1

Ad Hoc Networks: Study and Discussion of Performance

1.1. Introduction

Ad hoc networks are wireless networks that form spontaneously and organize themselves automatically without requiring a pre-existing infrastructure. An *ad hoc* network is a collection of hosts (nodes) equipped with antennas that can communicate with one another without administrative centralization based on the topology of wireless communication. Unlike wired networks in which only a few nodes called routers are responsible for delivering data, in an *ad hoc* network, each node acts as a terminal mode, and possibly as a link to relay messages when recipients are not within radio range of the transmitters. In an *ad hoc* network, a node can communicate directly with another node in point-to-point mode when the two nodes are located in the same transmission zone, while communication with a node in another zone is carried out via several intermediary nodes in multi-hop mode.

Initially of military origin [JUB 87], due to several of the benefits they offer, *ad hoc* networks are of confirmed interest for circumstances characterized by a total lack of pre-existing infrastructures. From an applicative point of view, *ad hoc* wireless networks are useful in situations that require the deployment of a rapid local network or one lacking infrastructure, such as reaction to a crisis, conferences, military applications and possibly household and office

networks. *Ad hoc* networks could, for example, enable medical personnel and officials to better coordinate their efforts during large-scale emergency situations that result in the complete destruction of network infrastructures, as was the case with the September 11th attacks and the 2003 blackout in the northeastern United States [JUR 07]. The principal advantages of *ad hoc* networks can be summed up in the following characteristics:

– Deployment: the deployment of an *ad hoc* network is quick, simple and low-cost.

– Organization: *ad hoc* networks are organized automatically via local collaboration between nodes with no centralization required.

– Adaptability: in the event of the breakdown of some nodes or links, *ad hoc* networks ensure continuity of operation and rebuild themselves locally using efficient access and routing techniques.

– Robustness: *ad hoc* networks ensure continuous operation and override critical states by executing local repairs without outside intervention.

Of the applications in which these types of networks are involved, we would note environmental applications such as the prevention of natural disasters and forest surveillance, as well as medical applications including the detection of physiological signals sensed on different parts of patients' bodies, etc. In these applications, the nodes are generally mobile and require specific power sources to be able to cover their energy needs while they are active. The issue of energy consumption is central here, and in the last part of this chapter, we will discuss it in detail, explaining the various solutions proposed in the literature, whose objective is to optimize and/or minimize energy consumption.

1.2. Concepts specific to *ad hoc* networks

1.2.1. *Topology*

An *ad hoc* network can be represented mathematically by an undirected graph $G = (V\ E)$, where V designates the group of nodes

and $E \subseteq V^2$ denotes the group of arcs corresponding to the direct communications possible. If u and v are two nodes of V, l'arc $(,)u\ v$ exists if and only if u can send a message directly to v; in this case, we say that v is neighbors with u. Pairs belonging to E are dependent on the position of the nodes and their communication ranges. If we assume that all of the nodes have an identical range R and if $d\ u\ v(,)$ designates the distance between nodes u and v, then [CHE 98]:

$$E = \left\{(u,v) \in V^2 \,\middle|\, d(u,v) \leq R\right\} \qquad [1.1]$$

The graph thus obtained is called a unitary graph. If a node reduces its communication range, its relationships with the other nodes will be changed. This operation leads to a modification of the network's topology, which is expressed by a change of the unitary graph by another derived graph. The derived graph is extracted on the basis of known models, for example the unit disk graph (UDG) model with the relative neighborhood graphing (RNG) technique, the objective of which is to remove the largest arcs to obtain graph $G_{NRG} = (V\ E,\ _{NRG)}$. E_{NRG} can be constructed using arcs *(u,v)* of the unitary graph as follows:

$$(u,v) \notin E_{NRG} \Leftrightarrow \exists w(d(u,w) < d(u,v) \wedge d(v,w) < d(u,v)) \qquad [1.2]$$

Thus, nodes u and v have a common neighbor, w, and so we can remove arc *(u,v)*.

1.2.2. *Connectivity*

DEFINITION 1.1.– *A graph G = (V E,) is connected if and only if, for every pair of distinct peaks, a chain exists that links them*.

The concept of a path is necessary to define connectivity.

DEFINITION 1.2.– *A path is a series of nodes with the form:* $s = u, a_1, ..., a_n, v \,|\, u, v, a_i \in V$ *, where u is the message source and v is the destination*.

DEFINITION 1.3.– *A graph* $G=(V,E)$ *is* λ*-connected if and only if* $\forall u,v\in V,\exists s\left|P(s)\geq\lambda\right.$*, where P(s) is the probability that v will receive the message sent by u over path s*.

Two different methods can be used to calculate $P(s)$:

– Either we consider that v can receive the message from u via any node on path s:

$$P(s)=P((a_1,v)\cup\ldots\cup(a_n,v))=\sum_{k=1}^{n}-1^{k+1}\sum_{1\leq i1<\ldots<ik\leq n}P((a_{i1},v)\cap\ldots\cap(a_{ik},v)) \qquad [1.3]$$

– Or we estimate that the message can only be received at v via a_n, giving us:

$$P(s)=P(a_1,a_2)\cap\ldots\cap P(a_{n-1},a_n)=\prod_{i=0}^{n-1}P(a_i,a_{i+1}) \qquad [1.4]$$

1.2.3. *Mobility*

A node in an *ad hoc* network is able to move, and it can join or leave the network at any time, which causes the appearance or disappearance, respectively, of links. The movement of nodes is generally random; highly developed routing protocols are needed to control node movement. On the one hand, node movement is an important factor in the provision of major *ad hoc* network services; on the other hand, however, it causes difficulties with regard to routing function, network connectivity and energy optimization. Moreover, in a mobile environment, routing protocols require that a route has been completely mapped before any possible data transmission. It means that a transmitter node may have to wait a long time for the new path to be found. Simulations in [CAR 99] on routing protocols in an *ad hoc* network demonstrate the impact of mobility on the transmission of packets. As mobility increases, the number of packets transmitted drops. The additional cost in terms of quantity of control information exchanged increases, thus reducing the bandwidth available for data

transfer. Some protocols are more sensitive than others to mobility; however, the fact that a node moves does not necessarily cause major reductions in the performance if the graph's connectivity is not affected (or it is only slightly affected) when it is necessary to transmit packets. Modeling mobility is among the difficulties encountered during *ad hoc* network simulations [ROH 14], but several models currently exist that include node mobility, including Random Walk, referenced point group mobility (RPGM), Manhattan and Freeway.

1.2.4. *Networks: wireless mesh network (WMN), wireless sensor networks (WSN) and mobile ad hoc network (MANET)*

A node in an *ad hoc* network is able to move; it can join or leave the network at any time, which causes the appearance or disappearance, respectively, of links. The movement of nodes is generally random; highly-developed routing protocols are needed to control node movement. On the one hand, node movement is an important factor in the provision of major *ad hoc* network services; on the other hand, however, it causes difficulties with regard to routing function, network connectivity and energy optimization. Moreover, in a mobile environment, routing protocols require that a route be completely mapped before a possible data transmission. It means that a transmitter node may have to wait a long time for the new path to be found. Simulations in [CAR 99] on routing protocols in an *ad hoc* network demonstrate the impact of mobility on the transmission of packets. As the mobility increases, the number of packets transmitted drops. The additional cost in terms of quantity of control information exchanged increases, thus reducing the bandwidth available for data transfer. Some protocols are more sensitive than others to mobility; however, the fact that a node moves does not necessarily cause major reductions in the performance if the graph's connectivity is not affected (or it is only slightly affected) when it is necessary to transmit packets. Modeling mobility is among the difficulties encountered during *ad hoc* network simulations [ROH 14], but several models currently exist that include node mobility, including Random Walk, RPGM, Manhattan and Freeway.

