1

Introduction

Wireless mobile ad hoc networks consist of mobile nodes interconnected by wireless multi-hop communication paths. Unlike conventional wireless networks, ad hoc networks have no fixed network infrastructure or administrative support. The topology of such networks changes dynamically as mobile nodes join or depart the network or radio links between nodes become unusable. In this chapter, I will introduce wireless ad hoc networks, and discuss their inherent vulnerable nature. Considering the inherent vulnerable nature of ad hoc networks, a set of security requirements is subsequently presented. The chapter also introduces the quality of service issues that are relevant for ad hoc networks.

1.1 Ad hoc networking

Conventional wireless networks require as prerequisites a fixed network infrastructure with centralized administration for their operation. In contrast, so-called (wireless) mobile ad hoc networks, consisting of a collection of wireless nodes, all of which may be mobile, dynamically create a wireless network amongst themselves without using any such infrastructure or administrative support [1,2]. Ad hoc wireless networks are self-creating, self-organizing, and self-administering. They come into being solely by interactions among their constituent wireless mobile nodes, and it is only such interactions that are used to provide the necessary control and administration functions supporting such networks.

Mobile ad hoc networks offer unique benefits and versatility for certain environments and certain applications. Since no fixed infrastructure, including base stations, is prerequisite, they can be created and used "any time, anywhere." Such networks could be intrinsically fault-resilient, for they do not operate under the limitations of a fixed topology. Indeed, since all nodes are allowed to be mobile, the composition of such networks is necessarily time varying. Addition and deletion of nodes occur only by interactions with other

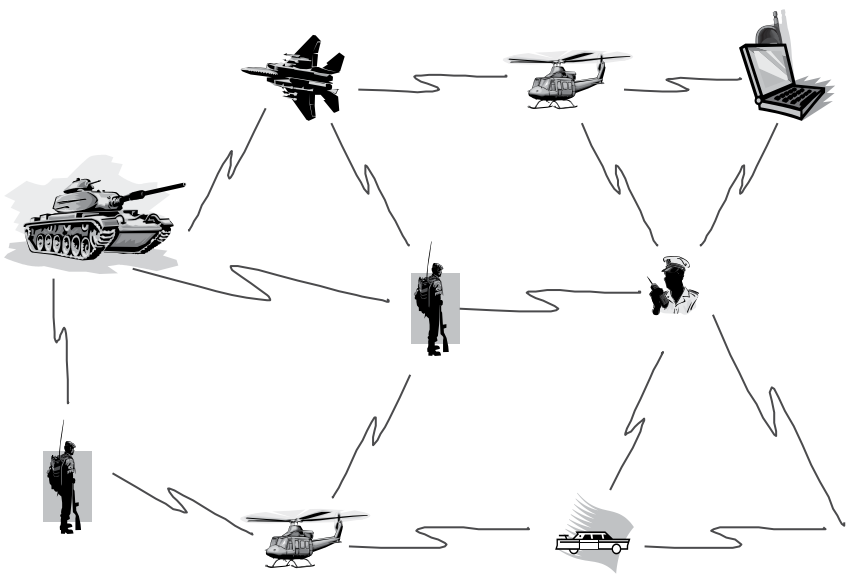

Figure 1.1 Conceptual representation of a mobile ad hoc network

nodes; no other agency is involved. Such perceived advantages elicited immediate interest in the early days among military, police, and rescue agencies in the use of such networks, especially under disorganized or hostile environments, including isolated scenes of natural disaster and armed conflict. See Fig. 1.1 for a conceptual representation. In recent days, home or small-office networking and collaborative computing with laptop computers in a small area (e.g., a conference or classroom, single building, convention center, etc.) have emerged as other major areas of application. These include commercial applications based on progressively developing standards such as Bluetooth [3], as well as other frameworks such as Piconet [4], HomeRF Shared Wireless Access Protocol [5], etc. In addition, people have recognized from the beginning that ad hoc networking has obvious potential use in all the traditional areas of interest for mobile computing.

Mobile ad hoc networks are increasingly being considered for complex multimedia applications, where various quality of service (QoS) attributes for these applications must be satisfied as a set of predetermined service requirements. As a minimum, the QoS issues pertaining to delay and bandwidth management are of paramount interest. In addition, because of the use of the ad hoc networks for military or police use, and of increasingly common commercial applications, various security issues need to be addressed. Cost-effective resolution of these issues at appropriate levels is essential for widespread general use of ad hoc networking.

