Lecture Notes

on

Wireless Sensor Networks

for

Bachelor of Technology (RA15)

(IV Year I Semester)

Prepared by

Mr. D. Mahesh

Assistant Professor

SR Engineering College

Innovation . Creativity . Entrepreneurship

Department of Computer Science and Engineering

S R Engineering College

Ananthasagar (V), Hasanparthy (M)

Warangal Urban, Telangana-506371.

SYLLABUS

UNIT – I

Introduction: Fundamentals of wireless communication technology, the electromagnetic spectrum radio propagation, characteristics of wireless channels, modulation techniques,multiple access techniques, wireless LANs, PANs, WANs, and MANs, Wireless Internet.

UNIT – II

Introduction to Ad-hoc / Sensor Networks: Key definitions of adhoc/sensor networks, unique constraints and challenges, advantages of ad-hoc/sensor network, driving applications, issues in adhoc wireless networks, issues in design of sensor network, sensor network architecture, data dissemination and gathering.

UNIT –III

MAC Protocols : Issues in designing MAC protocols for adhoc wireless networks, design goals, classification of MAC protocols, MAC protocols for sensor network, location discovery, quality, other issues, S-MAC, IEEE 802.15.4.

UNIT – IV

Routing Protocols: Issues in designing a routing protocol, classification of routing protocols, table-driven, on-demand, hybrid, flooding, hierarchical, and power aware routing protocols.

UNIT – V

QoS and Energy Management : Issues and Challenges in providing QoS, classifications, MAC, network layer solutions, QoS frameworks, need for energy management, classification, battery, transmission power, and system power management schemes.

TEXT BOOKS:

1. C. Siva Ram Murthy and B. S. Manoj, “Ad-hoc Wireless Networks”, Pearson Education, 2008.
2. Carlos De Morais Cordeiro and Dharma Prakash Agrawal, “Ad-hoc and Sensor Networks: Theory and Applications”, World Scientific Publishing Company, 2006.

REFERENCE BOOKS:

1. Holger Karl and Andreas willig, “Protocols and Architectures for Wireless Sensor Networks”, John Wiley & Sons, Inc., 2005.
2. C.K.Toh, “Ad-hoc Mobile Wireless Networks”, Pearson Education, 2002.
3. Erdal Cayirci and Chunming Rong, “Security in Wireless Ad Hoc and Sensor Networks”, John Wiley and Sons, 2009.
4. Charles E.Perkins, “Ad-hoc Networking”, Pearson Education, 2001.
5. Shih-Lin Wu and Yu-Chee Tseng, “Wireless Ad-hoc Networking”, Auerbach Publications, Taylor & Francis Group, 2007.

UNIT –I

Introduction to wireless communication technology

Fundamentals of wireless communication technology:

Wireless Communication is a broad and dynamic field that has spurred tremendous excitement and technological advance over the last few decades. Wireless Communication is, by any measure, the fastest growing segment of the communications industry. As such as it has captured the attention of the media and the imagination of the public. Cellular systems have experienced exponential growth over the last decade and there are currently about two billion users worldwide.

Evolution of Wireless Communication

There are several smaller steps that take place in leading up to the development of a new technology . Tracing the development of these earlier discoveries in brief can help us better understand how this technology actually functions and contributes towards what could be the next development. A brief review of the history of wireless communications covering radio, television, radar, satellite, wireless and mobile cellular and other wireless networks are presented in the following paragraph.

Radio and Television Communications

In 1874, Marconi performed simple experiments to send signals using electromagnetic waves at short distances of only about 100 meters. At that time scientists and experts believed that electromagnetic waves could only be transmitted in a straight line, and the main obstacle to radio transmission was the curvature of the earth's surface. Finally Marconi successfully experimented to prove that electromagnetic wave transmission was possible between two distant points even through obstacles in between.

Radar Communication

Radar has been recognized as one of the greatest scientific developments of the first half of the 20th century. The first practical radar system was produced in 1935 by the British physicist Robert Watson-Watt.

Radar is an active remote-sensing system that operates on the principle of echoes. A Radar display shows a map like picture of the area being scanned. The centre of the picture corresponds to the radar antenna and the radar echoes are shown as bright spots on the screen.

Satellite Communication

A satellite is an object that orbits or revolves around another object. Satellites can be sent into space through a variety of launch vehicles. Sir Isaac Newton in the 1720s was probably the first person to conceive the idea of a satellite. In 1945, Arthur C Clarke a science fiction envisioned a network of a communication satellite. Three satellites would be able to transmit signals around the world by transmitting in a line-of-sight direction with other orbiting satellites.

Cellular Communication

In 1946, American Telephone & Telegraph (AT&T) introduced the first American commercial mobile radio telephone service to private customers. It consisted of a central transmitter with one antenna which could serve a wide area.

1.1 The electromagnetic spectrum radio propagation:

Wireless communication is based on the principle of broadcast and reception of electromagnetic waves. These waves can be characterized by their frequency (f) or their wavelength (λ). Frequency is the number of cycles (oscillations) per second of the wave and is measured in Hertz (Hz). The speed of propagation of these waves (c) varies from medium to medium, except in a vacuum where all electromagnetic waves travel at the same speed, the speed of light. The relation between the above parameters can be given as

$$c = \lambda \times f$$

where c is the speed of light ($3 \times 108 m/s$), f is the frequency of the wave in Hz, and λ is its wavelength in meters.

Table 1 **Frequency bands and their common uses**

Band Name	Frequency	Wavelength	Applications
Extremely Low Frequency (ELF)	30 to 300 Hz	10,000 to 1,000 Km	Powerline frequencies
Voice Frequency (VF)	300 to 3,000 Hz	1,000 to 100 Km	Telephone communications
Very Low Frequency (VLF)	3 to 30 KHz	100 to 10 Km	Marine communications
Low Frequency (LF)	30 to 300 KHz	10 to 1 Km	Marine communications
Medium Frequency (MF)	300 to 3,000 KHz	1,000 to 100 m*	AM broadcasting
High Frequency (HF)	3 to 30 MHz	100 to 10 m	Long-distance aircraft/ship communications
Very High Frequency (VHF)	30 to 300 MHz	10 to 1 m	FM broadcasting
Ultra High Frequency (UHF)	300 to 3,000 MHz	100 to 10 cm	Cellular telephone
Super High Frequency (SHF)	3 to 30 GHz	10 to 1 cm	Satellite communications, microwave links
Extremely High Frequency (EHF)	30 to 300 GHz	10 to 1 mm	Wireless local loop
Infrared	300 GHz to 400 THz	1 mm to 770 nm	Consumer electronics
Visible Light	400 THz to 900 THz	770 nm to 330 nm	Optical communications

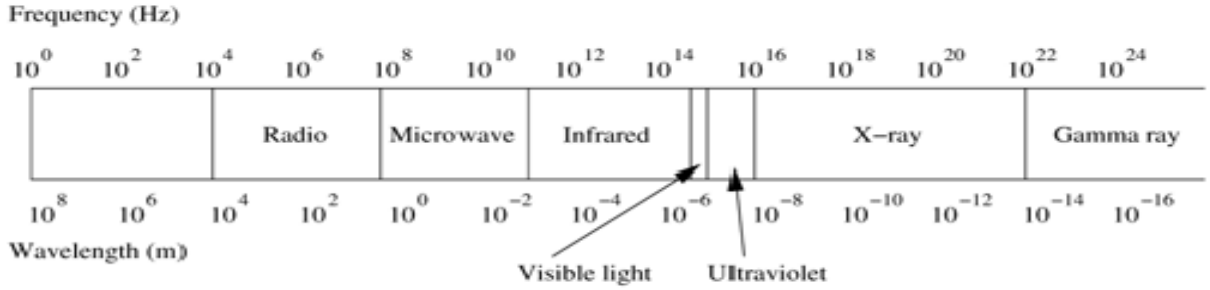


Figure 1. The electromagnetic spectrum

The low-frequency bands comprised of the radio, microwave, infrared, and visible light portions of the spectrum can be used for information transmission by modulating the amplitude, frequency, or the phase of the waves. Radio waves are easy to generate and are widely used for both indoor and outdoor communication due to properties such as their ability to pass through buildings and ability to travel long distances.

1) Radio waves

- Easy to generate and are widely used for both indoor and outdoor communication due to properties such as their ability to pass through buildings and ability to travel long distances.
- Radio transmission is omnidirectional.

2) Propagation waves

- VLF, LF, and MF bands the propagation of waves, also called as ground waves, follows the curvature of the Earth.
- The maximum transmission ranges of these waves are of the order of a few hundred kilometers.
- The HF and VHF band transmissions are absorbed by the atmosphere near the Earth's surface. However, a portion of the radiation, called the sky wave radiates outward and upward to the ionosphere in the upper atmosphere.

3) Microwave

- Microwave transmissions (in the SHF band) tend to travel in straight lines and hence can be narrowly focused.
- Microwaves were widely used for long-distance telephony, before they got replaced by fiber optics.
- They are also widely used for mobile phones and television transmission. Since the energy is concentrated.

4) Infrared waves

- Infrared waves and waves in the EHF band (also known as millimeter waves) are used for short range communication.
- They are widely used in television, VCR, and stereo remote controls.

5) Visible light

- The visible light part of the spectrum is just after the infrared portion.
- Unguided optical signaling using visible light provides very high bandwidth at a very low cost.
- But the main disadvantage here is that it is very difficult to focus a very narrow unidirectional laser beam, which limits the maximum distance between the transmitter and receiver.

In the VLF, LF, and MF bands the propagation of waves, also called as ground waves, follows the curvature of the Earth. The maximum transmission ranges of these waves are of the order of a few hundred kilometers. They are used for low bandwidth transmissions such as amplitude modulated.

1.2 Radio Propagation Mechanisms

Radio waves generally experience the following three propagation mechanisms:

• Reflection:

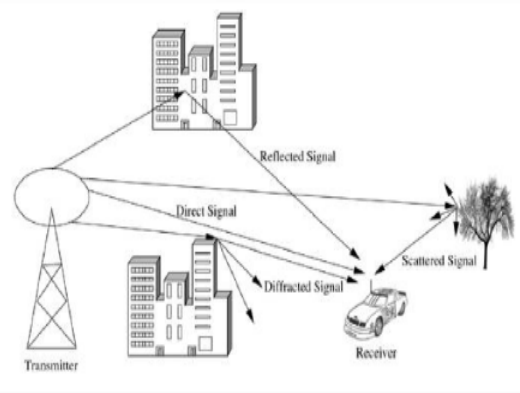
When the propagating radio wave hits an object which is very large compared to its wavelength (such as the surface of the Earth, or tall buildings), the wave gets reflected by that object. Reflection causes a phase shift of 180 degrees between the incident and the reflected rays.

• Diffraction:

This propagation effect is undergone by a wave when it hits an impenetrable object. The wave bends at the edges of the object, thereby propagating in different directions. This phenomenon is termed as diffraction. The dimensions of the object causing diffraction are comparable to the wavelength of the wave being diffracted. The bending causes the wave to reach places behind the object which generally cannot be reached by the line-of-sight transmission. The amount of diffraction is frequency-dependent, with the lower frequency waves diffracting more.

• Scattering:

When the wave travels through a medium, which contains many objects with dimensions small when compared to its wavelength, scattering occurs. The wave gets scattered into several weaker outgoing signals. In practice, objects such as street signs, lamp posts, and foliage cause scattering.



1.3 Characteristics of wireless channels:

1. **Path Loss**

- Path loss can be expressed as the ratio of the power of the transmitted signal to the power of the same signal received by the receiver, on a given path. It is a function of the propagation distance.
- Path loss is dependent on a number of factors such as the radio frequency used and the nature of the terrain.
- So, several models are required to describe the variety of transmission environments. There are two path loss model,
  - i. Free propagation model
  - ii. Two ray model or two path model

i. Free propagation model

- The simplest path loss model in which there is a direct-path signal between the transmitter and the receiver, with no atmospheric attenuation or multipath components.

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2$$

- The relationship between the transmitted power *Pt* and the received power *Pr* is given by Where *Gt* and *Gr* are the transmitter and receiver antenna gains, 1 respectively, in the direction from the transmitter to the receiver, *d* is the distance between the transmitter and receiver, and $\lambda = c/f$ (is the wavelength of the signal).

ii. Two ray model

The signal reaches the receiver through two paths, one a line-of sight path, and the other the path through which the reflected (or refracted, or scattered) wave is received. According to the two-path model, the received power is given by

$$P_r = P_t G_t G_r \left(\frac{h_t h_r}{d^2}\right)^2$$

Where *Pt* is the transmitted power, *Gt* and *Gr* represent the antenna gains at the transmitter and the receiver, respectively, *d* is the distance between the transmitter and receiver, and *ht* and *hr* are the heights of the transmitter and the receiver, respectively**.**

2. Fading

Fading refers to the fluctuations in signal strength when received at the receiver. Fading can be classified into two types:

i. Fast fading/small-scale fading
ii. Slow fading/large-scale fading.

i. **Fast fading** refers to the rapid fluctuations in the amplitude, phase, or multipath delays of the received signal, due to the interference between multiple versions (copies) of the same transmitted signal arriving at the receiver at slightly different times.
   - The time between the reception of the first version of the signal and the last echoed signal is called delay spread. The multipath propagation of the transmitted signal, which causes fast fading.
   - The multipath propagation of the transmitted signal, which causes fast fading. The multiple signal paths may sometimes add constructively or sometimes destructively at the receiver, causing a variation in the power level of the received signal.

ii. **Slow fading** occurs when objects that partially absorb the transmissions lie between the transmitter and receiver.
   - Slow fading is so called because the duration of the fade may last for multiple seconds or minutes.

- Slow fading may occur when the receiver is inside a building and the radio wave must pass through the walls of a building, or when the receiver is temporarily shielded from the transmitter by a building.
- Slow fading is also referred to as ***shadow fading*** since the objects that cause the fade, which may be large buildings or other structures, block the direct transmission path from the transmitter to the receiver.

Some common measures to overcome the fading effect are

(a) Diversity (b)Adaptive modulation

(a) Diversity Mechanism

- Based on the fact that independent paths between the same transmitter and receiver nodes experience independent fading effects. By providing multiple logical channels between the transmitter and receiver, and sending parts of the signal over each channel, the error effects due to fading can be compensated.

i. ***Time diversity mechanisms*** aim at spreading the data over time so that the effects of burst errors are minimized.

ii. ***Frequency diversity mechanisms*** spread the transmission over a wider frequency spectrum, or use multiple carriers for transmitting the information.

iii. ***Space diversity*** involves the use of different physical transmission paths.

(b) Adaptive Modulation Mechanisms

- The channel characteristics are estimated at the receiver and the estimates are sent by the receiver to the transmitter through a feedback channel.
- The transmitter adapts its transmissions based on the received channel estimates in order to counter the errors that could occur due to the characteristics of the channel. Adaptive techniques are usually very complex to implement.

3. Interference

- Wireless transmissions have to counter interference from a wide variety of sources. Two main forms of interference are adjacent channel interference and co-channel interference.

i. ***Adjacent channel interference*** case, signals in nearby frequencies have components outside their allocated ranges. These components may interfere with on-going transmissions in the adjacent frequencies. It can be avoided by carefully introducing guard bands2 between the allocated frequency ranges.

ii. ***Co-channel interference***, sometimes also referred to as narrow-band interference, is due to other nearby systems the same transmission frequency. Narrow-band interference due to frequency reuse in cellular systems can be minimized with the use of multiuser detection.

1.4 Modulation Techniques

Data (whether in analog or in digital format) has to be converted into electromagnetic waves for transmission over a wireless channel. The techniques that are used to perform this conversion are called modulation techniques. The modulation process alters certain properties of a radio wave, called a carrier wave, whose frequency is the same as the frequency of the wireless channel being used for the transmission. Modulation schemes can be classified under two major categories: analog modulation schemes and digital modulation schemes. This classification is based on the nature of the data, analog or digital, to be transmitted. Some of the commonly used modulation techniques are discussed below.

1.4.1 Analog Modulation

As the name implies, analog modulation techniques are used for transmitting analog data. The analog data signal is superimposed on a carrier signal. This superimposition is aimed at altering a certain property (amplitude or frequency) of the carrier signal. Some of the commonly used analog modulation techniques are amplitude modulation, frequency modulation, and phase modulation. These techniques are described below.

Amplitude Modulation

Amplitude modulation (AM) is one of the simplest modulation schemes. It was the first method to be used for transmitting voice. The transmitter superimposes the information signal x(t), also called the modulating signal, on the carrier signal c(t). The result of AM of the carrier signal of Figure 1.1 (b), by the information/modulating signal of Figure 1.1 (a), is shown in Figure 1.1 (c). It can be seen that the frequency of the modulated wave remains constant, while its amplitude varies with that of the information signal. When a cosine carrier wave is used, the AM wave can be mathematically represented as below:

$$s(t) = (1 + n_a x(t))\ cos(2\pi f_c t)$$

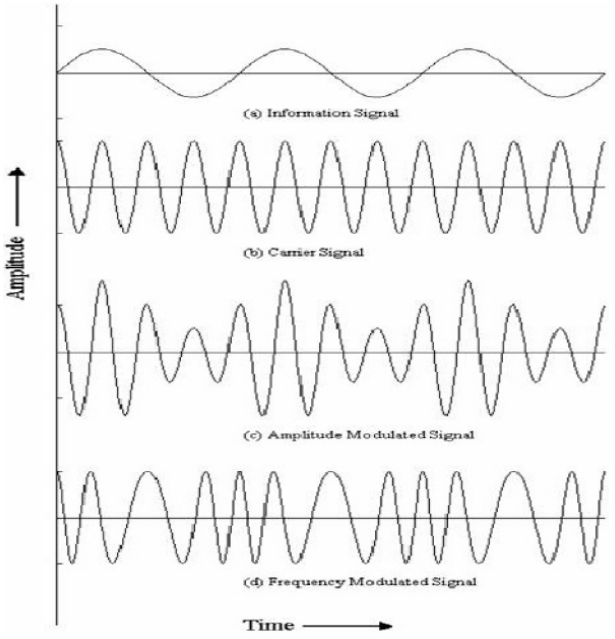


Fig 1.1. Analog modulation schemes.

Where *n*a, known as the modulation index, is the ratio of the amplitude of the information signal to that of the carrier signal, *fc* is the frequency of the carrier signal, $x(t)$ is the information signal, and $c(t) = cos(2\pi fct)$ is the carrier signal.

Angle Modulation

Frequency modulation and phase modulation come under this category. The angle modulated signal can be mathematically represented as

$$s(t) = A_c\left[cos(2\pi f_c t) + \Phi(t)\right]$$

Where *Ac* is the amplitude and *fc* is the frequency of the carrier signal. Frequency modulation and phase modulation are special cases of angle modulation. Changing the frequency of a wave also changes its phase, and vice versa. The angle modulated signal has constant amplitude, and as a consequence the transmitter operates at full power constantly, which also maximizes its range.

Frequency Modulation

In frequency modulation (FM), the amplitude of the modulated signal is kept constant, while the instantaneous frequency is altered to reflect the information signal that is being transmitted. With the condition that the derivative of the phase, that is, $\Phi'(t)$, is proportional to the information signal $x(t)$. Since frequency can also be defined as the rate of change of phase of the signal, here $\Phi'(t)$ represents the deviation of the instantaneous frequency of the modulated signal from that of the carrier signal. In FM, this $\Phi'(t)$ is directly proportional to the information signal/modulating signal. It can be represented as

$$\Phi'(t) = n_f x(t)$$

Where *nf* is a constant, known as the frequency modulation index. The instantaneous frequency of the carrier wave is changed according to the amplitude of the information signal, resulting in the stretching or the compressing of the carrier wave depending on the value of the modulating voltage. Some common applications where FM is used are radio broadcasts and first-generation cellular phones.

Phase Modulation

In phase modulation (PM), the phase of the modulated signal $\Phi(t)$ is directly proportional to the information signal $x(t)$. It is represented as

$$\Phi(t) = n_p x(t)$$

where *np* is a constant, known as the phase modulation index. The instantaneous phase deviation of the modulated signal from the carrier signal is $\Phi(t)$. In PM, this $\Phi(t)$ is proportional to the information signal/modulating signal.

Digital Modulation

Digital modulation schemes are used for transmitting digital signals that consist of a sequence of 0 and 1 bits. As in analog modulation, digital modulation also changes a certain property of the carrier signal. The main difference between analog and digital modulation is that while the changes occur in a continuous manner in analog modulation, they occur at discrete time intervals

in digital modulation. Some of the basic digital modulation techniques such as amplitude shift keying, frequency shift keying, and phase shift keying are described below.

Amplitude Shift Keying

In amplitude shift keying (ASK), when a bit stream is transmitted, a binary 1 is represented by the presence of the carrier signal $c(t)$ for a specified interval of time, and a binary 0 is represented by the absence of the carrier signal for the same interval of time. Mathematically, ASK can be represented as

$$s(t) = \begin{cases} A_c \cos(2\pi f_c t), & \text{for binary 1} \\ 0, & \text{for binary 0} \end{cases}$$

where *Ac* is the amplitude of the carrier signal and *fc* is its frequency. The result of ASK when applied to the bit pattern shown in Figure 1.2 (a) is shown in Figure 1.2 (b).

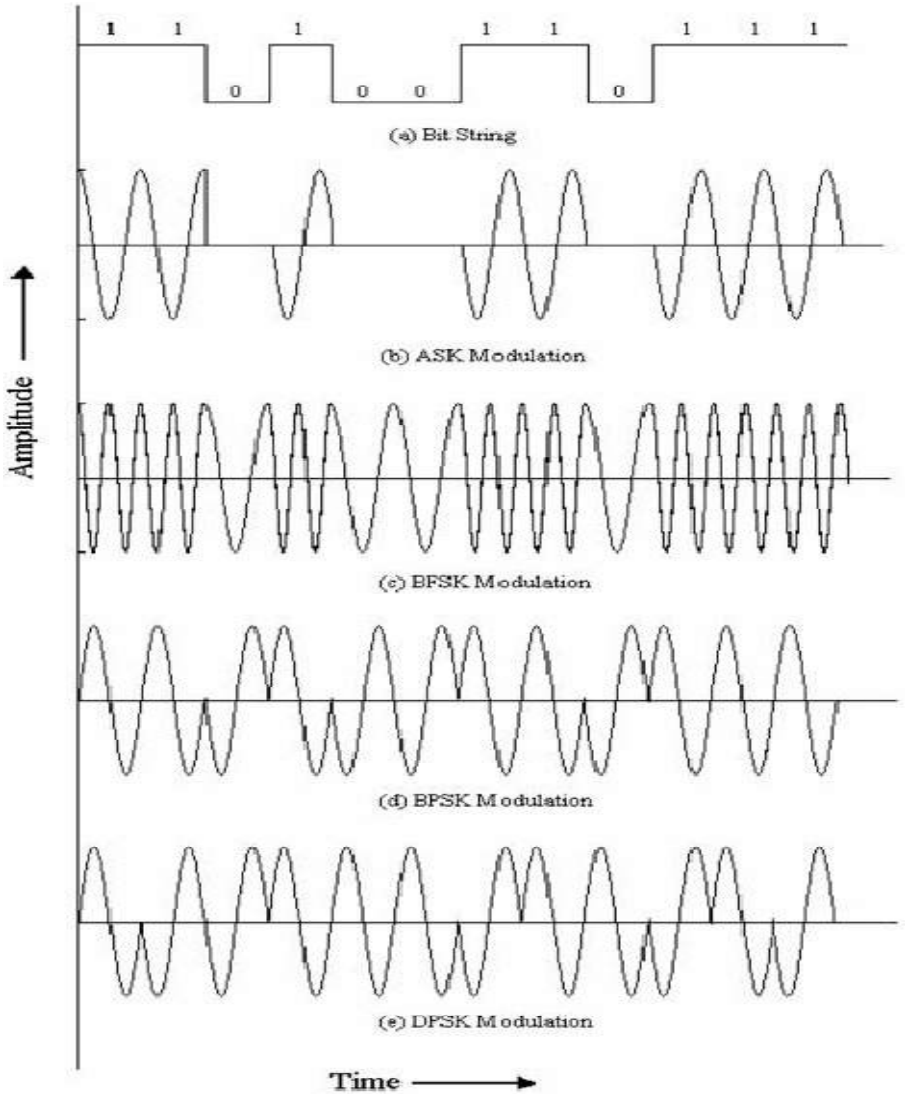


Figure 1.2. Digital modulation schemes.

Frequency Shift Keying

Frequency shift keying (FSK) is done as follows. If *fc* is the frequency of the carrier signal and if *k* is a small frequency offset, then transmission of a binary 1 is represented by the presence of a carrier signal of frequency *fc* + *k* for a specified interval of time. Since two frequency levels are used in this technique, it is also known as two-level FSK or binary FSK (BFSK). FSK can be mathematically represented as follows:

$$s(t) = \begin{cases} A_c\ cos(2\pi(f_c + k)t), & \text{for binary 1} \\ A_c\ cos(2\pi(f_c - k)t), & \text{for binary 0} \end{cases}$$

where *Ac* and *fc* are the amplitude and frequency, respectively, of the cosine carrier signal. FSK when applied to the bit pattern of Figure 2.2 (a) is shown in Figure 2.2 (c).

Phase Shift Keying

In phase shift keying (PSK), change in phase of the carrier signal is used to represent the 0 and 1 bits. The transmission of bit 0 is represented by the presence of the carrier for a specific interval of time, while the transmission of bit 1 is represented by the presence of a carrier signal with a phase difference of π radians for the same interval of time. PSK using a cosine carrier wave with amplitude *Ac* and frequency *fc* can be mathematically represented as

$$s(t) = \begin{cases} A_c\ cos(2\pi f_c t + \pi), & \text{for binary 1} \\ A_c\ cos(2\pi f_c t), & \text{for binary 0} \end{cases}$$

This technique is also known as binary PSK (BPSK) or two-level PSK since a single phase difference is used for representing 0 and 1 bits. Figure 2.2 shows the BPSK modulation of the bit pattern of Figure 2.2 (a).

Just as multiple frequency levels are used in FSK, multiple phase deviations can be used in PSK. This enables encoding of multiple bits by each phase representation. Quadrature PSK (QPSK), for example, uses four different phases each separated by π/2 radians. This would enable transmission of two bits per phase shift. The mathematical representation of QPSK is given below.

$$s(t) = \begin{cases} A_c\ cos\left(2\pi f_c t + \frac{\pi}{4}\right), & \text{for binary 10} \\ A_c\ cos\left(2\pi f_c t + \frac{3\pi}{4}\right), & \text{for binary 11} \\ A_c\ cos\left(2\pi f_c t + \frac{5\pi}{4}\right), & \text{for binary 01} \\ A_c\ cos\left(2\pi f_c t + \frac{7\pi}{4}\right), & \text{for binary 00} \end{cases}$$

π/4 shifted QPSK (π/4-QPSK) is a QPSK mechanism where the maximum phase deviation is limited to ± 135 degrees. The main advantage of π/4 shifted QPSK is that it can be received non-coherently, that is, the receiver need not lock to the phase of the transmitted signal, which simplifies the receiver design. It provides the bandwidth efficiency of QPSK along with lesser fluctuations in amplitude.

1.5 Multiple Access Techniques:

Multiple access techniques are based on the orthogonalization of signals, each signal represented as a function of time, frequency, and code. Hence, multiplexing can be performed with respect to one of these three parameters; the respective techniques are termed frequency division multiple access, time division multiple access, and code division multiple access. A brief discussion of each of the above techniques is presented below.

1.5.1 Frequency Division Multiple Access

The frequency division multiple access (FDMA) mechanism operates as below. The available bandwidth is divided into multiple frequency channels/bands. A transmitter–receiver pair uses a single dedicated frequency channel for communication. Figure 1.3 depicts the principle behind the operation of FDMA. The frequency spectrum is in effect divided into several frequency sub-bands. Transmissions on the main band of a channel also result in the creation of additional signals on the side bands of the channel. This is the main disadvantage of FDMA. FDMA has been widely adopted in analog systems for portable telephones and automobile telephones.

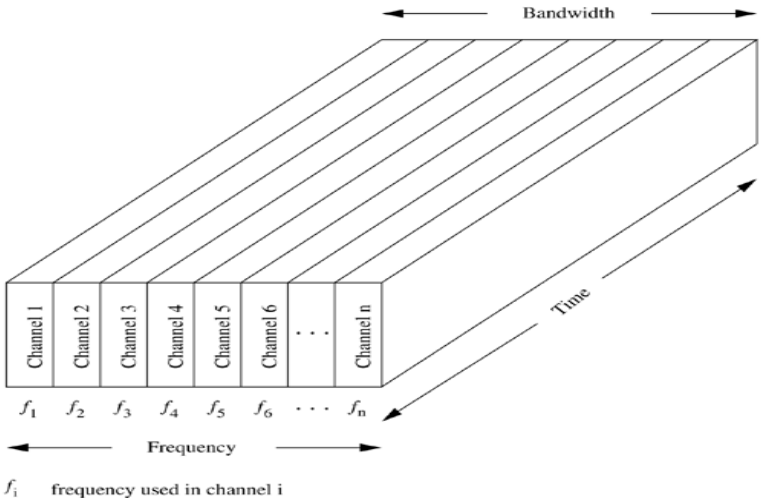


Figure 1.3. Illustration of FDMA.

In a cellular network, a central controller known as the base station (BS) dynamically allocates a different carrier frequency to each node, known as the mobile station (MS). This system, used for two-way communication between a pair of stations (MS and BS here), is called *frequency division duplexing* (FDD). Since high frequency transmissions suffer greater attenuation when compared to low frequency transmissions, high transmission power is required for high frequency channels for compensating the transmission losses.

1.5.2 Orthogonal Frequency Division Multiplexing

Orthogonal frequency division multiplexing (OFDM) is a multi-carrier transmission mechanism. It resembles FDMA in that both OFDM and FDMA split the available bandwidth into a number of frequency channels. OFDM is sometimes also referred to as discrete multi-tone (DMT) modulation. OFDM is currently used in several applications such as wireless local area networks (WLANs) and digital broadcasting.

1.5.3 Time Division Multiple Access

Time division multiple access (TDMA) shares the available bandwidth in the time domain. Each frequency band is divided into several time slots (channels). A set of such periodically repeating time slots is known as the TDMA frame. Each node is assigned one or more time slots in each frame, and the node transmits only in those slots. Figure 1.4 depicts the concept behind TDMA.

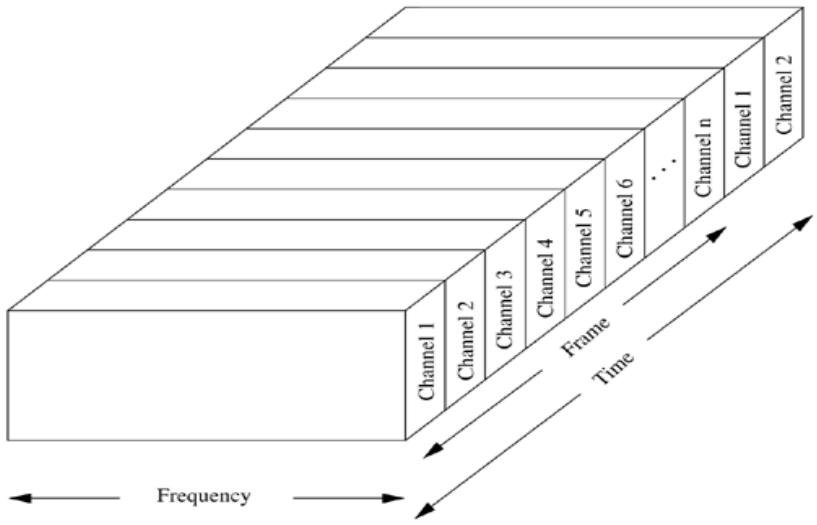


Figure 1.4. Illustration of TDMA.

FDMA requires the device to have the capability of simultaneously receiving and transmitting signals, which leads to increased cost. But when TDMA is used, the device can switch between slots and hence use the same transmitter for receiving also. Hence, the equipment cost in TDMA is less. TDMA is widely used in second-generation cellular systems such as GSM, etc.

1.5.4 Code Division Multiple Access

Unlike other systems such as TDMA and FDMA, code division multiple access (CDMA) does not assign a specific frequency to each user. Instead, every channel uses the entire spectrum. Individual conversations are encoded with a pseudorandom digital sequence. Two types of spread spectrum systems are widely in use today, namely, *frequency hopping spread spectrum* and *direct sequence spread spectrum,* which are described below.

- ***Frequency Hopping Spread Spectrum***

Frequency hopping spread spectrum (FHSS) is a simple technique in which the transmission switches across multiple narrow-band frequencies in a pseudorandom manner, that is, the sequence of transmission frequencies is known both at the transmitter and the receiver, but appears random to other nodes in the network.

The first transmission (darker shade in figure) uses the hopping sequence *f*4 *f*7 *f*2 *f*1 *f*5 *f*3 *f*6 *f*2 *f*3 and the second transmission uses the hopping sequence *f*1 *f*3 *f*6 *f*2 *f*4 *f*2 *f*7 *f*1 *f*5. Frequency hopped systems are limited by the total number of frequencies available for hopping. FHSS can be classified into two types: fast FHSS and slow FHSS. In fast FHSS, the dwell time on each frequency is very small, that is, the rate of change of frequencies is much higher than the information bit rate, resulting in each bit being transmitted across multiple frequency hops. In slow FHSS, the dwell time on each frequency is high, hence multiple bits are transmitted on each frequency hop. FHSS is now used mainly for short-range radio signals, particularly those in the unlicensed bands.

