. Assuming that both transmitters use the same power Pt, the corresponding signal powers received at **B**, represented by Pr_A and Pr_C, respectively, depend on the path loss characteristics and the corresponding path lengths. The probability of error at the receiver depends on the total *signal-to-interference-plus-noise ratio (SINR)* of the corresponding packet at the receiver. The receiver noise is usually a constant parameter. The interference power is calculated by adding the powers of all radio signals at the receiver other than the power of the packet in question. The probability of bit error and consequently the packet error probability increases with decreasing values of the SINR. The minimum SINR required to correctly receive a packet depends on the radio technology, such as modulation, demodulation, coding, etc. A given radio usually has a specified *minimum SINR threshold* $SINR_{min}$ for correctly receiving a packet. For instance, **B** will be able to receive the packet from **A** correctly if

$$\frac{Pr_A}{Pr_C + N} > SINR_{min}$$

where N is the receiver noise at **B**.

For a given transmitter and corresponding receiver, the *transmission range* is defined as the maximum distance at which the received SINR is equal to $SINR_{min}$ in the absence of any interference, i.e. the maximum distance at which reception will be error-free without interference. Usually radio channels

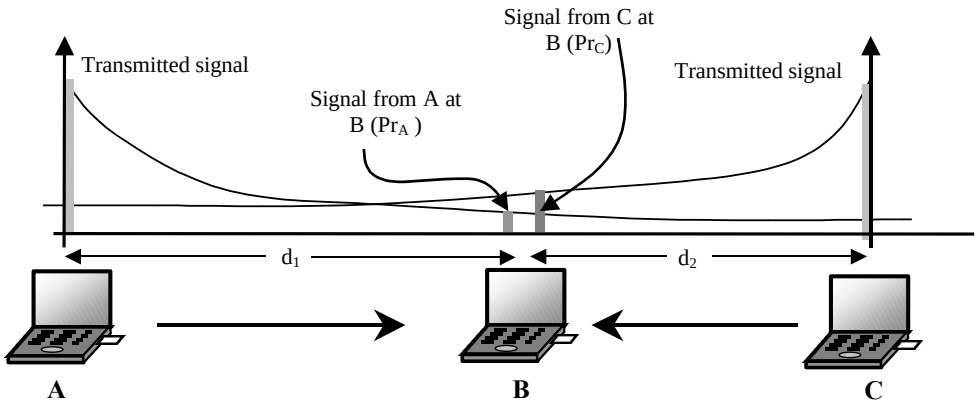

Fig. 6. Propagation path loss in the wireless channel.

are bidirectional, and hence a receiver will be able to receive a packet from a transmitter that is located within the transmission range (alternatively called the radio range).

A node determines the busy/idle state of channel by comparing the strength of the carrier power to a predetermined *carrier-sense threshold* T_{CS}. Typically, this threshold is chosen such that the carrier sensing range, i.e. the distance within which all transmissions are detected, is at least as much as the transmission range of the nodes. A lower value of T_{CS} increases the carrier sensing range, but it also reduces frequency reuse by making larger number of nodes wait for their transmissions around a given transmitting node.

It is important to note that correct packet reception is not guaranteed whenever the receiver is within the transmission range of a transmitter. It also depends on the total amount of other interfering signals present. Typically, the interference power from a transmitter that is located at a distance less than the transmission range is expected to preclude the reception of any other packet without errors. Hence two simultaneous transmissions from nodes that are within range of a receiving node are said to have met with a "collision". The term "collision" has been borrowed from wireline networks, where any two simultaneously transmitted packets are lost irrespective of the location of the transmitters. In wireless networks, it relates to packet loss due to interference.

In wireless networks, packets may be lost due to interference from even those transmitters that are located outside the radio range of a receiver. An example is shown in Figure 7, where the combined interference from several transmitting nodes, all of which are out of range from node **B**, disrupts the reception of the packet from **A** to **B**.

Capture: Another concept used in wireless packet networks is packet capture, which refers to the mechanism where a receiver can receive one of two simultaneously arriving packets if their received powers allow it [31]. For instance, in Figure 6, even if both **A** and **C** are transmitting at the same time, **B** can receive the packet from **A** as long as its power exceeds that from **C** by a sufficient margin. This is especially beneficial to the network performance under heavy traffic conditions when there are a large

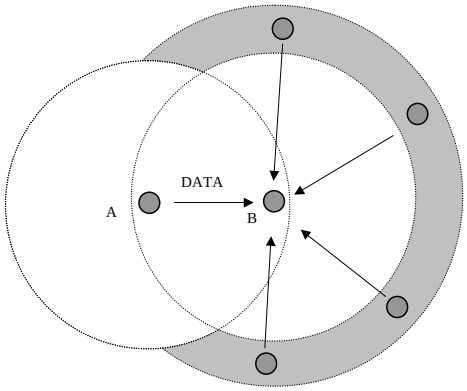

Fig. 7. Illustration of packet loss due to combined interference from transmissions outside the range of a receiver.

number of packet collisions, and some of the collided packets are received successfully. A possible negative effect of the capture phenomena is that it can lead to unfair sharing of the channel. This can be seen in Figure 6 where transmitted packets from **A** will never be successful as long as **C** is transmitting, whereas the packets sent from **C** will always be captured at **B**.

1) Other MAC Protocols: Several solutions to these known problems have been suggested by various researchers. In the following, some of the notable concepts for new MAC protocols are summarized.

Collision avoidance techniques: The principle cause for packet loss in ad hoc networks is due to collisions, or interference caused by transmissions from hidden terminals. Several MAC protocols have been suggested that have features to avoid such collisions. One such technique is the transmission of a *busy tone* to indicate an ongoing data exchange process, which was first suggested in [55]. Here, any node that hears an ongoing data transmission emits an out-of-band tone. A node hearing the busy tone will refrain from transmission, thereby increasing the distance of carrier sensing by a factor of 2. Two other MAC protocols, the *Dual Busy Tone Multiple Access* [9] and the *Receiver Initiated Busy Tone Protocol* [59] also use this concept to avoid collisions. These schemes require additional complexity of narrowband tone detection and the use of separate channels.