Mobile ad hoc networking emerged from studies on extending traditional Internet services to the wireless mobile environment. All current works, as well as this presentation, consider the ad hoc networks as a wireless extension to the Internet, based on the ubiquitous IP networking mechanisms and protocols. Today's Internet possesses an essentially static infrastructure where network elements are interconnected over traditional wire-line technology, and these elements, especially the elements providing the routing or switching functions, do not move. In a mobile ad hoc network, by definition, all the network elements move. As a result, numerous more stringent challenges must be overcome to realize the practical benefits of ad hoc networking. These include effective routing, medium (or channel) access, mobility management, power management, and security issues, all of which affect the quality of the service experienced by the user.

The absence of a fixed infrastructure for ad hoc networks means that the nodes communicate directly with one another in a peer-to-peer fashion. The mobility of these nodes imposes limitations on their power capacity, and hence, on their transmission range; indeed, these nodes must often satisfy stringent weight limitations for portability. Mobile hosts are no longer just end systems; to relay packets generated by other nodes, each node must be able to function as a router as well. As the nodes move in and out of range with respect to other nodes, including those that are operating as routers, the resulting topology changes must somehow be communicated to all other nodes, as appropriate. In accommodating the communication needs of the user applications, the limited bandwidth of wireless channels and their generally hostile transmission characteristics impose additional constraints on how much administrative and control information may be exchanged, and how often. Ensuring effective routing is one of the great challenges for ad hoc networking.

The lack of fixed base stations in ad hoc networks means that there is no dedicated agency for managing the channel resources for the network nodes. Instead, carefully designed distributed medium access techniques must be used for channel resources, and, hence, mechanisms must be available to recover efficiently from the inevitable packet collisions. Traditional carrier sensing techniques cannot be used, and the hidden terminal problem [6,7] may significantly diminish the transmission efficiency [8]. An effectively designed protocol for medium access control (MAC) is essential to the quest for QoS.

1.2 The ad hoc wireless network: operating principles

I start with a description of the basic operating principles of a mobile ad hoc network. Figure 1.2 depicts the peer-level multi-hop representation of such a

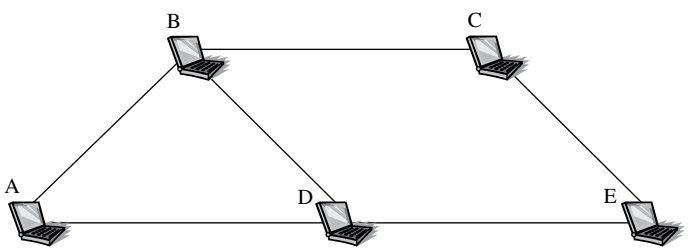

Figure 1.2 Example of an ad hoc network

network. Mobile node A communicates with another such node B directly (single-hop) whenever a radio channel with adequate propagation characteristics is available between them. Otherwise, multi-hop communication is necessary where one or more intermediate nodes must act as a relay (router) between the communicating nodes. For example, there is no direct radio channel (shown by the lines) between A and C or A and E in Fig. 1.2. Nodes B and D must, therefore, serve as intermediate routers for communication between A and C, and A and E, respectively. Indeed, a distinguishing feature of ad hoc networks is that all nodes must be able to function as routers on demand. To prevent packets from traversing infinitely long paths, an obvious essential requirement for choosing a path is that the path must be loop-free. A loop-free path between a pair of nodes is called a route.

An ad hoc network begins with at least two nodes broadcasting their presence (beaconing) with their respective address information. As discussed later, they may also include their location information, obtained, for example, by using a system such as the Global Positioning System (GPS), for more effective routing. If node A is able to establish direct communication with node B in Fig. 1.2, verified by exchanging suitable control messages between them, they both update their routing tables. When a third node, C, joins the network with its beacon signal, two scenarios are possible. The first is where both A and B determine that single-hop communication with C is feasible. In the second scenario, only one of the nodes, say B, recognizes the beacon signal from C and establishes the availability of direct communication with C. The distinct topology updates, consisting of both address and route updates, are made in all three nodes immediately afterwards. In the first case, all routes are direct. For the other, shown in Fig. 1.3, the route update first happens between B and C, then between B and A, and then again between B and C, confirming the mutual reachability between A and C via B.