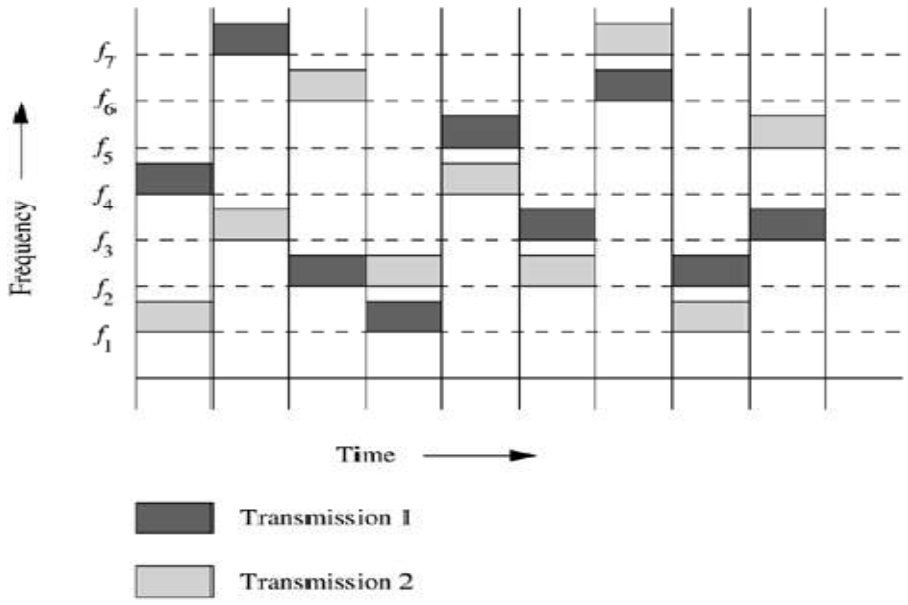
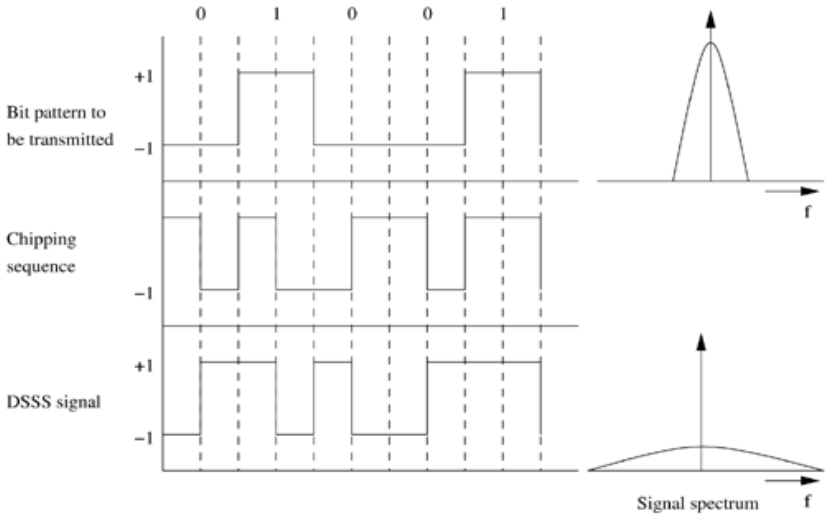


Figure 1.5. Illustration of FHSS.

- ***Direct Sequence Spread Spectrum***

The principle behind direct sequence spread spectrum (DSSS) can be explained easily by the following analogy. Suppose conversations in several languages are occurring in a room. People who understand only a particular language listen to and follow the conversation taking place in that language alone. Each node transmits using its code. At the receiver, the transmission is received and information is extracted using the transmitter's code. For transmitting a binary 1, the sender transmits its code; for a binary 0, the one's complement of the code is transmitted. Hence, transmission of a signal using CDMA occupies *n* times the bandwidth that would be required for a narrow-band transmission of the same signal.



Complementary code keying (CCK) is a modulation technique used in conjunction with DSSS. In CCK, a set of 64 8-bit code words is used for encoding data for the 5.5 Mbps and 11 Mbps data rates in the 2.4 GHz band of the IEEE 802.11 wireless networking standard

1.5.5 Space Division Multiple Access

The fourth dimension in which multiplexing can be performed is space. Instead of using omnidirectional transmissions (as in FDMA, TDMA, and CDMA) that cover the entire circular region around the transmitter, space division multiple access (SDMA) uses directional transmitters/antennas to cover angular regions. Figure 1.6 shows how SDMA could be used for satellite communication.

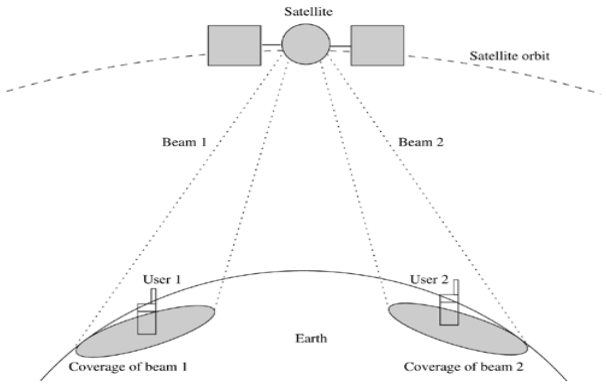


Figure 1.6 Illustration of SDMA.

1.6 Wireless LANs:

A wireless local area network (WLAN) is a wireless distribution method for two or more devices that use high-frequency radio waves and often include an access point to the Internet. A WLAN allows users to move around the coverage area, often a home or small office, while maintaining a network connection.

Every component that connects to a WLAN is considered a station and falls into one of two categories: access points (APs) and clients. APs transmit and receive radio frequency signals with devices able to receive transmitted signals; they normally function as routers. Clients may include a variety of devices such as desktop computers, workstations, laptop computers, IP phones and other cell phones and Smartphones. All stations able to communicate with each other are called basic service sets (BSSs), of which there are two types: independent and infrastructure. Independent BSSs (IBSS) exist when two clients communicate without using APs, but cannot connect to any other BSS. Such WLANs are called a peer-to-peer or an ad-hoc WLANs. The second BSS is called an infrastructure BSS. It may communicate with other stations but only in other BSSs and it must use APs.

1.7 Wireless PANs:

A WPAN (wireless personal area network) is a personal area network - a network for interconnecting devices centered around an individual person's workspace - in which the connections are wireless. Typically, a wireless personal area network uses some technology that permits communication within about 10 meters - in other words, a very short range. One such technology is Bluetooth, which was used as the basis for a new standard, IEEE 802.15.

A key concept in WPAN technology is known as *plugging in*. In the ideal scenario, when any two WPAN-equipped devices come into close proximity (within several meters of each other) or within a few kilometers of a central server, they can communicate as if connected by a cable. Another important feature is the ability of each device to lock out other devices selectively, preventing needless interference or unauthorized access to information.

1.8 WANS (Wide Area Networks)

A wide area network (WAN) is a network that exists over a large-scale geographical area. A WAN connects different smaller networks, including local area networks (LANs) and metro area networks (MANs). This ensures that computers and users in one location can communicate with computers and users in other locations. WAN implementation can be done either with the help of the public transmission system or a private network.

A WAN connects more than one LAN and is used for larger geographical areas. WANs are similar to a banking system, where hundreds of branches in different cities are connected with each other in order to share their official data. A WAN works in a similar fashion to a LAN, just on a larger scale. Typically, TCP/IP is the protocol used for a WAN in combination with devices such as routers, switches, firewalls and modems.

1.9 WMANs (Wireless metropolitan area networks)

A *Wireless Metropolitan Area Network* (WMAN) is also known as a Wireless Local Loop (*WLL*). WMANs are based on the *IEEE 802.16* standard. Wireless local loop can reach effective transfer speeds of 1 to 10 Mbps within a range of 4 to 10 kilometres, which makes it useful mainly for telecommunications companies.

The best-known wireless metropolitan area network is WiMAX, which can reach speeds on the order of 70 Mbps over a radius of several kilometers.

1.10 Wireless Internet

Wireless Internet refers to the extension of the services offered by the Internet to mobile users, enabling them to access information and data irrespective of their location. The inherent problems associated with wireless domain, mobility of nodes, and the design of existing protocols used in the Internet; require several solutions for making the wireless Internet a reality.

The major issues that are to be considered for wireless Internet are the following.

- Address mobility
- Inefficiency of transport layer protocols
- Inefficiency of application layer protocols

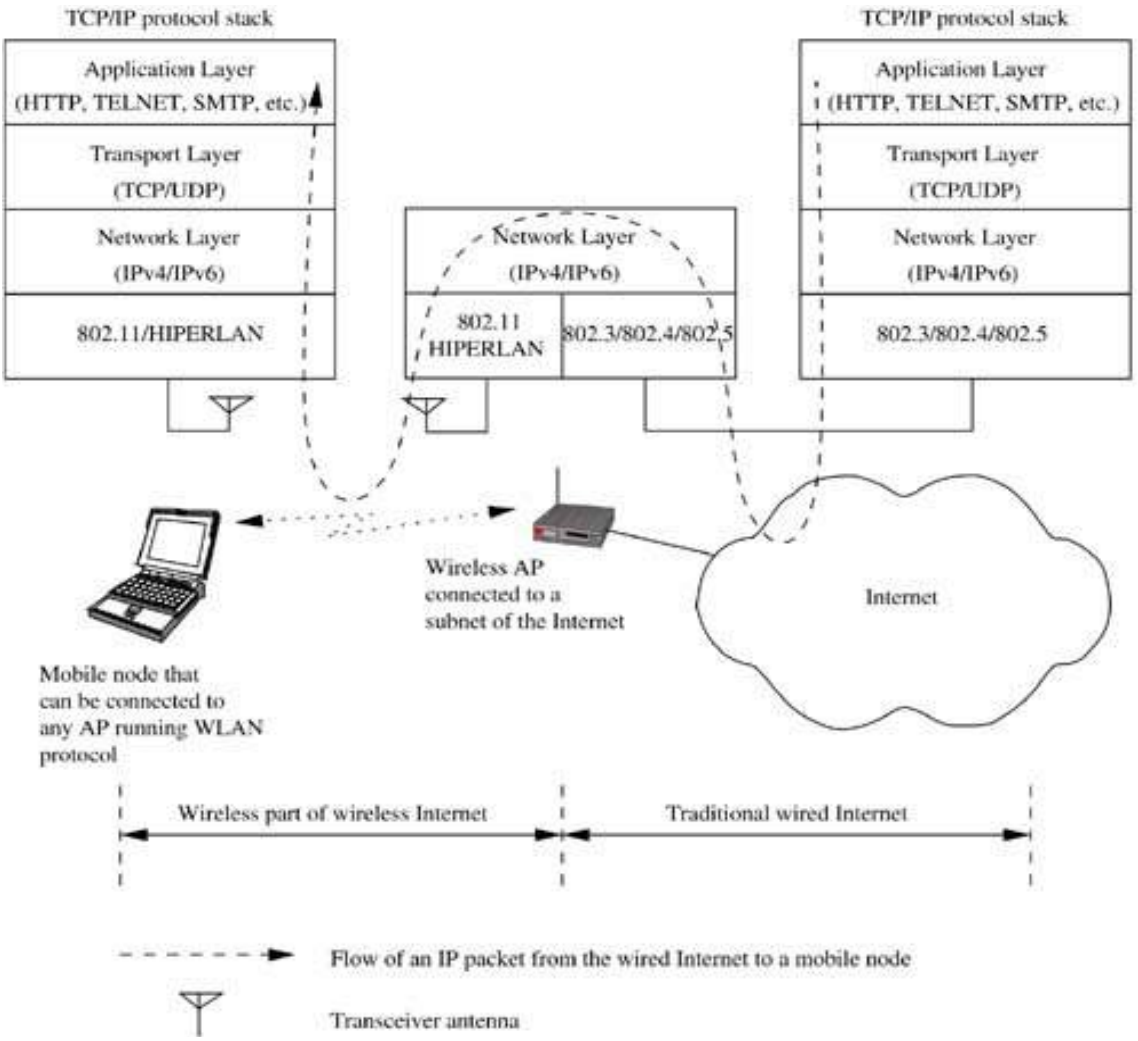


Figure 1.7. An illustration of wireless Internet.

Address Mobility

The network layer protocol used in the Internet is Internet protocol (IP) which was designed for wired networks with fixed nodes. IP employs a hierarchical addressing with a globally unique 32-bit address1 which has two parts, network identifier and host identifier, as shown in Figure 1.8 (a). The network identifier refers to the subnet address to which the host is connected. This addressing scheme was used to reduce the routing table size in the core routers of the Internet, which uses only the network part of the IP address for making routing decisions. This addressing scheme may not work directly in the wireless extension of the Internet, as the mobile hosts may move from one subnet to another, but the packets addressed to the mobile host may be delivered to the old subnet to which the node was originally attached, as illustrated in Figures 1.8 (b) and 1.8 (c).

Figure 1.8 shows the mobility of a node (with IP address 10.6.6.1) attached to subnet A (subnet address 10.6.6.x) moving over to another subnet B with address 10.6.15.x. In this case, the packets addressed to the node will be routed to the subnet A instead of the subnet B, as the network part in the mobile node's address is 10.6.6.x (see Figure 1.8 (c)). MobileIP2 is a solution that uses an address redirection mechanism for this address mobility issue in wireless Internet.

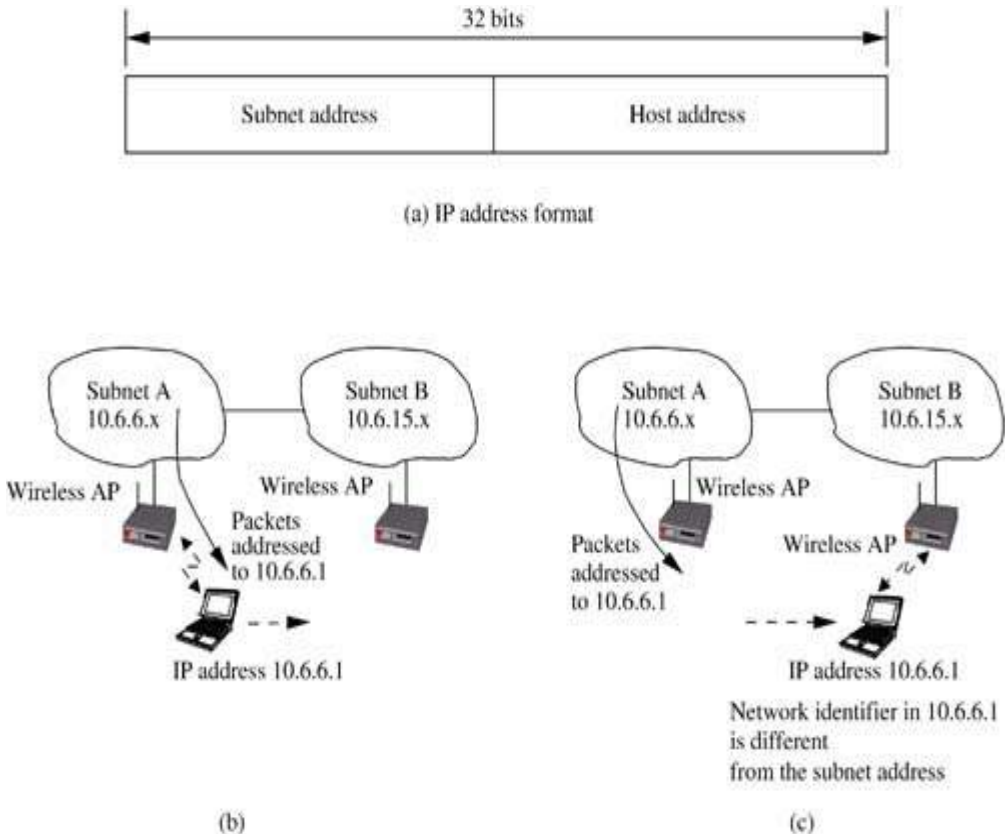


Figure 1.8. The address mobility problem.

Inefficiency of Transport Layer Protocols

The transport layer is very important in the Internet as it ensures setting up and maintaining end-to-end connections, reliable end-to-end delivery of data packets, flow control, and congestion control.TCP is the predominant transport layer protocol for wired networks, even though UDP, a connectionless unreliable transport layer protocol, is used by certain applications.

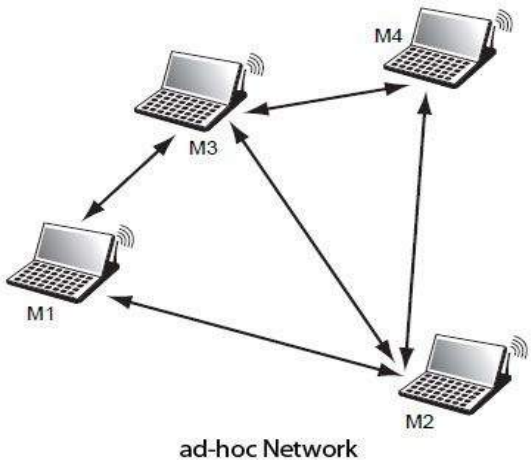
Inefficiency of Application Layer Protocols

Traditional application layer protocols used in the Internet such as HTTP,3 TELNET, simple mail transfer protocol (SMTP), and several markup languages such as HTML were designed and optimized for wired networks. Many of these protocols are not very efficient when used with wireless links. The major issues that prevent HTTP from being used in wireless Internet are its stateless operation, high overhead due to character encoding, redundant information carried in the HTTP requests, and opening of a new TCP connection with every transaction.

Unit-II

Ad-hoc Sensor Networks

Adhoc Network: An Ad hoc network will typically have a dynamic topology, which will have profound effects on network characteristics. Network nodes will often be battery powered, which limits the capacity of CPU, memory, and bandwidth. Ad hoc networks must also support communication between nodes that are only indirectly connected by a series of wireless hops through other nodes.



ad-hoc Network

Sensor Network: A Wireless sensor network can be defined as a network of devices that can communicate the information gathered from a monitored field through wireless links. The data is forwarded through multiple nodes, and with a gateway, the data is connected to other networks like wireless Ethernet.

2.1 Unique constraints and challenges

The ad hoc networks are self-forming, self-maintaining, self-healing architecture. The challenges are, no fixed access point, dynamic network topology, contrary environment and irregular connectivity. Ad hoc network immediately forms and accommodate the modification and limited power. Finally, ad hoc have no trusted centralized authority. Due to the dynamic changing property, the ad hoc faces some challenges which are listed in the below sections.

- **Quality of Service (QoS)**

The ad hoc network is dynamically creating the organization whenever the node wants to communicate with their neighbor node. Due the dynamic changing topology in ad hoc network, providing QoS is a tedious task. QoS are essential because of rapid development in mobile technology and real time applications like multimedia, voice. Providing QoS in ad hoc network is necessary to maintain best-effort-of service.

- Limited resources: Due to the dynamic changing of link flow, ad hoc network provide varying resources.
- Sufficient admission control: The admission control take decision about whether the available bandwidth is enough for link flow in available resources. Ad hoc networks providing finite bandwidth capacity may affect the end quality of service.

- Highly dynamic: Ad hoc network characteristics are dynamic changing topology and this dynamic changing occurs due to radio transmission and mobility.

- **Scalability**

The scalability problem occurs in ad hoc networks due to the nature of the multi-hop. The scalability in ad hoc network depends on the network size and forwarded packet capacity in the network.

Advantages of ad-hoc network:

The rapid development in ad hoc technology is widely used in portable computing such as laptop, mobile phone used to access the web services, telephone calls when the user are in travelling. Development of self-organizing network decrease the communication cost.

The growth of 4G technology enhances anytime, anywhere, anyhow communication in ad hoc network. Ad hoc network is simple to design and install. The advantages of an ad hoc network include:

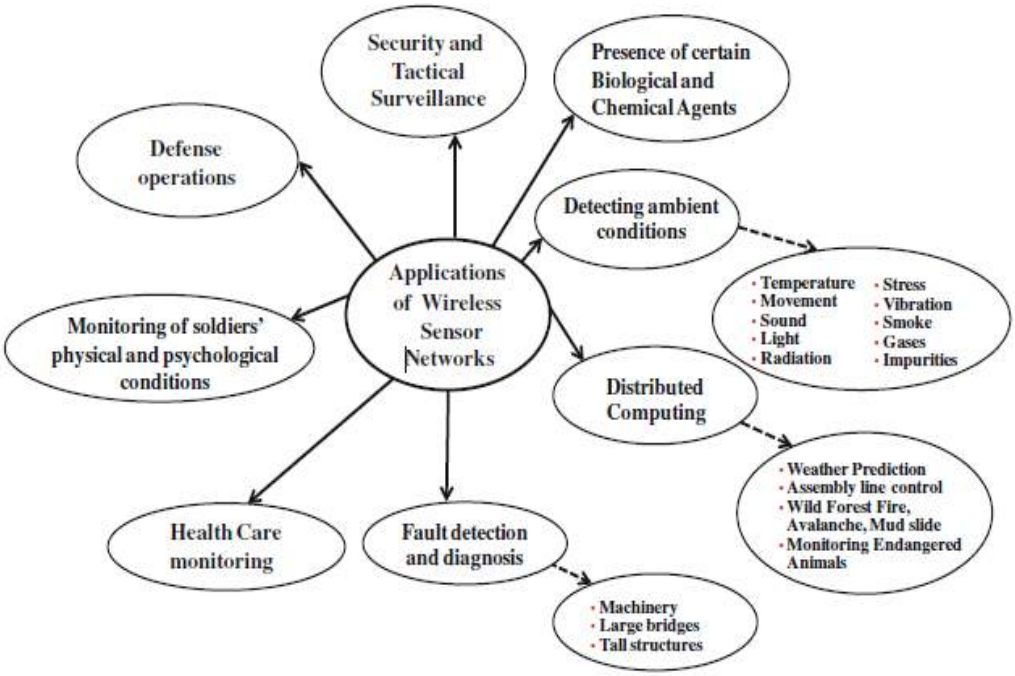
- Separation from central network administration.
- Self-configuring nodes are also routers.
- Self-healing through continuous re-configuration.
- Scalability incorporates the addition of more nodes.
- Mobility allows ad hoc networks created on the fly in any situation where there are multiple wireless devices.
- Flexible ad hoc can be temporarily setup at anytime, in any place.
- Lower getting-started costs due to decentralized administration. The nodes in ad hoc network need not rely on any hardware and software. So, it can be connected and communicated quickly.

Advantages of sensor network:

- Network setups can be carried out without fixed infrastructure.
- Suitable for the non-reachable places such as over the sea, mountains, rural areas or deep forests.
- Flexible if there is random situation when additional workstation is needed.
- Implementation pricing is cheap.
- It avoids plenty of wiring.
- It might accommodate new devices at any time.
- It's flexible to undergo physical partitions.
- It can be accessed by using a centralized monitor.

Driving applications:

Numerous applications of WSNs have emerged, and a broad classification has been given in Table 1. It is rather hard to organize them in a systematic way, and some overlap is unavoidable. Figure 1 shows many areas of applications. However, they do not represent any chronological progression of development nor a complete list but a comprehensive classification of different areas.



WSNs bring multitude of benefits to our daily lives.

Application	Benefit
Measure microclimates on farms	Increase crop yield per square km
Monitor traffic on road systems	Steer traffic away from jams, accidents, and construction zones; alert emergency services
Detect human presence in homes and offices	Reduce wasted power in HVAC and lighting
Electrical/gas/water metering	Optimize utility distribution systems and reduce inefficiencies

Defense Applications of WSNs:

WSNs were introduced for defense purpose, and we start with those types of applications. The idea here is to deploy SNs from low-flying airplanes or drones, and when SNs land on surface of land, they collect information from the surrounding area of war zone and send data to a powerful base station (BS) or sink node located inside the plane. Data are gathered and analyzed by the BS and determine strategic information such as type and number of tanks in the battlefield, number of soldiers, elevation of terrain, and types of hiding places such as bunkers.

Civilian Applications:

Many civilian applications have been suggested for WSNs. These can be divided into four main categories as shown in Fig. 2.1. These are discussed in the following paragraphs.

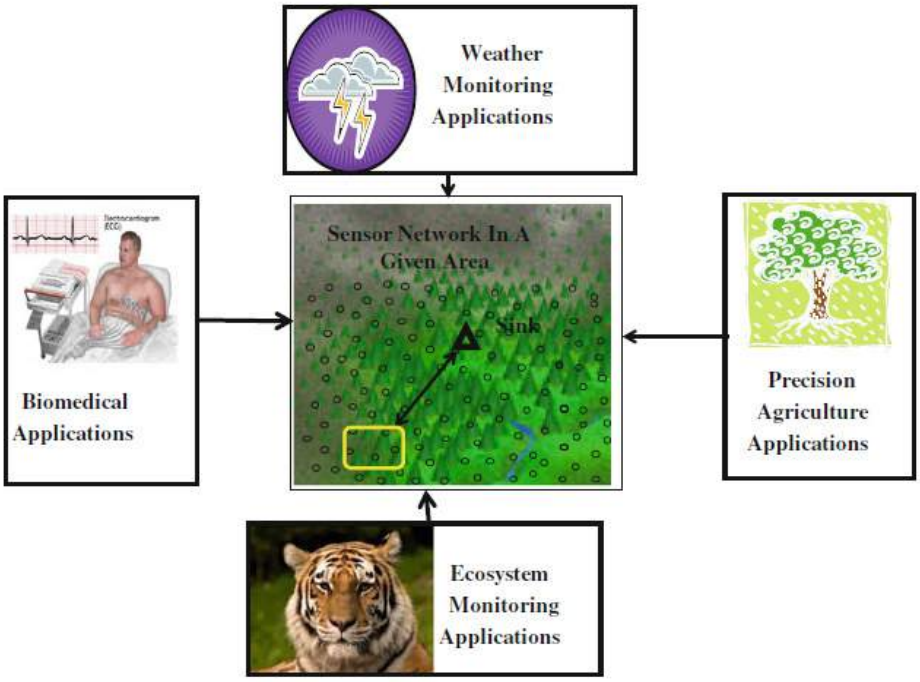


Figure 2.1. Categorization of WSNs' civilian applications

- Weather Monitoring Applications
- Precision Agriculture Applications
- Echo System Monitoring Applications
- Biomedical Applications

2.2 Issues in Ad hoc wireless Networks

The major issues that affect the design, deployment, and performance of an ad hoc wireless system are as follows:

- Medium access scheme
- Routing
- Multicasting
- Transport layer protocol
- Pricing scheme
- Quality of service provisioning
- Self-organization
- Security
- Energy management
- Addressing and service discovery
- Scalability
- Deployment considerations

2.2.1 Medium Access Scheme:

The primary responsibility of a medium access control (MAC) protocol in ad hoc wireless networks is the distributed arbitration for the shared channel for transmission of packets. The

major issues to be considered in designing a MAC protocol for ad hoc wireless networks are as follows:

- **Synchronization:** The MAC protocol design should take into account the requirement of time synchronization. Synchronization is mandatory for TDMA-based systems for management of transmission and reception slots. Synchronization involves usage of scarce resources such as bandwidth and battery power.
- **Hidden terminals:** Hidden terminals are nodes that are hidden (or not reachable) from the sender of a data transmission session, but are reachable to the receiver of the session. In such cases, the hidden terminal can cause collisions at the receiver node.
- **Exposed terminals:** Exposed terminals, the nodes that are in the transmission range of the sender of an on-going session, are prevented from making a transmission. In order to improve the efficiency of the MAC protocol, the exposed nodes should be allowed to transmit in a controlled fashion without causing collision to the on-going data transfer.
- **Throughput:** The MAC protocol employed in ad hoc wireless networks should attempt to maximize the throughput of the system. The important considerations for throughput enhancement are minimizing the occurrence of collisions, maximizing channel utilization, and minimizing control overhead.
- **Access delay:** The access delay refers to the average delay that any packet experiences to get transmitted. The MAC protocol should attempt to minimize the delay.

2.2.2 Routing:

The major requirements of a routing protocol in ad hoc wireless networks are the following:

- **Minimum route acquisition delay:** The route acquisition delay for a node that does not have a route to a particular destination node should be as minimal as possible. This delay may vary with the size of the network and the network load.
- **Quick route reconfiguration:** The unpredictable changes in the topology of the network require that the routing protocol be able to quickly perform route reconfiguration in order to handle path breaks and subsequent packet losses.
- **Support for time-sensitive traffic:** Tactical communications and similar applications require support for time-sensitive traffic. The routing protocol should be able to support both hard realtime and soft real-time traffic.
- **Security and privacy:** The routing protocol in ad hoc wireless networks must be resilient to threats and vulnerabilities. It must have inbuilt capability to avoid resource consumption, denial-of- service, impersonation, and similar attacks possible against an ad hoc wireless network.

2.2.3 Multicasting

It plays important role in emergency search & rescue operations & in military communication. Use of single-link connectivity among the nodes in a multicast group results in a tree-shaped multicast routing topology.

The major issues in designing multicast routing protocols are as follows:

1. ***Robustness :***
    - The multicast routing protocol must be able to recover & reconfigure quickly from potential mobility-induced link breaks thus making it suitable for use in high dynamic environments.
2. ***Efficiency :***
    - A multicast protocol should make a minimum number of transmissions to deliver a data packet to all the group members.
3. ***Control overhead :***
    - The scarce bandwidth availability in ad hoc wireless networks demands minimal control overhead for the multicast session.

2.2.4 Transport Layer Protocol

The main objectives of the transport layer protocols include:

- Setting up & maintaining end-to-end connections,
- Reliable end-to-end delivery of packets,
- Flow control & Congestion control.

2.2.5 Pricing Scheme

- Assume that an optimal route from node A to node B passes through node C, & node C is not powered on.
- Then node A will have to set up a costlier & non-optimal route to B.
- The non-optimal path consumes more resources & affects the throughput of the system.

2.2.6 Quality of Service Provisioning (QoS)

- QoS is the performance level of services offered by a service provider or a network to the user.
- QoS provisioning often requires
- Negotiation between host & the network.
- Resource reservation schemes.
- Priority scheduling & Call admission control.

2.2.7 Self-Organization

- One very important property that an ad hoc wireless network should exhibit is organizing & maintaining the network by itself.
- The major activities that an ad hoc wireless network is required to perform for self-organization are,
- Neighbour discovery.
- Topology organization & Topology reorganization (updating topology information)

2.2.8 Energy Management

- Energy management is defined as the process of managing the sources & consumers of energy in a node or in the network for enhancing the lifetime of a network.
- Features of energy management are :

  → shaping the energy discharge pattern of a node's battery to enhance battery life.

  → finding routes that consumes minimum energy. →Using distributed scheduling schemes to improve battery life. →Handling the processor & interface devices to minimize power consumption.

2.2.9 Deployment Considerations

The deployment of a commercial ad hoc wireless network has the following benefits when compared to wired networks

a) ***Low cost of deployment:***

- The use of multi-hop wireless relaying eliminates the requirement of cables & maintenance in deployment of communication infrastructure.
- The cost involved is much lower than that of wired networks.

b) Incremental deployment:

- Deployment can be performed incrementally over geographical regions of the city.
- The deployed part of the network starts functioning immediately after the minimum configuration is done.

c) ***Short deployment time:***

- Compared to wired networks, the deployment time is considerably less due to the absence of any wired links.

2.3 Issues in design of sensor network

Sensor networks pose certain design challenges due to the following reasons:

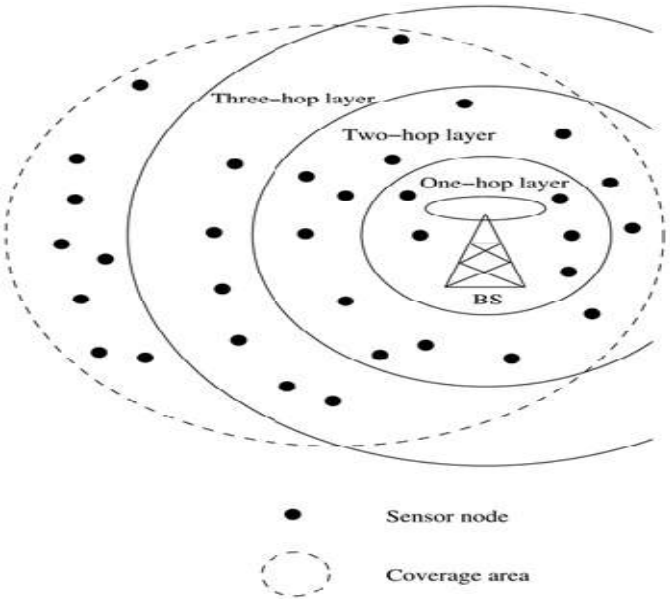
- Sensor nodes are randomly deployed and hence do not fit into any regular topology. Once deployed, they usually do not require any human intervention. Hence, the setup and maintenance of the network should be entirely autonomous.
- Sensor networks are infrastructure-less. Therefore, all routing and maintenance algorithms need to be distributed.
- An important bottleneck in the operation of sensor nodes is the available energy. Sensors usually rely only on their battery for power, which in many cases cannot be recharged or replaced. Hence, the available energy at the nodes should be considered as a major constraint while designing protocols. For instance, it is desirable to give the user an option to trade off network lifetime for fault tolerance or accuracy of results.
- Hardware design for sensor nodes should also consider energy efficiency as a primary requirement. The micro-controller, operating system, and application software should be designed to conserve power.

- Sensor nodes should be able to synchronize with each other in a completely distributed manner, so that TDMA schedules can be imposed and temporal ordering of detected events can be performed without ambiguity.
- A sensor network should also be capable of adapting to changing connectivity due to the failure of nodes, or new nodes powering up. The routing protocols should be able to dynamically include or avoid sensor nodes in their paths.
- Real-time communication over sensor networks must be supported through provision of guarantees on maximum delay, minimum bandwidth, or other QoS parameters.
- Provisions must be made for secure communication over sensor networks, especially for military applications which carry sensitive data.

2.4 Sensor network architecture

The design of sensor networks is influenced by factors such as scalability, fault tolerance, and power consumption. The two basic kinds of sensor network architecture are layered and clustered.

1. **Layered architecture:**



Layered architectures have been used with in-building wireless backbones, and in military sensor-based infrastructure, such as the multi-hop infrastructure network architecture (MINA).

Unified Network Protocol Framework (UNPF)

UNPF is a set of protocols for complete implementation of a layered architecture for sensor networks. UNPF integrates three operations in its protocol structure: network initialization and maintenance, MAC, and routing protocols.

(i) Network Initialization and Maintenance Protocol

The network initialization protocol organizes the sensor nodes into different layers, using the broadcast capability of the BS. The BS can reach all nodes in a one-hop communication over a

common control channel. The BS broadcasts its identifier (ID) using a known CDMA code on the common control channel. All nodes which hear this broadcast then record the BS ID.

(ii) MAC Protocol:

Network initialization is carried out on a common control channel. During the data transmission phase, the distributed TDMA receiver oriented channel (DTROC) assignment MAC protocol is used. Each node is assigned a reception channel by the BS, and channel reuse is such that collisions are avoided.