There are three types of *ad hoc* networks: wireless mesh networks (WMNs), wireless sensor networks (WSNs) and mobile *ad hoc* networks (MANETs). These networks have several similarities, as well as certain differences that logically involve different solutions.

1.2.4.1. *Wireless Mesh Network*

The mesh network is an emerging new technology that constitutes a particular case of an *ad hoc* network. It combines the benefits of *ad hoc* networks previously detailed with the benefits of terrestrial networks, notably in terms of available output due to the organization of the network with reduced-mobility nodes devoted to routing and mobile client nodes. WMN networks are composed of two essential elements: mesh routers and mesh clients. Mesh routers are links, while the mesh client connects to the closest mesh router and uses the *ad hoc* infrastructure to access the services of the *ad hoc* network. The difference between a classic *ad hoc* network and a mesh network is the restriction of routing functionalities to a subgroup of the network formed of mesh routers. WMNs diversify the capacities of an *ad hoc* network by introducing a hierarchy into the network [AKY 05].

1.2.4.2. *WSN sensor network*

A sensor network is an *ad hoc* network that is composed of nodes equipped with control units and measurement units. These units are characterized by a reduced processing and storage capacity due to their miniature sizes (on the order of 1 cm^3). The sensor nodes periodically send their acquired data to a special node called a sink node, which is responsible not only for the collection of reports but for the broadcasting of requests for the types of data required to the sensors via request messages [AKY 02, PIL 14, KAR 14]. The rapid deployment, reduced cost, self-organization and breakdown tolerance of WSN networks are characteristics conducive to their use in various domains, including military (chemical radiation, battlefield analysis), environment (temperature, humidity, seismic activity), medical (internal body imaging) and security (intrusion, surveillance, heating, etc.).

1.2.4.3. *MANET mobile network*

MANETs are characterized by a strong node dynamic with topology that is highly variable due to the frequent changes in node positions. To reach its destination, a message passes through several relay nodes. Unlike WSNs and WMNs, the nodes in MANETs are all mobile, and communication can take place between any of the nodes in a network. It means that the failure of any node is significant and must be handled quickly by specialized algorithms. Routing, topology control, and self-organization are basic techniques for the functioning of MANETs.

1.2.5. *Routing*

Routing is a function that consists of determining the route of each packet from a known source toward one or more destinations. Routing can also be defined as the task of transporting data from source nodes to destination nodes [XIA 08, TOU 99]. If a single destination is involved in the communication, this task is known as unicast routing, but when all the nodes in the network, or just a group of nodes, are receiving data, then we speak of broadcast and multicast routing, respectively [TAV 06]. The objective of routing algorithms is to find the shortest path between nodes, either in terms of number of hops, or in terms of link length (time). The algorithm functions in terms of time when links are dropped or re-established, or during a change in traffic conditions within the network. There are networks in which the packet can use different routes (this is generally the case with the Internet networks); these types of networks are called packet-switched networks and are always equipped with dynamic routing control functions. On the other hand, in networks based on packet switching architecture (such as traditional telephone networks and asynchronous transfer mode (ATM) networks), the routing decision is made on each connection, and all connection packets use the same path. The routing will be more complicated, and more complex problems may appear, when the network's users are mobile, and in a more complicated manner when the nodes are mobile themselves; this is the case with the satellite networks that use satellites in low orbit, called LEO, and *ad hoc* networks as well.

In *ad hoc* networks, decentralized routing is already used; therefore, the nodes will be more sensitive to changes in network topology. The instability of the wireless communication medium, the limitations of energy and bandwidth and the mobility of nodes introduce more difficulty and complexity during the design of routing protocols for mobile *ad hoc* networks. In MANETs, depending on the manner of establishing and maintaining routes during the transport of data from mobile nodes, we distinguish three principal categories of routing protocols: proactive, reactive and hybrid.

1.2.5.1. *Proactive protocols*

Proactive protocols are based on classic link-state and distance-vector algorithms. Each node holds routing information that concerns all the nodes in the network. This information is stored in routing tables that are updated with each topological change in the mobile *ad hoc* network in order to reconstruct the routes. Among the most widely used proactive routing protocols are: destination-sequenced distance vector (DSDV), Wireless Routing Protocol (WRP), global state routing (GSR) and optimized link state routing (OLSR) [COR 99, CAR 03].

1.2.5.2. *Reactive protocols*

Reactive routing protocols are also called on-demand protocols. They create and maintain routes according to the communications' needs in the network. When a transmitter node has need of a route, it launches a route discovery procedure [MAR 00]. Reactive protocols can be divided into two subcategories: source routing and hop-by-hop routing. The advantage of reactive protocols is that they offer greater adaptability to the topological changes of an *ad hoc* network [BAD 03, CHI 05] following the use of very recent "fresh" data for routing. Dynamic source routing (DSR), *ad hoc* on-demand distance vector (AODV) and core extraction distributed *ad hoc* routing (CEDAR) are the routing protocols that have been used with great frequency in recent years [PER 03].

1.2.5.3. *Hybrid protocols*

This category combines the first two types of routing protocols to achieve a shorter response time by taking advantage of the benefits of

proactive and reactive protocols. In a hybrid protocol, the network is broken down into small zones where routing inside each zone is ensured by the proactive protocol, while routing between the different zones is based on the reactive protocol. Zone Routing Protocol (ZRP) and zone-based hierarchical link state (ZHLS) are among the best-known hybrid protocols [TOU 99].

1.2.6. *Weak security*

Ad hoc networks are notable for their weak security against various attacks, whether internal or external. An *ad hoc* network can be attacked in its basic functions such as routing, which is vital for the network to function properly. Attacks such as black holes, identity spoofing and wormholes disrupt routing protocols via multiple tactics to prevent them from functioning correctly. The Sybille attack is effective against routing algorithms, data aggregation, the equitable distribution of resources and the detection of malevolent nodes. Sybille has aroused particular interest in the scientific community, as much because of its originality as of the difficulty in finding countermeasures at a reasonable cost in *ad hoc* networks [DOU 02]. Security issues in *ad hoc* networks are therefore quite complicated, as we seek to authorize new nodes to participate in the network while avoiding nodes that will reroute or disrupt the routing function [ZHU 14, NOV 14].

1.2.7. *Access to the environment*

The major challenge in *ad hoc* networks consists of knowing who has permission to transmit at a given time, which needs the design of protocols to manage this type of situation. Medium access control (MAC) protocols are expected to accomplish this task. The MAC layer contains random access protocols that are generally characterized by low output due to several factors that can influence the network's quality of service (QoS). Among the limitations faced by these MAC protocols during their operation are collisions, successive retransmissions, delivery times, error rates, etc. In addition, the power supply to mobile nodes in MANETs is dependent on power sources of limited capacity. Thus, in order to ensure the functioning of

MANETs for a sufficient period of time, it is necessary to apply techniques to optimize and/or save energy by considering all the sources of energy overconsumption or waste adopted by MAC access protocols. MAC access in *ad hoc* networks, and in wireless networks generally, has been defined by the IEEE work group according to standard 802.11 [MÜH 02, HAR 04]. The main objective of the MAC layer according to this standard consists of providing reliable data services for upper layer protocols.