Channel reservation techniques: The *Multiple Access with Collision Avoidance* (MACA) uses channel reservation based on the exchange of RTS and CTS control packets before transmission of the data packet [27]. This scheme was incorporated in the IEEE 802.11 standard with the addition of a positive acknowledgement packet to indicate successful packet reception. Later, other protocols such as MACAW [3], FAMA [17], and CARMA [18] also adopted the reservation scheme employing different variations of control packets.

Multiple channel MAC: Even though all nodes in an ad hoc network are required to be able to share the same medium, the concept of dividing the common medium into multiple orthogonal channels has been explored to reduce contention [38], [35], [24], [57]. When multiple channels are available, several concurrent transmissions are possible in the same neighborhood between distinct pairs of senders and

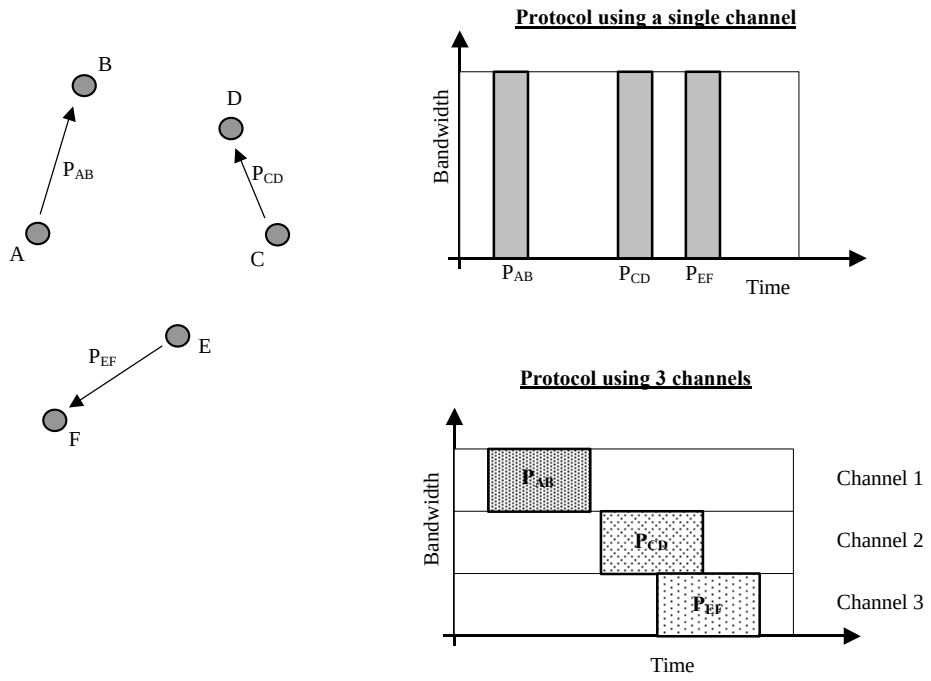

Fig. 8. Illustration of single-channel and multi-channel MAC protocols for concurrent transmissions of 3 packets: P_{AB} from **A** to **B**, P_{CD} from **C** to **D**, and P_{EF} from **E** to **F**.

receivers (Figure 8). If the same bandwidth is divided into N channels, either by frequency division or by using orthogonal CDMA codes, the traffic can be distributed over N channels. However, the transmission rate in each channel will also drop by a factor of N. It has been shown that such multi-channel schemes can achieve a higher throughput by using an appropriate channel selection algorithm which allows each node to select *the best* channel available in its neighborhood. Several schemes for channel selection based on the exchange of RTS and CTS packets and carrier sensing over all channels have been explored [38], [35], [24], [57]. The use of multiple channels increases the hardware complexity, but it improves the throughput performance in the network by distributing the traffic over time as well as over bandwidth.

Use of directional antennas: Traditional ad hoc networks use omnidirectional antennas, as the direction for transmission and reception is variable. However, use of directional transmission provides several benefits for improving the link performance between a pair of communicating nodes:

- Firstly, a directional transmission can reduce the amount of interference to neighboring nodes. This can lead to a higher amount of frequency reuse and packet success probability.
- Secondly, a directional antenna can be used for receiving from a desired direction, reducing the amount of interference at the receiving node from adjacent transmitters. This further reduces the packet error probability.
- Thirdly, directional antennas have a higher gain due to their directivity. This can allow the trans-

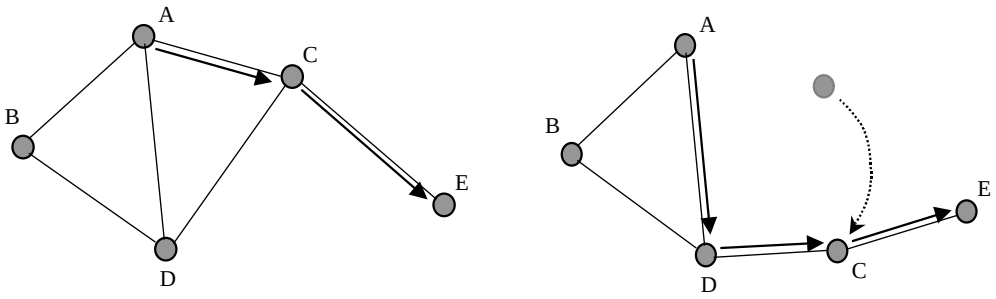

Fig. 9. Illustration of the change in the route from **A** to **E** due to the movement of node **C**.

mitters to operate at a smaller transmission power and still maintain adequate SINR at the receiver. It will also reduce the average power consumption in the nodes [36].

Despite these advantages, the usage of directional antennas in mobile ad hoc networks has additional design challenges. A mechanism for determining the direction for transmission and reception is required so that the mobile nodes can use directional antennas. Moreover, since all ad hoc networking protocols are traditionally designed for omnidirectional antennas, these protocols need to be adapted appropriately for proper functioning and maximizing the advantages that can be derived from directional transmissions and receptions. Many MAC and routing protocols that utilize directional antennas in ad hoc networks have been proposed in recent years [30], [37], [36]. A comprehensive discussion on the various aspects of using directional antennas in ad hoc networks is given in [48]. A central issue that concerns the applicability of directional antennas in mobile ad hoc networks is the comparatively larger size and cost of beamforming antennas that are ideal for such applications. With advancements in technology and the possibility of shifting towards higher frequency bands (such as the 5.8 GHz ISM band), it may be possible to design smaller as well as cheaper directional antennas. Hence, there is a growing interest towards utilizing directional antennas in ad hoc networks.