The mobility of nodes may cause the reachability relations to change in time, requiring route updates. Assume that for some reason, the link between B and

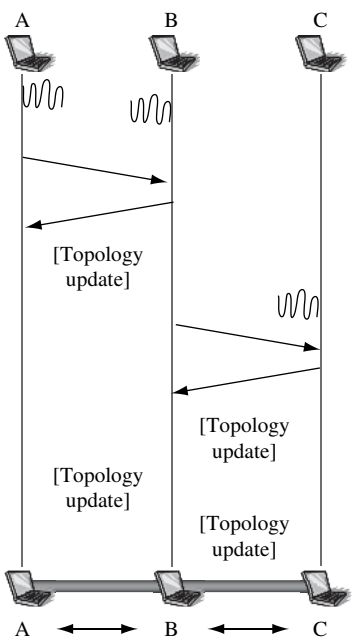

Figure 1.3 Bringing up an ad hoc network

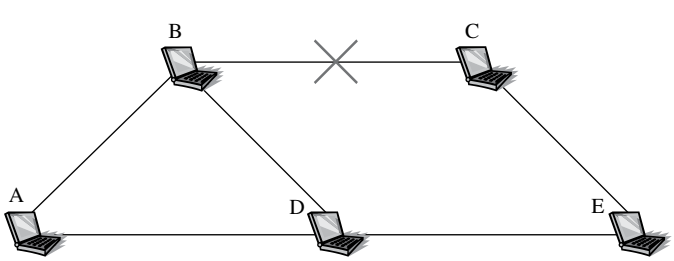

Figure 1.4 Topology update owing to a link failure

C is no longer available, as shown in Fig. 1.4. Nodes A and C can still reach each other, although this time only via nodes D and E. Equivalently, the original loop-free route $\langle A \leftrightarrow B \leftrightarrow C \rangle$ is now replaced by the new loop-free route $\langle A \leftrightarrow D \leftrightarrow E \leftrightarrow C \rangle$. All five nodes in the network are required to update their routing tables appropriately to reflect this topology change, which will be first detected by nodes B and C, then communicated to A and E, and then to D.

The reachability relation among the nodes may also change for other reasons. For example, a node may wander too far out of range, its battery may be depleted, or it may suffer a software or hardware failure. As more nodes join the network or some of the existing nodes leave, the topology

updates become more numerous, complex, and, usually, more frequent, thus diminishing the network resources available for exchanging user information.

Finding a loop-free path as a legitimate route between a source–destination pair may become impossible if the changes in network topology occur too frequently. Here, "too frequently" means that there was not enough time to propagate to all the pertinent nodes all the topology updates arising from the last network topology changes, or worse, before the completion of determining all loop-free paths accommodating the last topology changes. The ability to communicate degrades with accelerating rapidity as the knowledge of the network topology becomes increasingly inconsistent. Given a specific time-window, we call (the behavior of) an ad hoc network combinatorially stable if, and only if, the topology changes occur sufficiently slowly to allow successful propagation of all topology updates as necessary. Clearly, combinatorial stability is determined not only by the connectivity properties of the networks, but also by the complexity of the routing protocol in use and the instantaneous computational capacity of the nodes, among other factors. Combinatorial stability is an essential consideration for attaining QoS objectives in an ad hoc network, as we shall see below. I address the general issue of routing in mobile ad hoc networks separately in the next section.

The shared wireless environment of mobile ad hoc networks requires the use of appropriate medium access control (MAC) protocols to mitigate the medium contention issues, allow efficient use of limited bandwidth, and resolve so-called hidden and exposed terminal problems. These are basic issues, independent of the support of QoS; the QoS requirements add extra complexities for the MAC protocols, mentioned later in Chapter 5. The issues of efficient use of bandwidth and the hidden/exposed terminal problem have been studied exhaustively and are well understood in the context of accessing and using any shared medium. I briefly discuss the "hidden-terminal" problem [6] as an issue especially pertinent for the wireless networks.

Consider the scenario of Fig. 1.5, where a barrier prevents node B from receiving the transmission from D, and vice versa, or, as usually stated, B and D cannot "hear" each other. The "barrier" does not have to be physical; a large enough distance separating two nodes is the most commonly occurring "barrier" in ad hoc networks. Node C can "hear" both B and D. When B is transmitting to C, D, being unable to "hear" B, may transmit to C as well, thus causing a collision and exposing the *hidden-terminal* problem. In this case, B and D are "hidden" from each other. Now consider the case when C is transmitting to D. Since B can "hear" C, B cannot risk initiating a transmission to A for fear of causing a collision at C. Here is an example of the *exposed terminal* problem, where B is "exposed" to C.