1.3. MAC protocols in mobile *ad hoc* networks

Currently, mobile *ad hoc* networks use the distributed coordination function (DCF) protocol or its improvement, the enhanced DCF (EDCF) protocol, which is structured on the carrier sense multiple access with collision avoidance (CSMA/CA), when accessing the environment. Random access methods are generally grouped into two main families: ALOHA and its derivatives, and CSMA and its derivatives [BEN 07]. In the next sections, we will introduce the classic random access methods in order to set down the basic concepts of the new environmental access techniques [GAJ 14, MUK 14, MOC 12, DUA 14, SEN 14, JAC 14, MOK 14].

1.3.1. *ALOHA*

The name of this method comes from experiments conducted at the University of Hawaii to link computer centers scattered across several islands. In it, nodes transmit packets unconditionally as soon as they receive them. There is no support listening before transmission. In addition, the propagation time of signals on the satellite channel is a limiting factor, since nodes are warned of a collision only 20 ms after data transmission. In the event that a data transmission has not been executed correctly, the node will retransmit the packets after a random period of time. This access method therefore has a low satellite channel use rate, approaching 20%; techniques exist that are similar but have been modified to achieve better performances [ALT 87]. Transmission in this protocol is completely decentralized. The basic principle is as follows:

If you have a message to transmit, transmit it.

If the message interferes with other transmissions, try to send it later.

The probability that n packets will arrive in two different packet times is given by:

$$P(n) = \frac{(2\lambda)^n e^{-2\lambda}}{n!} \quad [1.5]$$

where λ is the traffic load.

The probability $P(0)$ that a packet will be successfully received without collision is:

$$P(0) = e^{-2\lambda} \quad [1.6]$$

So, the output Th is given as follows:

$$Th = \lambda.P(0) = \lambda.e^{-2\lambda} \quad [1.7]$$

Hence, the maximum output value of this technique:

$$Th_{\max} = \frac{1}{2.e} \approx 0.184 \quad [1.8]$$

1.3.1.1. *Slotted ALOHA (SALOHA)*

An improvement has been added to the original ALOHA protocol is called ALOHA with temporal segmentation. This protocol introduces user synchronization. Time is cut into fixed-duration intervals called slots. Users cannot start transmission until the beginning of each slot. The probability of collision is thus reduced, and the maximum output is doubled. Thus, if there is not another user in transmission mode at the start of the time slot (TS), the probability function Th_i will be equal to:

$$Th_i = \frac{P_i}{(1-P_i)} \prod_{i=1}^{n} (1-P_i) \quad [1.9]$$

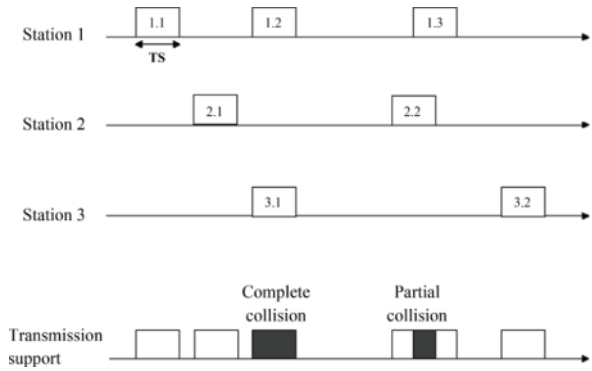

Figure 1.1. *Collision in the ALOHA protocol*

If Th is the traffic output and λ is the traffic, we can write:

$$Th_i = \frac{Th}{n} \text{ and } P_i = \frac{\lambda}{n}$$

$$\frac{Th}{n} = \frac{\lambda}{n} \cdot \left[\frac{1}{(1-\frac{\lambda}{n})} \right] \cdot \prod_{i=1}^{n} (1-\frac{\lambda}{n}) \qquad [1.10]$$

which means that the maximum output of SALOHA is: $Th_{\max} = \frac{1}{e} \approx 0.37 \text{ (Paq / TS)}$

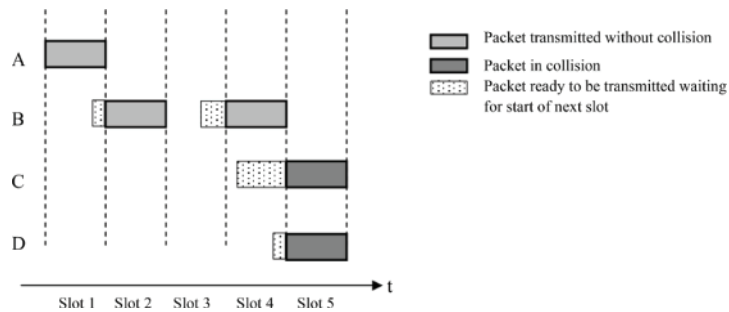

Figure 1.2. *Avoidance of collisions in SALOHA. For a color version of the figure, see www.iste.co.uk/benslama/adhocnetworks.zip*

1.3.1.2. *Multi-copy ALOHA*

When *m* copies of a packet are sent in the SALOHA (multi-copy) technique, the probability of successful transmission for this packet, or the probably that one packet of the *m* copies sent, will not enter into a collision will be higher than when a single copy is sent (S-ALOHA). This is true only when the other packets are sent in a single copy or when traffic in the channel remains stable without disruptions. In order to maximize the probability of successful transmission, we will assume that all users are transmitting the same number of copies (*m*) [MUH 04]. Multi-copy ALOHA is generally designed for satellite systems offering a higher probability of successful transmission, multichannel ALOHA systems or ALOHA systems with reservations, with a very high probability of success.

To assess output in this algorithm, we consider that the arrival of packets is a Poisson process. For simple SALOHA, we assume that the average retransmission period is larger than five slots, while the average period value for the *m* copies including the first transmission must be as large as five slots. So:

$$\lambda = \lambda_1 + \lambda_2 + \lambda_3 + ...\lambda_n = \sum_{i=1}^{n} \lambda_i \qquad [1.11]$$

The average number of copies per packet is:

$$N = \lambda^{-1}.\sum_{i=1}^{n} i.\lambda_i \qquad [1.12]$$

The probability P_i that packet *i* will be successfully received is:

$$P_i = 1 - prob[all\ copies\ are\ in\ collision] = 1 - \left(1 - e^{-N\lambda}\right)^i \qquad [1.13]$$

So, for *K* copies, we have:

$$Th_k = \lambda.P_k = \lambda\left[1-(1-e^{-k.\lambda})^k\right] \qquad [1.14]$$

To maximize the probability of successful transmission, a single copy must be transmitted when traffic in the channel is greater than 0.48, and when $\lambda \in [0.28, 0.48]$, we can only transmit two copies. The same figure shows that when λ is small, the probability P will reach the maximum value ($P = 1$) using a larger number of copies.

