Chapter-1

1. Introduction

Wireless technology is the method of delivering data from one point to another without using physical wires, and includes radio, cellular, infrared, and satellite. A historic perspective will provide you with a general understanding of the substantial evolution that has taken place in this area. The common wireless networks of today originated from many evolutionary stages of **wireless communications and telegraph and radio applications.** Although some discoveries occurred in the early 1800s, much of the evolution of wireless communication began with the emergence of the electrical age and was affected by modern economics as much as by discoveries in physics.

1.1. History of Wireless Network

In the 19th century, numerous inventors and scientists, including Michael Faraday, James Clerk Maxwell, Heinrich Rudolf Hertz, Nikola Tesla, David Edward Hughes, Thomas Edison, and Guglielmo Marconi, began to experiment with wireless communications. These innovators discovered and created many theories about the concepts of electrical magnetic *radio frequency (RF)*.

Wireless networking technology was first used by the U.S. military during World War II to transmit data over an RF medium using classified encryption technology, to send battle plans across enemy lines. The *spread spectrum* radio technologies often used in today's WLANs were also originally patented during the era of World War II, although they were not implemented until almost two decades later.

In 1970, the University of Hawaii developed the first wireless network, called ALOHAnet, to wirelessly communicate data between the Hawaiian Islands. The network used a LAN communication Open Systems Interconnection layer 2 protocols called ALOHA on a wireless shared medium in the 400 MHz frequency range. The technology used in ALOHAnet is often credited as a building block for the Medium Access Control technologies of Carrier Sense Multiple Access with Collision Detection (CSMA/CD) used in Ethernet and Carrier Sense

Multiple Access with Collision Avoidance (CSMA/CA) used in 802.11 radios.

In the 1990s, commercial networking vendors began to produce low-speed wireless data networking products, most of which operated in the 900 MHz frequency band. The Institute of Electrical and Electronics Engineers (IEEE) began to discuss standardizing WLAN technologies in 1991. In 1997, the IEEE ratified the original 802.11 standard that is the foundation of the WLAN technologies.

Legacy 802.11 technology was deployed between 1997 and 1999 mostly in warehousing and manufacturing environments for the use of low-speed data collection with wireless barcode scanners. In 1999, the IEEE defined higher data speeds with the 802.11b amendment. The introduction of data rates as high as 11 Mbps along with price decreases, ignited the sales of wireless home networking routers in the small office, home office (SOHO) marketplace. Home users soon became accustomed to wireless networking in their homes and began to demand that their employers also provide wireless networking capabilities in the workplace. After initial resistance to 802.11 technology, small companies, medium-sized businesses, and corporations began to realize the value of deploying 802.11 wireless networking in their enterprises.

If you ask the average user about their 802.11 wireless networks, they may give you a strange look. The name that many people recognize for the technology is *Wi-Fi*. Wi-Fi is a marketing term, recognized worldwide by millions of people as referring to 802.11 wireless networking.

What Does the Term *Wi-Fi* Mean?

Many people mistakenly assume that *Wi-Fi* is an acronym for the phrase *wireless fidelity* (much like *hi-fi* is short for *high fidelity*), but Wi-Fi is simply a brand name used to market 802.11 WLAN technology. Ambiguity in IEEE framework standards for wireless communications allowed manufacturers to interpret the 802.11 standard in different ways. As a result, multiple vendors could have IEEE 802.11–compliant devices that did not interoperate with each other. The organization Wireless Ethernet Compatibility Alliance (WECA) was created to further define the IEEE standard in such a way as to force interoperability between vendors. WECA, now known as the Wi-Fi

Alliance, chose the term *Wi-Fi* as a marketing brand. The Wi-Fi Alliance champions enforcing interoperability among wireless devices. To be Wi-Fi compliant, vendors must send their products to a Wi-Fi Alliance test lab that thoroughly tests compliance to the Wi-Fi certification.