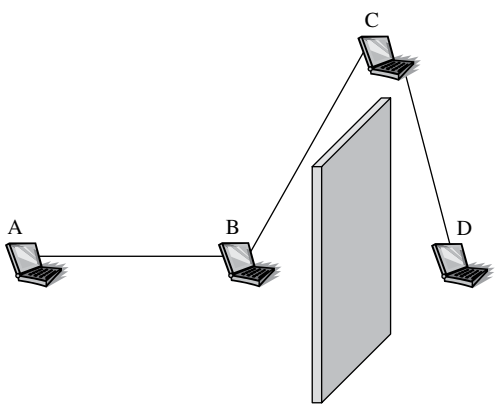

Figure 1.5 Example of hidden/exposed terminal problem

A simple message exchange protocol solves both problems. When D wishes to transmit to C, it first sends a request-to-send (RTS) message to C. In response, C broadcasts a clear-to-send (CTS) message that is received by both B and D. Since B has received the CTS message unsolicited, B knows that C is granting permission to send to a hidden terminal and hence refrains from transmitting. Upon receiving the CTS message from C in response to its RTS message, D transmits its own message.

Not only does the above (crude and deliberately simplified outline of the) dialogue solve the hidden terminal problem, but it solves the exposed terminal problem as well, for after receiving an unsolicited CTS message, B refrains from transmitting and cannot cause a collision at C. After an appropriate interval, determined by the attributes of the channel (i.e., duration of a time slot, etc.), B can send its own RTS message to C as the prelude to a message transmission.

Limitation on the battery power of the mobile nodes is another basic issue for ad hoc networking. Limited battery power restricts the transmission range (hence the need for each node to act as a router) as well as the duration of the active period for the nodes. Below some critical thresholds for battery power, a node will not be able to function as a router, thus immediately affecting the network connectivity, possibly isolating one or more segments of the network. Fewer routers almost always mean fewer routes and, therefore, increased likelihood of degraded performance in the network. Indeed, QoS obviously becomes meaningless if a node is not even able to communicate, owing to low battery power. Since exchange of messages necessarily means power consumption, many ad hoc networking mechanisms, especially routing and security protocols, explicitly include minimal battery power consumption as a design objective.

1.3 Ad hoc networks: vulnerabilities

There are various reasons why wireless ad hoc networks are at risk, from a security point of view. I next discuss the characteristics that make these networks vulnerable to attacks. Attacks are procedures that are launched by unauthorized entities or nodes within the networks to disrupt the normal operation of the enterprise.

The wireless links between nodes are highly susceptible to link attacks, which include passive eavesdropping, active interfering, leaking secret information, data tampering, impersonation, message replay, message distortion, and denial of service. Eavesdropping might give an adversary access to secret information, violating confidentiality. Active attacks might allow the adversary to delete messages, to inject erroneous messages, to modify messages, and to impersonate a node, thus violating availability, integrity, authentication, and nonrepudiation (these and other security needs are discussed in the next section).

Ad hoc networks do not have a centralized piece of machinery such as a name server or a base station, which could lead to a single point of failure and, thus, make the network that much more vulnerable. On the flipside, however, the lack of support infrastructure leads to prevention of application of standard techniques such as key management (discussed later in the book) to secure the network. This gives rise to the need for new schemes to ensure key agreement.

An additional problem that arises in ad hoc networks is the accurate detection of a compromised node. Usually compromised nodes are detected by monitoring their behavior. But in a wireless environment it is often difficult to distinguish between a truly misbehaving node and a node that appears to be misbehaving because of poor link quality. The presence of compromised nodes has the potential to cause Byzantine failures, which are encountered within mobile ad hoc network (MANET) routing protocols, wherein a set of the nodes could be compromised in such a way that the incorrect and malicious behavior cannot be directly noted at all. The compromised nodes may seemingly operate correctly, but, at the same time, they may make use of the flaws and inconsistencies in the routing protocol to distort the routing fabric of the network. In addition, such malicious nodes can also create new routing messages and advertize non-existent links, provide incorrect link state information and flood other nodes with routing traffic, thus inflicting Byzantine failures on the system. Such failures are especially severe because they may come from seemingly trusted nodes, whose malicious intentions have not yet been noted. Even if the compromised nodes were noticed and prevented from performing incorrect actions, the erroneous information generated by the Byzantine failures could have already been propagated through the network.