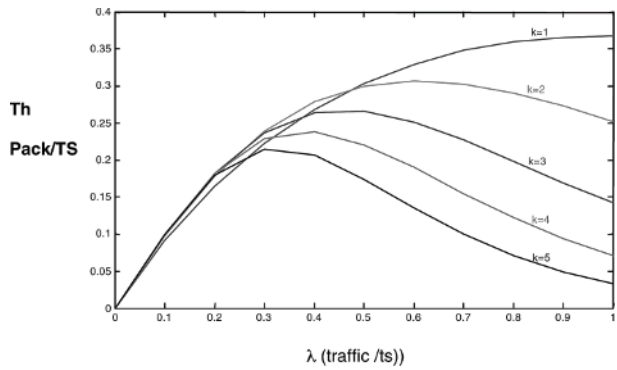

Figure 1.3. *Output in multi-copy ALOHA. For a color version of the figure, see www.iste.co.uk/benslama/adhocnetworks.zip*

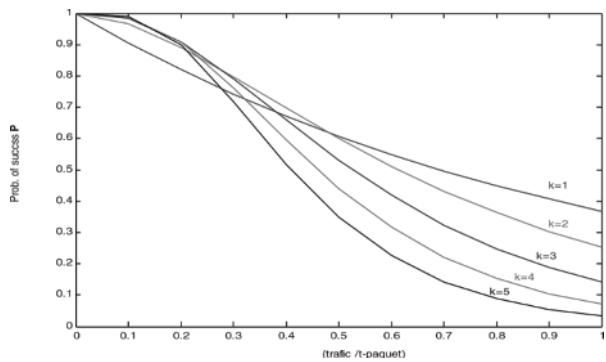

Figure 1.4. *Probability of success in multi-copy ALOHA. For a color version of the figure, see www.iste.co.uk/benslama/adhocnetworks.zip*

Consequently, random access methods such as ALOHA, SALOHA and multi-copy ALOHA have relatively modest performances and a loss rate that is too high for satellite or *ad hoc* contexts. In fact, communication systems in satellite networks cannot function using

access methods with a high collision rate, as retransmission necessarily introduces excessive delays. Therefore, the use of these access methods is limited to the conveyance of signals, identification messages or small control and acknowledgment packets.

1.3.2. *CSMA*

In CSMA, mobile devices transmit only when the channel is free in order to avoid collisions. The principle of CSMA can be explained as follows: a node wishing to transmit in a channel first listens to the communication environment (mesh with bus). If the environment is free it transmits; if not, it waits for a specific amount of time. If the transmitter has not received the information after a given time, it supposes that a collision has taken place. After collision, the node waits for a random period and then retransmits [UYA 14]. There are several variants of CSMA, each of which possesses different behavior in the event of an occupied environment [LES 12, YAN 13]; thus, we distinguish between non-persistent CSMA, persistent CSMA and P-persistent CSMA. In the case of non-persistent CSMA, if the channel is occupied, it waits for a random amount of time and then transmits, while in the case of persistent CSMA, when the channel is occupied, the node continues listening to the channel until it becomes free, and then begins the transmission. In the case of *P*-persistent CSMA, the node acts differently; when the channel is free, the node transmits with probability P and waits for a period with probability $(1-P)$, but when it finds the channel occupied, it continues listening to the channel until it becomes free. If the transmission has taken a long time, the node begins listening to the transmission channel again.

1.3.2.1. *CSMA with collision detection (CSMA/CD)*

In CSMA/CD, before any attempt at transmission, the node makes sure that the channel is not already being used (carrier detection), and when the channel is free, the node rechecks the channel after a random period; if the channel is still free, the node sends its packets. However, this does not confirm that the packets have been successfully received. In reality, one or more nodes may send their packets simultaneously following this procedure, causing a collision, in particular when the

network is crowded. This procedure is currently used mostly in wired networks, where collision detection is based on the type of electromagnetic propagation on a cable.

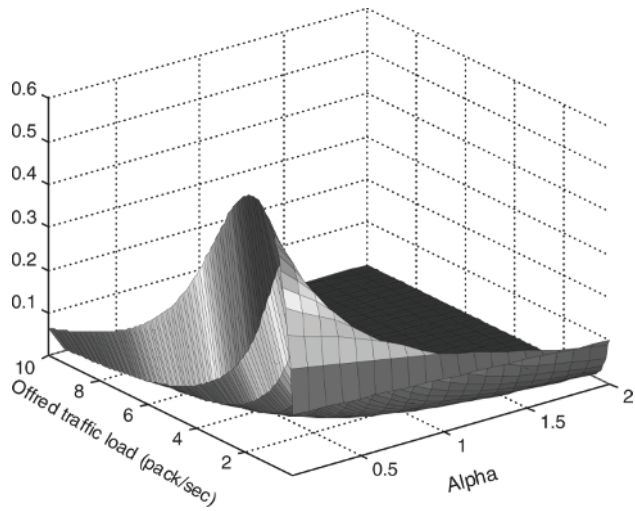

Figure 1.5. *Non-persistent CSMA output in 3D in relation to* λ *and* α*. Output is maximized when* $\alpha \in [0.1, 0.5]$ *and* $\lambda \in [1.6, 2.5]$*. For a color version of the figure, see www.iste.co.uk/benslama/adhocnetworks.zip*

1.3.2.2. *Standard 802.11 and the DCF algorithm*

The IEEE 802.11 standard was created in 1997 by the IEEE group. It describes the physical (PHY) layer and the data link layer, which is divided on two sublayers: the medium access control (MAC) layer and the logical link control (LLC) layer [RAM 14] of wireless networks characterized by a theoretical output of between 1 and 54 Mbit/sec, and a range that varies from 1 to 100 m depending on velocity and protocols used. Initially, frequency hopping spread spectrum (FHSS), direct sequence spread spectrum (DSSS), infrared (IR) and orthogonal frequency division multiplexing (OFDM) standards were adapted to the PHY layer, in which each of them used a different technique to spread the spectrum. The major problem with these standards was the lack of compatibility between them. However, many improvements have been made to these protocols, with new

versions appearing to eliminate the drawbacks of the old versions and add benefits for the proper functioning of wireless networks, particularly impacting output and QoS. Three new layers have been designed for the PHY layer: 802.11b (wireless fidelity (WiFi)), 802.11a (WiFi5) and 802.11g, with the latter modified by 802.11n [GIB 92]. Figure 1.6 illustrates the architecture of the PHY layer and the MAC layer according to standard 802.11 compared to that of the open systems interconnection (OSI) standard. The data link layer is subdivided into two sublayers: LLC sublayer and MAC sublayer. The Data Link layer of standard 802.11 uses the DCF algorithm for non-centralized wireless networks, a category to which *ad hoc* networks belong. The DCF can be used by all nodes and offers equitable access to the radio channel without any centralization of access management in a fully distributed mode.

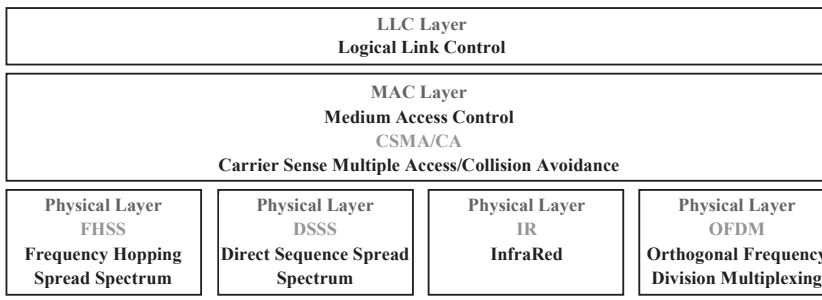

Figure 1.6. *PHY layer and data link layer in IEEE 802.11 [HAR 99]*

The first characteristic of the MAC layer consists of using acknowledgments to detect collisions and allow the retransmission of lost packets. In standard 802.11, the node can send in unicast mode toward a specific destination, or in broadcast mode toward multiple destinations. In the second case, there is no acknowledgment, and packets may be lost in collisions. The 802.11 standard in the MAC layer is based mainly on the use of Acknowledgment (ACK) frames in addition to request to send (RTS) and clear to send (CTS) information signals, which reduce errors due to interference and collisions on the radio channel in order to guarantee data integrity in the data link layer.