Wi-Fi radios are used for numerous enterprise applications and can also be found in laptops, mobile phones, cameras, televisions, printers, and many other consumer devices. More than 350 million Wi-Fi chipsets were shipped in 2010, with current estimates of annual sales of over one billion Wi-Fi chipsets within the next few years.

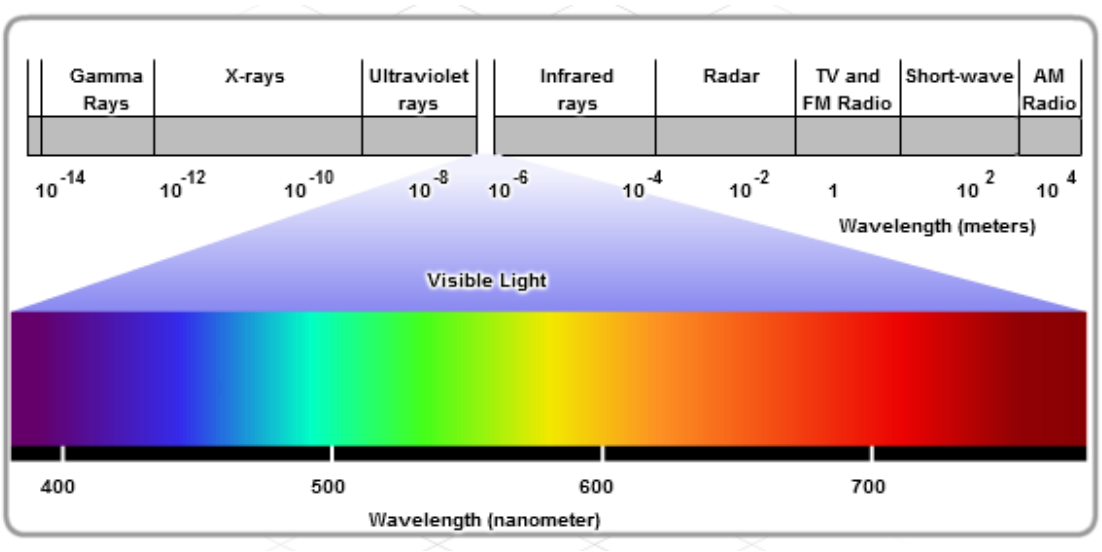
According to the Wi-Fi Alliance, the billionth Wi-Fi chipset was sold in 2009. Since the original standard was created in 1997, 802.11 technologies have grown to enormous proportions; Wi-Fi has now become part of our worldwide culture.

1.2. Wireless Technologies and Devices

In addition to the wired network, various technologies exist that allow the transmission of information between hosts without cables. These are known as wireless technologies.

Wireless technologies use electromagnetic waves to carry information between devices. An electromagnetic wave is the same medium that carries radio signals through the air.

The electromagnetic spectrum includes such things as radio and television broadcast bands, visible light, x-rays and gamma-rays. Each of these has a specific range of wavelengths and associated energies as shown in the diagram.

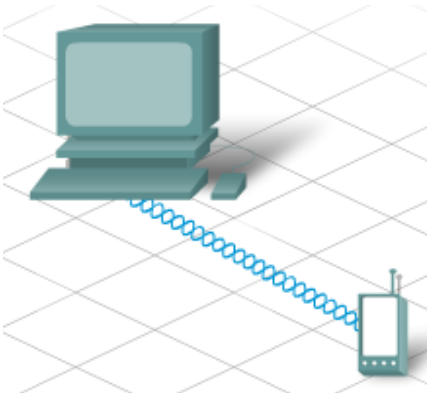


Some types of electromagnetic waves are not suitable for carrying data. Other parts of the spectrum are regulated by governments and licensed to various organizations for specific applications. Certain areas of the spectrum have been set aside to allow public use without the restriction of having to apply for special permits. The most common wavelengths used for public wireless communications include the Infrared and part of the Radio Frequency (RF) band.

Infrared

Infrared (IR) is relatively low energy and cannot penetrate through walls or other obstacles. However, it is commonly used to connect and move data between devices such as Personal Digital Assistants (PDAs) and PCs. A specialized communication port known as an Infrared Direct Access (IrDA) port uses IR to exchange information between devices. IR only allows a one-to-one type of connection.