III. Routing in Ad Hoc Networks

Movements of nodes in a mobile ad hoc network cause the nodes to move in and out of range from one another. As the result, there is a continuous making and breaking of links in the network, making the network connectivity (topology) to vary dynamically with time. Since the network relies on multihop transmissions for communication, this imposes major challenges for the network layer to determine the multihop route over which data packets can be transmitted between a given pair of source and destination nodes. Figure 9 demonstrates how the movement of a single node (**C**) changes the network topology rendering the existing route between **A** and **E** (i.e. **A**–**C**–**E**) unusable. The network needs to evaluate the changes in the topology caused by this movement and establish a new route from **A** to **E** (such as **A**–**D**–**C**–**E**).

Because of this time-varying nature of the topology of mobile ad hoc networks, traditional routing

techniques, such as the shortest-path and link-state protocols that are used in fixed networks, cannot be directly applied to ad hoc networks. A fundamental quality of routing protocols for ad hoc networks is that they must *dynamically* adapt to variations of the network topology. This is implemented by devising techniques for efficiently tracking changes in the network topology and rediscovering new routes when older ones are broken. Since an ad hoc network is infrastructureless, these operations are to be performed in a *distributed* fashion with the collective cooperation of all nodes in the network. Some of the desirable qualities of dynamic routing protocols for ad hoc networks are:

- **Routing overhead:** Tracking changes of the network topology requires exchange of control packets amongst the mobile nodes. These control packets must carry various types of information, such as node identities, neighbor lists, distance metrics, etc., which consume additional bandwidth for transmission. Since wireless channel bandwidth is at a premium, it is desirable that the routing protocol minimizes the number and size of control packets for tracking the variations of the network.
- **Timeliness:** Since link breakages occur at random times, it is hard to predict when an existing route will expire. The timeliness of adaptation of the routing protocol is crucial. A broken route causes interruption in an ongoing communication until a new route is established. Often the newly rediscovered route may be largely disjoint from the older route, which creates problems in rerouting the packets that were already transferred along the route and could not be delivered to the destination. Ideally, a new route should be determined before the existing one is broken, which may not be possible. Alternatively, a new route should be established with minimum delay.
- **Path optimality:** With constraints on the routing overhead, routing protocols for mobile ad hoc networks are more concerned with avoiding interruptions of communication between source and destination nodes rather than the optimality of the routes. Hence, in order to avoid excess transmission of control packets, the network may be allowed to operate with suboptimal (which are not necessarily the shortest) routes until they break. However, a good routing protocol should minimize overhead as well as the path lengths. Otherwise, it will lead to excessive transmission delays and wastage of power.
- **Loop freedom:** Since the routes are maintained in a distributed fashion, the possibility of loops within a route is a serious concern. The routing protocol must incorporate special features so that the routes remain free of loops.
- **Storage complexity:** Another problem of distributed routing architectures is the amount of storage space utilized for routing. Ad hoc networks may be applied to small portable devices, such as sensors, which have severe constraints in memory and hardware. Hence, it is desirable that the routing protocol be designed to require low storage complexity.
- **Scalability:** Routing protocols should be able to function efficiently even if the size of the network becomes large. This is not very easy to achieve, as determining an unknown route between a pair of mobile nodes becomes more costly in terms of the required time, number of operations, and expended bandwidth when the number of nodes increases.

Because of its many challenges, routing has been a primary focus of researchers in mobile ad hoc networks. The MANET working group in the IETF has been working on the issue of standardizing an IP based routing standard for mobile ad hoc networks. Consequently, a large number of dynamic

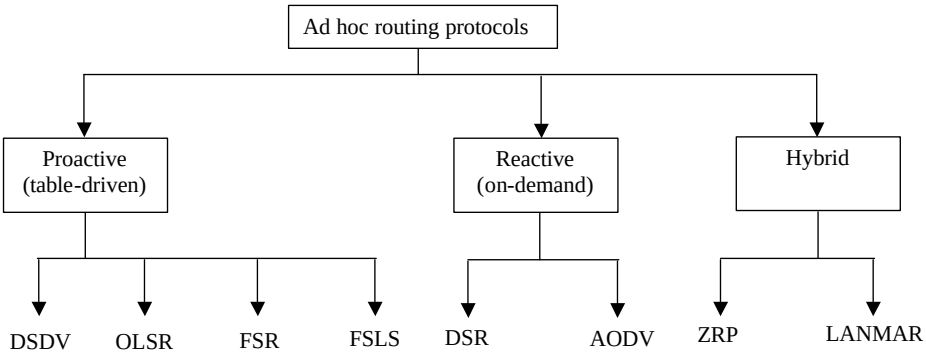

Fig. 10. Classification and examples of ad hoc routing protocols.

routing protocols applicable to mobile ad hoc networks have been developed. Reviews of prominent routing protocols for mobile ad hoc networks may be found in [4], [8], [50], [23].

Based on when routing activities are initiated, routing protocols for mobile ad hoc networks may be broadly classified into three basic categories: (a) *proactive* or *table-driven* protocols, (b) *reactive* or *on-demand* routing protocols, and (c) *hybrid* routing protocols. Some representative examples of each class are shown in Figure 10.

A. PROACTIVE ROUTING PROTOCOLS

Proactive protocols perform routing operations between all source destination pairs periodically, irrespective of the need of such routes. These protocols stem from conventional link state or distance vector routing algorithms, and attempt to maintain shortest-path routes by using periodically updated views of the network topology. These are typically maintained in routing tables in each node and updated with the acquisition of new information. Proactive protocols have advantages of providing lower latency in data delivery and the possibility of supporting applications that have quality-of-service constraints. Their main disadvantage is due to the wastage of bandwidth in sending update packets periodically even when they are not necessary, such as when there are no link breakages, or when only a few routes are needed.