1.3.2.3. *CSMA/CA*

1.3.2.3.1. ACK principle

A node wishing to transmit listens to the channel; if it is occupied, the transmission is postponed. If the channel is free for a fixed period of time called distributed inter-frame space (DIFS), the node starts transmission after a random amount of time called backoff. The receiving node waits for the cyclic redundancy check (CRC) of the packet received and sends back an acknowledgment of receipt (ACK). Reception of the ACK indicates to the transmitter that no collision has taken place. If the transmitter does not receive an ACK, it retransmits the fragment until it obtains one, or abandons the attempt after a certain number of retransmissions.

To monitor network activity, the MAC sublayer works in collaboration with the PHY layer, which uses the clear channel detection (CCA) algorithm to assess the availability of the channel. To find out whether the channel is free, the PHY layer measures the power of the signal received by the antenna, called a received signal strength indicator (RSSI). The PHY layer determines if the channel is free by comparing the RSSI value with a predefined threshold, and then transmits a free-channel indicator to the MAC layer. In the opposite case, the transmission is postponed.

1.3.2.3.2. Definitions of IFS

According to standard 802.11, we differentiate four types of IFS: short IFS (SIFS), point coordination function (PCF) IFS (PIFS), DCF IFS (DIFS) and extended IFS (EIFS), listed in decreasing order of duration as follows:

– *SIFS* is the shortest distance between the four types of IFS. It separates a data packet from its acknowledgment. It is used for the transmission of ACK and CTS frames, responses to polling and barrages of frames sent by a single node. SIFS value is set by the PHY layer and is calculated so that the transmitting node is capable of switching in reception mode in order to decode the incoming packet.

– *PIFS* is used in PCF mode (Ethernet) and allows PCF transmissions to access the medium via the use of a smaller IFS than the one used for the transmission of frames in DCF. *PIFS* = *SIFS* + *TS*. *TS* = minimum time required to determine the state of the channel + round-trip time + propagation time. It corresponds to the minimum interval between two PHY carrier-detection operations. This value is dependent on the characteristics of the PHY layer being considered and is a constant specified by the standard for a given PHY layer.

– *DIFS* is most frequently used with SIFS. It is used in DCF mode as the minimum wait time before transmission. *DIFS* = *SIFS* + *2*TS*.

– *EIFS* is the longest of the IFS. It is used when a collision is detected and inhibits subsequent collisions [RAM 14].

IFS is used to define the degrees of priority of frames. When several nodes wish to transmit at the same time, the node with the highest-priority frames (acknowledgments, for example) begins transmitting first. This role will fall next to the node judged to have the highest priority, carrying frames related to network administration or traffic, which have time constraints. Finally, the least important frames concerning asynchronous traffic will be transmitted after a longer waiting period.

1.3.2.3.3. Backoff

When a node wishes to transmit its data packets, it listens to the channel for a DIFS; if the channel is free, the node begins sending its data. It can happen that two nodes detect the channel that is free at the same time after the DIFS and begin their transmissions at the same time, which leads to a collision. To reduce the probability of such a collision, after the DIFS, the node waits for a random amount of time called backoff time (BT), which is composed of a number of TS included between 0 and contention window (CW). After the DIFS, the node decrements its backoff by a TS step until it reaches the value of zero. If the node detects that the channel is occupied during BT, it starts over and waits for the channel to become free during the DIFS;

this time, if the channel is still free when the backoff reaches a value of zero, the node starts its transmission immediately. After each successful transmission, the receiving node sends an acknowledgment frame to the transmitter after an SIFS. However, if the transmission is unsuccessful, the node enters a double-length backoff [0 – 2*CW] in order to reduce the number of collisions in the event that multiple nodes wish to access the transmission channel. Here, note that not all the nodes that have postponed their access to the channel during the BT must start a new backoff, but they continue to decrement their last backoff counter in order to have priority during subsequent retransmission attempts.

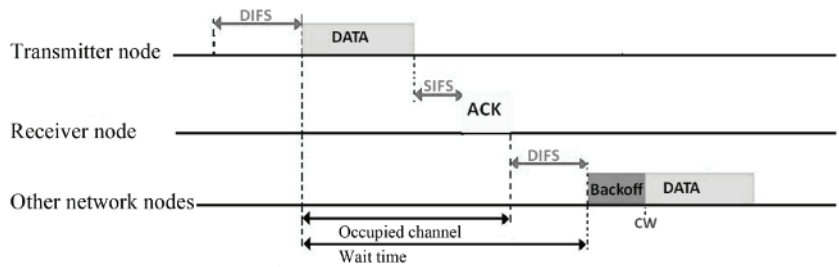

Figure 1.7. *IFS and backoff in CSMA/CA mode*

The size of the CW is initialized at a minimum value CWmin, which increases exponentially by doubling its size with each transmission failure. The CW size is reinitialized at CWmin after successful transmission or when it reaches the CWmax value after a limited number of retransmissions called retry limit (RL).

After a successful transmission, the transmitter node must conduct a backoff called post-backoff in order to separate successive frames.

Despite the integration of various techniques – backoff, IFS and ACK – the CSMA/CA protocol still experiences collisions. Hidden nodes and long frames are considered to be major sources of collisions. To limit collisions due to long frames and their multiple

transmissions, data frames can be divided into fragments that can be transmitted sequentially as individual data frames. The advantage of fragmentation is that in the event of transmission failure, the error is detected sooner and there is less data to retransmit. The disadvantage is the overload introduced by the increase in the CW with the addition of additional acknowledgment frames. To solve the problem of hidden nodes associated with CSMA/CA, DCF defines a control mechanism called RTS/CTS.

1.3.2.3.4. RTS/CTS

This mechanism is also called virtual carrier listening. It is preventive against problems of hidden nodes and long frames, which introduce retransmissions that are costly in terms of time and spectral resources.

A node wishing to transmit first sends a small RTS control packet to request authorization from its receiver to begin data transmission. The RTS contains information about the source, destination and duration necessary to complete transmission. The RTS is received by the receiving node and by the other nodes in the network. The receiving node then responds by sending a CTS control packet that contains the same information as the RTS to inform the transmitter that the channel is free and it can begin data transmission immediately. At this time, the other nodes in the network, which have received at least one of the two control packets (RTS/CTS), each send a virtual network allocation vector (NAV) carrier listening indicator for a certain amount of time, delaying their transmissions until the expiration of the NAV's timer. If a hidden node has not received the RTS, it can then receive the CTS in order to update its NAV, which can minimize the problem of hidden nodes. NAV updating is ensured by the reception of RTS or CTS signals. The NAV informs each of the nodes not involved in the transmission of the length of time during which the channel will have to be occupied. At this time, the nodes may switch to power-saving mode (PSM) for this period of time in order to conserve energy.

An RTS frame is of 20 bytes and a CTS frame is of 14 bytes in size; these are therefore short frames with a low probability of

collision. However, in the case of long frames, the collision rate is high due to multiple transmissions. To solve this problem, data frames can be divided into fragments that can then be transmitted sequentially and individually. The advantage of fragmentation is that, in the event of transmission failure, the error is detected sooner and there is less data to retransmit. However, an overload will be introduced by the increase in contention and also by the addition of more acknowledgment frames.