(iii) Routing Protocol:

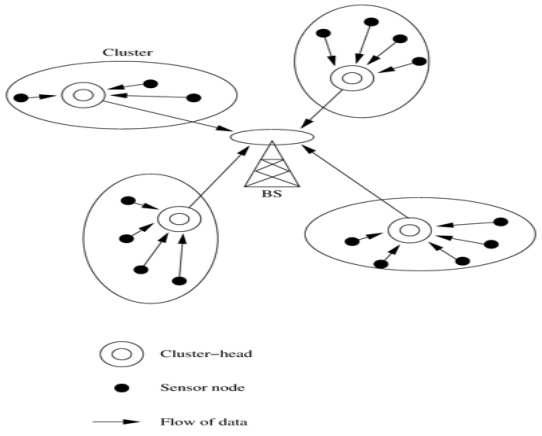
Downlink from the BS is by direct broadcast on the control channel. The layered architecture enables multi-hop data forwarding from the sensor nodes to the BS. A modification to the UNPF protocol set termed the UNPF-R has been proposed. It makes the sensor nodes adaptively vary their transmission range so that network performance can be optimized.

For a transmission range *R,* the objective function is $f(R) = \frac{c \times d}{n/N}$ where *N* is the total number of sensors in the system; *n* is the number of nodes in layer one; € is the energy consumption per packet; and *d* is the average packet delay. The BS selects a new transmission range *R′* as follows. If no packet is received by the BS from any sensor node for some interval of time, the transmission range is increased by Δr, a predefined increment. Otherwise, the transmission range is either decreased by Δr with probability 0.5 × (*n/N*), or increased by Δr with probability [1 - 0.5 × (*n/N*)]. The objective function is reevaluated with the new transmission range.

If $f(R') < f(R)$, then the transmission range *R′* is adopted. Otherwise, *R* is modified to *R′* with probability $e^{\frac{(f(R)-f(R'))\times(n/N)}{T}}$, where *T* is the temperature parameter, as in simulated annealing. The advantage of the UNPF-R is that it minimizes the energy × delay metric, and maximizes the number of nodes which can connect to the BS. The minimization of the energy × delay metric ensures that transmission should occur with minimum delay and with minimum energy consumption. The two conflicting objectives are together optimized by minimizing their product.

2. Clustered Architecture:

A clustered architecture organizes the sensor nodes into clusters, each governed by a cluster-head. The nodes in each cluster are involved in message exchanges with their respective cluster-heads, and these heads send messages to a BS, which is usually an access point connected to a wired network.



This is achieved through network layer protocols such as the low-energy adaptive clustering hierarchy (LEACH).

Low-Energy Adaptive Clustering Hierarchy (LEACH):

LEACH is a clustering-based protocol that minimizes energy dissipation in sensor networks. LEACH randomly selects nodes as cluster-heads and performs periodic reelection, so that the high-energy dissipation experienced by the cluster-heads in communicating with the BS is spread across all nodes of the network. Each iteration of selection of cluster-heads is called a round. The operation of LEACH is split into two phases: set-up and steady.

During the set-up phase, each sensor node chooses a random number between 0 and 1. If this is lower than the threshold for node *n, T*(*n*), the sensor node becomes a cluster-head. The threshold *T*(*n*) is calculated as,

$$T(n) = \begin{cases} \frac{P}{1-P[r \times mod(1/P)]} & \text{if } n \in G \\ 0 & \text{otherwise,} \end{cases}$$

where *P* is the desired percentage of nodes which are cluster-heads, *r* is the current round, and *G* is the set of nodes that has not been cluster-heads in the past 1/*P* rounds. The cluster-heads then assign a TDMA schedule for their cluster members.

The steady phase is of longer duration in order to minimize the overhead of cluster formation. During the steady phase, data transmission takes place based on the TDMA schedule, and the cluster-heads perform data aggregation/fusion through local computation. After a certain period of time in the steady phase, cluster-heads are selected again through the set-up phase.

2.5 Data Dissemination

Data dissemination is the process by which queries or data are routed in the sensor network. The data collected by sensor nodes has to be communicated to the BS or to any other node interested in the data. The node that generates data is called a *source* and the information to be reported is called an *event.* A node which is interested in an event and seeks information about it is called a *sink.* Traffic models have been developed for sensor networks such as the data collection and data dissemination (diffusion) models. In the data collection model, the source sends the data it collects to a collection entity such as the BS.

Flooding:

In flooding, each node which receives a packet broadcasts it if the maximum hop-count of the packet is not reached and the node itself is not the destination of the packet. But flooding has the following disadvantages:

- Implosion: This is the situation when duplicate messages are sent to the same node. This occurs when a node receives copies of the same message from many of its neighbors.
- Overlap: The same event may be sensed by more than one node due to overlapping regions of coverage. This results in their neighbors receiving duplicate reports of the same event.
- Resource blindness: The flooding protocol does not consider the available energy at the nodes and results in many redundant transmissions. Hence, it reduces the network lifetime.

Gossiping:

Gossiping is a modified version of flooding, where the nodes do not broadcast a packet, but send it to a randomly selected neighbor. This avoids the problem of implosion, but it takes a long time for a message to propagate throughout the network.

Rumor Routing:

Rumor routing is an agent-based path creation algorithm. Agents, or "ants," are long-lived entities created at random by nodes. These are basically packets which are circulated in the network to establish shortest paths to events that they encounter. They can also perform path optimizations at nodes that they visit. When an agent finds a node whose path to an event is longer than its own, it updates the node's routing table.

Figure 2.2 illustrates the working of the rumor routing algorithm. In Figure 2.2 (a), the agent has initially recorded a path of distance 2 to event *E*1. Node *A*'s table shows that it is at a distance 3 from event *E*1 and a distance 2 from *E*2. When the agent visits node *A,* it updates its own path state information to include the path to event *E*2. The updating is with one hop greater distance than what it found in *A,* to account for the hop between any neighbor of *A* that the agent will visit next, and *A*. It also optimizes the path to *E*1 recorded at node *A* to the shorter path through node *B*. The updated status of the agent and node table is shown in Figure 2.2 (b).

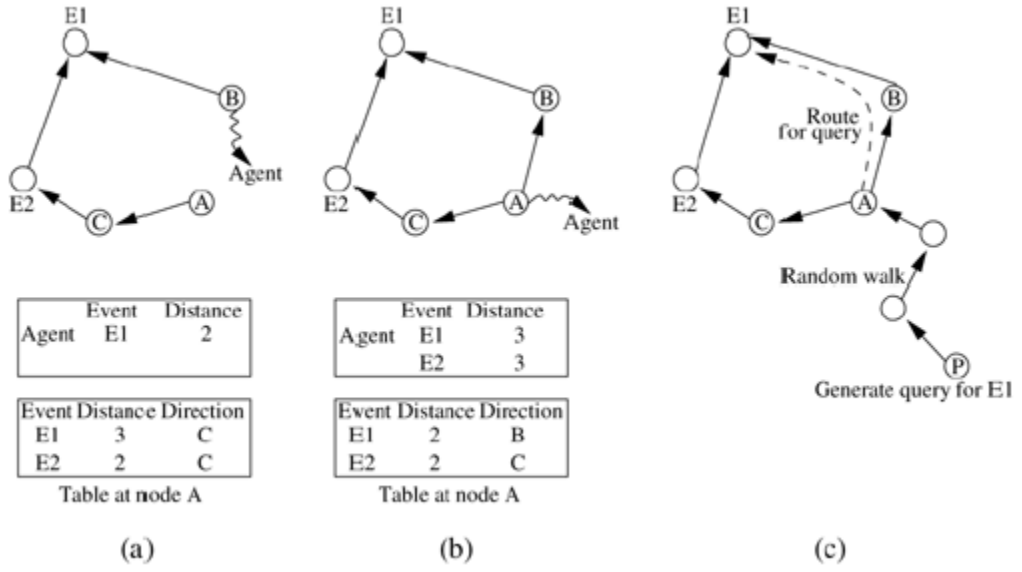
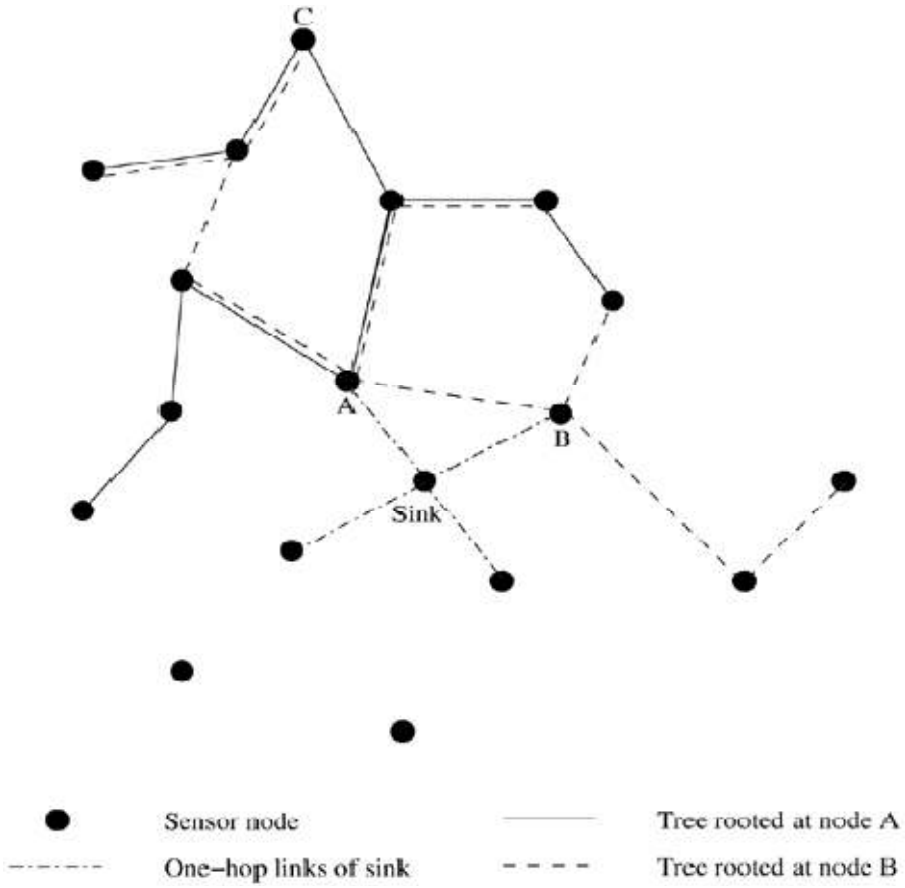


Figure 2.2. Rumor routing.

Sequential Assignment Routing:

The sequential assignment routing (SAR) algorithm creates multiple trees, where the root of each tree is a one-hop neighbor of the sink. Each tree grows outward from the sink and avoids nodes with low throughput or high delay. At the end of the procedure, most nodes belong to multiple trees.



The SAR algorithm chooses a path with high estimated energy resources, and provisions can be made to accommodate packets of different priorities. A weighted QoS metric is used to handle prioritized packets, which is computed as a product of priority level and delay. The routing ensures that the same weighted QoS metric is maintained.

Directed Diffusion:

They generate requests/queries for data sensed by other nodes, instead of all queries arising only from a BS. Hence, the sink for the query could be a BS or a sensor node. The directed diffusion routing protocol improves on data diffusion using interest gradients. The diffusion model allows nodes to cache or locally transform (aggregate) data. This increases the scalability of communication and reduces the number of message transmissions required.

Sensor Protocols for Information via Negotiation:

A family of protocols called sensor protocols for information via negotiation (SPIN) is proposed in.SPIN uses negotiation and resource adaptation to address the deficiencies of flooding. Negotiation reduces overlap and implosion, and a threshold-based resource-aware operation is used to prolong network lifetime.

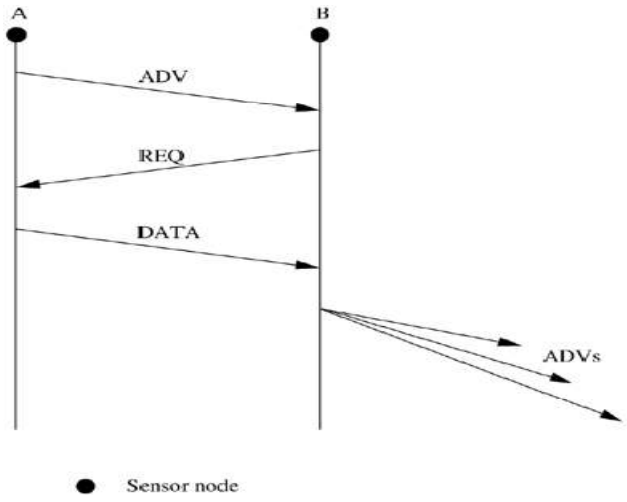
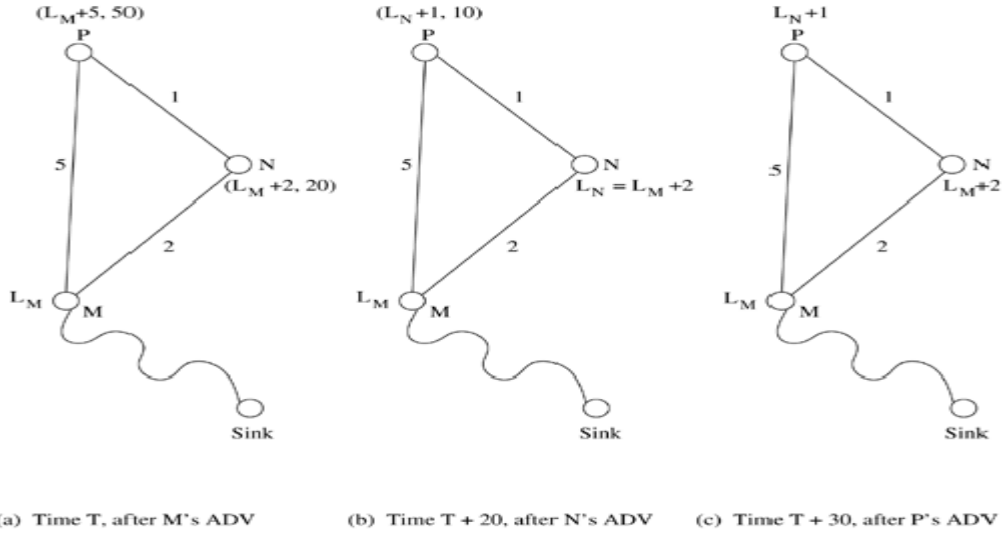


Figure 2.3. SPIN Protocol

Cost-Field Approach:

The cost-field approach considers the problem of setting up paths to a sink. It is a two-phase process, the first phase being to set up the cost field, based on metrics such as delay, at all sensor nodes, and the second being data dissemination using the costs.

Phase 1 sets up a cost field starting from the sink node. A sink broadcasts an ADV packet with its own cost as 0. When a node *N* hears an ADV message from node *M,* it sets its own path cost to *min*(*LN, LM + CNM),* where *LN* is the total path cost from node *N* to sink, *LM* represents the cost of node *M* to sink, and *CNM* is the cost from node *N* to *M*. If *LN* was updated, the new cost is broadcast through another ADV.



(a) Time T, after M's ADV (b) Time T + 20, after N's ADV (c) Time T + 30, after P's ADV

Phase 2 is the data dissemination process. Once the cost field is established, a source sends its message to sink *S* with cost *CS*. The message also contains a cost-so-far field, initially set to 0. Each intermediate node forwards the packet if the cost recorded in the packet plus its own cost equals the original source-to-sink cost. This ensures that the original optimal path is used whenever a packet is routed. While forwarding, the intermediate nodes also update the cost-so-far field.

Geographic Hash Table:

Geographic hash table (GHT) is a system based on data-centric storage, inspired by Internet- cale distributed hash table (DHT) systems such as Chord and Tapestry. The routing protocol used is greedy perimeter stateless routing (GPSR), which again uses geographical information to route the data and queries.GHT is more effective in large sensor networks, where a large number of events are detected but not all are queried.

Small Minimum Energy Communication Network:

Small minimum energy communication network (SMECN) is a protocol proposed in to construct a sub-network from a given communication network. If the entire sensor network is represented by a graph *G*, the subgraph *G′* is constructed such that the energy usage of the network is minimized. The power required to transmit data between two nodes *u* and *v* is modeled As

$$p(u, v) = t \times d(u, v)^n$$

where *t* is a constant, *n* is the path loss exponent indicating the loss of power with distance from the transmitter, and *d*(*u, v*) is the distance between *u* and *v*. Let the power needed to receive the data be *c*. Since the transmission power increases exponentially with distance, it would be more economical to transmit data by smaller hops. Suppose the path between *u* (*i.e., u*0) and *v* (*i.e., uk*) is represented by *r1*= (*u*0 , *u1 , ...uk*), such that each (*ui , ui*+ 1) is an edge in the subgraph *G′*, then the total power consumed for the transmission is

$$C(r) = \sum_{i=0}^{k-1} (p(u_i, u_{i+1}) + c)$$

The path *r* is the ME path if *C*(*r*) ≤ *C*(*r′*) for all paths *r′* between *u* and *v* in the graph *G*. The subgraph*G′* is said to have the ME property if there exists a path *r* in *G′* which is an ME path in *G*, for all node pairs (*u, v*). SMECN uses only the ME paths from *G′* for data transmission, so that the overall energy consumed is minimized.

2.6 Data Gathering

The objective of the data-gathering problem is to transmit the sensed data from each sensor node to a BS. One round is defined as the BS collecting data from all the sensor nodes once. This scheme performs poorly with respect to the energy × delay metric.

Power-Efficient Gathering for Sensor Information Systems:

Power-efficient gathering for sensor information systems (PEGASIS) is a data-gathering protocol based on the assumption that all sensor nodes know the location of every other node, that is, the topology information is available to all nodes.

The goals of PEGASIS are as follows:

- Minimize the distance over which each node transmits
- Minimize the broadcasting overhead
- Minimize the number of messages that need to be sent to the BS
- Distribute the energy consumption equally across all nodes

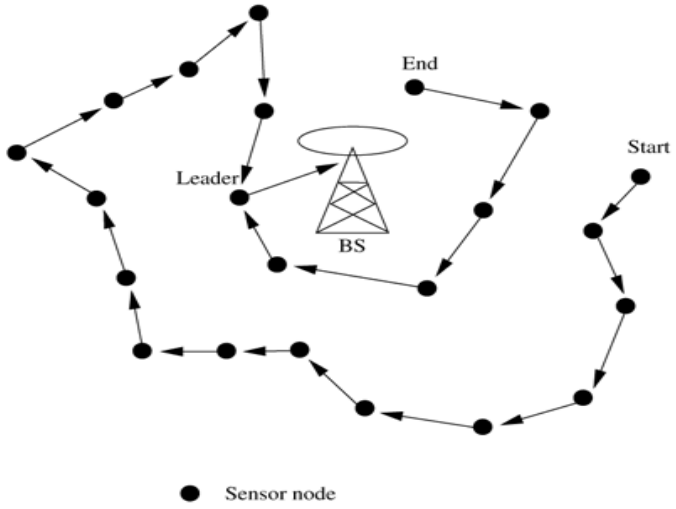
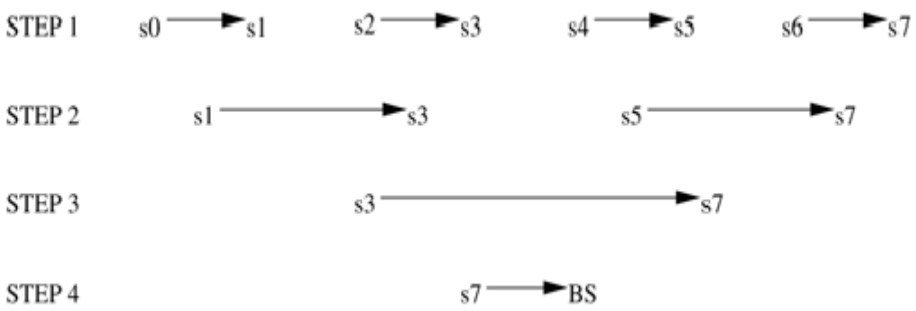


Figure2.4. Data gathering with PEGASIS.

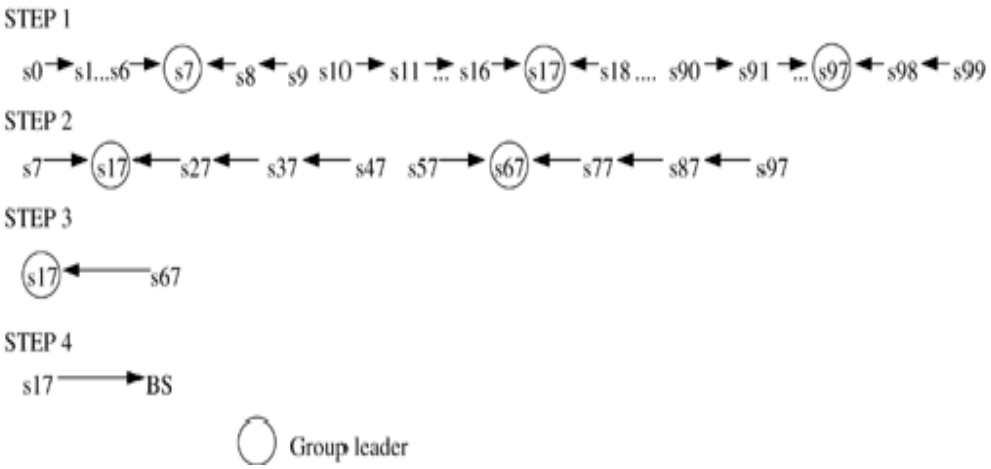
Binary Scheme:

This is also a chain-based scheme like PEGASIS, which classifies nodes into different levels. All nodes which receive messages at one level rise to the next. The number of nodes is halved from one level to the next. The number of nodes is halved from one level to the next. For instance, consider a network with eight nodes labeled *s*0 to *s*7. This scheme is possible when nodes communicate using CDMA, so that transmissions of each level can take place simultaneously.



Chain-Based Three-Level Scheme:

For non-CDMA sensor nodes, a binary scheme is not applicable. The chain based three-level scheme addresses this situation, where again a chain is constructed as in PEGASIS. The chain is divided into a number of groups to space out simultaneous transmissions in order to minimize interference. One node out of each group aggregates data from all group members and rises to the next level. The index of this leader node is decided *a priori.* In the second level, all nodes are divided into two groups, and the third level consists of a message exchange between one node from each group of the second level.



Finally, the leader transmits a single message to the BS. The working of this scheme is illustrated in Figure. The network has 100 nodes, and the group size is ten for the first level and five for the second level. Three levels have been found to give the optimal energy × delay through simulations.

UNIT – III

MAC Protocols

3.1 Issues in designing MAC protocols for adhoc wireless networks

The main issues in designing MAC protocol for ad hoc wireless network are: Bandwidth efficiency.

- Bandwidth must be utilized in efficient manner.
- Minimal Control overhead.
- BW = ratio of BW used for actual data transmission to the total available BW.

3.1.1 Quality of service support

- Essential for supporting time-critical traffic sessions.
- They have resource reservation mechanism that takes into considerations the nature of wireless channel and the mobility of nodes.

3.1.2 Synchronization

- MAC protocol must consider synchronization between nodes in the network.
- Synchronization is very important for BW (time slot) reservation by nodes.
- Exchange of control packets may be required for achieving time synchronisation among nodes.

3.1.3 Hidden and exposed terminal problems

- The hidden terminal problem refers to the collision of packets at a receiving node due to the simultaneous transmission of those nodes that are not within the direct transmission range of the sender but are within the transmission range of the receiver.
- Collision occurs when both nodes transmit packets at the same time without knowing about the transmission of each other.

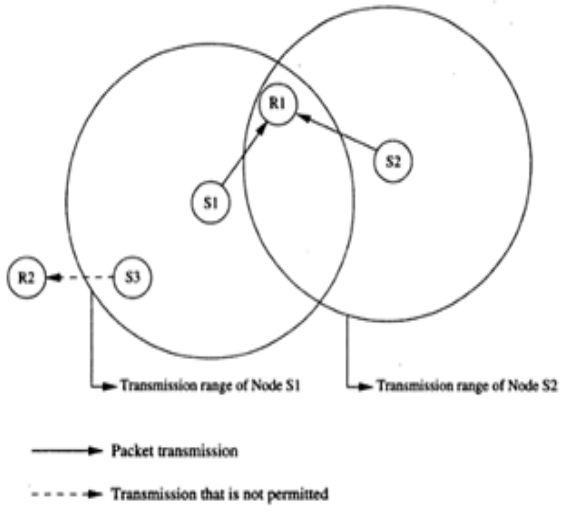


Figure | Hidden and exposed terminal problems.

- S1 and S2 are hidden from each other & they transmit simultaneously to R1 which leads to collision.
- The exposed terminal problem refers to the inability of a node, which is blocked due to transmission by a nearby transmitting node, to transmit to another node.
- If S1 is already transmitting to R1, then S3 cannot interfere with on-going transmission & it cannot transmit to R2.
- The hidden & exposed terminal problems reduce the throughput of a network when traffic load is high.

3.1.4 Error-prone shared broadcast channel

- When a node is receiving data, no other node in its neighbourhood should transmit.
- A node should get access to the shared medium only when its transmission do not affect any ongoing session.
- MAC protocol should grant channel access to nodes in such a manner that collisions are minimized.
- Protocol should ensure fair BW allocation.

3.1.5 Distributed nature/lack of central coordination

Do not have centralized coordinators.

- Nodes must be scheduled in a distributed fashion for gaining access to the channel.
- MAC protocol must make sure that additional overhead, in terms of BW consumption, incurred due to this control information is not very high.

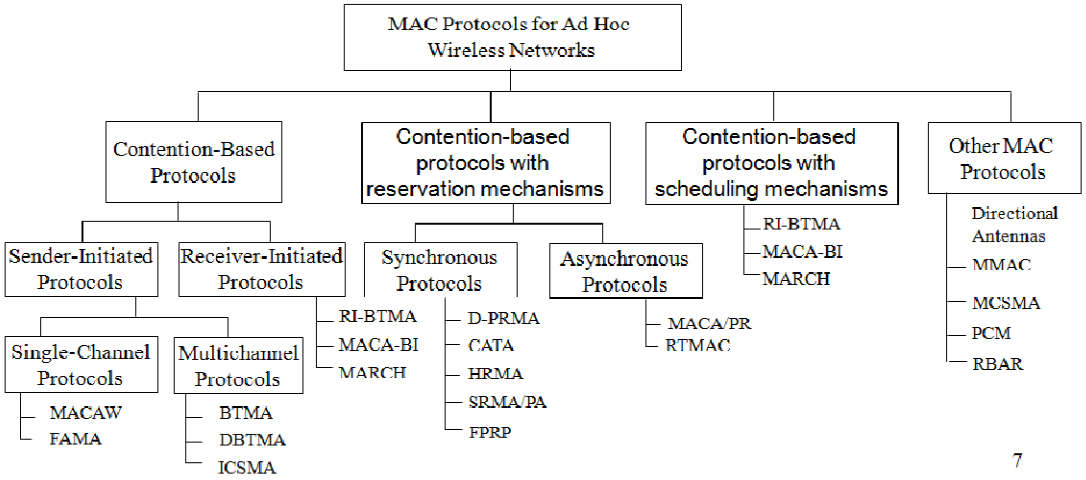
3.1.6 Mobility of nodes

- Nodes are mobile most of the time
- The protocol design must take this mobility factor into consideration so that the performance of the system is not affected due to node mobility.

3.2 Classification of MAC Protocols

Ad hoc network MAC protocols can be classified into three basic types:

i. Contention-based protocols
ii. Contention-based protocols with reservation mechanisms
iii. Contention-based protocols with scheduling mechanisms
iv. Other MAC protocols [protocols which do not fall under above 3 categories]



Contention-based protocols

• ***Sender-initiated protocols***: Packet transmissions are initiated by the sender node.

• *Single-channel sender-initiated protocols*: A node that wins the contention to the channel can make use of the entire bandwidth.

• *Multichannel sender-initiated protocols:* The available bandwidth is divided into multiple channels.

• ***Receiver-initiated protocols*:** The receiver node initiates the contention resolution protocol.

Contention-based protocols with reservation mechanisms

• ***Synchronous protocols***: All nodes need to be synchronized. Global time synchronization is difficult to achieve.

• ***Asynchronous protocols:*** These protocols use relative time information for effecting reservations.

Contention-based protocols with scheduling mechanisms

• Node scheduling is done in a manner so that all nodes are treated fairly and no node is starved of bandwidth.

• Scheduling-based schemes are also used for enforcing priorities among flows whose packets are queued at nodes.

• Some scheduling schemes also consider battery characteristics.

Other protocols are those MAC protocols that do not strictly fall under the above categories.

Other MAC protocols

Similarly to ad hoc routing protocols discussed earlier, our discussion on MAC protocol issues is also far from being exhaustive. There are many other issues to be considered such as fairness. Fairness has many meanings and one of them might say that stations should receive equal bandwidth. With the basic approach of IEEE 802.11, this fairness is not easy to accomplish since unfairness will eventually occur when one node backs off much more than some other node. MACAW's solution to this is to append the contention window value (CW) to packets a node transmits, so that all nodes hearing that CW use it for their future transmissions. Since CW is an

indication of the level of congestion in the vicinity of a specific receiver node, MACAW proposes maintaining a CW independently for each receiver. There are also other proposals such as Distributed Fair Scheduling and Balanced MAC.

3.2.1 Contention-based protocols

Contention-based protocols do not have any bandwidth reservation mechanisms. All ready nodes contend for the channel simultaneously, and the winning node gains access to the channel. Since nodes are not guaranteed bandwidth, these protocols cannot be used for transmitting real-time traffic, which requires QoS guarantees from the system.

MACAW: A Media Access Protocol for Wireless LANs

MACA was proposed due to the shortcomings of CSMA protocols when used for wireless networks. In what follows, a brief description on why CSMA protocols fail in wireless networks is given. This is followed by detailed descriptions of the MACA protocol and the MACAW protocol.

MACA Protocol

The MACA protocol was proposed as an alternative to the traditional carrier sense multiple access (CSMA) protocols used in wired networks. In CSMA protocols, the sender first senses the channel for the carrier signal. If the carrier is present, it retries after a random period of time. Otherwise, it transmits the packet. CSMA senses the state of the channel only at the transmitter.

MACA does not make use of carrier-sensing for channel access. It uses two additional signaling packets: the request-to-send (RTS) packet and the clear-to-send (CTS) packet. When a node wants to transmit a data packet, it first transmits an RTS packet. The receiver node, on receiving the RTS packet, if it is ready to receive the data packet, transmits a CTS packet. Once the sender receives the CTS packet without any error, it starts transmitting the data packet. This data transmission mechanism is depicted in Figure 3.1. Hence, the exposed terminal problem is also overcome in MACA. But MACA still has certain problems, which was why MACAW, described below, was proposed.

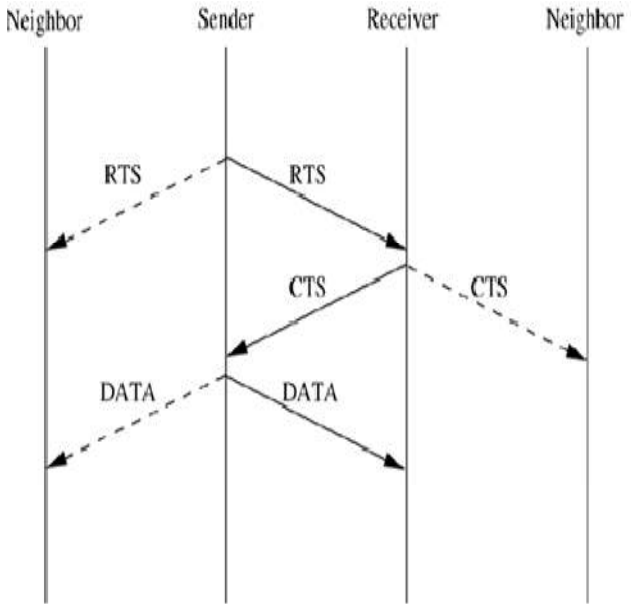


Figure 3.1 Packet transmission in MACA.

MACAW Protocol

The binary exponential back-off mechanism used in MACA at times starves flows. For example, consider Figure 3.2. Here both nodes S1 and S2 keep generating a high volume of traffic. The node that first captures the channel (say, node S1) starts transmitting packets. The packets transmitted by the other node S2 get collided, and the node keeps incrementing its back-off window according to the BEB algorithm. As a result, the probability of node S2 acquiring the channel keeps decreasing, and over a period of time it gets completely blocked. To overcome this problem, the back-off algorithm has been modified in MACAW.

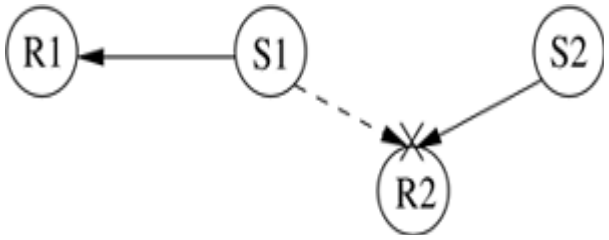


Figure 3.2. Example topology.

In MACAW another modification related to the back-off mechanism has been made. MACAW implements per flow fairness as opposed to the per node fairness in MACA. This is done by maintaining multiple queues at every node, one each for each data stream, and running the backoff algorithm independently for each queue. A node that is ready to transmit packets first determines how long it needs to wait before it could transmit an RTS packet to each of the destination nodes corresponding to the top-most packets in the node's queues. It then selects the packet for which the waiting time is minimal.

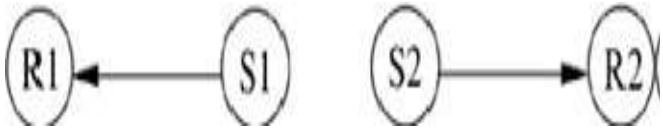


Figure 3.3. Example topology.

The MACAW protocol uses one more control packet called the request-for-request-to-send (RRTS) packet. The following example shows how this RRTS packet proves to be useful. Consider Figure 3.3. Here assume transmission is going on between nodes S1 and R1. Now node S2 wants to transmit to node R2. But since R2 is a neighbor of R1, it receives CTS packets from node R1, and therefore it defers its own transmissions. Node S2 has no way to learn about the contention periods during which it can contend for the channel, and so it keeps on trying, incrementing its back-off counter after each failed attempt. Hence the main reason for this problem is the lack of synchronization information at source S2.