Destination-Sequenced Distance-Vector Routing (DSDV): DSDV [44] is based on the classical Bellman-Ford algorithm [2] with adaptations that are specifically targeted for mobile networks. The Bellman-Ford algorithm uses the distance vector approach, where every node maintains a routing table that records the "next hop" for every reachable destination along the shortest route, and the minimum distance (number of hops). Whenever there is any change in this minimum distance, the information is reported to neighboring nodes and the tables are updated if required.

To make this algorithm adequate for mobile ad hoc networks, DSDV added a *sequence number* with each distance entry to indicate the *freshness* of that entry. A sequence number is originated at the destination node, and is incremented by each node that sends an update to its neighbors. Thus, a newer routing table update for the same destination will have a higher sequence number. Routing table updates are periodically transmitted throughout the network, with each node updating its routing table entries based on the latest sequence number corresponding to that entry. If two updates for the same destination have identical sequence numbers but different distances, then the shorter distance is recorded. The addition of sequence numbers remove the possibility of long lived loops and also the *counting-to-infinity* problem, where it takes a large number of update messages to ascertain that a node is not reachable [44]

Optimized Link State Routing Protocol (OLSR): OLSR is a comparatively newer proactive routing protocol [15]. It is an adaptation of conventional link-state routing in which each node tries to maintain information about the network topology. Each node determines the link costs to each of its neighbors by broadcasting HELLO messages periodically. Whenever there is a change in the link costs, the node broadcasts this information to all other nodes. In classical link-state algorithms, this is done by each node *flooding* the whole network with update packets containing updated link costs. Nodes use this information to apply a shortest path algorithm (such as Dijkstra's shortest path algorithm [11]) to determine the best route to a specific destination.

OLSR optimizes the link-state protocol in two ways. First, it reduces the size of the update packets sent during the broadcasts by including only a subset of links to its neighbors. These are the links to a select set of neighbors known as the *multipoint relays (MPR)*. The set of MPRs of a node consist of the minimum set of one hop neighbors of that node so that the node can reach all of its two hop neighbors by using these nodes as relay points. Each node computes its MPR set from the exchange of neighborhood information with all its neighbors. Second, instead of every neighbor broadcasting the update packets sent out by a node, only the MPR nodes participate in broadcasting of these packets in OLSR. This minimizes the traffic of control packets during flooding. However, the savings of bandwidth achieved using these two techniques come at a cost of propagating incomplete topology information in the network. The updates include only MPR sets and not the sets of all neighbors of the broadcasting nodes. Hence, a shortest path algorithm based on this partial topology information will generate routes containing the MPR nodes only. When the network is dense, i.e. when each node has many neighbors, OLSR will work out to be efficient due to the reduction of control traffic for updates in the network.

Issues in proactive routing: The key characteristic of proactive routing protocols is that updates are sent periodically irrespective of need. Another issue is that they are table-driven. These two properties cause serious problems for making proactive routing protocols scale with network size. However, these protocols work well under heavy traffic and high mobility conditions as they try to maintain fresh routing information continuously.

Several new approaches have been proposed to make proactive protocols more scalable. One example is *Fisheye State Routing (FSR)* [41], which is also an adaptation of link-state routing to ad hoc networks. FSR tries to limit routing load by avoiding flooding the network with routing information. Entire link state information is only transmitted to the first hop neighbors. In addition, it uses lower update rates for nodes that are located further away. Hence, FSR maintains accurate route information on nodes

that are close by but the accuracy degrades with increasing distance of the destination from the source. Overall, this technique saves the volume and size of routing traffic. A similar approach is adopted in the *Fuzzy Sighted Link State algorithm (FSLS)* [51]. As discussed above, OLSR reduces routing load by broadcasting incomplete topology information. In general, these sacrifices lead to increased scalability of proactive routing protocols.

B. REACTIVE ROUTING PROTOCOLS

Reactive protocols are designed to minimize routing overhead. Instead of tracking the changes in the network topology to continuously maintain shortest path routes to all destinations, these protocols determine routes only when necessary. Typically, these protocols perform a *route discovery* operation between the source and the desired destination when the source needs to send a data packet and the route to the destination is not known. As long as a route is live, reactive routing protocols only perform *route maintenance* operations and resorts to a new route discovery only when the existing one breaks. The advantage of this *on-demand* nature of operation is that it usually has a much lower average routing overhead in comparison to proactive protocols. However, it has the disadvantage that a route discovery may involve *flooding* the entire network with query packets. Flooding is wasteful, which can be required quite frequently in case of high mobility or when there are a large number of active source-destination pairs. Moreover, route discovery adds to the latency in packet delivery as the source has to wait till the route is determined before it can transmit. Despite these drawbacks, on-demand protocols receive comparatively more attention than proactive routing protocols, as the bandwidth advantage makes them more scalable.

Dynamic Source Routing (DSR): DSR is a reactive routing protocol that uses a concept called *source routing* [25]. Each node maintains a *route cache* where it lists the complete routes to all destinations for which the routes are known. A source node includes the route to be followed by a data packet in its header. Routes are discovered on demand by a process known as *route discovery*. When a node does not have a route cache entry for the destination to which it needs to send a data packet, it initiates a route discovery by broadcasting a route REQUEST or QUERY message seeking a route to the destination. The REQUEST packet contains the identities of the source and the desired destination. Any node that receives a REQUEST packet first checks its route cache for an existing entry to the desired destination. If it does not have such an entry, the node adds its identity to the header of the REQUEST packet and transmits it. Eventually, the REQUEST packet will flood the entire network by traversing to all the nodes tracing all possible paths. When a REQUEST packet reaches the destination, or a node that has a known route to the destination, a REPLY is sent back to the source following the same route that was traversed by that REQUEST packet in the reverse direction. This is done by simply copying the sequence of node identities obtained from the header of REQUEST packet. The REPLY packet contains the entire route to the destination, which is recorded in the source node's route cache.

When an existing route breaks, it is detected by the failure of forwarding data packets on the route. Such a failure is observed by the absence of the link layer acknowledgement expected by the node where the link failure has occurred. On detecting the link failure, the node sends this information back an ERROR packet to the source. All nodes that receive the ERROR packet, including the source, delete all existing

routes from their route caches that contain the specified link. If a route is still needed, a fresh route discovery is initiated.