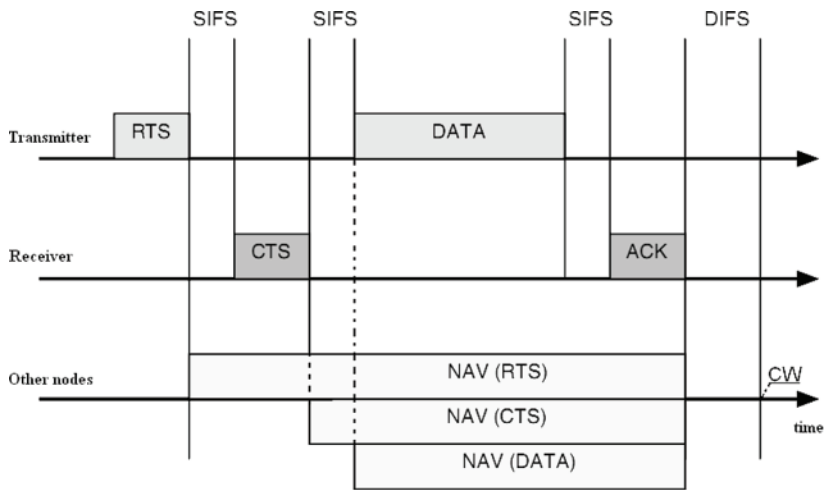

Figure 1.8. *NAV update*

In conditions like this, where there are significant overloads and inefficient use of bandwidth, an initiation threshold is implemented to limit the use of radio support. If the length of the data to be transmitted is lower than this threshold, transmission will take place without RTS/CTS; otherwise, the RTS/CTS mechanism is used.

It should be mentioned here that in broadcast mode, the RTS/CTS mode is inoperative. In this case, the frame is broadcasted toward all nodes, which means that there are multiple recipients and thus we can have multiple CTS simultaneously, which causes more collisions.

1.3.2.3.5. MAC frames in 802.11

802.11 frames are generally formed of four main parts: preamble, header, data and CRC. Each of these parts is characterized by a limited number of bits and a function that is well determined during the transmission of frames in an *ad hoc* network. The figure below explains the components of a MAC 802.11 frame.

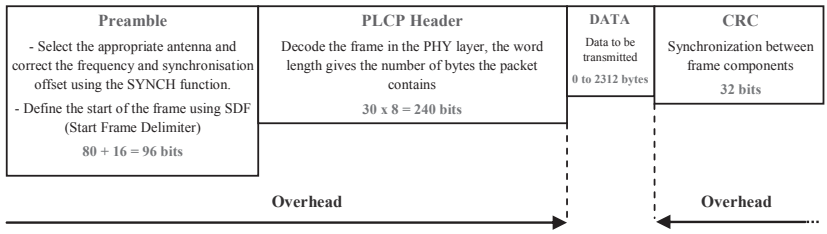

Figure 1.9. *Usual format of an 802.11 frame*

A physical layer convergence protocol–protocol data unit (PLCP-PDU) frame is composed of a PLCP header and of data from the MAC layer. The PLCP-PDU header contains two fields: preamble and header. Two types of preamble are defined: long one (192 bits) and short one (132 bits). A long preamble secures the connection to the network and thus the transmission. The word length of the PLCP_PDU gives the number of bytes the packet contains, which helps the PHY layer to correctly detect the end of the packet. In addition, the error checking header field is the CRC error detection field formed in 16 bits. MAC frames are sent with outputs ranging from 1 to 2, 5.5 or 11 Mbits with regard to the 802.11b [SHA 14, BAI 12].

– *ACK frame*

RA: Address of the receiving node (receiver address). This is the address indicated in the *Address 2* field of the frame preceding the ACK frame.

Duration is equal to zero or the value of the preceding *Duration* field minus the time requested to transmit the ACK frame and SIFS interval.

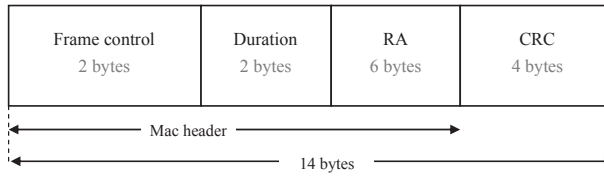

Figure 1.10. *ACK frame*

– *RTS frame*

TA: Address of transmitter node (transmitter address).

Duration is equal to the time necessary for the transmission of the management frame or of subsequent data, plus a CTS frame, plus an ACK frame and plus three SIFS intervals.

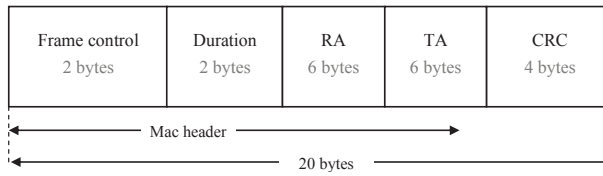

Figure 1.11. *RTS frame*

– *CTS frame*

RA: Address of the receiver of the CTS frame, directly copied from the TA field of the RTS.

Duration is equal to the RTS duration minus the transmission time of the CTS frame and one SIFS interval.

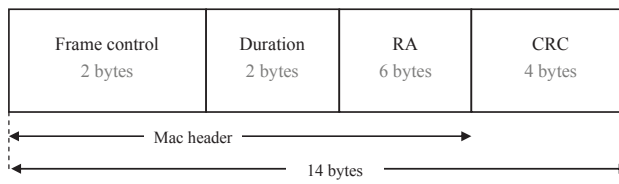

Figure 1.12. *CTS frame*

The frames used in standard 802.11 [ZHA 13a, PAV 14, LEE 14, SZO 14, SOR 14, SWA 14] usually follow the format shown in Figure 1.9; they are divided into three main types, specifically:

- *Data* frames for data transmission,

- *Control* frames to control access to the support (for example ACK, RTS and CTS),

- *Management* frames to exchange management information but without transmission to the upper layers.

At the time of transmission, a MAC header and a PHY header are added to the data generated by the upper layers. The transmission velocities of some of these parts can vary. The PHY header is sent at a constant velocity of 1 Mbit/sec, while the MAC header can be sent at a velocity that may reach 12 Mbit/sec. Control and acknowledgment packets are sent at different velocities.

An example of a 1,032-byte frame sent at 11 Mbit/sec can be broken down as follows:

– A DIFS with 50 μsec, a backoff with 0 to 31 TS of length 20 usec (0 to 620 μsec),

– *Data* frame of 987.7 μsec with:

  - *PHY header* of 192 μsec at 1 Mbit/sec,

  - *MAC header* of 24.7 μsec at 11 Mbit/sec,

  - *Useful data* of 771 μsec at 11 Mbit/sec.

– A *SIFS* of 10 μsec and the 192-μsec *PHY header acknowledgment* at 1 Mbit/sec,

– *MAC header acknowledgment* of 56 μsec at 2 Mbit/sec.

Theoretically, the maximum possible outputs can be calculated according to the size of the packets used during transmission.

1.4. Energy consumption in *ad hoc* networks

Energy sources in *ad hoc* networks are provided mainly by cells or batteries, which power the nodes during their operations in the

network. These batteries are of limited capacity and can cover node activity for a reduced period of time in terms of several parameters: number and type of operations carried out; types of transmitters and protocols used in the network; and network mobility distance. Nodes in *ad hoc* networks generally consume energy [FEE 01] when they transmit data toward a pre-determined destination (transmission) or when they receive data (reception), as well as when they listen to the channel and during hibernation. Without energy, the nodes cannot function, and in this event, the network remains inactive. Since the energy supplied to mobile nodes is practically limited by sources (batteries) with a short lifespan, the saving, control and optimization of this energy in *ad hoc* network are of vital importance and currently pos one of the greatest challenges facing scientific researchers.