Floor Acquisition Multiple Access Protocols(FAMA)

The floor acquisition multiple access (FAMA) protocols are based on a channel access discipline which consists of a carrier-sensing operation and a collision-avoidance dialog between the sender and the intended receiver of a packet. Floor acquisition refers to the process of gaining control of

the channel. At any given point of time, the control of the channel is assigned to only one node, and this node is guaranteed to transmit one or more data packets to different destinations without suffering from packet collisions. Carrier-sensing by the sender, followed by the RTS-CTS control packet exchange, enables the protocol to perform as efficiently as MACA in the presence of hidden terminals, and as efficiently as CSMA otherwise.

Busy Tone Multiple Access Protocols (BTMA)

Busy Tone Multiple Access

The busy tone multiple access (BTMA) protocol is one of the earliest protocols proposed for overcoming the hidden terminal problem faced in wireless environments. The transmission channel is split into two: a data channel and a control channel. The data channel is used for data packet transmissions, while the control channel is used to transmit the busy tone signal.

When a node is ready for transmission, it senses the channel to check whether the busy tone is active. If not, it turns on the busy tone signal and starts data transmission; otherwise, it reschedules the packet for transmission after some random rescheduling delay. Any other node which senses the carrier on the incoming data channel also transmits the busy tone signal on the control channel. Thus, when a node is transmitting, no other node in the two-hop neighborhood of the transmitting node is permitted to simultaneously transmit.

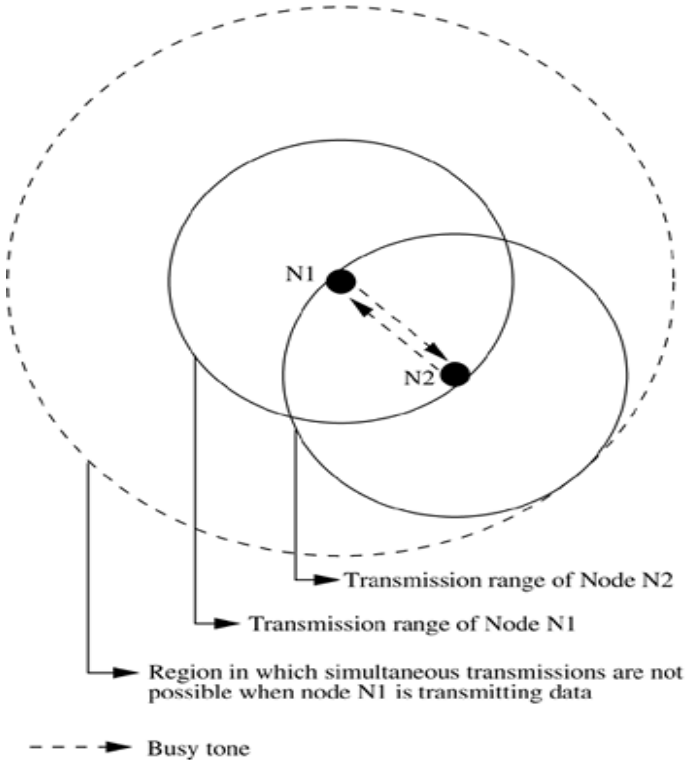


Figure 3.4. Transmission in BTMA.

Dual Busy Tone Multiple Access Protocol

The dual busy tone multiple access protocol (DBTMA) is an extension of the BTMA scheme. Here again, the transmission channel is divided into two: the data channel and the control

channel. As in BTMA, the data channel is used for data packet transmissions. The control channel is used for control packet transmissions (RTS and CTS packets) and also for transmitting the busy tones. DBTMA uses two busy tones on the control channel, *BTt* and *BTr* . The *BTt* tone is used by the node to indicate that it is transmitting on the data channel. The *BTr* tone is turned on by a node when it is receiving data on the data channel. The two busy tone signals are two sine waves at different well-separated frequencies.

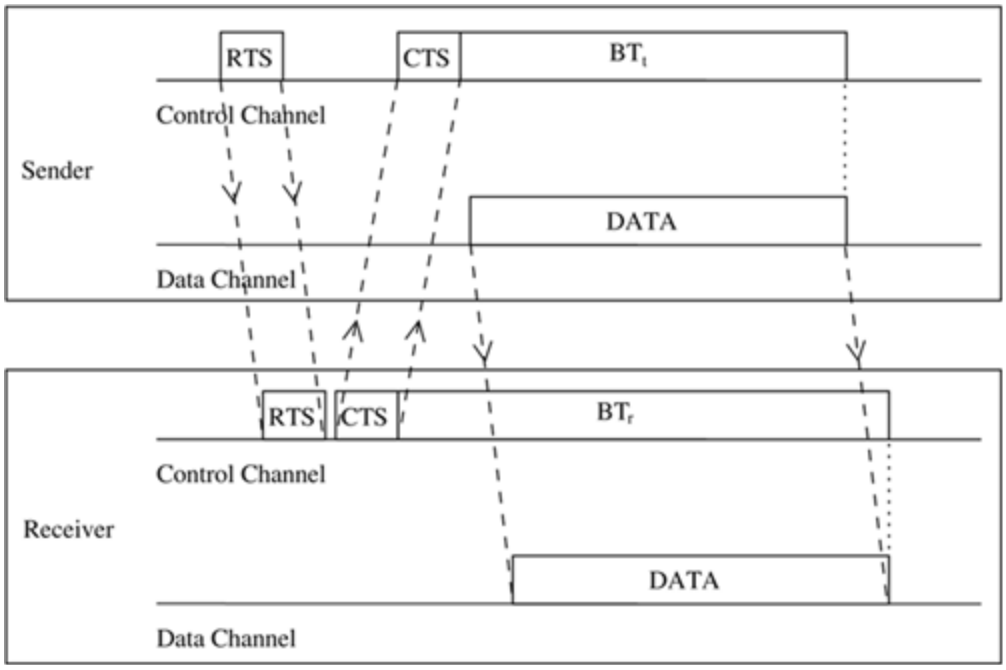


Figure 3.5. Packet transmission in DBTMA.

When compared to other RTS/CTS-based medium access control schemes (such as MACA and MACAW), DBTMA exhibits better network utilization. This is because the other schemes block both the forward and reverse transmissions on the data channel when they reserve the channel through their RTS or CTS packets. But in DBTMA, when a node is transmitting or receiving, only the reverse (receive) or forward (transmit) channels, respectively, are blocked. Hence the bandwidth utilization of DBTMA is nearly twice that of other RTS/CTS-based schemes.

MACA-By Invitation

MACA-by invitation (MACA-BI) is a receiver-initiated MAC protocol. It reduces the number of control packets used in the MACA protocol. MACA, which is a sender-initiated protocol, uses the three-way handshake mechanism, where first the RTS and CTS control packets are exchanged, followed by the actual DATA packet transmission. MACA-BI eliminates the need for the RTS packet. In MACA-BI the receiver node initiates data transmission by transmitting a ready to receive (RTR) control packet to the sender (Figure 3.6). If it is ready to transmit, the sender node responds by sending a DATA packet. Thus data transmission in MACA-BI occurs through a two-way handshake mechanism.

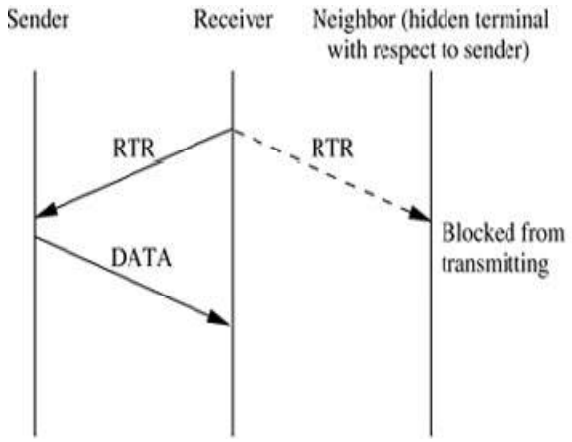


Figure 3.6. Packet transmission in MACA-BI.

The receiver node may not have an exact knowledge about the traffic arrival rates at its neighboring sender nodes. It needs to estimate the average arrival rate of packets. For providing necessary information to the receiver node for this estimation, the DATA packets are modified to carry control information regarding the backlogged flows at the transmitter node, number of packets queued, and packet lengths.

However, the hidden terminal problem still affects the control packet transmissions. This leads to protocol failure, as in certain cases the RTR packets can collide with DATA packets. One such scenario is depicted in Figure 3.7. Here, RTR packets transmitted by receiver nodes R1 and R2 collide at node A. So node A is not aware of the transmissions from nodes S1 and S2. When node A transmits RTR packets, they collide with DATA packets at receiver nodes R1 and R2.

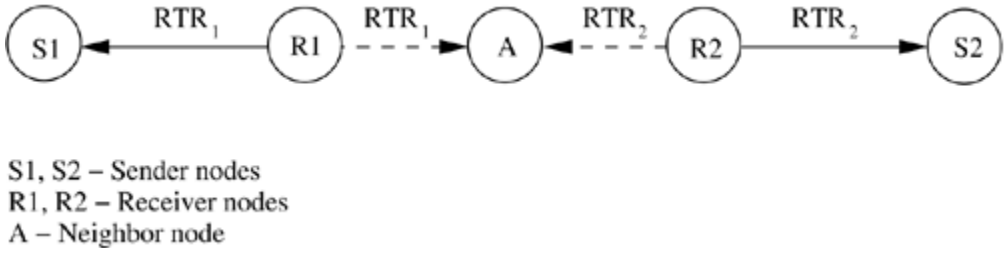


Figure 3.7. Hidden terminal problem in MACA-BI.

The efficiency of the MACA-BI scheme is mainly dependent on the ability of the receiver node to predict accurately the arrival rates of traffic at the sender nodes.

3.2.2 contention-based protocols with reservation mechanisms

Protocols described in this section have certain mechanisms that aid the nodes in effecting bandwidth reservations. Though these protocols are contention-based, contention occurs only during the resource (bandwidth) reservation phase. Once the bandwidth is reserved, the node gets exclusive access to the reserved bandwidth. Hence, QoS support can be provided for real-time traffic.

Distributed Packet Reservation Multiple Access Protocol

The distributed packet reservation multiple access protocol (D-PRMA) extends the earlier centralized packet reservation multiple access (PRMA) scheme into a distributed scheme that can be used in ad hoc wireless networks. PRMA was proposed for voice support in a wireless LAN with a base station, where the base station serves as the fixed entity for the MAC operation. D-PRMA extends this protocol for providing voice support in ad hoc wireless networks.

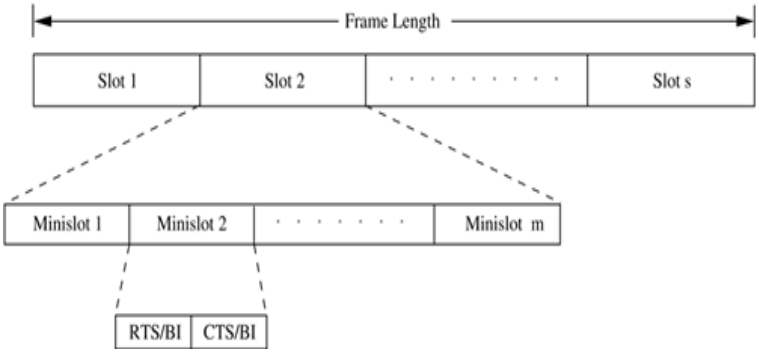


Figure 3.8. Frame structure in D-PRMA.

In order to prioritize nodes transmitting voice traffic (voice nodes) over nodes transmitting normal data traffic (data nodes), two rules are followed in D-PRMA. According to the first rule, the voice nodes are allowed to start contending from minislot 1 with probability $p = 1$; data nodes can start contending only with probability $p < 1$. For the remaining $(m - 1)$ minislots, both the voice nodes and the data nodes are allowed to contend with probability $p < 1$. This is because the reservation process for a voice node is triggered only after the arrival of voice traffic at the node; this avoids unnecessary reservation of slots.

Collision Avoidance Time Allocation Protocol

The collision avoidance time allocation protocol (CATA) is based on dynamic topology dependent transmission scheduling. Nodes contend for and reserve time slots by means of a distributed reservation and handshake mechanism. CATA supports broadcast, unicast, and multicast transmissions simultaneously. The operation of CATA is based on two basic principles:

- The receiver(s) of a flow must inform the potential source nodes about the reserved slot on which it is currently receiving packets. Similarly, the source node must inform the potential destination node(s) about interferences in the slot.
- Usage of negative acknowledgments for reservation requests, and control packet transmissions at the beginning of each slot, for distributing slot reservation information to senders of broadcast or multicast sessions.

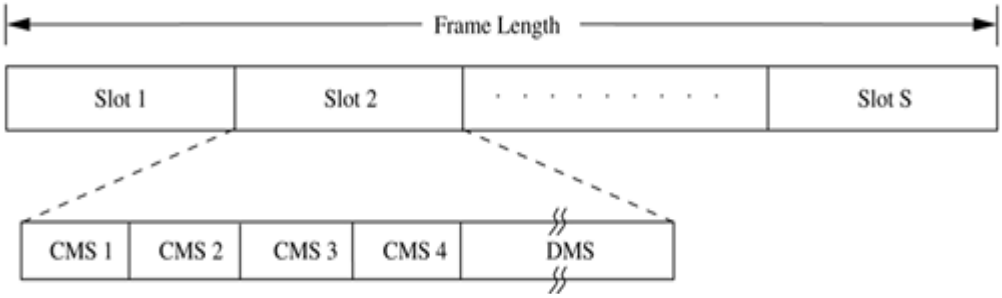


Figure 3.9. Frame format in CATA.

Each node that receives data during the DMS of the current slot transmits a slot reservation (SR) packet during the CMS1 of the slot. This serves to inform other neighboring potential sender nodes about the currently active reservation. The SR packet is either received without error at the neighbor nodes or causes noise at those nodes, in both cases preventing such neighbor nodes from attempting to reserve the current slot.

Five-Phase Reservation Protocol(FPRP)

The five-phase reservation protocol (FPRP) is a single-channel time division multiple access (TDMA)-based broadcast scheduling protocol. Nodes use a contention mechanism in order to acquire time slots. The protocol is fully distributed, that is, multiple reservations can be simultaneously made throughout the network.

Time is divided into frames. There are two types of frames: reservation frame (RF) and information frame (IF). Each RF is followed by a sequence of IFs. Each RF has *N* reservation slots (RS), and each IF has *N* information slots (IS). In order to reserve an IS, a node needs to contend during the corresponding RS. Based on these contentions, a TDMA schedule is generated in the RF and is used in the subsequent IFs until the next RF. The structure of the frames is shown in Figure 3.10. Each RS is composed of *M* reservation cycles (RC).

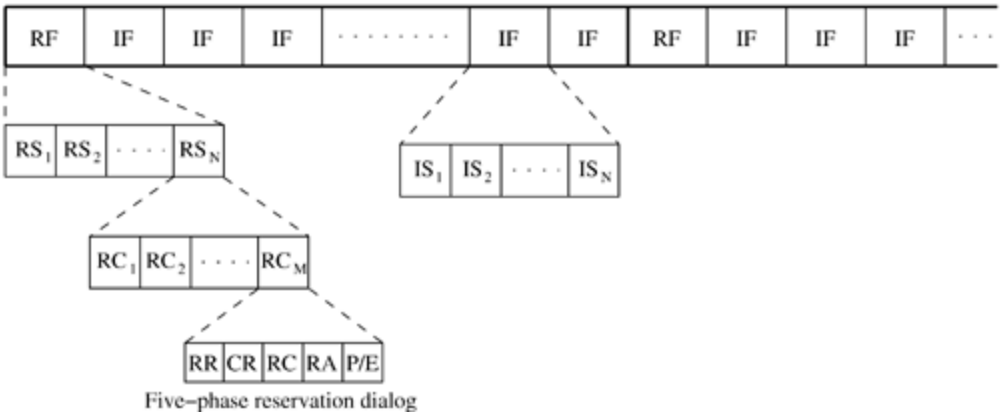


Figure 3.10. Frame structure in FPRP.

The protocol assumes the availability of global time at all nodes. Each node therefore knows when a five-phase cycle would start. The five phases of the reservation process are as follows:

1. Reservation request phase: Nodes that need to transmit packets send reservation request (RR) packets to their destination nodes.

2. Collision report phase: If a collision is detected by any node during the reservation request phase, then that node broadcasts a collision report (CR) packet. The corresponding source nodes, upon receiving the CR packet, take necessary action.
3. Reservation confirmation phase: A source node is said to have won the contention for a slot if it does not receive any CR messages in the previous phase. In order to confirm the reservation request made in the reservation request phase, it sends a reservation confirmation (RC) message to the destination node in this phase.
4. Reservation acknowledgment phase: In this phase, the destination node acknowledges reception of the RC by sending back a reservation acknowledgment (RA) message to the source. The hidden nodes that receive this message defer their transmissions during the reserved slot.
5. Packing and elimination (P/E) phase: Two types of packets are transmitted during this phase: packing packet and elimination packet. The details regarding the use of these packets will be described later in this section.

Each of the above five phases is described below.

Reservation request phase:

In this phase, each node that needs to transmit packets sends an RR packet to the intended destination node with a contention probability *p,* in order to reserve an IS. Such nodes that send RR packets are called requesting nodes (RN). Other nodes just keep listening during this phase.

Collision report phase:

If any of the listening nodes detects collision of RR packets transmitted in the previous phase, it broadcasts a collision report (CR) packet. By listening for CR packets in this phase, an RN comes to know about collision of the RR packet it had sent. If no CR is heard by the RN in this phase, then it assumes that the RR packet did not collide in its neighborhood. It then becomes a transmitting node (TN).

Reservation confirmation phase:

An RN that does not receive any CR packet in the previous phase, that is, a TN, sends an RC packet to the destination node. Each neighbor node that receives this packet understands that the slot has been reserved, and defers its transmission during the corresponding information slots in the subsequent information frames until the next reservation frame.

Reservation acknowledgment phase:

On receiving the RC packet, the intended receiver node responds by sending an RA packet back to the TN. This is used to inform the TN that the reservation has been established. In case the TN is isolated and is not connected to any other node in the network, then it would not receive the RA packet, and thus becomes aware of the fact that it is isolated.

Packing/elimination (P/E) phase:

In this phase, a packing packet (PP) is sent by each node that is located within two hops from a TN, and that had made a reservation since the previous P/E phase. A node receiving a PP understands that there has been a recent success in slot reservation three hops away from it, and because of this some of its neighbors would have been blocked during this slot.

3.2.3 Contention-based MAC protocols with scheduling mechanisms

Protocols that fall under this category focus on packet scheduling at the nodes and transmission scheduling of the nodes. Scheduling decisions may take into consideration various factors such as delay targets of packets, laxities of packets, traffic load at nodes, and remaining battery power at nodes. In this section, some of the scheduling-based MAC protocols are described.

Distributed Priority Scheduling and Medium Access in Ad Hoc Networks

This work presents two mechanisms for providing quality of service (QoS) support for connections in ad hoc wireless networks. The first technique, called distributed priority scheduling (DPS), piggy-backs the priority tag of a node's current and head-of-line packets on the control and data packets. By retrieving information from such packets transmitted in its neighborhood, a node builds a scheduling table from which it determines its rank (information regarding its position as per the priority of the packet to be transmitted next) compared to other nodes in its neighborhood.

Distributed Priority Scheduling

The distributed priority scheduling scheme (DPS) is based on the IEEE 802.11 distributed coordination function. DPS uses the same basic RTS-CTS-DATA-ACK packet exchange mechanism. The RTS packet transmitted by a ready node carries the priority tag/priority index for the current DATA packet to be transmitted. The priority tag can be the delay target for the DATA packet. On receiving the RTS packet, the intended receiver node responds with a CTS packet. The receiver node copies the priority tag from the received RTS packet and piggybacks it along with the source node id, on the CTS packet. Neighbor nodes receiving the RTS or CTS packets (including the hidden nodes) retrieve the piggy-backed priority tag information and make a corresponding entry for the packet to be transmitted, in their scheduling tables (STs). Each node maintains an ST holding information about packets, which were originally piggy-backed on control and data packets.

Distributed Wireless Ordering Protocol

The distributed wireless ordering protocol (DWOP) consists of a media access scheme along with a scheduling mechanism. It is based on the distributed priority scheduling scheme. DWOP ensures that packets access the medium according to the order specified by an ideal reference scheduler such as first-in-first-out (FIFO), virtual clock, or earliest deadline first. In this discussion, FIFO is chosen as the reference scheduler. In FIFO, packet priority indices are set to the arrival times of packets. Similar to DPS, control packets are used in DWOP to piggy-back priority information regarding head-of-line packets of nodes. As the targeted FIFO schedule would transmit packets in order of the arrival times, each node builds up a scheduling table (ST) ordered according to the overheard arrival times. The key concept in DWOP is that a node is made eligible to contend for the channel only if its locally queued packet has a smaller arrival time compared to all other arrival times in its ST(all other packets queued at its neighbor nodes), that is, only if the node finds that it holds the next region-wise packet in the hypothetical FIFO schedule.

3.3 MAC protocols for sensor network

MAC protocols in sensor networks must create a network infrastructure to establish communication links among the thousands of randomly scattered sensors. It must also ensure fair and efficient sharing of communication resources among the nodes, so that the overall lifetime of the network can be maximized. The challenges posed by sensor network MAC protocols make them distinct from other wireless based networks.

There are three basic kinds of MAC protocols used in sensor networks: fixed-allocation, demand-based, and contention-based. Fixed-allocation MAC protocols share the common medium through a predetermined assignment. They are appropriate for sensor networks that continuously monitor and generate deterministic data traffic, since all nodes which have been allotted the channel can make use of their slot in each round. Demand-based MAC protocols are used in such cases, where the channel is allocated according to the demand of the node. Though they require the additional overhead of a reservation process, variable rate traffic can be efficiently transmitted using demand-based MAC protocols. Finally, the contention-based

MAC protocols involve random-access-based contention for the channel when packets need to be transmitted. They are again suitable for bursty traffic, but there is a possibility of collisions and no delay guarantees can be provided.

3.3.1 Self-Organizing MAC for Sensor Networks and Eavesdrop and Register

Self-organizing MAC for sensor (SMACS) networks and eavesdrop and register (EAR) are two protocols which handle network initialization and mobility support, respectively. SMACS is a distributed protocol for network initialization and link-layer organization. In this protocol, neighbor discovery and channel assignment take place simultaneously in a completely distributed manner. A communication link between two nodes consists of a pair of time slots, at a fixed frequency, which is randomly chosen at the time of establishing the link.

The EAR protocol enables seamless connection of nodes under mobile and stationary conditions. This protocol makes use of certain mobile nodes, besides the existing stationary sensor nodes, to offer service to maintain connections.

3.3.2 Hybrid TDMA/FDMA

This is a centrally controlled scheme which assumes that nodes communicate directly to a nearby BS. A pure TDMA scheme minimizes the time for which a node has to be kept on, but the associated time synchronization costs are very high. A pure FDMA scheme allots the minimum required bandwidth for each connection. If the transmitter consumes more power, a TDMAscheme is favored, since it can be switched off in idle slots to save power. On the other hand, the scheme favors FDMA when the receiver consumes greater power. This is because, in FDMA, the receiver need not expend power for time synchronization by receiving during the guard band between slots, which becomes essential in a TDMA scheme.

3.3.3. CSMA-Based MAC Protocols

Traditional CSMA-based schemes are more suitable for point-to-point stochastically distributed traffic flows. On the other hand, sensor networks have variable but periodic and correlated traffic. An adaptive transmission rate control (ARC) is also used, which balances originating and route-through traffic in nodes. This ensures that nodes closer to the BS are not favored over farther nodes. ARC uses linear increase and multiplicative decrease of originating traffic in a node. The penalty for dropping route-through traffic is higher, since energy has already been

invested in making the packets reach until that node. Hence, CSMA based MAC protocols are contention-based and are designed mainly to increase energy efficiency and maintain fairness.

3.4 Location Discovery

The location information of sensors has to be considered during aggregation of sensed data. This implies each node should know its location and couple its location information with the data in the messages it sends. A low-power, inexpensive, and reasonably accurate mechanism is needed for location discovery. A global positioning system (GPS) is not always feasible because it cannot reach nodes in dense foliage or indoors. It also consumes high power and makes sensor nodes bulkier. Two basic mechanisms of location discovery are now described.

Indoor Localization

Indoor localization techniques use a fixed infrastructure to estimate the location of sensor nodes. Fixed beacon nodes are strategically placed in the field of observation, typically indoors, such as within a building. The randomly distributed sensors receive beacon signals from the beacon nodes and measure the signal strength, angle of arrival, and time difference between the arrival of different beacon signals.

Sensor Network Localization

In situations where there is no fixed infrastructure available and prior measurements are not possible, some of the sensor nodes themselves act as beacons. They have their location information, using GPS, and these send periodic beacons to other nodes. In the case of communication using RF signals, the received signal strength indicator (RSSI) can be used to estimate the distance, but this is very sensitive to obstacles and environmental conditions.

Localization algorithms require techniques for location estimation depending on the beacon nodes' location. These are called multi-lateration (ML) techniques.

Some simple ML techniques are described in what follows

Atomic ML: If a node receives three beacons, it can determine its position by a mechanism similar toGPS. This is illustrated in Figure 1.

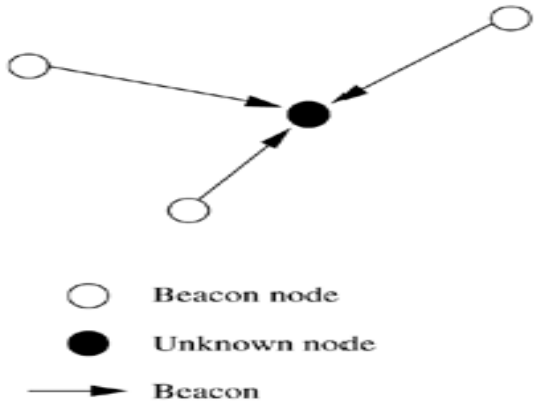
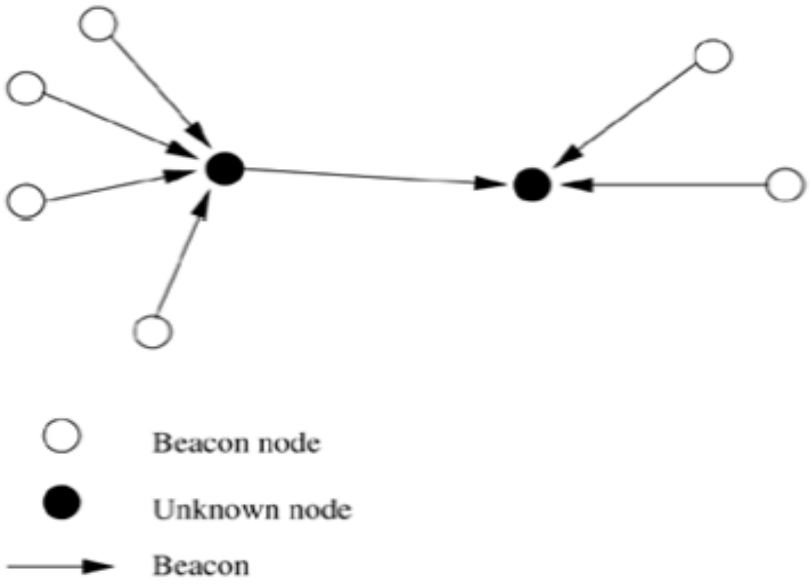
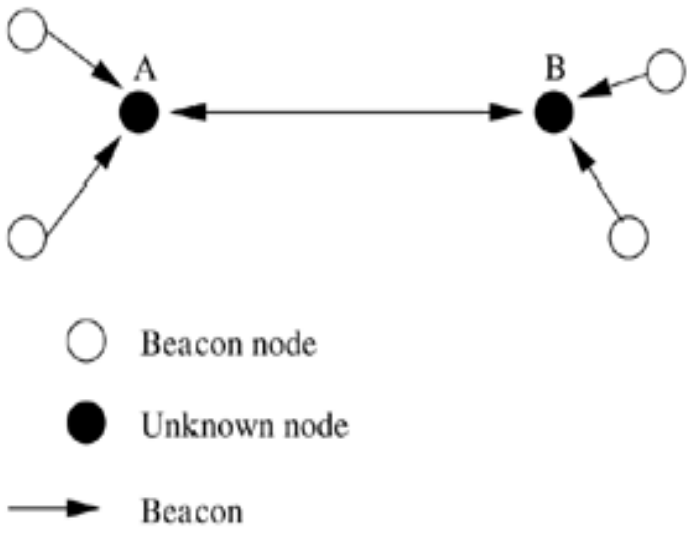


Figure 3.1. Atomic multi-lateration.

• **Iterative ML**: Some nodes may not be in the direct range of three beacons. Once a node estimates its location, it sends out a beacon, which enables some other nodes to now receive at least three beacons. Iteratively, all nodes in the network can estimate their location.



Collaborative ML: When two or more nodes cannot receive at least three beacons each, they collaborate with each other. As shown in Figure, node *A* and node *B* have three neighbors each. Of the six participating nodes, four are beacons, whose positions are known. Hence, by solving a set of simultaneous quadratic equations, the positions of *A* and *B* can be determined.



A directionality-based localization approach has been proposed, this assumes that beacon nodes have broadcast capability to reach all nodes of the network, and a central controller rotates the beacons with a constant angular velocity ω radians/s.

A mathematical technique called multidimensional scaling (MDS), an O(n3) algorithm (where n is the number of sensors), is used to assign locations to nodes such that the distance constraints are satisfied. The obtained picture of the network could be a rotated or flipped version

of the actual network. If the actual positions of any three nodes in the network are known, then the entire network can be normalized (rotated or flipped) to obtain a very accurate localization of all other nodes.

3.5 Quality of a Sensor Network

The purpose of a sensor network is to monitor and report events or phenomena taking place in a particular area. Hence, the main parameters which define how well the network observes a given area are "coverage" and "exposure."

3.5.1 Coverage

Coverage is a measure of how well the network can observe or cover an event. Coverage depends upon the range and sensitivity of the sensing nodes, and the location and density of the sensing nodes in the given region. The *worst-case* coverage defines areas of breach, that is, where coverage is the poorest. This can be used to determine if additional sensors need to be deployed to improve the network. The *best-case* coverage, on the other hand, defines the areas of best coverage. A path along the areas of best coverage is called a maximum support path or maximum exposure path.

A mathematical technique to solve the coverage problem is the Voronoi diagram. It can be proved that the path *PB* will be composed of line segments that belong to the Voronoi diagram corresponding to the sensor graph. In two dimensions, the Voronoi diagram of a set of sites is a partitioning of the plane into a set of convex polygons such that all points inside a polygon are closest to the site enclosed by the polygon, and the polygons have edges equidistant from the nearby sites. A Voronoi diagram for a sensor network, and a breach path from *I* to *F,* are shown in Figure 3.2.

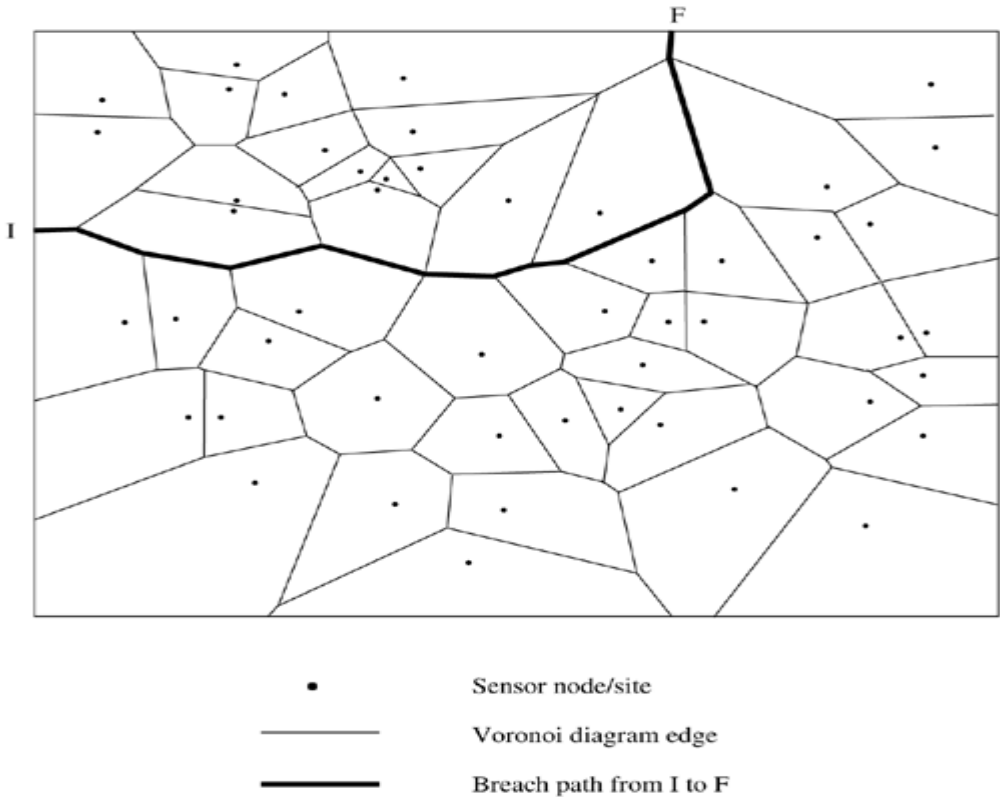


Figure 3.2. Voronoi diagram

The algorithm to find the breach path *PB* is:

- Generate the Voronoi diagram, with the set of vertices *V* and the set of edges *E.* This is done by drawing the perpendicular bisectors of every line segment joining two sites, and using their points of intersection as the vertices of the convex polygons.
- Create a weighted graph with vertices from *V* and edges from *E,* such that the weight of each edge in the graph is the minimum distance from all sensors in *S.* The edge weights represent the distance from the nearest sensor. Smaller edge weights imply better coverage along the edge.
- Determine the maximum cost path from *I* to *F,* using breadth-first search. The maximum cost implies least coverage. Hence, the required breach path is along this maximum-cost path determined from the Voronoi diagram. The breach path shows the region of maximum vulnerability in a sensor network, where the coverage provided by the sensors is the weakest.