Ad Hoc On Demand Distance Vector Routing (AODV): AODV [42] can be described as an on-demand extension of the DSDV routing protocol. Like DSDV, each route maintains routing tables containing the next hop and sequence numbers corresponding to each destination. However, the routes are created on demand, i.e. only when a route is needed for which there is no "fresh" record in the routing table. In order to facilitate determination of the freshness of routing information, AODV maintains the time since when an entry has been last utilized. A routing table entry is "expired" after a certain predetermined threshold of time.

The mechanism for creating routes in AODV is somewhat different from that used in DSR. Here, when a node needs a route to some destination, it broadcasts a route REQUEST packet in which it includes the last known sequence number for that destination. The REQUEST packet is forwarded by all nodes that do not have a fresher route (determined by the sequence numbers) to the specified destination. While forwarding the REQUEST packet, each node records the earlier hop taken by the REQUEST packet in its routing table entry for the source (originator of the route discovery). Hence, a propagating REQUEST packet creates *reverse routes* to the source in the routing tables of all forwarding nodes. When the REQUEST packet reaches the desired destination or a node that knows a fresher route to it, it generates a route REPLY packet that is sent back along the same path that was taken by the corresponding REQUEST packet. The REPLY packet contains the number of hops to the destination as well as the most recent sequence number. Each node that forwards the REPLY packet enters the routing information for the destination node in its routing table, thus creating the *forward route* to the destination.

Routing table entries are deleted when an ERROR packet is received from one of the intermediate nodes on the route forwarding a data packet to the destination. When such an ERROR packet reaches the source, it may initiate a fresh route discovery to determine a fresh route to the destination.

Issues in reactive routing: Since reactive routing protocols only transmit routing packets when needed, these protocols are comparatively more efficient when there are fewer link breakages, such as under low mobility conditions. In addition, when there are only a few communicating nodes in the network, the routing functions are only concerned with maintaining the routes that are active. Because of these benefits, reactive or on-demand routing protocols have received more attention than proactive protocols for mobile ad hoc networks.

The main concern with reactive routing protocols is the need for flooding the entire network in search of a route when needed. Many optimizations have been suggested to reduce the excessive number of routing packets transmitted throughout the network during such flooding operations in reactive protocols. For instance, DSR has the option of broadcasting a *nonpropagating request packet* for route discovery, which in that case is broken into two phases. In the first phase, the source broadcasts a nonpropagating route request packet that only queries its first hop neighbors for a known route to the destination. These packets are not forwarded by the neighbors. If none of the neighbors return a route, the source then proceed to the second phase where a traditional propagating request packet is sent. The advantage of this scheme is that it avoids a network-wide flood of request packets when the route to the destination is known by one of the first-hop neighbors. A similar scheme is implemented in AODV using the concept

of an *expanding ring search*. Here, increasingly larger neighborhoods, controlled by either hop-wise or time-wise constrained request packets, are searched to find the route to the destination. Some other techniques that perform similar optimizations are: *salvaging*, where an intermediate node in DSR uses an alternative route from its own cache when the original route is broken; and *promiscuous listening*, in which a node that overhears a packet not addressed to itself finds that it has a shorter route to the same destination, and sends a *gratuitous reply* to the source with this new route. This increases the freshness of the route cache entries without additional route discoveries.

C. HYBRID ROUTING PROTOCOLS

The use of *hybrid routing* is an approach that is often used to obtain a better balance between the adaptability to varying network conditions and the routing overhead. These protocols use a combination of reactive and proactive principles, each applied under different conditions, places, or regions. For instance, a hybrid routing protocol may benefit from dividing the network into clusters and applying proactive route updates within each cluster and reactive routing across different clusters. Routing schemes that employ proactive route maintenance on top of reactive route discoveries have also been considered.

Zone Routing Protocol (ZRP): ZRP [22] divides the network into *zones* or clusters of nodes. The nodes within each zone maintain routing information for one another using a proactive algorithm such as a distance vector or link state protocol. Hence, all nodes maintain updated routing tables consisting of routes to all other nodes within the same zone (known as *intra-zone routing*). Each zone also identifies a set of *peripheral nodes* that are located at the edges of the zone for communication with other zones. When a packet is to be sent to a node for which the source does not have an entry in its routing table, it is assumed that the destination is located in another zone. In that case, the node requests the peripheral nodes to send out a route request packet to all other zones in the networks. This is known as *inter-zone routing*, which uses a process that is similar to DSR except that the request packets are only handled by the peripheral nodes in the network. When the request packet reaches a peripheral node of the zone that contains the destination, a reply is sent back to the source. The overhead of flooding in such a route discovery is limited due to the involvement of peripheral nodes only. The proactive protocol in this hybrid framework limits the spread of periodic update packets within each zone. ZRP is especially suitable for large networks, however the flooding of request packets during interzone route discoveries may still be a cause of concern.

Landmark Ad Hoc Routing Protocol (LANMAR): LANMAR is designed for ad hoc networks which have the characteristics of group mobility, such as a group of soldiers moving together in a battlefield. Each group dynamically identifies a specific node within the group to be a *landmark* node. A proactive link state routing protocol is used to maintain routing information within the group and a distance vector algorithm is used to do the same amongst all landmark nodes. Hence, each node has detailed topology information for all nodes within the group and distance and routing vector information to all landmarks. No routing information is maintained for non-landmark nodes belonging to other groups. Packets to be sent to such a destination are forwarded towards the corresponding landmark. When the packet reaches the nodes within the group containing the destination, it is forwarded to the destination, possibly without going through its landmark. This scheme reduces the size of routing tables as well as the overhead of

routing traffic forming a two-level routing hierarchy. Hence, it is expected to be more scalable than the so called *flat* routing protocols.

D. Other Concepts in Ad Hoc Routing

There is an increasing list of new ideas and protocols for routing in mobile ad hoc networks. The MANET working group in the IETF publishes all significant developments and discussions by the group online in its mailing list [20], which is the most comprehensive source of up-to-date information on research on ad hoc routing protocols. In addition to the representative protocols in the three broad categories of routing protocols described above, it is worthwhile to look at some of the other concepts that have been applied to routing in mobile ad hoc networks.