Sources of energy consumption in *ad hoc* networks can be linked to two principal operations: consumption related to communications and consumption related to the processing and analysis of information. Communication practically requires the use of transmitters at the source, en route and upon arrival at the receiving node. The objective of the transmitter is to generate original packets, control the route and redirect packets toward another recipient node. The objective of the receiver is to receive data, control the packets received and transfer packets toward other destinations. The quantity of energy consumed varies therefore according to the type of transmitter used. An example is given in [MAK 07] of a Proxim RangeLAN2 2.4 GHz 1.6 Mbps Personal Computer Memory Card International Association (PCMCIA) card, which requires 1.5 W for transmission, 0.75 W for reception and 0.01 W in hibernation mode; thus, switching between transmission and reception for a node occurs in between 6 and 30 μsec. In addition, there is an example of the Lucent 15 dBm 2.4 GHz PCMCIA card, which consumes 1.82 W in transmission, 1.80 W in reception and 0.18 W in hibernation mode. Though not nil, power consumption in hibernation mode can be disregarded in comparison to other consumption values. Figure 1.13 shows the consumption of some transmission modules from the firm Ericsson, used to establish a high-frequency (HF) link in a global system for mobile communication (GSM) cell site.

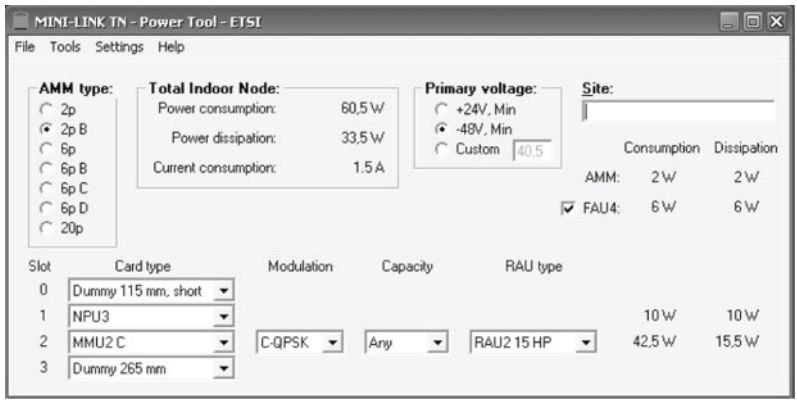

Figure 1.13. *Variation in energy consumption according to types of transmission modules*

Consumption related to the processing and analysis of data before transmission and after reception can generally be summed up in operations of calculation, sampling, modulation, encoding, decoding, filtering, compression, A/N conversion, etc. This is demonstrated in [MAK 07]: an msp430 micro-controller can consume between 14 and 23 mW depending on family, while the quasi-delay insensitive (QDI) 8-bit CISC MICA asynchronous microcontroller, which functions at 23.8 Mips (2.5 V), can consume up to 28 mW. Consequently, the more radio communications take place, the more the microcontroller must make calculations and the higher the proportion of energy consumed by it on the overall energy tally. The technique of data compression, which is used to reduce packet length, causes additional energy consumption due to the increased number of processing and calculation operations. Therefore, a relationship exists between the two sources of energy consumption, and protocols intended to diminish communications consumption may cause an increase in consumption related to data processing and analysis. In order to optimize energy consumption, we must control the balance between the two sources of this energy consumption.

An example of an underwater *ad hoc* network is given in [JUR 07]. The curve shows the influence of transmitter–receiver distance and frequency on the average lifespan of batteries. Battery life is

constantly reduced with the increase in distance. When the distances between nodes are small and the nodes can transmit at low frequencies, the impact of the medium's absorption will be negligible, and the majority of the energy consumption is due to signal attenuation. On the other hand, transmission at high frequencies over long distances greatly reduces battery life.

1.4.1. *Energy overconsumption and/or waste*

Though energy consumption in *ad hoc* networks generally occurs via various types of operations and in all the layers of the network, the largest amount of consumption is due to communications. MAC protocols include sources of energy overconsumption and sometimes waste in their details, and these sources should be reduced or eliminated.

– Collisions in the MAC layer [KRA 98, YE 02] are responsible for retransmission phenomena that cause an overconsumption of energy, resulting in delays causing considerable losses in both energy and time (output). In reality, we cannot definitively eliminate retransmissions that occur because of errors made during transmission and during collisions. However, several solutions and techniques for reducing them are suggested in the literature [STE 97, JON 01, HAC 03, SIV 00, SIN 98, LI 01, SCH 01].

– We have seen that nodes in inactive or idle mode consume energy without carrying out any operations other than listening to the channel. This results in a loss of energy that can be minimized via practical techniques such as occasionally putting nodes into hibernation mode or turning off the transmitter for a pre-determined amount of time. The power aware multi access protocol with signaling (PAMAS), IEEE 802.11 PSM, sparse topology and energy management (STEM) and sensor-MAC (S-MAC) protocols, which are explained in [SIN 98, GAD 04, KAR 90], respectively, have developed this approach.

– In addition, large amounts of time and energy are gratuitously lost during the switching of mobile radios between Tx mode and Rx mode, and vice versa; this is the case with protocols based on slot-by-slot technology. An algorithm is suggested in [GAD 04] to avoid this

energy loss, in which case reception and transmission slots are considered separately.

– Poor quality of the communication medium can cause many transmission errors and thus a high error rate, while the packets involved in these error transmission become useless, which means that the energy used during their transmission will be lost. In similar channel conditions, transmissions can be avoided until the channel is reestablished in order to avoid energy waste [ZOR 97]. Error control protocols based on automatic repeat request (ARQ) and forward error correction (FEC) techniques can also be used to conserve energy; these protocols are explaned in [LET 97, WOO 01, CAN 00].

– Most of these routing protocols rely on the number of hops as a metric in the choice of paths, while other protocols such as associativity-based routing (ABR) and signal-stability adaptive routing (SSA) rely mainly on link quality [JUB 87]. Unfortunately, these routing metrics have a negative impact on the lifespan of nodes and of the network in general, due to energy overconsumption for some nodes in favor of others.

– Choosing the wrong type of routing protocol can cause an undesirable overconsumption of energy. In [CHE 01], comparisons between several routing protocols based on experimental measurements have shown that the DSDV and temporally ordered routing algorithm (TORA) protocols increase energy consumption by 51.8% compared to the DSR and AODV protocols, which have a low energy consumption rate and greater stability in the network.

The lack of cohesion between MAC and routing protocols results in a risk of producing cutoffs and ruptures in network connectivity due to the inequitable use of nodes, as the routing protocol may call upon low-energy nodes that the MAC protocol has put into hibernation mode, and the MAC protocol may put high-energy nodes into hibernation mode that might be selected by the routing protocol to transport information. A solution is examined in [MES 11] that consists of establishing a sort of switching between the different protocols to achieve better energy conservation (up to 14%).