No part of the network is dedicated to support any specific network functionality. All nodes are expected to contribute to routing (topology discovery, data forwarding). The examples of functions that rely on a central service, and which are also of high relevance, are naming services, certification authorities, directory, and other administrative services. In ad hoc networks, nodes cannot rely on such a service. Even if such services were assumed, their availability would not be guaranteed, either due to the dynamically changing topology that could easily result in a partitioned network, or due to congested links close to the node acting as a server.

The absence of infrastructure and the consequent absence of authorization facilities impede the usual practice of establishing a line of defence, distinguishing nodes as trusted and non-trusted. Such a distinction would have been based on a security policy, the possession of the necessary credentials and the ability of nodes to validate them. In the case of wireless ad hoc networks, there may be no grounds for such a priori node classification, since all nodes are required to cooperate in supporting the network operation, while no prior security association can be assumed for all the network nodes.

Additionally, freely roaming nodes form transient associations with their neighbors; they join and leave sub-domains independently and without notice. Thus, it may be difficult, in most cases, to have a clear picture of the ad hoc network membership at a given time. Consequently, especially in the case of a large network, no form of established trust relationships among the majority of nodes can be assumed.

In such an environment, there is no guarantee that a path between two nodes would be free of malicious nodes. There is a possibility that a path consisting of malicious nodes may not comply with the rules of the protocol employed and can attempt to disrupt the network operation. The mechanisms currently incorporated in ad hoc routing protocols cannot cope with disruptions due to malicious behavior. For example, any node could claim that it is one hop away from the sought destination, causing all routes to the destination to pass through itself. Alternatively, a malicious node could corrupt any in-transit route request (reply) packet and cause data to be misrouted.

The presence of even a small number of adversarial nodes could result in repeatedly compromised routes, and, as a result, the network nodes would have to rely on cycles of timeout and new route discoveries to communicate. This would incur arbitrary delays before the establishment of a non-corrupted path, while successive broadcasts of route requests would impose excessive transmission overhead. In particular, intentionally falsified routing messages would result in a denial-of-service (DoS) experienced by the end nodes.

The dynamic and transient nature of an ad hoc network can result in constant changes in trust among nodes. This can create problems, for example, with key management, if cryptography is used in the routing protocol. It must not be trivial, for example, to recover private keys from the device. Evidence that tampering has occurred would be required so as to distinguish a tampered node from the rest. Standard security solutions would not be good enough since they are essentially for statically configured systems. This gives rise to the need for security solutions, which adapt to the dynamically changing topology and movement of nodes in and out of the network.

Moreover, the battery-powered operation of ad hoc networks gives attackers ample opportunity to launch a denial-of-service attack by creating additional transmissions or expensive computations to be carried out by a node in an attempt to exhaust its batteries.

In addition, sensor networks (a form of wireless ad hoc network) are made up of devices that tend to have limited computational abilities. For example, the working memory of a sensor node is insufficient even to hold the variables (of sufficient length to ensure security) that are required in asymmetric cryptographic algorithms, let alone perform operations on them. This may exclude techniques such as frequent public key cryptography during normal operation. A particular challenge is that of broadcasting authenticated data to the entire sensor network. Current proposals for authenticated broadcast rely on asymmetric digital signatures for the authentication, and these are impractical for many reasons (e.g., long signatures with high communication overheads of 50–1000 bytes per packet; very high overheads to create and verify the signature) for sensor networks.

Lastly, scalability is another issue, which has to be addressed when security solutions are being thought of, for the simple reason that an ad hoc network may consist of hundreds or even thousands of nodes. Many ad hoc networking protocols are applied in conditions where the topology must scale up and down efficiently, e.g., because of network partitions or mergers. The scalability requirements here refer to the scalability of individual security services such as key management for example.

The above discussion makes it clear that ad hoc networks are inherently insecure, more so than their wireline counterparts, and need robust security schemes that take into consideration the inherently susceptible nature of these networks. Coming up with a security scheme, in general, necessitates the discussion of the fundamental components that make up security. In the next section, I take a look at the essential security needs of such networks. By this, I mean the factors that ought to be taken into consideration when designing a security scheme.