Geographic position aided routing: The fundamental problems of routing in ad hoc networks arise due to the random movements of the nodes. Such movements make topological information stale, and hence, when an on demand routing protocol needs to find the route, it often has to flood the entire network looking for the destination. One of the ways of reducing the wastage of bandwidth in transmitting route request packets to every node in the network is to confine the search using geographical location information. *Geographical Positioning Systems (GPS)* can detect the physical location of a terminal using universal satellite-transmitted wireless signals. In recent times, GPS systems have become smaller, more versatile, as well as cost effective. Hence, several protocols have been proposed which assumes the presence of a GPS receiver in each node and utilizes the location information in routing [39], [29], [54], [1].

One of the approaches for utilizing geographic location information in routing is to *forward data packets in the direction* of the location of the destination node, as proposed in [39], [1], [54]. It may be required to define geographic location specific addresses instead of logical node addresses to do that [39].

An alternative concept is proposed in the *Location Aided Routing (LAR)* protocol [29], which uses location information in on-demand routing *to limit the spread of request packets* for route discoveries. LAR uses information such as the last known location and speed of movements of a destination to determine a REQUEST ZONE, which is defined as a restricted area within which the REQUEST packets are forwarded in order to find the destination. Two different ways of defining REQUEST ZONES have been proposed. The idea is to allow route request packets to be forwarded by only those nodes that lie within the REQUEST ZONE, specified by the source. This limits the overhead of routing packets for route discovery, which would normally be flooded over the whole network.

A related protocol that uses *spatial locality* based on hop counts to confine the spread of request packets was proposed in [6]. This protocol uses the concept that once an existing route is broken, a new route can be determined within a certain distance (measured in number of hops) from the old route. The protocol confines the spread of route request packets while searching for a new route to replace one that is freshly broken. For a new route discovery where no earlier routes were in record, the protocol still uses traditional flooding. However, this *query localization* technique for rediscovering routes still saves routing overhead.

Stability based routing: A different approach to improve the performance of routing in mobile ad hoc

networks is based on using routes that are selected on the basis of their *stability*. The *Associativity Based Routing (ABR)* protocol [56] maintains an *association stability metric* that measures the duration of time for which a link has been stable. While discovering a new route, the protocol selects paths that have a high aggregate association stability. This is done with the idea that a long-lived link is likely to be stable for a longer interval than a link that has been relatively shortlived.

Signal Stability Based Routing (SSR) [12] uses signal strengths to determine stable links. It allows the discrimination between "strong" and "weak" links when a route request packet is received by a node. The request packet is forwarded by the node if it has been received over a strong link. This allows the selection of routes that are expected to be stable for a longer time.

Multipath routing: On demand or reactive routing protocols suffer from the disadvantage that data packets cannot be transmitted until the route discovery is completed. This delay can be significant under heavy traffic conditions when the REQUEST and/or the REPLY packet may take a considerable amount of time in traversing its path. This characteristic, along with the fact that each route discovery process consumes additional bandwidth for the transmission of REQUEST and REPLY packets motivate us to find ways to reduce the frequency of route discoveries in on-demand protocols. One way of doing that is to maintain multiple alternate routes between the same source-destination pair such that when the primary route breaks, the transmission of data packets can be switched over to the next available path in the memory. Under the assumption that multiple paths do not break at the same time, which is most often true if the paths are sufficiently disjoint, the source may delay a fresh route discovery if the alternate paths are usable. As a result, many routing protocols have been designed to maintain multiple paths or routes for each pair of source and destination nodes.

The Temporally Ordered Routing Algorithm (TORA) [40] provides multiple alternate paths by maintaining a "destination oriented" directed acyclic graph from the source. The DSR protocol also has an option of maintaining multiple routes for each destination in the route cache, so that an alternate route can be used upon failure of the primary route. Two multipath extensions of DSR were proposed in [34] which aggressively determine multiple disjoint paths for each destination. Here, two different schemes for selecting alternative routes were considered, both benefiting from reducing the frequency of route discoveries caused by link breakages. Several other multipath routing protocols that derive benefits using the same principle have also been proposed [45], [16], [32].

Preemptive routing: A purely reactive routing protocol typically does not avoid a multihop communication from being interrupted *before* the route breaks due to a link failure. Most reactive routing protocols initiate a fresh route discovery when an ERROR packet is received at the source due to a link breakage. This introduces a pause in the communication until a new route is found. The goal of *preemptive routing* protocols is to avoid such pauses by triggering a route discovery and switching to a new (and hopefully better) route before the existing route breaks. Such protocols can be viewed as a combination of proactive and reactive routing, where the route maintenance is performed proactively but the basic routing framework is reactive.

The crucial design issue in such protocols is to detect when to initiate a preemptive route discovery to find a "better" route. The protocol proposed in [19] uses the technique of determining this by observing when the signal strength falls below a predetermined threshold. If the wireless channel is relatively

static, then this correctly detects the initiation of link failure due to increasing distance between the two nodes in the link. However, multipath fading and shadowing effects might lead to false alarms while using this technique. Alternatively, using a time-to-live parameter was proposed in [33]. In this protocol, a preemptive route discovery is initiated when a route has been in use for a predetermined threshold of time. The preemption obviously makes the route discoveries more frequent than what would be observed in a purely reactive scheme. To keep the routing overhead low, the preemptive routing protocol presented in [33] proposes the use of query localization in the preemptive searches.