1.4.2. *Toward more efficient energy consumption*

Controlling energy consumption in *ad hoc* networks is of vital importance given the limited energy sources available, which are usually in the form of batteries or cells. During their activities, nodes consume energy in a non-uniform manner, and thus energy distribution in the network is inequitable. This is due to the quality of operating policies followed in the different layers of the network. Therefore, targeting efficient energy consumption calls on all of the layers in an *ad hoc* network. However, the MAC layer has the most influence on the energetic behavior of the network. We saw in the last section how MAC protocols can be responsible for energy overconsumption and even waste due to node behavior that is not well structured. Moreover, routing protocols play a large role in the loss or saving of energy, and cohesion between these two protocols leads to a balanced use of the nodes in *ad hoc* networks.

There are currently three axes of development for the optimization of energy consumption by communications in *ad hoc* networks:

– Energy savings in the case of the problem of energy loss in inactive mode. This is a matter of maximizing the duration of the nodes' hibernation mode.

– Control of transmission power, which consists of increasing network capacity and transporting data at minimal energy cost by allowing nodes to determine the minimum transmission power sufficient to maintain network connectivity.

– Load distribution, the principal objective of which is to balance energy consumption among mobile nodes.

In the next part of this chapter, we will examine the various solutions contributed for each network layer (according to IEEE standard 802.11) for efficient energy consumption in MANETs.

1.4.2.1. *Data link layer*

1.4.2.1.1. MAC sublayer

The MAC sublayer is responsible for providing reliability to the upper layers of the network in point-to-point connections established

by the PHY layer. The objective of this layer is to minimize simultaneous accesses that cause collisions. MAC protocols are grouped into two major categories according to the method of access in each protocol; there are fixed and random protocols.

Energy management in the IEEE 802.11 protocol

The idea behind the 802.11 protocols is to wait for a random period of BT when the channel becomes free before starting transmission. The backoff mechanism limits the risk of collision but does not eliminate it completely, and when a collision occurs, a new backoff will be automatically initiated. However, on each consecutive collision, the size of the window will be doubled in order to reduce the chance that such collisions will occur again. This approach reduces the number of collisions, and thus conserves node energy; in addition, it introduces additional delays and a considerable lowering of output in the *ad hoc* network.

PAMAS protocol

This protocol is a development of an older protocol, MACA [KAR 90], with the introduction of a separate channel designed for control messages (RTS, CTS and busy tone). It is intended to save energy in *ad hoc* networks. In PAMAS, before a node begins data transmission, it must send an RTS via the control channel and await the CTS response of the receiving node. If it does not receive the CTS, the node enters a backoff period. However, if it receives the CTS, the node transmits its packets via the data channel. At this time, the receiver node sends a busy tone message through the control channel to inform the nodes monitoring the control channel that the data channel is occupied. At that point, any nodes unable to transmit or receive are instructed to turn off their radio interfaces to save energy.

In PAMAS, the use of an independent control channel allows nodes to determine when and for how long the radios will remain turned off. A node must turn off its radio interface either when it has no data to transmit and is not concerned by a transmission from a neighboring node or when it has packets to transmit but a transmission is in the process of being executed by its neighboring nodes. Each

node can determine the amount of time during which it must turn off its radio using the probe protocol, as explained in [SIN 98]. In summary, PAMAS functions based on the principle that when a node is free/empty (free medium), no energy will be consumed. Research [AKY 05] has shown that the PAMAS protocol reduces energy consumption by at least 50% in networks with large communication loads (0.5–3 packets/sec/node). Numerous routing algorithms have been suggested for *ad hoc* networks with the objective of saving energy; some energy-optimizing routing protocols are presented below.

Geographic adaptive fidelity (GAF) protocol

GAF is a routing protocol that uses the node localization technique in *ad hoc* networks via the use of GPS. This protocol consists of creating virtual grids based on the area involved in routing by dividing this area into small zones such that for two adjacent grids Gx and Gy, all the nodes in Gx can communicate with all the nodes in Gy in order to ensure permanent network connectivity. One condition is necessary in this protocol, which is that a single node must be active in each grid, and that the other nodes in the same grid must be in hibernation mode for a fixed period of time. This technique saves energy and confirms the fidelity of the network. However, in some environments where nodes are highly mobile, routing fidelity may be reduced if an active node leaves the grid, which can cause data loss.

Switched port analyzer (SPAN) protocol

In the SPAN protocol [GAD 04], routing is done with the aid of coordinators. The selection of a coordinator node is based on the energy level it possesses and the number of neighboring node pairs it can connect. Coordinators remain permanently active in order to ensure multi-hop routing in the network, while the other nodes remain in hibernation mode and verify periodically whether they should activate to become a coordinator, or not yet. Via this process, SPAN ensures energy savings for nodes in hibernation mode on the one hand, and equitable energy distribution via the tactic of alternating nodes to replace the coordinator on the other hand, which results in a considerable extension of the lifespan of the network. Experiments

have shown that the SPAN protocol gives better results than the GAF protocol, though GAF is simpler to implement than SPAN.

1.4.2.1.2. LLC sublayer

This layer is responsible for the security of communications via controlling errors in transmission, data encryption/decryption and packet retransmission. There are two main families of techniques used for error control: ARC technique [ATE 08] and FEC technique. These techniques are known for non-optimized use of bandwidth and energy due to successive packet retransmissions and the time elapsed during error correction. Great care must be taken when using these techniques in a wireless connection, where the error rate is higher due to noise, fading and disconnections caused by node mobility. In addition, equilibrium in this layer must be maintained between the various characteristics of the network in order to improve output, security and energy efficiency; for example, the improvement of channel quality via an encoding system may reduce output through the addition of redundant bits in packet transmission. In addition, if transmission power is increased to avoid interference, the batteries are exhausted and the duration of operation of MANETs is reduced. For this reason, recent research has suggested alternative techniques and new protocols for error control with more moderate energy consumption.

Adaptive error control with ARQ

A new protocol is proposed in [HEC 05, RAZ 14] for error control with optimized energy consumption. This protocol is based on the following three principles:

– Avoiding persistence during data retransmission,

– controlling the number of retransmissions according to probability of success,

– avoiding transmissions in mediocre channel conditions.

In this protocol, ARQ functions normally until an error is detected in the control channel following the absence of ACK signals; at that point, the protocol enters probing mode, in which a probing packet is sent to each t slot. This packet contains only the input bit and not the data bits so that it consumes less energy. Probing mode continues until

the reception of a correct ACK signal, indicated that the channel has been reestablished, and the protocol reverts to normal mode and continues the transmission from the cutoff point. The results obtained in [HAN 04] show that in a channel with fading, the developed ARQ protocol is better than the classic ARQ in terms of energy consumption and number of packets transmitted.

1.5. Conclusion

Every day, *ad hoc* networks demonstrate their importance in daily life via the benefits they bring to services in different areas of economic, social and cultural activity. However, medium access and energy control remain critical points that are preventing us from benefiting from the maximum capacity of these networks. In this chapter, we have introduced *ad hoc* networks and their different characteristics, and we have discussed in detail the medium access phenomenon and the major problems present in the MAC layer. As we have emphasized, output and QoS currently constitute the major challenges of medium access protocols experiencing saturations and limitations with regard to multimedia applications and real-time services.

The energy factor is of primary importance in the management and control of *ad hoc* networks, notably in the evaluation of routing and medium access protocols. In this vein, we explained in the last section of this chapter the impact of energy on the functioning of *ad hoc* networks and the relationship between the effectiveness of these protocols and energy consumption. In addition, we have introduced the various solutions contributed by the MAC and LLC communication protocols attempting to guarantee more efficient energy consumption in dynamic MANET systems, which require additional energy in order to cover node mobility.