A related problem is that of finding the best-case coverage. The problem is formally stated as finding the path which offers the maximum coverage, that is, the maximum support path *PS* in *S*, from *I* to *F*. The solution is obtained by a mathematical technique called Delaunay triangulation, shown in Figure 3.3. This is obtained from the Voronoi diagram by connecting the sites whose polygons share a common edge. The best path *PS* will be a set of line segments from the

Delaunay triangulation, connecting some of the sensor nodes. The algorithm is again similar to that used to find the maximum breach path, replacing the Voronoi diagram by the Delaunay triangulation, and defining the edge costs proportional to the line segment lengths. The maximum support path is hence formed by a set of line segments connecting some of the sensor nodes.

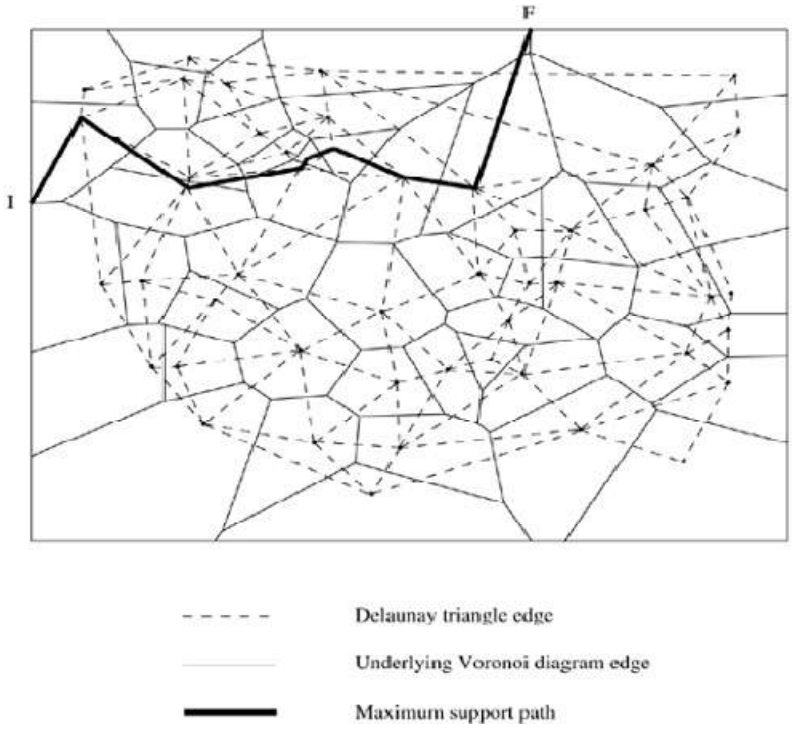


Figure 3.3. Delaunay triangulation.

3.5.2 Exposure

Exposure is defined as the expected ability of observing a target in the sensor field. It is formally defined as the integral of the sensing function on a path from source node *Ps* to destination node *Pd* . The sensing power of a node *s* at point *p* is usually modeled as

$$S(s,p) = \frac{\lambda}{[d(s,p)]^k}$$

where λ and *k* are constants, and *d*(*s, p*) is the distance of *p* from *s*. Consider a network with sensors *s*1 ,*s*2 , ..., *s*n . The total intensity at point *p,* called the all sensor field intensity, is given by

$$I_A(F,p) = \sum_{i=1}^{n} S(s_i,p)$$

The closest- sensor field intensity at *p* is

$$I_C(F,p) = S(s_{min},p)$$

where s_{min} is the closest sensor to *p*.

3.6 Other Issues

Some issues that are recently being explored in sensor networks, such as energy-efficient hardware and architecture, real-time communication on sensor networks, transport layer protocols, and security issues. Because these are mostly in the research stage, there are many improvements to be made on these fronts.

3.6.1 Energy-Efficient Design

As has been emphasized throughout the chapter, sensor nodes have a very stringent energy constraint. Energy optimization in sensor networks must prolong the life of a single node as well as of the entire network. Computation can be carried out in a power-aware manner using dynamic power management (DPM). One of the basic DPM techniques is to shut down several components of the sensor node when no events take place. The processor has a time-varying computational load, hence the voltage supplied to it can be scaled to meet only the instantaneous processing requirement. This is called dynamic voltage scaling (DVS).

3.6.2 Synchronization

Synchronization among nodes is essential to support TDMA schemes on multihop wireless networks. Also, time synchronization is useful for determining the temporal ordering of messages sent from sensors and the proximity of the sensors. Usually, sensor nodes are dropped into the environment from which data has to be collected, and their exact positions are not fixed before deployment.

Synchronization protocols typically involve delay measurements of control packets. The delays experienced during a packet transmission can be split into four major components: send time, access time, propagation time, and receive time. The send time is the time spent at the sender to construct the message.

Resynchronization is the process of synchronizing different network partitions that are independently synchronized to different clocks to a common clock. In dynamic networks such as sensor networks, frequent changes in topology make resynchronization an important issue. Resynchronization takes place in situations such as the merging of two partitions due to mobility, where all clocks in a partition may need to be updated to match the leader of the other partition, as shown in Figure 3.4.

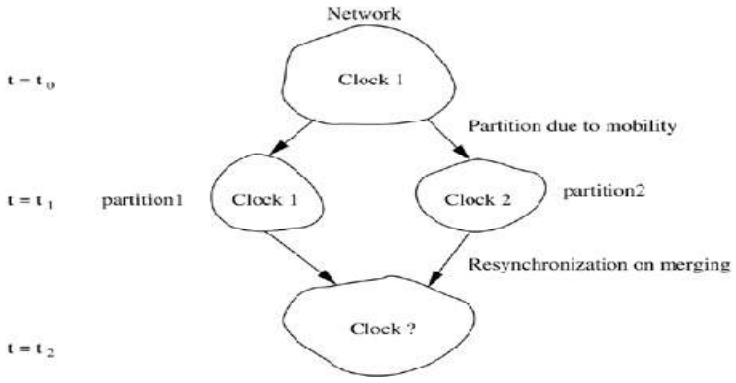


Figure 3.4. Resynchronization.

3.6.3 Transport Layer Issues

The major issue in transport layer protocols for sensor networks is the provision of reliable data delivery. This assumes special significance in the design of general-purpose sensor networks, where groups of nodes may need to be reconfigured or reprogrammed to suit an evolving application. This may require disseminating a code segment to some nodes, where loss of even a single line of code would render the retasking operation a failure. PSFQ consists of three functions: message relaying (pump), error recovery (fetch), and selective status reporting (report). The pump operation disseminates data to all target nodes, performs flow control, and localizes loss by ensuring caching at intermediate nodes. Hence, the errors on one link are rectified locally without propagating them down the entire path. When a receiver detects gaps in the received sequence numbers, a loss is indicated, and it goes into fetch mode. It requests a retransmission from neighbor nodes. An attempt is made to aggregate losses, that is, many message losses are batched into a single fetch operation, which is especially appropriate for bursty losses.

3.6.4 Security

Sensor networks, based on an inherently broadcast wireless medium, are vulnerable to a variety of attacks. Security is of prime importance in sensor networks because nodes assume a large amount of trust among themselves during data aggregation and event detection. From a set of sensor nodes in a given locality, only one final aggregated message may be sent to the BS, so it is necessary to ensure that communication links are secure for data exchange.

3.6.5 Real-Time Communication

Support for real-time communication is often essential in sensor networks which are used for surveillance or safety-critical systems. The communication delay between sensing an intrusion and taking appropriate action greatly affects the quality of tracking provided by a surveillance system. Two protocols which support real-time communication in sensor networks — SPEED and RAP — are discussed in this section.

SPEED

SPEED is a localized algorithm which provides real-time unicast, real-time area-multicast (multicast to all nodes in a particular region), and real-time anycast support for packet transmission. SPEED has minimal overheads, as it does not require routing tables. It is compatible with best-effort MAC layer, not requiring any special MAC support. It also distributes traffic and load equally across the network using nondeterministic forwarding.

RAP

RAP provides APIs for applications to address their queries. An application layer program in theBS can specify the kind of event information required, the area to which the query is addressed, and the deadline within which information is required. The underlying layers of RAP ensure that the query is sent to all nodes in the specified area, and the results are sent back to the BS. The protocol stack of RAP consists of location addressed protocol (LAP) in the transport layer, velocity monotonic scheduling (VMS) as the geographic routing protocol, and a contention-based MAC scheme that supports prioritization.

Self-Organizing MAC for Sensor Networks (S-MAC)

S-MAC is a medium-access control (MAC) protocol designed for wireless sensor networks. Wireless sensor networks use battery-operated computing and sensing devices. A network of these devices will collaborate for a common application such as environmental monitoring. We expect sensor networks to be deployed in an ad hoc fashion, with individual nodes remaining largely inactive for long periods of time, but then becoming suddenly active when something is detected. These characteristics of sensor networks and applications motivate a MAC that is different from traditional wireless MACs such as IEEE 802.11 in almost every way: energy conservation and self-configuration are primary goals, while per-node fairness and latency are less important.

S-MAC uses three novel techniques to reduce energy consumption and support self-configuration. To reduce energy consumption in listening to an idle channel, nodes periodically sleep. Neighboring nodes form *virtual clusters* to auto-synchronize on sleep schedules. Inspired by PAMAS, S-MAC also sets the radio to sleep during transmissions of other nodes. Unlike PAMAS, it only uses in-channel signaling. Finally, S-MAC applies *message passing* to reduce contention latency for sensor-network applications that require store-and-forward processing as data move through the network. We evaluate our implementation of S-MAC over a sample sensor node, the Mote, developed at University of California, Berkeley. The experiment results show that, on a source node, an 802.11-like MAC consumes 2--6 times more energy than S-MAC for traffic load with messages sent every 1-10s.

IEEE 802.15.4:

he IEEE 802.15.4 standard was developed to provide a framework and the lower levels for low cost, low power networks. It only provides the MAC and PHY layers, leaving the upper layers to be developed according to the market needs.

Now with technologies such as Zigbee being used in a large way, the use of IEEE 802.15.4 technology is corresponding increasing, and it is becoming an important standard. However with widespread marketing for Zigbee and other standards, IEEE 802.15.4 is less well known.

IEEE 802.15.4 basics

The IEEE 802.15.4 standard is aimed at providing the essential lower network layers for a wireless personal area network (WPAN). The chief requirements are low-cost, low-speed ubiquitous communication between devices. It does not aim to compete with the more commonly used end user-oriented systems such as IEEE 802.11 where costs are not as critical and higher speeds are demanded. Instead, IEEE 802.15.4 provides for very low cost communication of nearby devices with little to no underlying infrastructure.

IEEE 802.15.4 standard

The IEEE 802.15.4 standard has undergone a number of releases. In addition to this there are a number of variants of the IEEE 802.15.4 standard to cater for different forms of physical layer, etc. These are summarized below in the table.

IEEE 802.15.4 VERSION	DETAILS AND COMMENTS
IEEE 802.15.4 - 2003	This was the initial release of the IEEE 802.15.4 standard. It provided for two different PHYs - one for the lower frequency bands of 868 and 915 MHz, and the other for 2.4 GHz.
IEEE 802.15.4 - 2006	This 2006 release of the IEEE 802.15.4 standard provided for an increase in the data rate achievable on the lower frequency bands. This release of the standard updated the PHY for 868 and 915 MHz. It also defined four new modulation schemes that could be used - three for the lower frequency bands, and one for 2.4 GHz.
IEEE 802.15.4a	This version of the IEEE 802.15.4 standard defined two new PHYs. One used UWB technology and the other provided for using chirp spread spectrum at 2.4 GHz.
IEEE 802.15.4c	Updates for 2.4 GHz, 868 MHz and 915 MHz, UWB and the China 779-787 MHz band.
IEEE 802.15.4d	2.4 GHz, 868 MHz, 915 MHz and Japanese 950 - 956 MHz band.
IEEE 802.15.4e	This release defines MAC enhancements to IEEE 802.15.4 in support of the ISA SP100.11a application.
IEEE 802.15.4f	This will define new PHYs for UWB, 2.4 GHz band and also 433 MHz
IEEE 802.15.4g	This will define new PHYs for smart neighbourhood networks. These may include applications such as smart grid applications for the energy industry. It may include the 902 - 928 MHz band.

Although new versions of the standard are available for use by any of the higher layer standards, Zigbee still uses the initial 2003 release of the IEEE 802.15.4 standard.

IEEE 802.15.4 applications

The IEEE 802.15.4 technology is used for a variety of different higher layer standards. In this way the basic physical and MAC layers are already defined, allowing the higher layers to be provided by individual system in use.

APPLICATION OR SYSTEM	DESCRIPTION OF THE IEEE 802.15.4 APPLICATION OR SYSTEM
Zigbee	Zigbee is supported by the Zigbee Alliance and provides the higher levels required for low powered radio system for control applications including lighting, heating and many other applications.
Wireless HART	WirelessHART is an open-standard wireless networking technology that has been developed by HART Communication Foundation for use in the 2.4 GHz ISM band. The system uses IEEE802.15.4 for the lower layers and provides a time synchronized, self-organizing, and self-healing mesh architecture.
RF4CE	RF4CE, Radio Frequency for Consumer Electronics has amalgamated with the Zigbee alliance and aims to provide low power radio controls for audio visual applications, mainly for domestic applications such as set to boxes, televisions and the like. It promises enhanced communication and facilities when compared to existing controls.
MiWi	MiWi and the accompanying MiWi P2P systems are designed by Microchip Technology. They are designed for low data transmission rates and short distance, low cost networks and they are aimed at applications including industrial monitoring and control, home and building automation, remote control and automated meter reading.
ISA100.11a	This standard has been developed by ISA as an open-standard wireless networking technology and is it described as a wireless system for industrial automation including process control and other related applications.
6LoWPAN	This rather unusual name is an acronym for "IPv6 over Low power Wireless Personal Area Networks" It is a system that uses the basic IEEE 802.15.4, but using packet data in the form of Ipv6.

While the IEEE 802.15.4 standard may not be as well known as some of the higher level standards and systems such as Zigbee that use IEEE 802.15.4 technology as the underpinning lower levels system, it is nevertheless very important. It spans a variety of different systems, and as such provides a new approach - providing only the lower layers, and allowing other systems to provide the higher layers which are tailored for the relevant application.

UNIT – IV

Routing Protocols

Introduction

A variety of routing protocols for ad hoc wireless networks has been proposed in the recent past. This chapter first presents the issues involved in designing a routing protocol and then the different classifications of routing protocols for ad hoc wireless networks. It then discusses the working of several existing routing protocols with illustrations.

4.1 Issues in designing a routing protocol

The major challenges that a routing protocol designed for ad hoc wireless networks faces are mobility of nodes, resource constraints, error-prone channel state, and hidden and exposed terminal problems. A detailed discussion on each of the following is given below.

4.1.1 Mobility

- Network topology is highly dynamic due to movement of nodes. hence, an ongoing session suffers frequent path breaks.
- Disruption occurs due to the movement of either intermediate nodes in the path or end nodes.
- Wired network routing protocols cannot be used in adhoc wireless networks because the nodes are here are not stationary and the convergence is very slow in wired networks.
- Mobility of nodes results in frequently changing network topologies.
- Routing protocols for ad hoc wireless networks must be able to perform efficient and effective mobility management.

4.1.2 Bandwidth Constraint

- Abundant bandwidth is available in wired networks due to the advent of fiber optics and due to the exploitation of wavelength division multiplexing (WDM) technologies.
- In a wireless network, the radio band is limited, and hence the data rates it can offer are much less than what a wired network can offer.
- This requires that the routing protocols use the bandwidth optimally by keeping the overhead as low as possible.
- The limited bandwidth availability also imposes a constraint on routng protocols in maintaining the topological information.

4.1.3 Error-prone shared broadcast radio channel

- The broadcast nature of the radio channel poses a unique challenge in ad hoc wireless networks.
- The wireless links have time-varying characteristics in terms of link capacity and link-error probability.
- This requires that the adhoc wireless network routing protocol interact with the MAC layer to find alternate routes through better-quality links.

- Transmissions in ad hoc wireless networks result in collisions of data and control packets.
- Therefore, it is required that ad hoc wireless network routing protocols find paths with less congestion.

4.1.4 Hidden and exposed terminal problems

- The hidden terminal problem refers to the collision of packets at a receiving node due to the simultaneous transmission of those nodes that are not within the direct transmission range of the receiver, but are within the transmission range of the receiver.
- Collision occurs when both nodes transmit packets at the same time without knowing about the transmission of each other.
- Ex: consider figure 4.1. Here, if both node A and node C transmit to node B at the same time, their packets collide at node B. This is due to the fact that both node A and C are hidden from each other, as they are not within the direct transmission range of each other and hence do not know about the presence of each other.

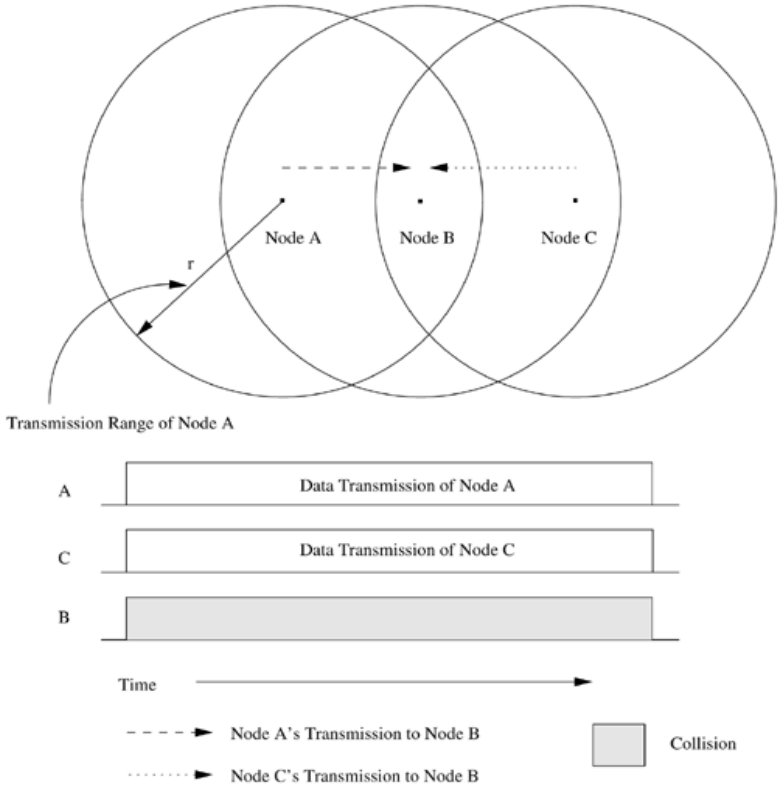


Figure 4.1. Hidden terminal problem.

The exposed terminal problem refers to the inability of a node which is blocked due to transmission by a nearby transmitting node to transmit to another node. Consider the example in Figure 4.3. Here, if a transmission from node B to another node A is already in progress, node C cannot transmit to node D, as it concludes that its neighbor, node B, is in transmitting mode and hence should not interfere with the on-going transmission. Thus, reusability of the radio spectrum is affected. For node C to transmit simultaneously when node B is transmitting, the transmitting frequency of node C must be different from its receiving frequency.

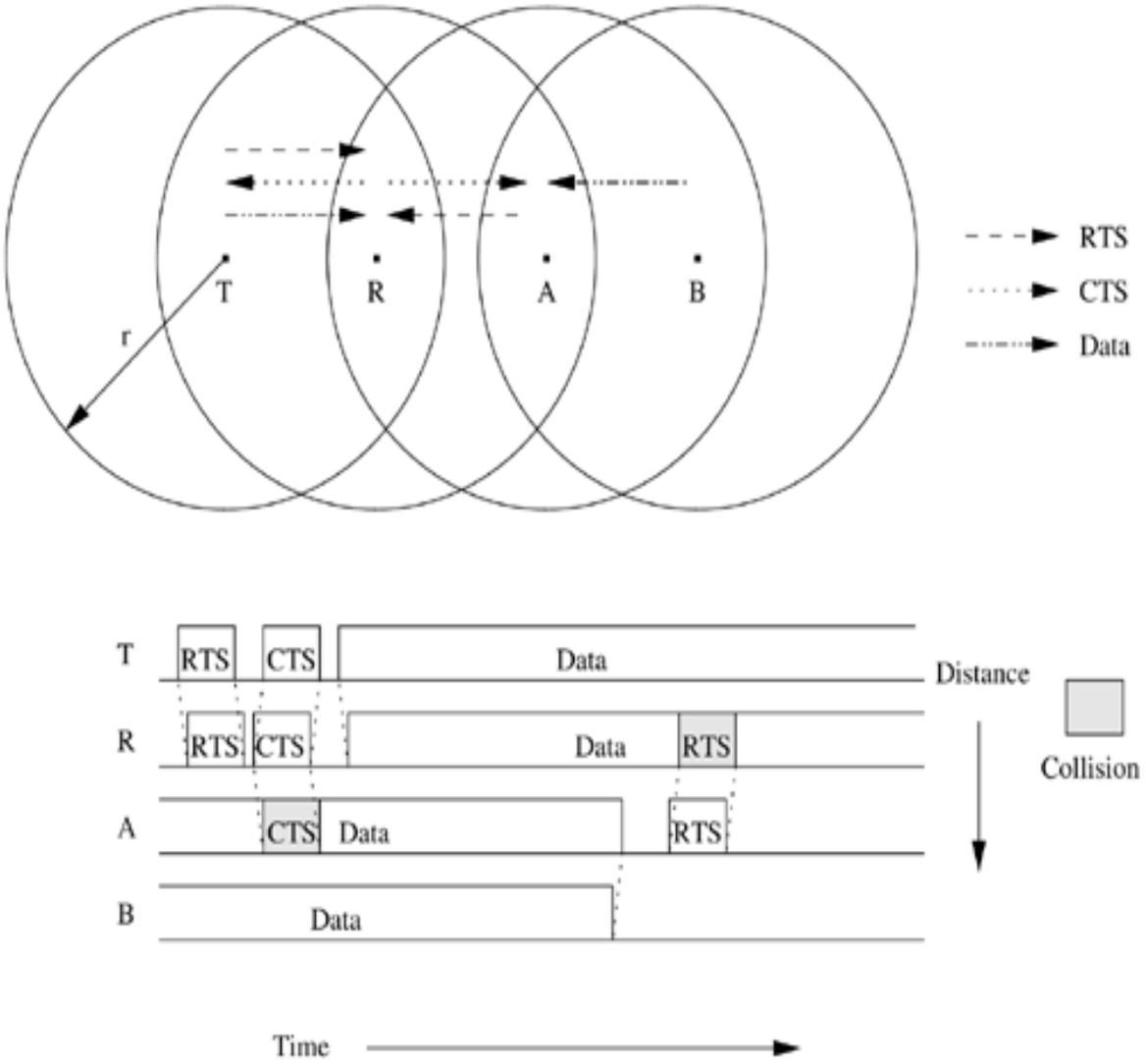


Figure 4.2. Hidden terminal problem with RTS-CTS-Data-ACK scheme.

Solution for this problem includes medium access collision avoidance (MACA):

- Transmitting node first explicitly notifies all potential hidden nodes about the forthcoming transmission by means of a two-way handshake control protocol called RTS-CTS protocol exchange.
- This may not solve the problem completely but it reduces the probability of collisions.

Medium access collision avoidance for wireless (MACAW):

- An improved version of MACA protocol.
- Introduced to increase the efficiency.
- Requires that a receiver acknowledges each successful reception of data packet.

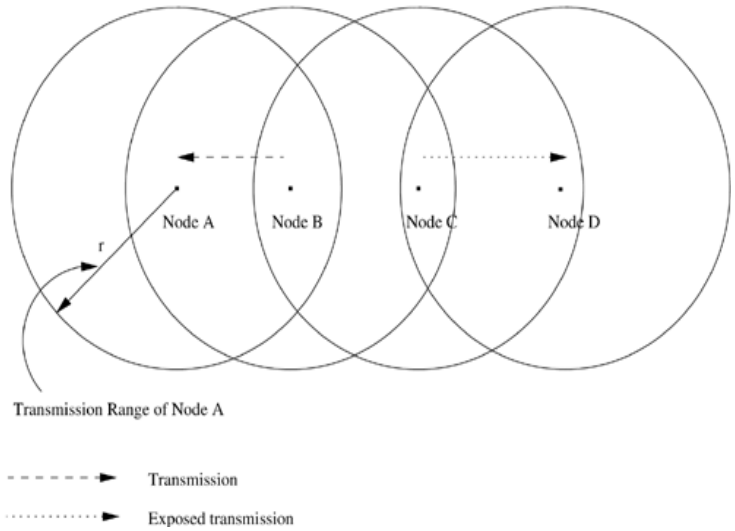


Figure 4.3. Exposed terminal problem**.**

Other solutions include floor acquisition multiple access (FAMA) and Dual busy tone multiple access (DBTMA).

- The exposed terminal problem refers to the inability of a node which is blocked due to transmission by a nearby transmitting node to transmit to another node.
- Ex: consider the figure 4.3. Here, if a transmission from node B to another node A is already in progress, node C cannot transmit to node D, as it concludes that its neighbor node B, is in transmitting mode and hence should not interfere with the on-going transmission. Thus, reusability of the radio spectrum is affected.

- **Resource Constraints**
  - Two essential and limited resources are battery life and processing power.
  - Devices used in adhoc wireless networks require portability, and hence they also have size and weight constraints along with the restrictions on the power source.
  - Increasing the battery power and processing ability makes the nodes bulky and less portable.

4.2 Characteristics of an Ideal Routing Protocol for Ad Hoc Wireless Networks

A routing protocol for ad hoc wireless networks should have the following characteristics:

- It must be fully distributed as centralized routing involves high control overhead and hence is not scalable.
- It must be adaptive to frequent topology changes caused by the mobility of nodes.
- Route computation and maintenance must involve a minimum number of nodes. Each node in the network must have quick access to routes, that is, minimum connection setup time is desired.
- It must be localized, as global state maintenance involves a huge state propagation control overhead.

- It must be loop-free and free from state routes.
- The number of packet collisions must be kept to a minimum by limiting the number of broadcasts made by each node. The transmissions should be reliable to reduce message loss and to prevent the occurrence of state routes.
- It must converge to optimal routes once the network topology becomes stable. The convergence must be quick.
- It must optimally use scarce resources such as bandwidth, computing power, memory, and battery power.
- Every node in the network should try to store information regarding the stable local topology only. Changes in remote parts of the network must not cause updates in the topology information maintained by the node.
- It should be able to provide a certain level of quality of service (QoS) as demanded by the applications, and should also offer support for time-sensitive traffic.

4.3 Classifications of Routing Protocols

The routing protocols for ad hoc wireless networks can be broadly classified into four categories based on

- Routing information update mechanism
- Use of temporal information for routing
- Routing topology
- Utilization of specific resources

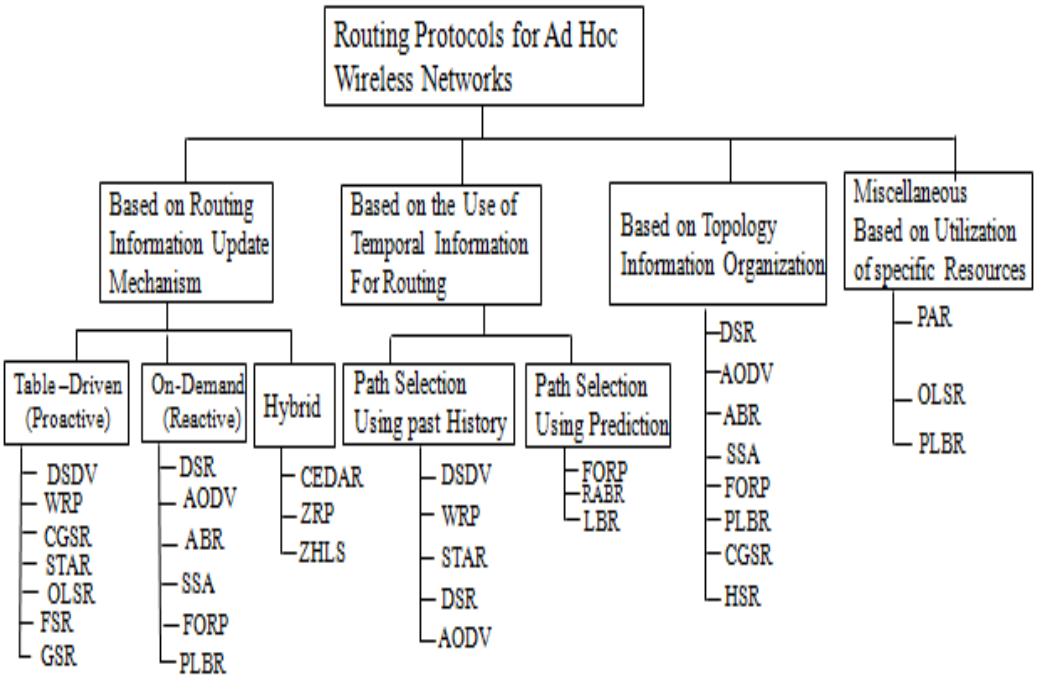


Figure 4.4. Classifications of routing protocols.

Based on the Routing Information Update Mechanism

Ad hoc wireless network routing protocols can be classified into 3 major categories based on the routing information update mechanism. They are:

- *Proactive or table-driven routing protocols* :
  - Every node maintains the network topology information in the form of routing tables by periodically exchanging routing information.
  - Routing information is generally flooded in the whole network.
  - Whenever a node requires a path to a destination, it runs an appropriate path-finding algorithm on the topology information it maintains.
- *Reactive or on-demand routing protocols*:
  - Do not maintain the network topology information.
  - Obtain the necessary path when it is required, by using a connection establishment process.
- *Hybrid routing protocols:*
  - Combine the best features of the above two categories.
  - Nodes within a certain distance from the node concerned, or within a particular geographical region, are said to be within the routing zone of the given node.
  - For routing within this zone, a table-driven approach is used.
  - For nodes that are located beyond this zone, an on-demand approach is used.

Based on the use of temporal information for routing

The protocols that fall under this category can be further classified into two types:

- *Routing protocols using past temporal information:*
  - Use information about the past status of the links or the status of links at the time of routing to make routing decisions.
- *Routing protocols that use future temporal information:*
  - Use information about the about the expected future status of the wireless links to make approximate routing decisions.
  - Apart from the lifetime of wireless links, the future status information also includes information regarding the lifetime of the node, prediction of location, and prediction of link availability.

Based on the routing topology

Ad hoc wireless networks, due to their relatively smaller number of nodes, can make use of either a flat topology or a hierarchical topology for routing.

- *Flat topology routing protocols:*
  - Make use of a flat addressing scheme similar to the one used in IEEE 802.3 LANs.

- It assumes the presence of a globally unique addressing mechanism for nodes in an ad hoc wireless network.

- *Hierarchical topology routing protocols:*
  - Make use of a logical hierarchy in the network and an associated addressing scheme.
  - The hierarchy could be based on geographical information or it could be based on hop distance.

Based on the utilization of specific resources

- *Power-aware routing:*
  - Aims at minimizing the consumption of a very important resource in the ad hoc wireless networks: the battery power.
  - The routing decisions are based on minimizing the power consumption either logically or globally in the network.
- *Geographical information assisted routing :*
  - Improves the performance of routing and reduces the control overhead by effectively utilizing the geographical information available.

4.3.1 Table-Driven Routing Protocols

- These protocols are extensions of the wired network routing protocols.
- They maintain the global topology information in the form of tables at every node.
- Tables are updated frequently in order to maintain consistent and accurate network state information

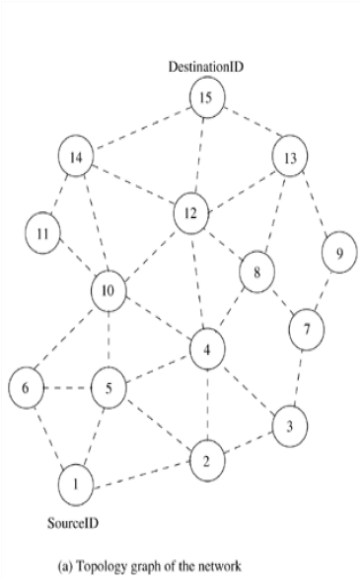
Ex: Destination sequenced distance vector routing protocol (DSDV), wireless routing protocol (WRP), source-tree adaptive routing protocol (STAR) and cluster-head gateway switch routing protocol (CGSR).

4.3.1.1 Destination sequenced distance-vector routing protocol

- It is an enhanced version of the distributed Bellman-Ford algorithm where each node maintains a table that contains the shortest distance and the first node on the shortest path to every other node in the network.
- It incorporates table updates with increasing sequence number tags to prevent loops, to counter the count-to-infinity problem, and for faster convergence.
- As it is a table-driven routing protocol, routes to all destinations are readily available at every node at all times.
- The tables are exchanged between neighbors at regular intervals to keep an up-to-date view of the network topology.
- The table updates are of two types:
  - ***Incremental updates:*** Takes a single network data packet unit (NDPU). These are used when a node does not observe significant changes in the local topology.

- ***Full dumps:*** Takes multiple NDPUs. It is done either when the local topology changes significantly or when an incremental update requires more than a single NDPU.

- Table updates are initiated by a destination with a new sequence number which is always greater than the previous one.
- Consider the example as shown in figure 4.5 (a). Here node 1 is the source node and node 15 is the destination. As all the nodes maintain global topology information, the route is already available as shown in figure 4.5 (b).
- Here the routing table node 1 indicates that the shortest route to the destination node is available through node 5 and the distance to it is 4 hops, as depicted in figure (b)
- The reconfiguration of a path used by an on-going data transfer session is handled by the protocol in the following way.
- The end node of the broken link initiates a table update message with the broken link's weight assigned to infinity (∞) and with a sequence number greater than the stored sequence number for that destination.
- Each node upon receiving an update with weight ∞, quickly disseminates it to its neighbors in order to propagate the broken-link information to the whole network.
- A node always assign an odd number to the link break update to differentiate it from the even sequence number generated by the destination.
- Figure 4.6 shows the case when node 11 moves from its current position.