IV. CONCLUSION

The mobile ad hoc network is one of the newest members in the family of wireless networks that span the planet. This chapter has aimed to provide the main issues and an overview of the developments in the MAC and routing protocols for mobile ad hoc networks. Although a vast amount of work has been done on it in the recent past, many questions still remain unanswered. Some of the issues that need further thought are presented below:

- *MAC*: How can we design improved and robust MAC schemes that would dynamically adjust to variations of the wireless link characteristics and simultaneously cater to the need for higher data rates, quality-of-service requirements, and power savings, that would be crucial in many future applications?
- *Routing*: By far the biggest issue in mobile ad hoc networking research is routing. With the rapid and diverse nature of growth of mobile ad hoc networks, the choice of the routing protocol is likely to depend in the network size, mobility, and application requirements. However, it will be interesting to see if an approach to generate a unified standard for ad hoc routing is achievable.
- *Transport*: The issues of transport layer protocols for mobile ad hoc networks require special attention. A discussion on these issues is outside the scope of this chapter. It is often said that optimizing ad hoc network performance requires a multi-layer approach, where design problems at different layers of the protocol stack are addressed together for a unified solution. How can we arrive at such a design solution?
- *Scalability*: Many applications are already being conceived where hundreds of thousands of nodes are being considered for ad hoc networking. How do we design protocols for these large scale networks?
- *Internet connectivity*: What is the best paradigm for extending the reach of the Internet to mobile terminals that form a mobile ad hoc network with access points to th Internet?
- *Security*: All wireless networks are susceptible to security problems such as eavesdropping and jamming. How can we provide security to mobile ad hoc networks?
- *Power*: One of the major limitations of portability arises from limitations of battery power. In addition to developing improved battery technology, future ad hoc networking protocols have to be made more power efficient so that the network can survive longer without replacement of batteries.

The above items are far from being a complete list of challenging research problems that ad hoc networking has posed before us. It is the hope of the author that this chapter has inspired the reader to look into some of them in more detail.

REFERENCES

[1] S. Basagni, I. Chlamtac, V. R. Syrotiuk, and B. A. Woodward. A distance routing effect algorithm for mobility (DREAM). In *Proceedings of the ACM MOBICOM'1998*, October 1998.

[2] D. Bertsekas and R. Gallager. *Data Networks*. Prentice-Hall, 1987.

[3] Vaduvur Bharghavan, Alan Demers, Scott Shenker, and Lixia Zhang. MACAW: A media access protocol for wireless LAN's. In *Proceedings of the SIGCOMM'94*, pages 212–225, August 1994.

[4] J. Broch, D. A. Maltz, D. B. Johnson, Y-C. Hu, and J. Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. In *Proceedings of the 4th International Conference on Mobile Computing and Networking (ACM MOBICOM'98)*, October 1998.

[5] F. Cali, M. Conti, and E. Gregori. IEEE 802.11 wireless LAN: Capacity analysis and protocol enhancement. In *Proceedings of IEEE INFOCOM'98*, pages 142–149, 1998.

[6] R. Castaneda and S.R. Das. Query localization techniques for on-demand routing protocols in ad hoc networks. *in Proceedings of the 1999 ACM Mobicom Conference*, Aug 1999.

[7] H. S. Chhaya and S. Gupta. Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol. In *Proc. of IEEE Personal Communications Conference*, pages 8–15, October 1996.

[8] S.R. Das, R. Castaneda, J. Yan, and R. Sengupta. Comparative performance evaluation of routing protocols for mobile, ad hoc networks. In *7th Int. Conf. on Computer Communications and Networks (IC3N)*, October 1998.

[9] Jing Deng and Zygmunt J. Haas. Dual busy tone multiple access (DBTMA): A new medium access control for packet radio networks. In *Proceedings of IEEE ICUPS'98*, October 1998.

[10] IEEE Standards Department. Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE standard 802.11–1997, 1997.

[11] E. W. Dijkstra. A note on two problems in connection with graphs. *Numerical Mathematics*, 1:269–271, Oct. 1959.

[12] R. Dube, C. D. Rais, K. Wang, and S. K. Tripathi. Signal stability based adaptive routing (SSA) for mobile ad hoc networks. *IEEE Personal Communication*, Feb. 1997.

[13] J. Haarsten et. al. Bluetooth: Vision, goals, and architecture. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2(4):38–45, Oct 1998.

[14] K. J. Negus et. al. HomeRF and SWAP: Wireless networking for the connected home. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2(4):28–37, Oct 1998.

[15] P. Jacquet et al. Optimized link state routing protocol. `draft-ietf-manet-olsr-05.txt`, 2000. IETF Internet Draft.

[16] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin. Highly-resilient, energy-efficient multipath routing in wireless sensor networks. In *Proceedings of ACM/SIGMOBILE MOBIHOC'2001*, October 2001.

[17] R. Garces and J. J. Garcia-Luna-Aceves. Floor acquisition multiple access with collision resolution. In *Proceedings of the ACM/IEEE Mobile Computing and Networking Conference*, pages 10–12, November 1996.

[18] R. Garces and J. J. Garcia-Luna-Aceves. Collision avoidance and resolution multiple access with transmission queues. *ACM Wireless Networks Journal*, 1998.

[19] T. Goff, N. B. Abu-Ghazaleh, D. S. Phatak, and R. Kahvecioglu. Preemptive routing in ad hoc networks. In *Proceedings of the ACM MOBICOM'2001*, 2001.

[20] IETF MANET Working Group. `http://www.ietf. org/html.charters/manet-charter.html`.

[21] Z. J. Haas. On the performance of a medium access control scheme for the reconfigurable wireless networks. In *Proceedings of IEEE MILCOM'97*, November 1997.

[22] Z. J. Haas and M. R. Pearlman. The performance of query control schemes for the zone routing protocol. *ACM/IEEE Trans. Net.*, 9:427–38, Aug. 2001.

[23] X. Hong, K. Xu, and M. Gerla. Scalable routing for mobile ad hoc networks. *IEEE Network*, 2002.

[24] N. Jain, S. R. Das, and A. Nasipuri. A multichannel mac protocol with receiver-based channel selection for multihop wireless networks. In *Proceedings of the IEEE IC3N'2001*, October 2001.

[25] D. Johnson and D. Maltz. Dynamic source routing in ad hoc wireless networks. In T. Imielinski and H. Korth, editors, *Mobile computing*. Kluwer Academic, 1996.

[26] John Jubin and Janet D. Tornow. The DARPA packet radio network protocols. *Proceedings of the IEEE*, 75(1):21–32, January 1987.

[27] P. Karn. MACA: A new channel access method for packet radio. In *Proceedings of ARRL/CRRL Amateur Radio* 9^{th} *Computer Networking Conference*, 1990.