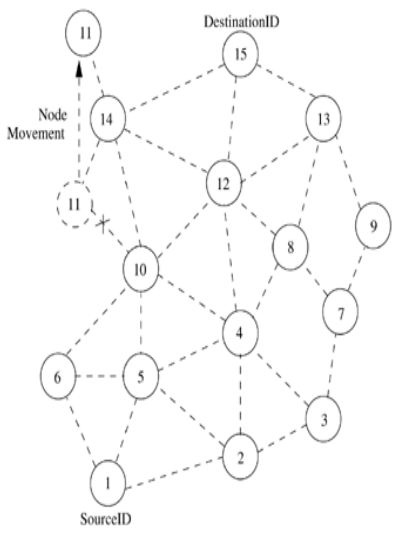


(a) Topology graph of the network

Dest	NextNode	Dist	SeqNo
2	2	1	22
3	2	2	26
4	5	2	32
5	5	1	134
6	6	1	144
7	2	3	162
8	5	3	170
9	2	4	186
10	6	2	142
11	6	3	176
12	5	3	190
13	5	4	198
14	6	3	214
15	5	4	256

(b) Routing table for Node 1

Figure 4.5. Route establishment in DSDV.



Routing Table for Node 1

Dest	NextNode	Dist	SeqNo
2	2	1	22
3	2	2	26
4	5	2	32
5	5	1	134
6	6	1	144
7	2	3	162
8	5	3	170
9	2	4	186
10	6	2	142
11	5	4	180
12	5	3	190
13	5	4	198
14	6	3	214
15	5	4	256

Figure 4.6. Route maintenance in DSDV**.**

Advantages

- Less delay involved in the route setup process.
- Mechanism of incremental update with sequence number tags makes the existing wired network protocols adaptable to ad hoc wireless networks.

- The updates are propagated throughout the network in order to maintain an up-to-date view of the network topology at all nodes.

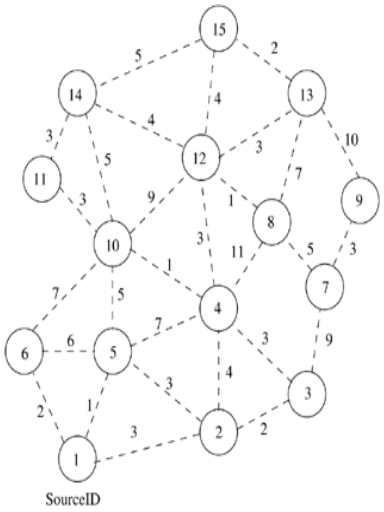
Disadvantages

- The updates due to broken links lead to a heavy control overhead during high mobility.
- Even a small network with high mobility or a large network with low mobility can completely choke the available bandwidth Suffers from excessive control overhead.
- In order to obtain information about a particular destination node, a node has to wait for a table update message initiated by the same destination node.
- This delay could result in state routing information at nodes.

4.3.1.2 Wireless Routing Protocol (WRP)

- WRP is similar to DSDV; it inherits the properties of the distributed bellman-ford algorithm.
- To counter the count-to-infinity problem and to enable faster convergence, it employs a unique method of maintaining information regarding the shortest distance to every destination node in the network and penultimate hop node on the path to every destination node.
- Maintains an up-to-date view of the network, every node has a readily available route to every destination node in the network.
- It differs from DSDV in table maintenance and in the update procedures.
- While DSDV maintains only one topology table, WRP uses a set of tables to maintain more accurate information.
- The table that are maintained by a node are:
    - Distance table (DT): contains the network view of the neighbors of a node. It contains a matrix where each element contains the distance and the penultimate node reported by the neighbor for a particular destination.
    - Routing table (RT): contains the up-to-date view of the network for all known destinations. It keeps the shortest distance, the predecessor/penultimate node, the successor node, and a flag indicating the status of the path. The path status may be a simplest (correct) path or a loop (error), or destination node not marked (null).
    - Link cost table (LCT): contains the cost of relaying messages through each link. The cost of broken link is ∞.it also contains the number of update periods passed since the last successful update was received from that link.
    - Message retransmission list (MRL): contains an entry for every update message that is to be retransmitted and maintains a counter for each entry.
    - After receiving the update message, a node not only updates the distance for transmitted neighbors but also checks the other neighbors' distance, hence convergence is much faster than DSDV.

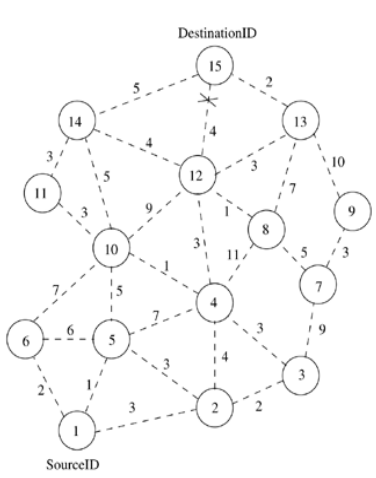
- Consider the example shown in figure below, where the source of the route is node 1 and destination is node 15. As WRP proactively maintains the route to all destinations, the route to any destination node is readily available at the source node.
- From the routing table shown, the route from node 1 to node 15 has the next node as node 2. The predecessor node of 15 corresponding to this route is route 12. The predecessor information helps WRP to converge quickly during link breaks.



Routing Entry at Each Node for DestinationID 15

Node	NextNode	Pred	Cost
15	15	15	0
14	15	14	5
13	15	13	2
12	15	12	4
11	14	14	8
10	4	12	8
9	13	13	12
8	12	12	5
7	8	12	10
6	10	12	15
5	10	12	13
4	12	12	10
3	4	12	7
2	4	12	11
1	2	12	14

Figure 4.7. Route establishment in WRP.



Routing Entry at Each Node for DestinationID 15

Node	NextNode	Pred	Cost
15	15	15	0
14	15	14	5
13	15	13	2
12	15	13	5
11	14	14	8
10	4	13	9
9	13	13	12
8	12	13	6
7	8	13	11
6	10	13	16
5	10	13	14
4	12	13	8
3	4	13	11
2	4	13	12
1	2	13	15

Figure 4.8. Route maintenance in WRP.

When a node detects a link break, it sends an update message to its neighbors with the link cost of the broken link set to ∞. After receiving the update message; all affected nodes update their minimum distances to the corresponding nodes. The node that initiated the update message then finds an alternative route, if available from its DT. Figure 4.8 shows route maintenance in WRP.

Advantages

- WRP has the same advantages as that of DSDV.
- It has faster convergence and involves fewer table updates.

Disadvantages

- The complexity of maintenance of multiple tables demands a larger memory and greater processing power from nodes in the adhoc wireless network.
- It is not suitable for highly dynamic and also for very large ad hoc wireless networks.

4.3.1.3 Cluster-Head Gateway Switch Routing Protocol (CGSR)

- Uses a hierarchical network topology, CGSR organizes nodes into clusters, with coordination among the members of each cluster entrusted to a special node named *cluster-head.* This cluster-head is elected dynamically by employing a least cluster change (LCC) algorithm.

- According to this algorithm, a node ceases to be a cluster-head only if it comes under the range of another cluster-head, where the tie is broken either using the lowest ID or highest connectivity algorithm.
- Clustering provides a mechanism to allocate bandwidth, which is a limited resource, among different clusters, thereby improving reuse.
- A token-based scheduling is used within a cluster for sharing the bandwidth among the members of the cluster.
- CGRS assumes that all communication passes through the cluster-head. Communication between 2 clusters takes place through the common member nodes that are members of both the cluster are called *gateways.*
- A gateway is expected to be able to listen to multiple spreading codes that are currently in operation in the clusters in which the node exist as a member.
- A gateway conflict is said to occur when a cluster-head issues a token to a gateway over spreading code while the gateway is tuned to another code.
- Gateways that are capable of simultaneously communicating over two interfaces can avoid gateway conflicts.
- The performance of routing is influenced by token scheduling and code scheduling that is handled at cluster-heads and gateways, respectively.
- Every member node maintains a routing table containing the destination cluster-head for every node in the network.
- In addition to the cluster member table, each node maintains a routing table which keeps the list of next-hop nodes for reaching every destination cluster.
- The cluster routing protocol is used here.
- Figure below shows the cluster head, cluster gateways, and normal cluster member nodes in an ad hoc wireless network.

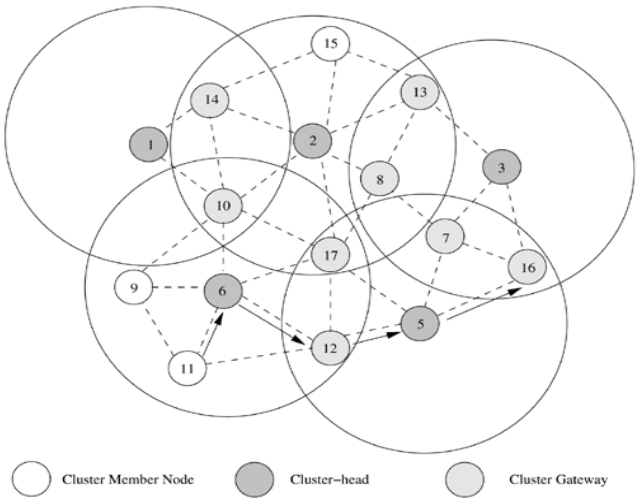


Figure 4.9. Route establishment in CGSR**.**

Advantages

- CGSR is a hierarchical routing scheme which enables partial coordination between nodes by electing cluster-heads.
- Better bandwidth utilization is possible.
- Easy to implement priority scheduling schemes with token scheduling and gateway code scheduling.

Disadvantages

- Increase in path length and instability in the system at high mobility when the rate of change of cluster-head is high.
- In order to avoid gateway conflicts, more resources are required.
- The power consumption at the cluster-head node is also a matter of concern.
- Lead to Frequent changes in the cluster-head, which may result in multiple path breaks.

4.3.1.4 Source-Tree Adaptive Routing Protocol (STAR)

Key concept: least overhead routing approach (LORA)

- This protocol attempts to provide feasible paths that are not guaranteed to be optimal Involves much less control overhead.
- In STAR protocol, every node broadcasts its source tree information.
- The source tree of a node consists of the wireless links used by the node.
- Every node builds a partial graph of the topology.
- During initialization, a node sends an update message to its neighbors.
- Each node will have a path to every destination node.
- The path would be sub-optimal; the data packet contains information about the path to be traversed in order to prevent the possibility of routing loop formation.
- In the presence of a reliable broadcast mechanism, STAR assumes implicit route maintenance.
- In addition to path breaks, the intermediate nodes are responsible for handling the routing loops.
- The RouteRepair packet contains the complete source tree of node k and the traversed path of the packet.
- When an intermediate node receives a RouteRepair update message, it removes itself from the top of the route repair path and reliably sends it to the head of the route repair path.

Advantages

- Very low communication overhead.
- Reduces the average control overhead.

4.3.2 On-Demand Routing Protocols

They execute the path-finding process and exchange routing information only when a path is required by a node to communicate with a destination

4.3.2.1 Dynamic Source Routing Protocol (DSR)

- Designed to restrict the bandwidth consumed by control packets in adhoc wireless networks by eliminating the periodic table update messages.
- It is beacon-less and does not require periodic hello packet transmissions.
- Basic approach is to establish a route by flooding RouteRequest packets in the network.
- Destination node responds by sending a RouteReply packet back to the source.
- Each RouteRequest carries a sequence number generated by the source node and the path it has traversed, a node checks the sequence number on the packet before forwarding it.
- The packet is forwarded only if it is not a duplicate RouteRequest.
- The sequence number on the packet is used to prevent loop formations and to avoid multiple transmissions.
- Thus, all nodes except the destination forward a RouteRequest packet during the route construction phase. In figure 4.10, source node 1 initiates a RouteRequest packet to obtain a path for destination node 15.
- This protocol uses a route cache that stores all possible information extracted from the source route contained in a data packet.
- During network partitions, the affected nodes initiate RouteRequest packets.
- DSR also allows piggy-backing of a data packet on the RouteRequest.
- As a part of optimizations, if the intermediate nodes are also allowed to originate RouteReply packets, then a source node may receive multiple replies from intermediate nodes.
- In fig 4.11, if the intermediate node 10 has a route to the destination via node 14, it also sends the RouteReply to the source node.
- The source node selects the latest and best route and uses that for sending data packets.
- Each data packet carries the complete path to its destination.
- If a link breaks, source node again initiates the route discovery process

All the intermediate nodes flood the *RouteRequest* packet if it is not redundant. For example, after receiving the *RouteRequest* packet from node 1 (refer to Figure 4.10), all its neighboring nodes, that is, nodes 2, 5, and 6, forward it. Node 4 receives the *RouteRequest* from both nodes 2 and 5. Node 4 forwards the first *RouteRequest* it receives from any one of the nodes 2 and 5 and discards the other redundant/duplicate *RouteRequest* packets. The *RouteRequest* is propagated till it reaches the destination which initiates the *RouteReply.* As part of optimizations, if the intermediate nodes are also allowed to originate *RouteReply* packets, then a source node may receive multiple replies from intermediate nodes. For example, in Figure 4.11, if the intermediate node 10 has a route to the destination via node 14, it also sends the *RouteReply* to the source

node. The source node selects the latest and best route, and uses that for sending data packets. Each data packet carries the complete path to its destination. A *RouteError* message is generated from the node adjacent to the broken link to inform the source node. The source node reinitiates the route establishment procedure. The cached entries at the intermediate nodes and the source node are removed when a*RouteError* packet is received. If a link breaks due to the movement of edge nodes (nodes 1 and 15), the source node again initiates the route discovery process.

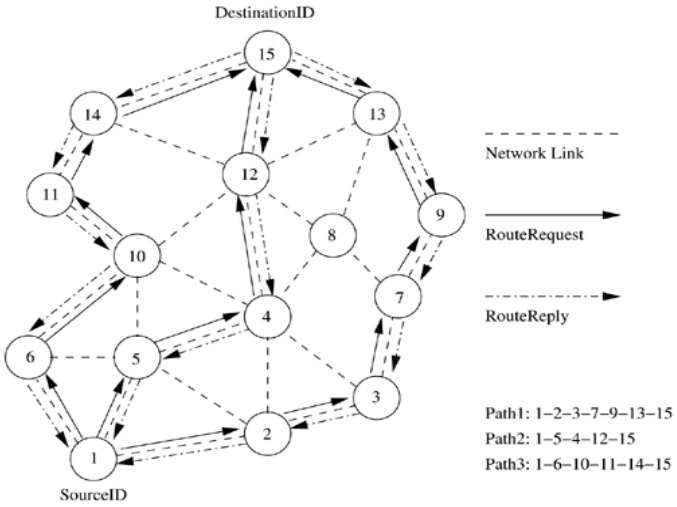


Figure 4.10. Route establishment in DSR.

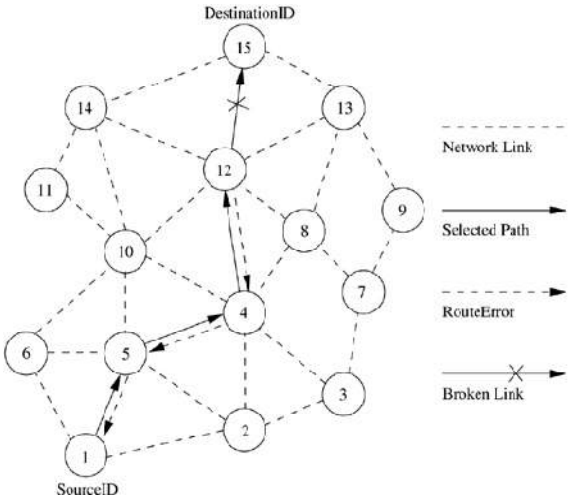


Figure 4.11. Route maintenance in DSR.

Advantages

- Uses a reactive approach which eliminates the need to periodically flood the network with table update messages.
- Route is established only when required.
- Reduce control overhead

Disadvantages

- Route maintenance mechanism does not locally repair a broken link.
- Stale route cache information could result in inconsistencies during route construction phase.
- Connection set up delay is higher.
- Performance degrades rapidly with increasing mobility.
- Routing overhead is more & directly proportional to path length

4.3.2.2 Ad Hoc On-Demand Distance Vector Routing Protocol (AoDV)

- Route is established only when it is required by a source node for transmitting data packets.
- It employs destination sequence numbers to identify the most recent path.
- Source node and intermediate nodes store the next hop information corresponding to each flow for data packet transmission.
- Uses DestSeqNum to determine an up-to-date path to the destination.

- A RouteRequest carries the source identifier, the destination identifier, the source sequence number, the destination sequence number, the broadcast identifier and the time to live field.
- DestSeqNum indicates the freshness of the route that is accepted by the source.
- When an intermediate node receives a RouteRequest, it either forwards it or prepares a RouteReply if it has a valid route to the destination. The validity of the intermediate node is determined by comparing the sequence numbers. If a RouteRequest is received multiple times, then duplicate copies are discarded.
- Every intermediate node enters the previous node address and its BcastID. A timer is used to delete this entry in case a RouteReply packet is not received. AODV does not repair a broken path locally, When a link breaks, the end nodes are notified. Source node re-establishes the route to the destination if required.

In this figure, source node 1 initiates a path-finding process by originating a *RouteRequest* to be flooded in the network for destination node 15, assuming that the *RouteRequest* contains the destination sequence number as 3 and the source sequence number as 1. When nodes 2, 5, and 6 receive the *RouteRequest* packet, they check their routes to the destination. In case a route to the destination is not available, they further forward it to their neighbors. Here nodes 3, 4, and 10 are the neighbors of nodes 2, 5, and 6. This is with the assumption that intermediate nodes 3 and 10 already have routes to the destination node, that is, node 15 through paths 10-14-15 and 3-7-9-13-15, respectively. If the destination sequence number at intermediate node 10 is 4 and is 1 at intermediate node 3, then only node 10 is allowed to reply along the cached route to the source. This is because node 3 has an older route to node 15 compared to the route available at the source node (the destination sequence number at node 3 is 1, but the destination sequence number is 3 at the source node), while node 10 has a more recent route (the destination sequence number is 4) to the destination. If the *RouteRequest* reaches the destination (node 15) through path 4-12-15 or any other alternative route, the destination also sends a *RouteReply* to the source. In this case, multiple *RouteReply* packets reach the source. All the intermediate nodes receiving a *RouteReply* update their route tables with the latest destination sequence number. They also update the routing information if it leads to a shorter path between source and destination.

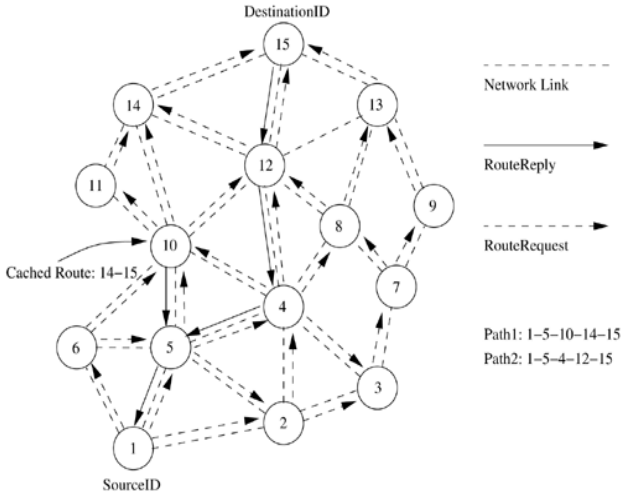


Figure 4.12. Route establishment in AODV.

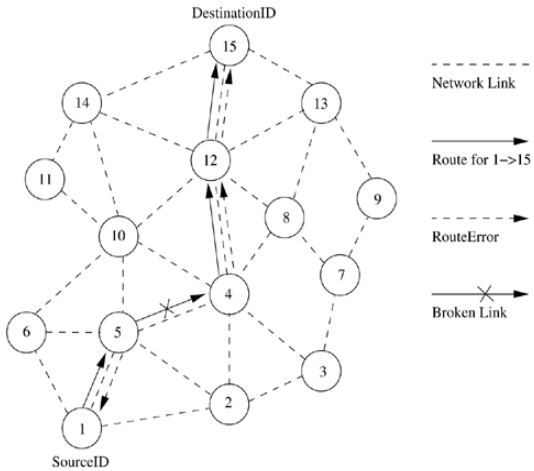


Figure 4.13. Route maintenance in AODV.

Advantages

- Routes are established on demand and DestSeqNum are used to find latest route to the destination.
- Connection setup delay is less.

Disadvantages

- Intermediate nodes can lead to inconsistent routes if the source sequence number is very old.
- Multiple RouteReply packets to single RouteRequest packet can lead to heavy control overhead.
- Periodic beaconing leads to unnecessary bandwidth consumption

4.3.2.3 Temporally Ordered Routing Algorithm (TORA)

Source-initiated on-demand routing protocol

- Uses a link reversal algorithm.
- Provides loop free multi path routes to the destination.
- Each node maintains its one-loop local topology information.
- Has capability to detect partitions.
- Unique property.
- Limiting the control packets to a small region during the reconfiguration process initiated by a path break.

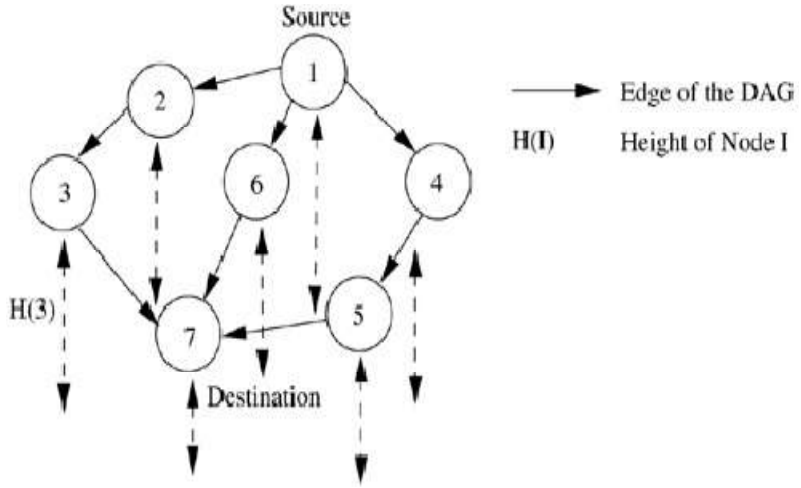


Figure 4.14. Illustration of temporal ordering in TORA**.**

TORA has 3 main functions: establishing, maintaining and erasing routes

- The route establishment function is performed only when a node requires a path to a destination but does not have any directed link.
- This process establishes a destination-oriented directed acyclic graph using a query/update mechanism.

- Once the path to the destination is obtained, it is considered to exist as long as the path is available, irrespective of the path length changes due to the re-configurations that may take place during the course of data transfer session.
- If the node detects a partition, it originated a clear message, which erases the existing path information in that partition related to the destination.

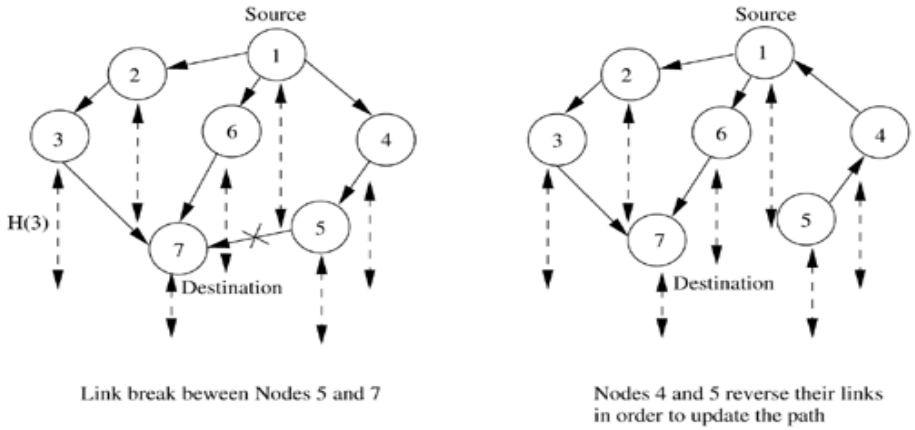


Figure 7.15. Illustration of route maintenance in TORA.

Advantages

- Incur less control overhead.
- Concurrent detection of partitions.
- Subsequent deletion of routes.

Disadvantages

- Temporary oscillations and transient loops.
- Local reconfiguration of paths result in non-optimal routes.

4.3.2.4 Associativity-Based Routing (ABR)

It is a distributed routing protocol that selects routes based on the stability of the wireless links.

- It is a beacon-based on-demand routing protocol.
- A link is classified as stable or unstable based on its temporal stability.
- The temporal stability is determined by counting the periodic beacons that a node receives from its neighbors.
- Each node maintains the count of its neighbor's beacons and classifies each link as stable or unstable.
- The link corresponding to a stable neighbor is termed as a stable link, while a link to an unstable neighbor is called an unstable link.
- A source node floods RouteRequest packets throughout the network if a route is not available in its route cache.
- All intermediate nodes forward the RouteRequest packet.

- A RouteRequest packet carries the path it has traversed and the beacon count for the corresponding nodes in the path.
- When the first RouteRequest reaches the destination, the destination waits for a time period T to receive multiple RouteRequests through different paths.
- If two paths have the same proportion of stable links, the shorter of them is selected.
- If more than one path is available, then a random path among them is selected as the path between source and destination, as depicted in Figure 4.18, the *RouteRequest* reaches the destination through three different routes. Route 1 is 1-5-10-14-15, route 2 is 1-5-4-12-15, and route 3 is 1-2-4-8-13-15. ABR selects route 3 as it contains the highest percentage of stable links compared to route 1 and route 2. ABR gives more priority to stable routes than to shorter routes. Hence, route 3 is selected even though the length of the selected route is more than that of the other two routes.

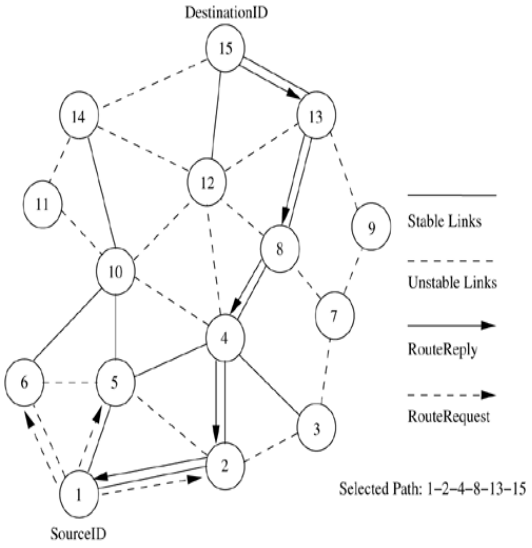


Figure 4.18. Route establishment in ABR.

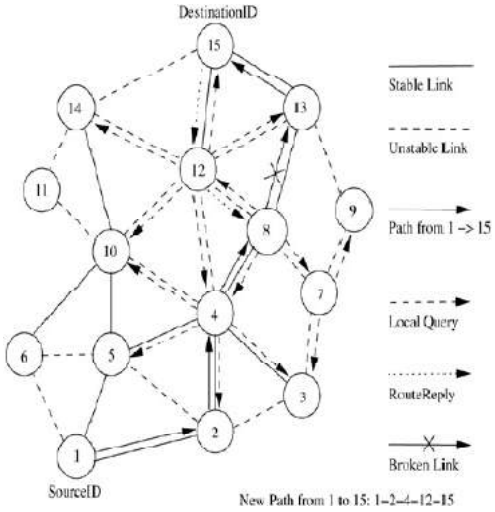


Figure 4.19. Route maintenance in ABR.

If a node fails to repair the broken link, then its uplink node (the previous node in the path which is closer to the source node) reinitiates the LQ broadcast. This route repair process continues along the intermediate nodes toward the source node until it traverses half the length of the broken path or the route is repaired. In the former case, the source node is informed, which initiates a new route establishment phase.

Advantages

- Stable routes have a higher preference compared to shorter routes.
- They result in fewer path breaks which, in turn, reduces the extent of flooding due to reconfiguration of paths in the network.

Disadvantages

- Chosen path may be longer than the shortest path between the source and destination because of the preference given to stable paths.
- Repetitive LQ broadcasts may result in high delays during route repairs.

4.4 Hybrid Routing Protocols

Here, each node maintains the network topology information up to *m* hops. The different existing hybrid protocols are presented below.

4.4.1 Core Extraction Distributed Ad Hoc Routing Protocol (CEDAR)

- CEDAR integrates routing and support for QoS.
- It is based on extracting core nodes (also called as Dominator nodes) in the network.
- Core nodes together approximate the minimum Dominating Set (DS).
- A DS of a graph is defined as a set of nodes such that every node in the graph is either present in the DS or is a neighbor of some node present in the DS.
- There exists at least one core node within every three hops.
- The nodes that choose a core node as their dominating node are called core member nodes of the core node concerned.
- The path between two core nodes is termed as virtual link. CEDAR employs a distributed Algorithm to select core nodes.
- The selection of core nodes represents the core extraction phase.
- CEDAR uses the core broadcast mechanism to transmit any packet throughout the network in the unicast mode, involving as minimum number of nodes as possible.
- Route Establishment in CEDAR: It is carried out in two phase.
- The first phase finds a core path from source to destination. The core path is defined as the path from dominator of the source node (source core) to the dominator of the destination node (destination core).
- In the second phase, a QoS feasible path is found over the core path.
- A node initiates a RouteRequest if the destination is not in the local topology table of its core node; otherwise the path is immediately established.
- For establishing a route, the source core initiates a core broadcast in which the RouteRequest is sent to all neighboring core nodes which inturn forwards it.
- A core node which has the destination node as its core member replies to the source core.
- Once the core path is established, a path with the requested QoS support is then chosen.
- A node after which the break occurred:
    - Sends a notification of failure.
    - Begins to find a new path from it to the destination.
    - Rejects every received packet till the moment it finds the new path to the destination.
- Meanwhile, as the source receives the notification message:
    - It stops to transmit.

- Tries to find a new route to the destination.
- If the new route is found by either of these two nodes, a new path from the source to the destination is established.

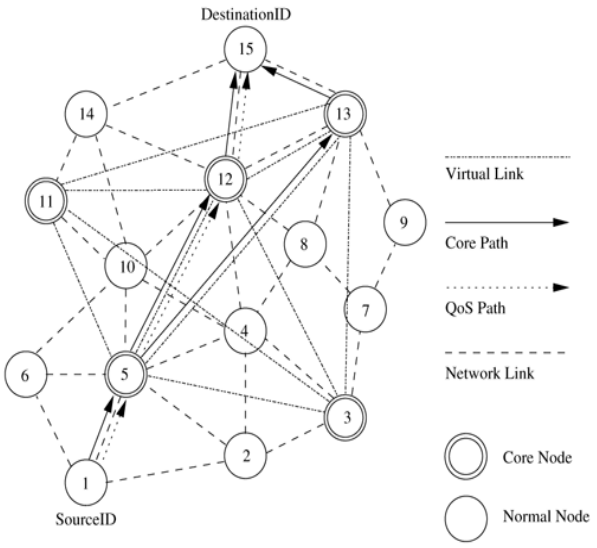


Figure 4.20 Route establishment in CEDAR.

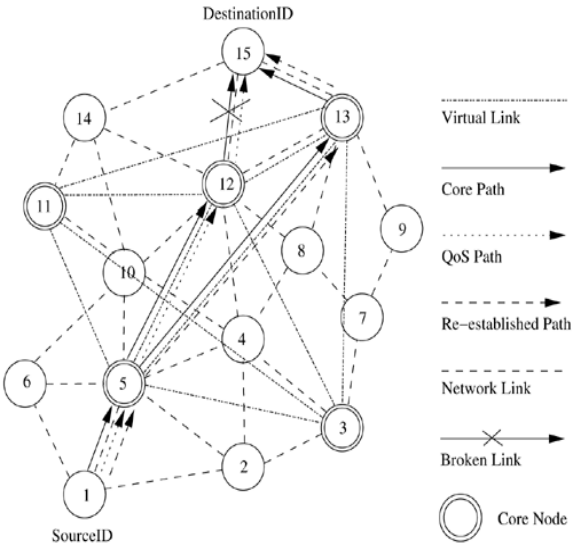


Figure 4.21 Route maintenance in CEDAR.

Advantages

- Performs both routing and QoS path computation very efficiently with the help of core nodes.
- Utilization of core nodes reduces traffic overhead.
- Core broadcasts provide a reliable mechanism for establishing paths with QoS support.

Disadvantages

- Since route establishment is carried out at core nodes, the movement of core nodes adversely affects the performance of the protocol.
- Core node update information causes control overhead.

4.4.2 Zone Routing Protocol (ZRP)

Effectively combines the best features of both Proactive and Reactive routing protocols.

- It use a Proactive routing scheme within a limited zone in the r-hop neighborhood of every node.
- Use a Reactive routing scheme for nodes beyond this. An Intra-Zone Routing Protocol (IARP) is used in the zone where a particular node employs proactive routing.
- The Reactive routing protocol used beyond this zone is referred to as Inter-Zone Routing Protocol (IERP).
- The routing zone of a given node is a subset of the network, within which all nodes are reachable within less than or equal to.

Route Establishment: When a node s (node 8 in the fig 4.22) has packets to be sent to a destination node d (node 15 in fig), it checks whether node d is within its zone.

- If the destination belongs to its own zone, then it delivers the packets directly.
- Otherwise, node s broadcasts the RouteRequest to its peripheral nodes (in fig, node 8 broadcasts RouteRequest to node 2, 3, 5, 7, 9, 10, 13, 14 and 15).
- If any peripheral node finds node d to be located within its routing zone, it sends a RouteReply back to node 8 indicating the path; otherwise, the node rebordercasts the RouteRequest packet to the peripheral nodes.
- This process continues until node d is located. During RouteRequest propagation, every node that forwards the RouteRequest appends its address to it.
- This information is used for delivering the RouteReply packet back to the source.
- The criteria for selecting the best path may be the shortest path, least delay path etc.
- When an intermediate node in an active path detects a broken link in the path, it performs a local path reconfiguration in which the broken link is bypassed by means of a short alternate path connecting the ends of the broken link
- A path update message is then sent to the sender node, this results in sub-optimal path between two end points.

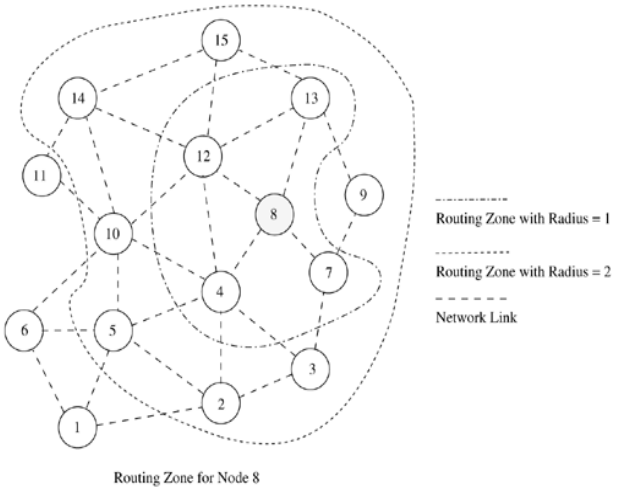


Figure 4.22. Routing zone for node 8 in ZRP.