[28] L. Kleinrock and F. A. Tobagi. Packet switching in radio channels: Part–i - carrier sense multiple access modes and their throughput-delay characteristics. *IEEE Transactions in Communications*, COM-23(12):1400–1416, 1975.

[29] Y. Ko and N. H. Vaidya. Location-aided routing (LAR) in mobile ad hoc networks. In *ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, pages 66–75, November 1998.

[30] Y. B. Ko, V. Shankarkumar, and N. H. Vaidya. Medium access control protocols using directional antennas in ad hoc networks. In *Proceedings of IEEE INFOCOM'2000*, Mar. 2000.

[31] C. T. Lau and C. Leung. Capture models for mobile packet radio networks. *IEEE Transactions on Communications*, 40(5):917–925, 1992.

[32] S-J Lee and M. Gerla. Split multipath routing with maximally disjoint paths in ad hoc networks. In *Proceedings of IEEE ICC'2001*, 2001.

[33] A. Nasipuri, R. Burleson, B. Hughes, and J. Roberts. Performance of a hybrid routing protocol for mobile ad hoc networks. *Proc. of IEEE International Conference of Computer Communication and Networks (ICCCN 2001)*, Oct. 2001.

[34] A. Nasipuri, R. Castaneda, and S. R. Das. Performance of multipath routing for on-demand protocols in mobile ad hoc networks. *ACM/Baltzer Mobile Networks and Applications (MONET) Journal*, 6:339–349, 2001.

[35] A. Nasipuri and S. R. Das. Multichannel CSMA with signal power-based channel selection for multihop wireless networks. *Proc. of IEEE Fall Vehicular Technology Conference (VTC 2000)*, Sept. 2000.

[36] A. Nasipuri, K. Li, and U. R. Sappidi. Power consumption and throughput in mobile ad hoc networks using directional antennas. In *Proceedings of the IEEE International Conference on Computer Communications and Networks (IC3N)*, Miami, October 2002.

[37] A. Nasipuri, S. Ye, J. You, and R. E. Hiromoto. A MAC protocol for mobile ad hoc networks using directional antennas. *Proc. of IEEE Wireless Communications and Networking Conference (WCNC 2000)*, Sept 2000.

[38] A. Nasipuri, J. Zhuang, and S. R. Das. A multichannel CSMA MAC protocol for multihop wireless networks. *Proc. of IEEE Wireless Communications and Networking Conference (WCNC'99)*, Sept 1999.

[39] J. C. Navas and T. Imielinski. Geographic addressing and routing. In *Proceedings of the ACM MOBICOM'97*, 1997.

[40] V. Park and S. Corson. Temporally ordered routing algorithm (TORA) version 1, functional specification. `http://www.ietf.org/ internet-drafts/ draft-ietf-manet -tora-spec-01. txt`, 1998. IETF Internet Draft.

[41] G. Pei, M. Gerla, and T-W Chen. Fisheye state routing: A routing scheme for ad hoc wireless networks. In *Proc. of the IEEE ICC*, June 2000.

[42] Charles Perkins and Elizabeth Royer. Ad hoc on demand distance vector (AODV) routing. `http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-02.txt`, November 1998. IETF Internet Draft.

[43] Charles E. Perkins. *Ad Hoc Networking*. Addison Wesley, 2002.

[44] Charles E. Perkins and Pravin Bhagwat. Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. In *Proceedings of the ACM SIGCOMM '94 Conference*, pages 234–244, August 1994.

[45] D. S. Phatak and T. Goff. A novel mechanism for data streaming across multiple IP links for improving throughput and reliability in mobile environments. In *Proceedings of the IEEE INFOCOM'2002*, 2002.

[46] David C. Plummer. An ethernet address resolution protocol: Or converting network protocol addresses to 48 bit ethernet addresses for transmission on ethernet hardware, November 1982.

[47] G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Communications of the ACM*, 43(5):51–58, May 2000.

[48] R. Ramanathan. On the performance of beamforming antennas in ad hoc networks. In *Proceedings of ACM/SIGMOBILE MOBIHOC'2001*, October 2001.

[49] J. Redi and B. Welsh. Energy conservation for tactical robot networks. In *Proc. IEEE MILCOM*, pages 1429–33, 1999.

[50] E. M. Royer and C. K. Toh. A reivew of current routing protocols for ad hoc mobile wireless networks. *IEEE Pers. Comm.*, pages 46–55, April 1999.

[51] C. Santivanez, R. Ramanathan, and I. Stavrakakis. Making link-state routing scale for ad hoc networks. In *Proc. of 2001 ACM Intl. Symp. Mobile Ad Hoc Net. Comp.*, Oct. 2001.

[52] N. Schacham and J. Westcott. Future directions in packet radio architectures and protocols. *Proceedings of the IEEE*, 75(1):83–99, Jan 1987.

[53] ETSI Secretariat. Hiperlan functional specification. Draft prETS 300 652, 1995.

[54] I. Stojmenovic and X. Lin. GEDIR: Loop-free location based routing in wireless networks. In *Proceedings of the Intl. Conf. on Parallel and Dist. Comp. Sys.*, Nov. 1999.

[55] F. A. Tobagi and L. Kleinrock. Packet switching in radio channels: Part–II - the hidden terminal problem in carrier sense multiple-access and the busy-tone solution. *IEEE Transactions in Communications*, COM-23(12):1417–1433, 1975.

[56] C.-K. Toh. Associativity-based routing for ad hoc mobile networks. *Wireless Personal Communications*, 4:103–139, 1997.

[57] Y-C Tseng, S-L Wu, C-Y Lin, and J-P Shen. A multichannel mac protocol with power control for multihop mobile ad hoc networks. In *Proceedings of the 21st Internl. Conf. Dist. Comp. Syst.*, April 2001.

[58] J. Weinmuller, M. Schlager, A. Festag, and A. Wolisz. Performance study of access control in wireless LANs - IEEE 802.11 DFWMAC and ETSI RES 10 hiperlan. *Mobile Networks and Applications*, 2:55–67, 1997.

[59] C. Wu and V. O. K. Li. On the performance of a medium access control scheme for the reconfigurable wireless networks. In *Proc. 6th WINLAB Workshop*, 1997.