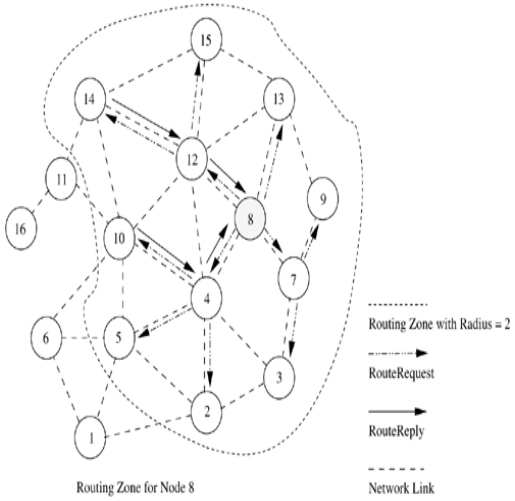


Figure 4.23. Path finding between node 8 and node 16.

Advantage

- Reduce the control overhead by combining the best features of Proactive and Reactive protocols.

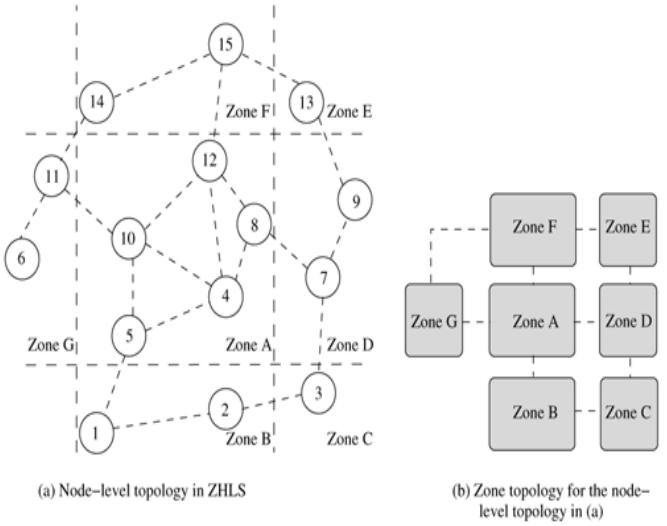
Disadvantage

- Control overhead may increase due to the large overlapping of nodes routing zones.

4.4.3 Zone Based Hierarchical Link State Routing Protocol (ZHLS)

- ZHLS uses the geographical location info of the nodes to form non-overlapping zones. A Hierarchical Addressing that consists of a zone ID and a node ID is employed.

- Similar to ZRP, ZHLS also employs a Proactive approach inside the geographical zone and a Reactive approach behind the zone.
- Every node requires GPS support for obtaining its own geographical location that is used to map itself into corresponding zone.
- The assignment of zone addresses to geographical areas is important and is done during a phase called the network design phase or network deployment phase.
  i. Each node maintains two link state packets: (LSP)
  ii. Node level LSP: list of connected neighbors.
  iii. Zone LSP: list of connected zones.
- Route Establishment, If a source node src wants to communicate with a destination node dest, src checks whether dest resides in its own zone.
- If dest belongs to same zone, then packets are delivered to the dest as per the Intra-Zone routing table.
- If dest does not belong to the same zone, then the src originates a location request packet containing the sender's and destination's information. This location info is forwarded to every other zone.
- The gateway node of a zone at which the location request packet is received verifies its routing table for the destination node.
- The gateway node that finds the destination node required by a location request packet originates a location response packet containing the zone information to the sender.



(a) Node–level topology in ZHLS

(b) Zone topology for the node–level topology in (a)

Table 4.1. Zone link state packets

Source Zone	Zone Link State Packet
A	B, D, F, and G
B	C and A
C	B and D
D	A, C, and E
E	A, D, and F
F	A, E, and G
G	A and F

Figure 4.24. Zone-based hierarchical link state routing protocol.

Route Maintenance

- If a given gateway node away causing a zone level connection failure, routing can still take place with the help of the other gateway nodes.
- This is due to the hierarchical addressing that makes use of zone ID and node ID.

4.5 Routing Protocols With Efficient Flooding Mechanisms

- Many protocols flood the network with RouteRequest packets in order to obtain a path to the destination.
- Flooding of control packets results in:
  - Wastage of bandwidth.
  - Increase in number of collisions.
- Protocols with efficient flooding mechanisms:
  - Preferred link-based routing (PLBR) protocol.
  - Optimized link state routing (OLSR) protocol.

Preferred Link Based Routing (PLBR) Protocols

- Use the preferred link approach in an implicit manner by processing a RouteRequest packet only if it is received through a strong link.
- Here a node selects a subset of nodes from its Neighbors List (NL). This subset is referred to as the Preferred List (PL) selection of this subset may be based on link or node characteristics.
- All neighbors receive RouteRequest packets because of the broadcast radio channel, but only neighbors present in the PL forward them further.
- Each node maintains information about its neighbors and their neighbors in a table called Neighbor's Neighbor Table (NNT). It periodically transmits a beacon containing the changed neighbor's information.

Route Establishment

- If dest is in src's NNT, the route is established directly. Otherwise, src transmits a RouteRequest packet containing
  - Source node's address (SrcID)
  - Destination node's address (DestID)
  - Unique sequence number (SeqNum)
  - Traversed path (TP)
  - PL
  - TTL flag
  - NoDelay flag
- A node is eligible for forwarding a RouteRequest only if it satisfies the following criteria:
- The node ID must be present in the received RouteRequest packet's PL.
- RouteRequest packet must not have been already forwarded by the node, and the TTL on the packet must be greater than zero.

- If the dest is in the eligible node's NNT, the RouteRequest is forwarded as a unicast packet to the neighbor.
- If the computed PLT is empty, the RouteRequest packet is discarded and marked as sent.
- If the RouteRequest reaches the destination, the route is selected by the route selection procedure given below.

Route selection

- When multiple Route Request packets reach dest,the route selection procedure selects the best route among them.
- The criterion for selecting the best route can be the shortest path, or the least delay path, or the most stable path.
- Dest starts a timer after receiving the first route request packet. The timer expires after a certain RouteSelectWait period, after which no more RouteRequest packets would be accepted.
- From the received Route Request packets, a route is selected as follows:
- For every RouteRequest i that reached Dest during the RouteSelectWait period, Max(Wmin) is selected, where is the min. Weight of the link in the path followed by i if two or more paths have the same value for the shortest path is selected.
- After selecting a route, all subsequent RouteRequest packets from the same src with a seqnum less than or equal to the seqnum of the selected RouteRequest are discarded.
- If the node delay flag is set, the route selection procedure is omitted and TP of the first RouteRequest reaching the Dest is selected as the route.

Algorithms for preferred links computation:

Neighbor-Degree-Based preferred link algorithm (NDPL) Weight Based preferred link algorithm (WBPL)

NDPL (Neighbor-Degree-Based preferred link algorithm)

Let d → node that calculates the preferred list table PLT. TP Traversed path. OLDPL → preferred list of the received RouteRequest packet. NNTd → NNT of the node d. N(i) → neighbors of node I and itself. INL→ include list, a set containing all reachable neighbors by transmitting the RouteRequest packet. EXL → Exclude list, a set containing all neighbors that are unreachable by transmitting the RouteRequest packet after execution of the algorithm.

Step 1: Node d marks the nodes that are not eligible for further forwarding the RouteRequest packet.

a) If a node i of TP is a neighbor of node d mark all neighbors of i as reachable i.e add N(i) to INL.

b) If a node i of OLDPL is a neighbor of node d and i < d, then include N(i) in INL.

c) If neighbor i of node d has a neighbor n present in TP, add N(i) to INL.

d) If neighbor i of node d has a neighbor n present in OLDPL and n < d, add N(i) to INL.

Step 2: If neighbor i of node d is not in INL, put i in PLT and mark all neighbor of i as reachable. If i is present in INL, mark the neighbors of i as unreachable by adding them to EXL.
Step 3: If neighbor i of d has a neighbor n present in EXL, put i in PLT and mark all neighbors of i as reachable. Delete all neighbors of i from EXL.

Step 4: Reduction steps are applied here in order to remove overlapping neighbors from PLT without compromising on reachability.

a) Remove each neighbor i from PLT if N (i) is covered by remaining neighbors of PLT. Here the minimum degree neighbor is selected every time.

b) Remove neighbor i from PLT whose N (i) is covered by node d itself.

Weight-Based Preferred Link Algorithm (WBPL)

In this algorithm, a node finds the preferred links based on stability, which is indicated by a weight, which in turn is based on its neighbors' temporal and spatial stability.

1. Let $BCnt_i$ be the count of *beacons* received from a neighbor i and TH_{bcon} is the number of beacons generated during a time period equal to that required to cover twice the transmission range $(TH_{bcon} = \frac{2 \times transmission\ range}{maximum\ velocity \times period\ of\ beacon})$. Weight given to i based on time stability (WT^i_{time}) is

$$WT_{time} = \begin{cases} 1 & \text{if } BCnt_i > TH_{bcon} \\ BCnt_i / TH_{bcon} & \text{otherwise.} \end{cases}$$

2. Estimate the distance (D^i_{Est}) to i from the received power of the last few packets using appropriate propagation models. The weight based on spatial stability is $WT^i_{spatial} = \frac{R - D_{Est}}{R}$.
3. The weight assigned to the link i is the combined weight given to time stability and spatial stability. $W_i = WT^i_{time} + WT^i_{spatial}$.
4. Arrange the neighbors in a non-increasing order of their weights. The nodes are put into the *PLT* in this order.
5. If a link is overloaded, delete the associated neighbor from *PLT*. Execute *Step* 1 of NDPL and delete $\forall i,\ i \in PLT \cap i \in INL$. Also, delete those neighbors from *PLT* that satisfy *Step* 4 of NDPL.

Advantages

- Minimizes broadcast storm problem. Hence, highly scalable.
- Reduction in control overhead results in decrease in the number of collisions and improvement in efficiency of the protocol.

Disadvantage

- Computationally more complex.

4.5.1 Optimized Link State Routing (OLSR)

- It is a proactive routing protocol that employs an efficient link state packet forwarding mechanism called multipoint relaying (MPR).
- This protocol optimizes the pure link state routing protocol.
- Optimizations are done in two ways:
  - By reducing the size of control packets.
  - By reducing the no. of links that are used for forwarding the link state packets.
- The subset of links or neighbors that are designated for link state updates and are assigned the responsibility of packet forwarding are called multipoint relays.
- The set consisting of nodes that are multipoint relays is referred to as MPRset.
- Each node(say, P) in the n/w selects an MPRset that processes and forwards every link state packet that node P originates.
- The neighbor nodes that do not belong to the MPRset process the link state packets originated by node P but do not forward them.
- Similarly, each node maintains a subset of neighbors called MPR selectors, which is nothing but the set of neighbors that have selected the node as a multipoint relay.
- In order to decide on the membership of the nodes in the MPRset, a node periodically sends *Hello* messages that contain:
  - List of neighbors with which the node has bidirectional links
  - List of neighbors whose transmission was received in the recent past but with whom bidirectional links have not yet been confirmed.
- The nodes that receive this Hello packet update their own two-hop topology tables.
- The selection of multipoint relays is also indicated in the Hello packet.
- The Data structure called neighbor table is used to store the list of neighbors, the two-hop neighbors, and the status of neighbor nodes.
- The neighbor nodes can be in one of the three possible link status states, i.e
  - Unidirectional
  - Bidirectional

Multipoint relay

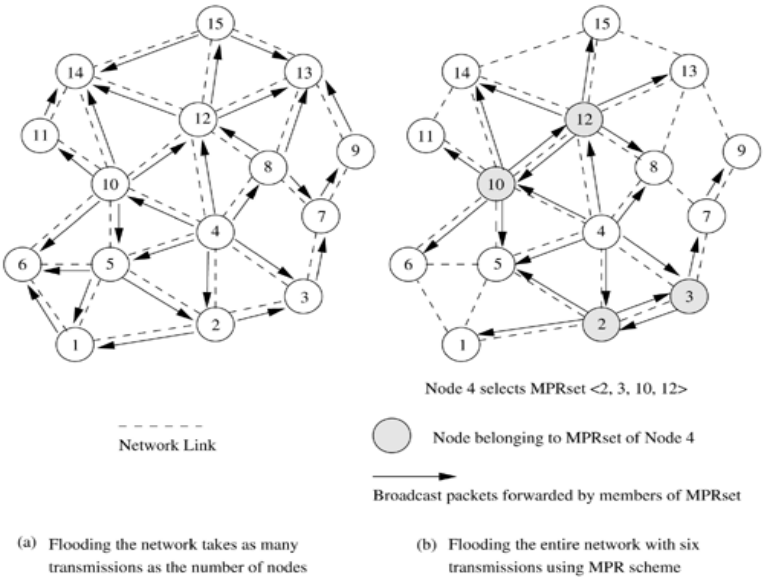


Figure 4.25. An example selection of MPRset in OLSR.

Selection of multipoint relay nodes (refer fig b) Ni(x) → ith hop neighbor set of node x MPR(x) → MPRset of node x.

Step1: MPR(x) ← Ø /* initializing empty MPRset */

Step2: MPR(x) ← {those nodes that belong to N1(x) and which are the only neighbors of nodes in N2(x)}

Step3: while there exists some node in N2(x) which is not covered by MPR(x)

a) For each node in N1(x),which is not in MPR(x),compute the maximum number of nodes that it covers among the uncovered nodes in the set N2(x).

b) Add to MPR(x) the node belonging to N1(x) for which this number is maximum.

Advantages:

- Reduces the routing overhead.
- Reduces the no. of broadcasts done.
- Hence low connection setup time and reduced control overhead.

4.6 Hierarchical Routing Protocols

The use of routing hierarchy has several advantages □□Reduction in size of routing tables and better scalability.

4.6.1 Hierarchical State Routing (HSR) protocol

- It is a distributed multi-level hierarchical routing protocol that employs clustering at different levels with efficient membership management at every level of clustering.
- Each cluster has its leader.

- Clustering is organized in levels:
  - **Physical**: between nodes that have physical wireless one-hop links between them.
  - **Logical**: based on certain relations.
  - Figure 4.26 illustrates the multilayer clustering defined by the HSR protocol. At the lowest level *(L* = 0), there are six cluster leaders (nodes 1, 2, 3, 4, 5, and 6). Nodes are classified as cluster leaders, or gateway nodes, or normal member nodes.
  - A cluster leader is entrusted with responsibilities such as slot/frequency/code allocation, call admission control, scheduling of packet transmissions, exchange of routing information, and handling route breaks. In Figure 4.27, node 5 is a clusterhead marked as *L*0-5, which refers to the level of clustering (*L* = 0) and node ID (5).
  - Similarly, each of the higher-level cluster leaders is also marked (*e.g., L*1 − 6, *L* − 2 − 6, and *L*3 − 6 refer to the same node 6, but acting as leader with the given leader IDs at levels 1, 2, and 3, respectively).
  - The spectrum reuse schemes, including spreading code assignment, can be used among the cluster leaders of the *L* = 0 clusters. For the nodes under the leadership of node 6 at level 0, the cluster members are nodes 9, 10, 11, 12, and 17.
  - Those nodes that belong to multiple clusters are referred to as cluster gateway nodes. For the level 0 cluster whose leader is node 6, the cluster gateways are nodes 10, 12, and 17.
  - The second level of clustering is done among the leaders of the first level, that is, the leaders of 0*th* level clusters, *L*0 - 1, *L*0 - 2, *L0* - 3,*L0* - 4, *L0* - 5, and *L0* - 6, form the members of the first-level cluster.

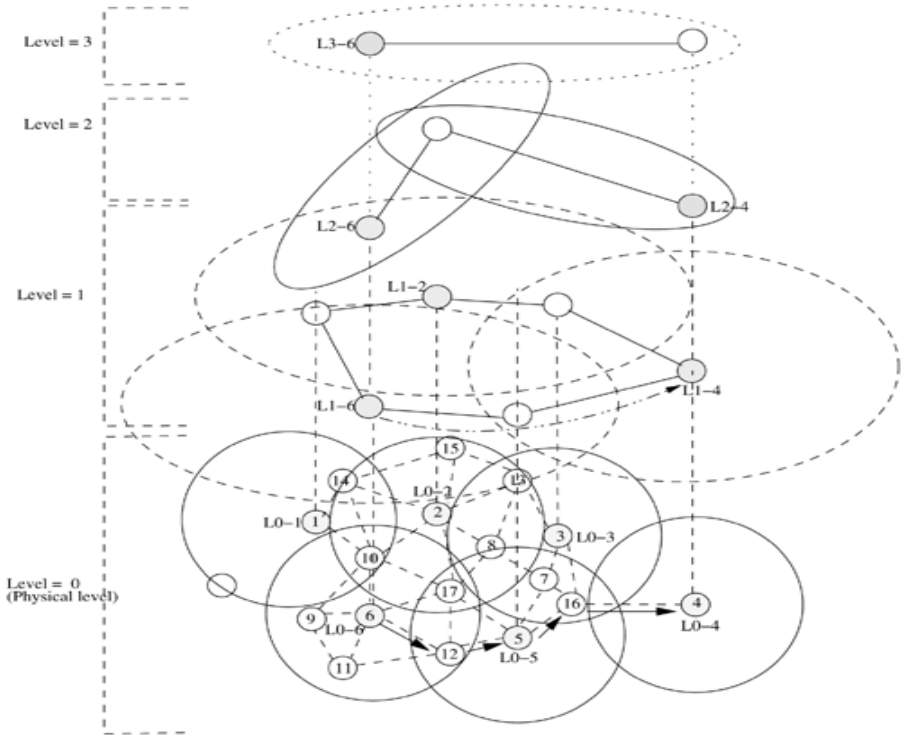


Figure 4.26. Example of HSR multi-level clustering.

Advantages

- Reduces routing table size storage required is O (n x m).
- For flat topology, it is O (nm)
  - n→ no. of nodes
  - m→ no. of levels

Disadvantage

- Process of exchanging information concerned all the levels of the hierarchy as well as the process of leader election in every cluster makes it quite problematic for adhoc networks.

4.6.2 Fish-Eye State Routing Protocol (FSR)

- It is a generalization of the GSR protocol.
- It uses Fisheye technique to reduce the routing overhead.
- Principle: Property of a fish's eye that can capture pixel information with greater accuracy near its eye's focal point.
- This accuracy decreases with an increase in the distance from the center of the focal point
- This property is translated to routing in adhoc wireless networks by a node
- Each node maintains accurate information about near nodes.
- Nodes exchange topology information only with their neighbors.
- A sequence numbering scheme is used to identify the recent topology changes
- This constitutes a link-level information exchange of distance vector protocols and complete topology information exchange of link state protocols.
- FSR defines routing scope, which is the set of nodes that are reachable in a specific no. of hops.
- The scope of a node at two hops is the set of nodes that can be reached in two hops fig 4.27 shows scope of node 5 with one hop and two hops.
- The routing overhead is significantly reduced

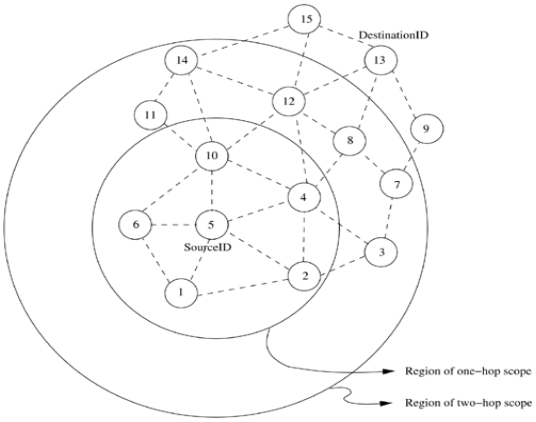


Figure 4.27. Fisheye state routing.

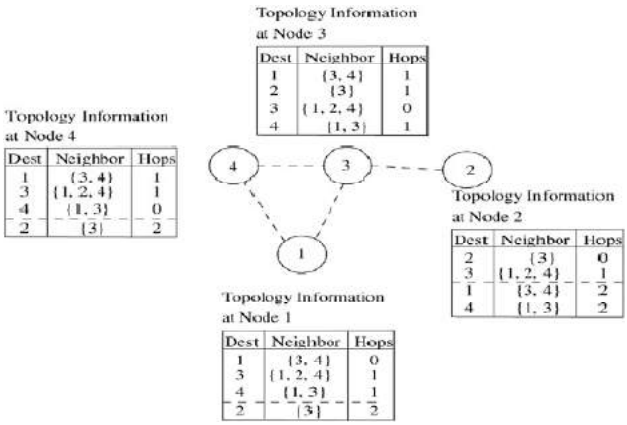


Figure 4.28. An illustration of routing tables in FSR.

- The link state info for the nodes belonging to the smallest scope is exchanged at the highest frequency. Frequency of exchanges decreases with an increase in scope.
- Fig 4.28 illustrates an example depicting the n/w topology information maintained at nodes in a n/w.
- Message size for a typical topology information update packet is significantly reduced.
- The routing information for the nodes that are one hop away from a node are exchanged more frequently than the routing information about nodes that are more than one hop away.
- Information regarding nodes that are more than one hop away from the current node are listed below the dotted line in the topology table.

Advantages

- Reduce bandwidth consumption by link state update packets.
- Suitable for large and highly mobile adhoc wireless network.

Disadvantage

- Very poor performance in small adhoc networks

4.7 Power-Aware Routing Protocols

Some of the Power-aware routing protocols are discussed below:

- ***Power-Aware Routing Metrics***

  The limitation on the availability of power for operation is a significant bottleneck. Hence, the use of routing metrics contributes to the efficient utilization of energy and increases the lifetime of the network

- ***Minimal energy consumption per packet***
  - This metric aims at minimizing the power consumed by a packet in traversing from source node to the destination node.
  - The energy consumed by a packet when traversing through a path is the sum of the energies required at every intermediate hop in that path.
  - This metric doesn't balance the load;
    - **Disadvantages**
      - Selection of path with large hop length.
      - Inability to measure the power consumption in advance.
      - Inability to prevent the fast discharging of batteries at some nodes
- ***Maximize network connectivity***
  - This metric attempt to balance the routing load among the cut set (the subset of the nodes in the network, the removal of which results in network partitions).
  - It is difficult to achieve a uniform battery draining rate for the cut set.

- ***Maximum variance in Node power levels***
  - This metric proposes to distribute the load among all nodes in the network so that the power consumption pattern remains uniform across them.
  - This problem is very complex when the rate and size of the data packets vary
- ***Minimum cost per packet***
  - In order to maximize the life of every node in the network, this routing metric is made as a function of the state of the node's battery.
  - A node's cost decreases with an increase in its battery change and vice versa.
  - Cost of node can be easily computed
    - **Advantage**
      - congestion handling & cost calculation
- ***Minimize maximum node cost***

- This metric minimizes the maximum cost per node for a packet after routing a number of packets or after a specific period.
- This delays the failure of a node, occurring due to higher discharge because of packet forwarding.

UNIT – V

QoS and Energy Management

5.1Introduction

The goal of QoS provisioning is to achieve a more deterministic network behavior, so that information carried by the network can be better delivered and network resources can be better utilized.

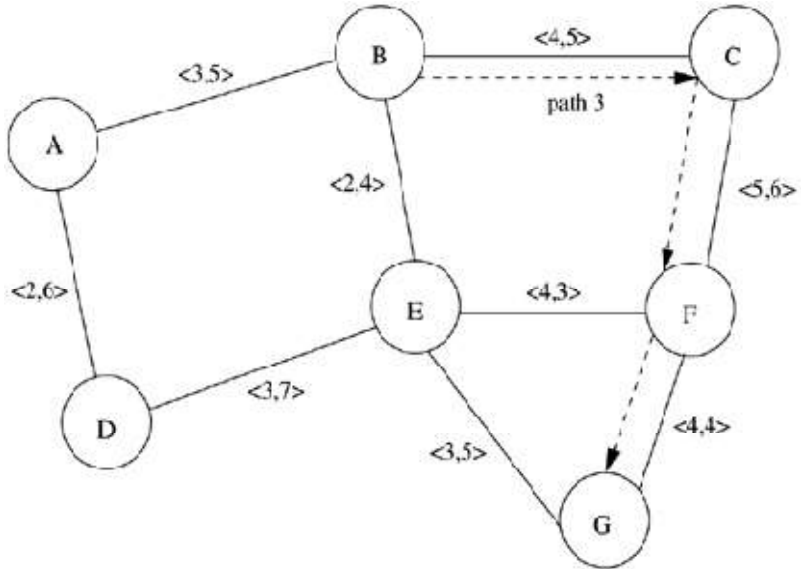


Figure 5.1. An example of QoS routing in ad hoc wireless network.

- For example, consider the network shown in Figure 5.1. The attributes of each link are shown in a tuple < *BW, D* >, where *BW* and *D*represent available bandwidth in Mbps and delay1 in milliseconds.
- Suppose a packet-flow from node*B* to node *G* requires a bandwidth guarantee of 4 Mbps. Throughout the chapter, the terms "node" and "station" are used interchangeably. QoS routing searches for a path that has sufficient bandwidth to meet the bandwidth requirement of the flow.
- Here, six paths are available between nodes *B* and *G* as shown in Table 10.1. QoS routing selects path 3 (*i.e., B* → *C* → *F* → G) because, out of the available paths, path 3 alone meets the bandwidth constraint of 4 Mbps for the flow.
- The end-to-end bandwidth of a path is equal to the bandwidth of the bottleneck link (*i.e.,* the link having minimum bandwidth among all the links of a path).
- The end-to-end delay of a path is equal to the sum of delays of all the links of a path. Clearly, path 3 is not optimal in terms of hop count and/or end-to-end delay parameters, while path 1 is optimal in terms of both hop count and end-to-end delay parameters.
- Hence, QoS routing has to select a suitable path that meets the QoS constraints specified in the service request made by the user.

Table 5.1. Available paths from node *B* to node *G*

No.	Path	Hop Count	End-to-end Bandwidth (Mbps)	End-to-end Delay (milliseconds)
1	$B \rightarrow E \rightarrow G$	2	2	9
2	$B \rightarrow E \rightarrow F \rightarrow G$	3	2	11
3	$B \rightarrow C \rightarrow F \rightarrow G$	3	4	15
4	$B \rightarrow C \rightarrow F \rightarrow E \rightarrow G$	4	3	19
5	$B \rightarrow A \rightarrow D \rightarrow E \rightarrow G$	4	2	23
6	$B \rightarrow A \rightarrow D \rightarrow E \rightarrow F \rightarrow G$	5	2	25

QoS provisioning often requires negotiation between host and network, call admission control, resource reservation, and priority scheduling of packets. QoS can be rendered in ad hoc wireless networks through several ways, namely, per flow, per link, or per node. In ad hoc wireless networks, the boundary between the service provider (network) and the user (host) is not defined clearly, thus making it essential to have better coordination among the hosts to achieve QoS.

5.2 Issues and Challenges in Providing QoS in Ad Hoc Wireless Networks

Some of the characteristics are dynamically varying network topology, lack of precise state information, lack of a central controller, error-prone shared radio channel, limited resource availability, hidden terminal problem, and insecure medium.

- **Dynamically varying network topology:** Since the nodes in an ad hoc wireless network do not have any restriction on mobility, the network topology changes dynamically. Hence, the admitted QoS sessions may suffer due to frequent path breaks, thereby requiring such sessions to be reestablished over new paths.
- **Imprecise state information:** In most cases, the nodes in an ad hoc wireless network maintain both the link-specific state information and flow-specific state information. The link-specific state information includes bandwidth, delay, delay jitter, loss rate, error rate, stability, cost, and distance values for each link. The flow-specific information includes session ID, source address, destination address, and QoS requirements of the flow (such as maximum bandwidth requirement, minimum bandwidth requirement, maximum delay, and maximum delay jitter).
- **Lack of central coordination:** Unlike wireless LANs and cellular networks, ad hoc wireless networks do not have central controllers to coordinate the activity of nodes. This further complicates QoS provisioning in ad hoc wireless networks.
- **Error-prone shared radio channel:** The radio channel is a broadcast medium by nature. During propagation through the wireless medium, the radio waves suffer from several impairments such as attenuation, multipath propagation, and interference.
- **Hidden terminal problem:** The hidden terminal problem is inherent in ad hoc wireless networks. This problem occurs when packets originating from two or more sender nodes, which are not within the direct transmission range of each other, collide at a common receiver node. It necessitates the retransmission of the packets, which may not be acceptable for flows that have stringent QoS requirements.

- **Insecure medium:** Due to the broadcast nature of the wireless medium, communication through a wireless channel is highly insecure. Therefore, security is an important issue in ad hoc wireless networks, especially for military and tactical applications.

5.3 Classifications of Qos Solutions

The QoS solutions can be classified in two ways. One classification is based on the QoS approach employed, while the other one classifies QoS solutions based on the layer at which they operate in the network protocol stack.

5.3.1 Classifications of QoS Approaches

As shown in Figure 5.2, several criteria are used for classifying QoS approaches. The QoS approaches can be classified based on the interaction between the routing protocol and the QoS provisioning mechanism, based on the interaction between the network and the MAC layers, or based on the routing information update mechanism. Based on the interaction between the routing protocol and the QoS provisioning mechanism, QoS approaches can be classified into two categories: *coupled* and *decoupled* QoS approaches. In the case of the coupled QoS approach, the routing protocol and the QoS provisioning mechanism closely interact with each other for delivering QoS guarantees. If the routing protocol changes, it may fail to ensure QoS guarantees. But in the case of the decoupled approach, the QoS provisioning mechanism does not depend on any specific routing protocol to ensure QoS guarantees.

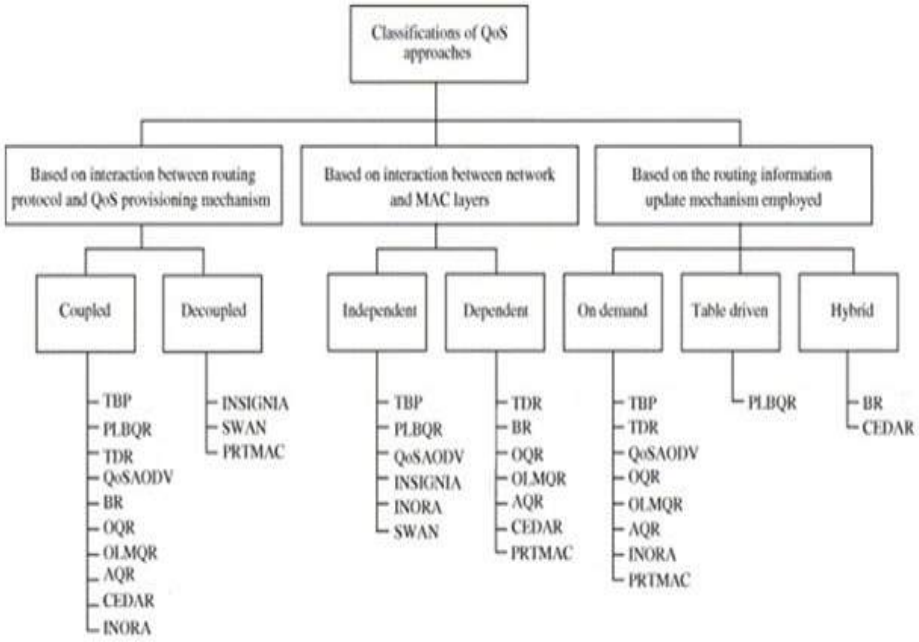


Figure 5.2. Classifications of QoS approaches.

5.3.2 Layer-Wise Classification of Existing QoS Solutions

The existing QoS solutions can also be classified based on which layer in the network protocol stack they operate in. Figure 5.3 gives a layer-wise classification of QoS solutions. The figure also shows some of the cross-layer QoS solutions proposed for ad hoc wireless networks. The following sections describe the various QoS solutions listed in Figure 5.3.

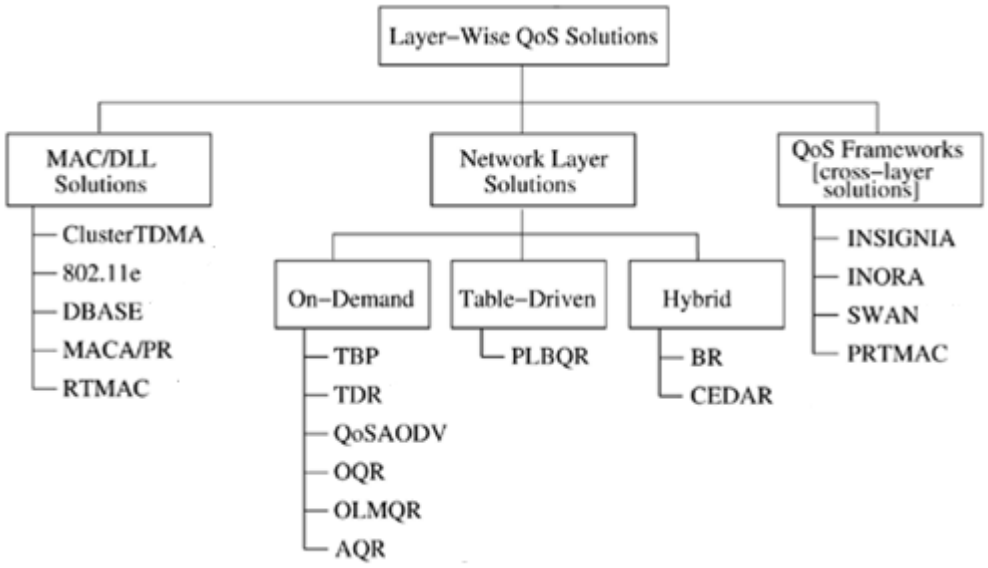


Figure 10.3. Layer-wise classification of QoS solutions.

5.4 MAC Layer Solutions

The MAC protocol determines which node should transmit next on the broadcast channel when several nodes are competing for transmission on that channel. The existing MAC protocols for ad hoc wireless networks use channel sensing and random back-off schemes, making them suitable for best-effort data traffic.

QoS support for applications in ad hoc wireless networks has been proposed. Some of these protocols are described below.

5.4.1 Cluster TDMA

- In this clustering approach, nodes are split into different groups. Each group has a cluster-head (elected by members of that group), which acts as a regional broadcast node and as a local coordinator to enhance the channel throughput.
- Every node within a cluster is one hop away from the cluster-head. The formation of clusters and selection of cluster-heads are done in a distributed manner.
- Clustering algorithms split the nodes into clusters so that they are interconnected and cover all the nodes.
- Three such algorithms used are lowest-ID algorithm, highest degree (degree refers to the number of neighbors which are within transmission range of a node) algorithm, and least cluster change (LCC) algorithm.
- In the lowest-ID algorithm, a node becomes a cluster-head if it has the lowest ID among all its neighbors.
- In the highest-degree algorithm, a node with a degree greater than the degrees of all its neighbors becomes the cluster-head.

- In the LCC algorithm, cluster-head change occurs only if a change in network causes two cluster-heads to come into one cluster or one of the nodes moves out of the range of all the cluster-heads.

5.4.2 IEEE 802.11e

The IEEE 802.11 MAC protocol is first described. Then, the recently proposed mechanisms for QoS support, namely, enhanced distributed coordination function (EDCF) and hybrid coordination function (HCF), defined in the IEEE 802.11e.

The time interval between the transmissions of two consecutive frames is called the inter-frame space (IFS). There are four IFSs defined in the IEEE 802.11 standard, namely, short IFS (SIFS), PCF IFS(PIFS), DCF IFS (DIFS), and extended IFS (EIFS). The relationship among them is as follows:

$$SIFS < PIFS < DIFS < EIFS$$

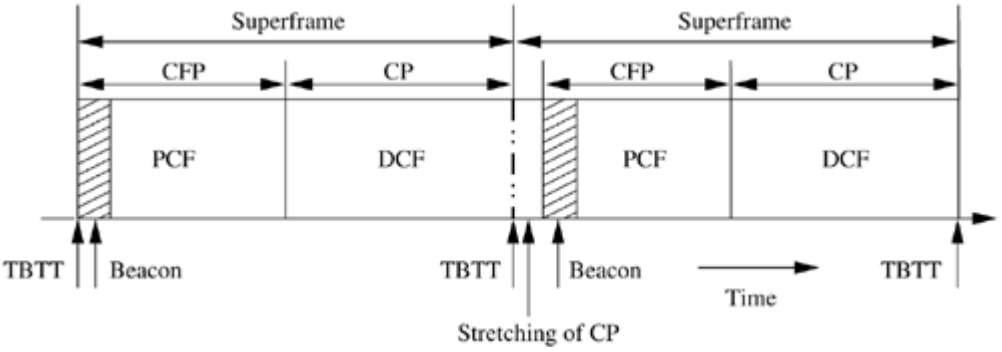


Figure 5.4. PCF and DCF frame sharing.

PCF has certain shortcomings which make it unsuitable for supporting real-time traffic. At TBTT, the PC has to sense the medium idle for at least PIFS before transmitting the beacon frame. If the medium is busy around TBTT, the beacon is delayed, thereby delaying the transmission of real-time traffic that has to be delivered in the following CFP. Further, polled stations' transmission durations are unknown to the PC. PCF is not scalable to support real-time traffic for a large number of users. Due to these reasons, several mechanisms have been proposed to enhance the IEEE 802.11 standard to provide QoS support.

5.4.3 DBASE

The distributed bandwidth allocation/sharing/extension (DBASE) protocol supports multimedia traffic [both variable bit rate (VBR) and constant bit rate (CBR)] over ad hoc WLANs. In an ad hoc WLAN, there is no fixed infrastructure (*i.e.,* AP) to coordinate the activity of individual stations. The stations are part of a single-hop wireless network and contend for the broadcast channel in a distributed manner.

The Access Procedure for Non Real-Time Stations The channel access method for *nrt*-stations is based on conventional DCF. An *nrt*-station with data traffic has to keep sensing the channel for an additional random time called data back-off time (DBT) after detecting the channel as being idle for a DIFS period. The DBT is given by

$$DBT = rand(a,\ b) \times slottime$$

The Access Procedure for Real-Time Stations Each *rt*-station maintains a virtual reservation table (RSVT). In this virtual table, the information regarding all *rt*-stations that have successfully reserved the required bandwidth is recorded. Before initiating an *rt*-session, the *rt*-station sends an RTS in order to reserve the required bandwidth. Before transmitting the RTS, a corresponding entry is made in the RSVT of the node.

5.5 Network Layer Solutions

The bandwidth reservation and real-time traffic support capability of MAC protocols can ensure reservation at the link level only, hence the network layer support for ensuring end-to-end resource negotiation, reservation, and reconfiguration is very essential.

5.5.1 QoS Routing Protocols

- QoS routing protocols search for routes with sufficient resources in order to satisfy the QoS requirements of a flow.
- The information regarding the availability of resources is managed by a resource management module which assists the QoS routing protocol in its search for QoS feasible paths.
- The QoS routing protocol should find paths that consume minimum resources. The QoS metrics can be classified as additive metrics, concave metrics, and multiplicative metrics.

5.5.2 Ticket-Based QoS Routing Protocol

Ticket-based QoS routing is a distributed QoS routing protocol for ad hoc wireless networks. This protocol has the following features:

- It can tolerate imprecise state information during QoS route computation and exhibits good performance even when the degree of imprecision is high.
- It probes multiple paths in parallel for finding a QoS feasible path. This increases the chance of finding such a path. The number of multiple paths searched is limited by the number of tickets issued in the probe packet by the source node. State information maintained at intermediate nodes is used for more accurate route probing. An intelligent hop-by-hop selection mechanism is used for finding feasible paths efficiently.
- The optimality of a path among several feasible paths is explored. A low-cost path that uses minimum resources is preferred when multiple feasible paths are available.
- A primary-backup-based fault-tolerant technique is used to reduce service disruption during path breaks that occur quite frequently in ad hoc wireless networks.

5.5.3 Predictive Location-Based QoS Routing Protocol

- The predictive location-based QoS routing protocol (PLBQR) is based on the prediction of the location of nodes in ad hoc wireless networks.
- The prediction scheme overcomes to some extent the problem arising due to the presence of stale routing information.
- No resources are reserved along the path from the source to the destination, but QoS-aware admission control is performed.
- The network does its best to support the QoS requirements of the connection as specified by the application.

Location and Delay Predictions

In establishing a connection to the destination *D,* the source *S* first has to predict the geographic location of node *D* and the intermediate nodes, at the instant when the first packet reaches the respective nodes. Hence, this step involves location prediction as well as propagation delay prediction.

Delay Prediction

The source node *S* has to predict the time instant *tf* at which a packet reaches the given destination node or intermediate node *D.* This can be known only if the end-to-end delay between nodes *S* and *D* is known. It is assumed that the end-to-end delay for a data packet from node *S* to node *D* is equal to the delay experienced by the latest update message received by node *S* from node *D.*

QoS Routing

Which is refreshed by means of update messages? Using this information, the source node performs source-routing. The network state information is maintained in two tables, namely, the *update table* and the *routing table.* When node *A* receives an update message from node *B*, node *A* updates the corresponding entry for node *B* in the update table. In that entry, node *A* stores the ID of node *B*, the time instant at which the update packet was sent, the time at which the update packet was received, the geographic coordinates, speed, resource parameters of node *B*, and optionally the direction of motion of node *B.*

5.6 QoS Frameworks for Ad Hoc Wireless Networks

A framework for QoS is a complete system that attempts to provide required/promised services to each user or application. All components within this system cooperate in providing the required services. The key component of any QoS framework is the QoS service model which defines the way user requirements are met. The key design issue here is whether to serve users on a per session basis or on a per class basis. Each class represents an aggregation of users based on certain criteria. The other key components of the framework are QoS routing which is used to find all or some of the feasible paths in the network that can satisfy user requirements, QoS signaling for resource reservation, QoS medium access control, call admission control, and packet scheduling schemes. The QoS modules, namely, routing protocol, signaling protocol, and the resource management mechanism, should react promptly to changes in the network state (topology changes) and flow state (change in the end-to-end view of the service delivered).

- *Routing protocol*: Similar to the QoS routing protocols, discussed earlier in this chapter, the routing protocol module in any QoS framework is used to find a path from the source to the destination and to forward the data packet to the next intermediate relay node. QoS routing describes the process of finding suitable path(s) that satisfy the QoS service requirements of an application. If multiple paths are available, *QoS resource reservation signaling*: Once a path with the required QoS is found, the next step is to reserve the required resources along that path. This is done by the resource reservation signaling protocol. For example, for applications that require certain minimum bandwidth guarantees, signaling protocol communicates with the medium access control subsystem to find and reserve the required bandwidth.

- *Admission control*: Even though a QoS feasible path may be available, the system needs to decide whether to actually serve the connection or not. If the call is to be served, the signaling protocol reserves the resources; otherwise, the application is notified of the rejection.
- *Packet scheduling*: When multiple QoS connections are active at the same time through a link, the decision on which QoS flow is to be served next is made by the scheduling scheme. For example, when multiple delay-constrained sessions are passing through a node, the scheduling mechanism decides on when to schedule the transmission of packets when packets belonging to more than one session are pending in the transmission queue of the node.

5.6.1 INSIGNIA

The INSIGNIA QoS framework was developed to provide adaptive services in ad hoc wireless networks. Adaptive services support applications that require only a minimum quantitative QoS guarantee (such as minimum bandwidth) called *base QoS*. Here user sessions adapt to the available level of service without explicit signaling between the source-destination pairs.

The INSIGNIA framework has the following key components for supporting adaptive real-time services:

- *Routing module*: The routing protocol finds a route from the source to the destination. It is also used to forward a data packet to the next intermediate relay node. The routing module is independent of other components and hence any existing routing protocol can be used. INSIGNIA assumes that the routing protocol provides new routes in case of topology changes.
- *Admission control*: This module allocates bandwidth to flows based on the maximum/minimum bandwidth requested. Once the bandwidth is reserved, the reservation must be refreshed periodically by a soft state mechanism. Typically, the reception of data packets refreshes the reservations done.
- *Packet scheduling*: Packets that are to be routed to other nodes are handled by the packet-scheduling module. The packets to be transmitted by a node are scheduled by the scheduler based on the forwarding policy. INSIGNIA uses a weighted round-robin service discipline.
- *Medium access control (MAC)*: The MAC protocol provides QoS-driven access to the shared wireless medium for adaptive real-time services. The INSIGNIA framework is transparent to any underlying MAC protocol.

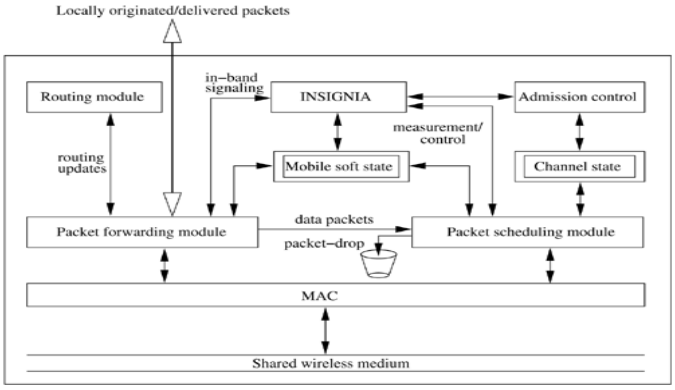


Figure 5.5. INSIGNIA QoS framework.

Route Maintenance

Due to host mobility, an on-going session may have to be rerouted in case of a path break. The flow restoration process must reestablish the reservation as quickly and efficiently as possible. During restoration, INSIGNIA does not preempt resources from the existing flows for admitting the rerouted flows.

INSIGNIA supports three types of flow restoration, namely, *immediate restoration,* which occurs when a rerouted flow immediately recovers to its original reservation; *degraded restoration,* which occurs when a rerouted flow is degraded for a period (T) before it recovers to its original reservation; and *permanent restoration,* which occurs when the rerouted flow never recovers to its original reservation.

5.7 Need for Energy Management in Ad Hoc Wireless Networks

The energy efficiency of a node is defined as the ratio of the amount of data delivered by the node to the total energy expended. Higher energy efficiency implies that a greater number of packets can be transmitted by the node with a given amount of energy reserve. The main reasons for energy management in ad hoc wireless networks are listed below:

- **Limited energy reserve:** The main reason for the development of ad hoc wireless networks is to provide a communication infrastructure in environments where the setting up of a fixed infrastructure is impossible. Ad hoc wireless networks have very limited energy resources.
- **Difficulties in replacing the batteries:** Sometimes it becomes very difficult to replace or recharge the batteries. In situations such as battlefields, this is almost impossible. Hence, energy conservation is essential in such scenarios.
- **Lack of central coordination:** The lack of a central coordinator, such as the base station in cellular networks, introduces multi-hop routing and necessitates that some of the intermediate nodes act as relay nodes.
- **Constraints on the battery source:** Batteries tend to increase the size and weight of a mobile node. Reducing the size of the battery results in less capacity which, in turn, decreases the active lifespan of the node. Hence, in addition to reducing the size of the battery, energy management techniques are necessary to utilize the battery capacity in the best possible way.

- **Selection of optimal transmission power:** The transmission power selected determines the reachability of the nodes. The consumption of battery charge increases with an increase in the transmission power. An optimal value for the transmission power decreases the interference among nodes, which, in turn, increases the number of simultaneous transmissions.

5.7.1 Classification of Energy Management Schemes

The need for energy management in ad hoc wireless networks, discussed in the previous section, points to the fact that energy awareness needs to be adopted by the protocols at all the layers in the protocol stack, and has to be considered as one of the important design objectives for such protocols. Energy conservation can be implemented using the following techniques:

- Battery management schemes
- Transmission power management schemes
- System power management schemes

The system power management approach can be further divided into the following categories:

  - Device management schemes
  - Processor power management schemes

Figure 5.6 provides an overview of some of the techniques at different layers of the protocol stack that fall into three categories: battery management, transmission power management, and system power management schemes. Though these schemes cannot be strictly classified under the different layers of the OSI protocol stack as they reside in more than one layer, the classification provided in this section is based on the highest layer in the protocol stack used by each of these protocols.

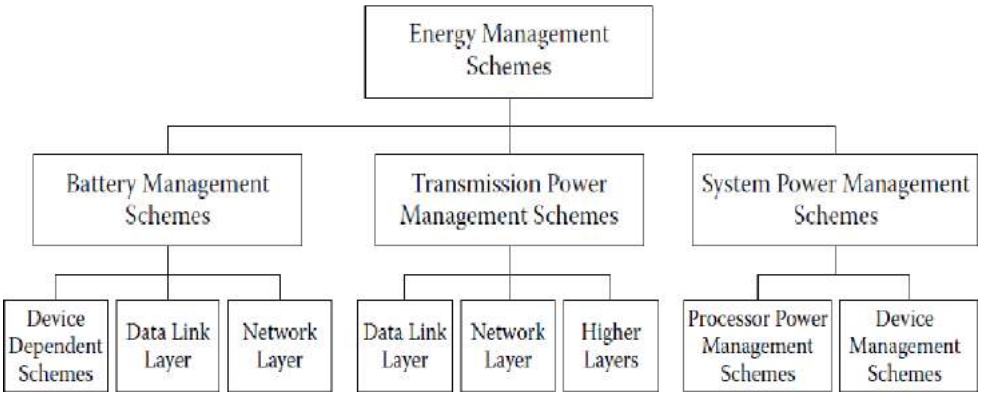


Figure 5.6. Classification of energy management schemes.

5.7.1.2 Battery Management Schemes

Battery-driven systems are those systems which are designed taking into consideration mainly the battery and its internal characteristics. They try to maximize the amount of energy provided by the power source by exploiting the inherent property of batteries to recover their charge when kept idle.

- ***Overview of Battery Characteristics***

  The major components of batteries are illustrated in Figure 5.7. A battery mainly consists of an anode, a cathode, an electrolyte medium, and a case. The anode is often a metal and the cathode a metallic oxide. The electrolyte is a salt solution that promotes the ion flow. The porous separator is used to prevent a short circuit between anode and cathode by keeping them from touching one another. The battery is contained in a structural support (case) that provides dimensional stability and a positive and a negative electrode for discharging (or recharging) the cell.

- ***Battery technologies***

  The most popular rechargeable battery technologies developed over the last two decades are comprised of nickel-cadmium, lithium ion, nickel metal-hydride, reusable alkaline, and lithium polymer. The main factors considered while designing a battery technology are the energy density (the amount of energy stored per unit weight of the battery), cycle life [the number of (re)charge cycles prior to battery disposal], environmental impact, safety, cost, available supply voltage, and charge/discharge characteristics.

- **Principles of battery discharge:** A battery typically consists of an array of one or more cells. Hence, in the subsequent sections, the terms "battery" and "cell" are used interchangeably. The three main voltages that characterize a cell are: (1) the open circuit voltage (*Voc*), that is, the initial voltage under a no-load condition of a fully charged cell, (2) the operating voltage (*Vi*), that is, the voltage under loaded conditions, and (3) the cut-off voltage (*Vcut*) at which the cell is said to be discharged. All the cells are defined by three main capacities:

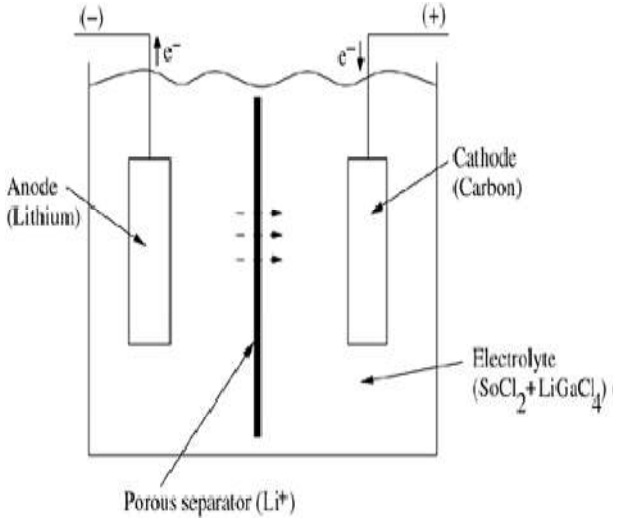


Figure 5.7. Basic structure of a lithium/thionyl chloride battery.

– Theoretical capacity: The amount of active materials (the materials that react chemically to produce electrical energy when the cell is discharged and restored when the cell is charged) contained in the cell refers to its theoretical capacity. A cell cannot exceed its theoretical capacity.

– Nominal (standard) capacity: This corresponds to the capacity actually available when discharged at a specific constant current. It is expressed in ampere-hours.

– Actual capacity: The energy delivered under a given load is said to be the actual capacity of the cell. A cell may exceed the actual capacity but not the theoretical capacity.

The constant current discharge behavior of lithium-manganese dioxide ($LiMnO2$) cells with $Voc = 3V$ and $Vcut = 1\ V$. The discharge curve is flat most of the time and a gradual slope is developed as the voltage reaches the cut-off voltage. The performance of a cell's discharge is measured using the following parameters:

– Discharge time: The time elapsed when a fully charged cell reaches its cut-off voltage and has to be replaced or recharged is called the discharge time of the cell.

– Specific power (energy): This is the power (energy) delivered by a fully charged cell under a specified discharge current. It is expressed in watt-per kilogram (watt-hour-per-kilogram).

Discharge current: There are mainly two models of battery discharge: constant current discharge and pulsed current discharge. In pulsed current discharge, the battery switches between short discharge periods and idle periods (rest periods).

- **Battery models:** Battery models depict the characteristics of the batteries used in real life. The pros and cons of following battery models are summarized: analytical models, stochastic models, electric circuit models, and electrochemical models.
- **Battery scheduling:** The use of multiple batteries in mobile nodes has become very common. The key aspect behind this kind of architecture is the property of charge recovery by the battery when it remains in idle state. A detailed description of the charge recovery property of the battery can be found in the next section.
- **Smart battery standard (SBS):** This is an emerging technology toward the development of batteries that consume low power. The main aim of SBS is to create standards by which the systems become aware of the batteries and interact with them in order to provide a better performance.

5.7.1.3 Transmission Power Management Schemes

The variation in transmission power greatly influences the reachability of a node. Increasing the transmission range not only increases coverage, but also the power consumption rate at the transmitter. This section deals with finding a trade-off between the two contradictory issues, that is, increasing the coverage of a node and decreasing its battery consumption.

Data Link Layer Solutions

Power control can be effected at the data link layer by means of topology control and constructing a power control loop. This section describes different power-based solutions at the data link layer. Some of the solutions proposed to calculate the optimum transmission range are as follows:

- Dynamic power adjustment policies
- Distributed topology control algorithms
- Constructing distributed power control loop
- Centralized topology control algorithm

Dynamic Power Adjustment

Based on the Link Affinity Ad hoc wireless networks are prone to constant link failures due to node mobility, hence the stability of routes cannot be assured in such situations. But frequent link failures lead to reduced throughput.

Distributed Topology Control Mechanisms

According to this algorithm, each node of the ad hoc wireless network independently runs a localized algorithm and decides the appropriate power level to be used by that node. A node increases the power directionally until it finds one node in all the directions. Then it tries to increase the lifetime of the nodes to a greater extent by reducing the transmission power and having less coverage of the nodes while guaranteeing the same connectivity as the one achieved when the nodes are maximally powered.

Constructing Distributed Power Control Loop

a power control loop which increases the battery lifetime by 10-15% and the throughput by around 15%. The algorithm is tested on the model that assumes mobility, group communication, and fading due to blockages such as manmade obstacles. The proposed algorithm works at the MAC layer in a distributed fashion. The main objective behind the algorithm is to reduce the energy cost of communication between the nodes and thereby increasing the battery lifetime and the effective bandwidth.

Centralized Topology Control Algorithm

Centralized algorithm which adjusts the power level of the nodes to create the desired topology. The problem is constrained as an optimization problem with power level as the optimization objective and the constraints are connectivity and biconnectivity.

The *connect* algorithm is similar to the minimum cost spanning tree algorithm. The basic idea used in this algorithm is to iteratively merge the connected components until only one component is left. The following steps are performed in order to carry out this algorithm:

Step 1: First, the connected node pairs are sorted in the increasing order of the mutual distance.

Step 2: If the nodes are in different network components, the power of the nodes are increased so as to reach the other nodes.

Step 3: Step 2 is repeated until the whole network becomes connected.

The *biconnect* algorithm attempts to discover a biconnected graph from the given graph *M* so as to satisfy the objectives and the constraints. The extension to the biconnected network from the algorithm *connect* can be done as follows:

Step 1: The biconnected components are identified in the graph induced by the algorithm *connect*based on the depth-first search method.

Step 2: The nodes are arranged in non-decreasing order of the connected node pairs as done in the previous algorithm.

Step 3: Nodes which are in different components of the network are connected by adjusting the power appropriately, and this step is repeated until the network becomes biconnected.

5.7.1.4 System Power Management Schemes

System power consists of the power used by all hardware units of the node. This power can be conserved significantly by applying the following schemes:

1. Processor power management schemes
2. Device power management schemes

- ***Processor Power Management Schemes***

Processor power management schemes deal with techniques that try to reduce the power consumed by the processor, such as reducing the number of calculations performed. In this section, we discuss some of the power management techniques that are applied at the hardware level when there is a request from the higher layers.

Power-Saving Modes

The nodes in an ad hoc wireless network consume a substantial amount of power even when they are in an idle state since they keep listening to the channel, awaiting request packets from the neighbors. In order to avoid this, the nodes are switched off during idle conditions and switched on only when there is an arrival of a request packet. This primarily has two advantages: reducing the wastage in power consumed when the node is in the listen mode, and providing idle time for the batteries of the node to recover charges.

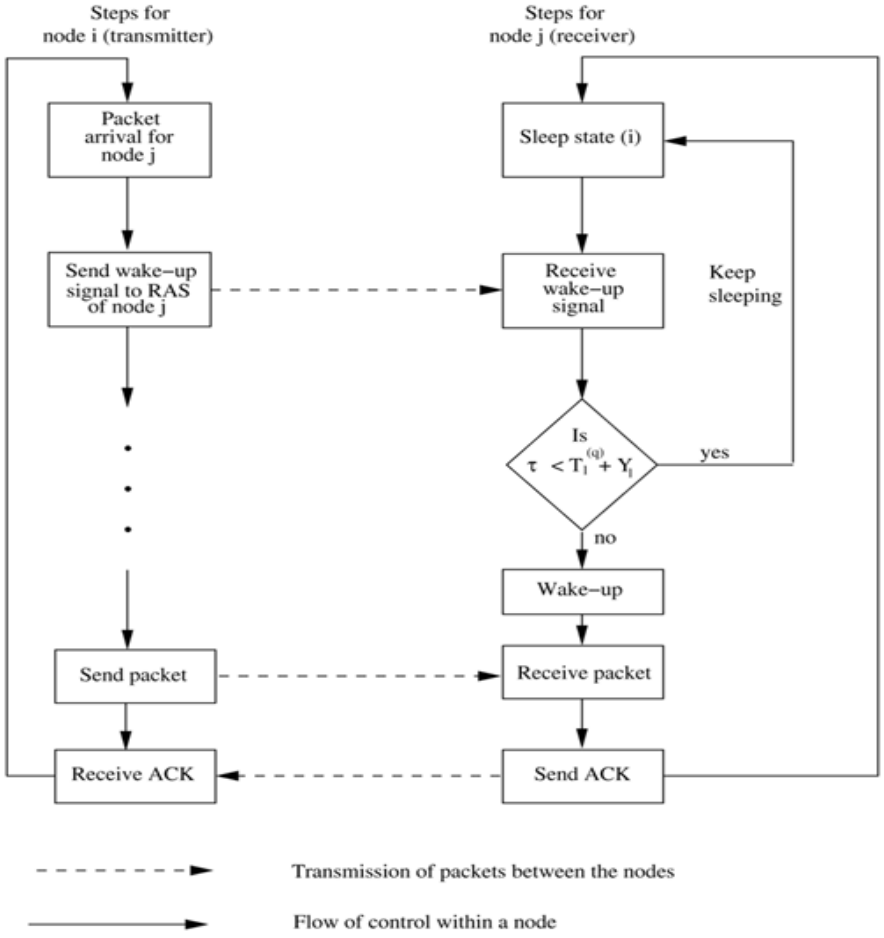


Figure 5.8. Power management scheme using remote activated switch.

Power-Aware Multi-Access Signaling

Power-aware multi-access signaling (PAMAS) is another approach for determining the time duration for which the node should be turned off. This scheme suggests the addition of a separate signaling channel in the MACA protocol. The RTS-CTS signaling takes place in this separate channel, which determines the time period for which the node has to be powered off. The algorithm is divided into two parts:

Addition of separate signaling channel: This can be explained through a state diagram as shown in Figure 5.9. A node can be in any one of the six states represented within the boxes. Initially, when a node neither transmits nor receives packets, it stays in the *idle* state.

- *Packet transmission:*
  - As soon as the node gets a packet for transmission, it transmits an RTS and enters the *Await CTS* state.
  - If it does not receive the CTS, it enters the binary exponential back-off (*BEB*) state. A node also enters the *BEB* state if it hears a busy tone when a neighboring node which is actively transmitting sends a busy tone in the control channel.
  - After receiving the CTS, it enters the *Transmit packet* state and starts transmitting the packet.
- *Packet reception:*
  - As soon as a node receives an RTS, it sends a CTS back to the sender and enters the *Await packet* state, only if no other neighboring nodes exist in the *Await CTS* or *Transmit packet* state.
  - If packets arrive on time, the node enters the *Receive packet* state and starts receiving the packets.
  - If the packet does not arrive on time, the node enters the *idle* state again.

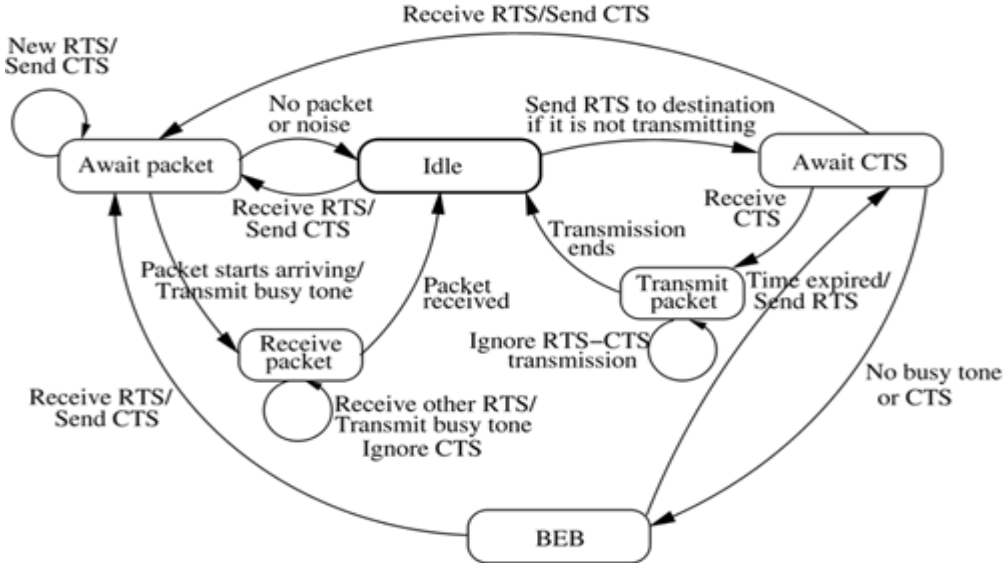


Figure 5.9. PAMAS protocol.

Powering off the radios: We now discuss the conditions under which the node enters the power-off mode:

- Condition 1: The node has no packets for transmission.
- Condition 2: A neighbor node is transmitting or receiving packets, that is, the channel is busy.

- ***Device Power Management Schemes***

Some of the major consumers of power in ad hoc wireless networks are the hardware devices present in the nodes. Various schemes have been proposed in the design of hardware that minimize the power consumption.

Low-Power Design of Hardware

Low-power design of hardware results in a significant improvement in the energy conservation. Some of the low-power design suggestions include varying clock speed CPUs, disk spin down, and flash memory. We now look into some of the sources of power consumption in the ad hoc wireless networks and the corresponding solutions to reduce power consumption

- Major sources of power consumption in ad hoc wireless networks are the transmitters and receivers of the communication module. The design of transceivers has a significant effect on the power consumption. Hence, great care must be taken while designing them.
- The main hardware of a mobile node, in general, consists of the LCD display, DRAM, CD ROM drive, CPU, wireless interface card (in the case of a computer), and I/O subsystems.

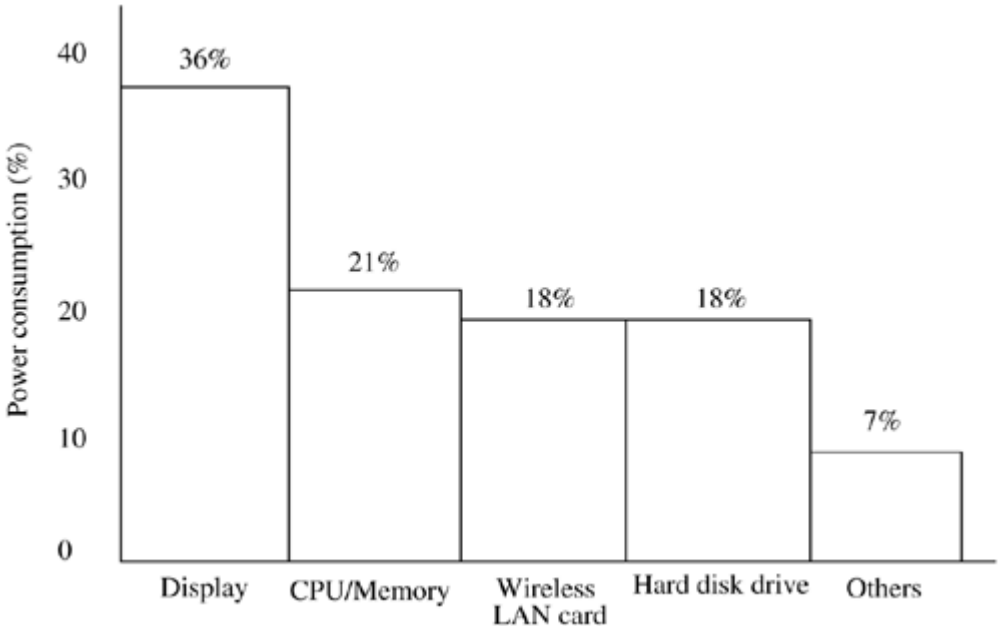


Figure 5.10. Power consumed by various units of hardware.

CPU Power Consumption

The energy required for the CPU operation depends largely on the clock frequency (F). As the clock rate increases, frequent switching of the logic gates between different voltage levels (V), that is, the ground voltage and the peak voltage, takes place, which leads to higher power

consumption. Another effect that significantly influences the CPU power consumption is the chaining of transistors. The larger the capacitance (*C*) of these transistors, the higher the energy required. Hence, the total power required by the CPU is proportional to *CV2F*. The solution suggested is as follows:

- The parameter *C* can be set during the chip design.
- The values of *F* and *V* can be set dynamically at run-time which, along with power-aware CPU scheduling policies, reduces the power consumption significantly.

Hard Disk Drive (HDD) Power Consumption

Various approaches have been suggested for turning off the drives and to bring down the speed of spinning. We now see how the spin-down can be performed on the disk drives.

- By using historical data: One method suggested is based on the traces of disk usage collected over a long period of time. By analyzing various spin-down thresholds, an optimal value for threshold has to be agreed upon, which acts as a balance between the two contradictory requirements of reducing power consumption and reducing the access delays.
- Spin-up/spin-down policies: Some of the policies used in deciding the time at which the hard disk speed has to be varied are given below.
- Optimal-optimal policy: According to this policy, by having a complete knowledge of the future, the optimal values for spin-down can be obtained. This tells when the disk has to be spun down to obtain maximum efficiency, and when the disk has to be spun up to get ready for the next disk operation. This is an unrealistic policy.
- Threshold-demand policy: This algorithm forces spin-down of the disk only after attaining a certain threshold value. But the spin-up takes place only if there exists a request for the disk.
- Predictive-predictive policy: Both the spin-up and spin-down time values are predicted based on the past values.