IR is also used for remote control devices, wireless mice, and wireless keyboards. It is generally used for short-range, line-of-sight, communications. However, it is possible to reflect the IR signal off objects to extend the range. For greater ranges, higher frequencies of electromagnetic waves are required.



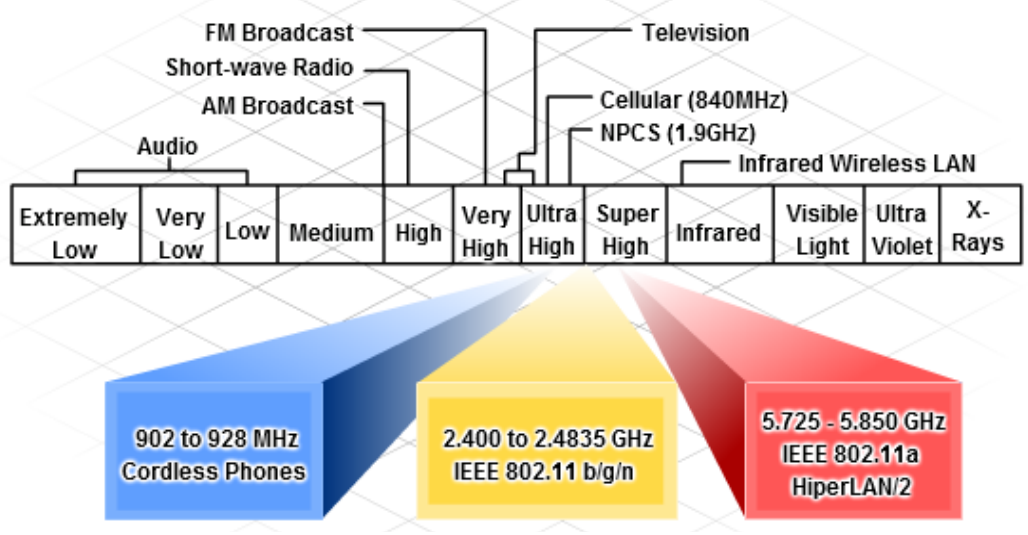
Radio Frequency (RF)

RF waves can penetrate through walls and other obstacles, allowing a much greater range than IR.

Certain areas of the RF bands have been set aside for use by unlicensed devices such as wireless LANs, cordless phones and computer peripherals. This includes the 900 MHz, 2.4 GHz and the 5 GHz frequency ranges. These ranges are known as the Industrial Scientific and Medical (ISM) bands and can be used with very few restrictions.

Bluetooth is a technology that makes use of the 2.4 GHz band. It is limited to low-speed, short-range communications, but has the advantage of communicating with many devices at the same time. This one-to-many communications has made Bluetooth technology the preferred method over IR for connecting computer peripherals such as mice, keyboards and printers.

Other technologies that make use of the 2.4 GHz and 5GHz bands are the modern wireless LAN technologies that conform to the various IEEE 802.11 standards. They are unlike Bluetooth technology in that they transmit at a much higher power level, which gives them a greater range.



1.3. Benefits and Limitations of Wireless Technology

Wireless technology offers many advantages compared to traditional wired networks.

One of the main advantages is the ability to provide anytime, anywhere connectivity. The widespread implementation of wireless in public locations, known as hotspots, allows people to easily connect to the Internet to download information and exchange emails and files.

Wireless technology is fairly easy and inexpensive to install. The cost of home and business wireless devices continues to decrease. Yet, despite the decrease in cost, the data rate and capabilities of these devices have increased, allowing faster, more reliable wireless connections.

Wireless technology enables networks to be easily expanded, without the limitations of cabled connections. New and visiting users can join the network quickly and easily.

Generally the following are the *benefits* of wireless technology:

- ***Mobility*** – allows for easy connection of both stationary and mobile clients
- ***Scalability***- can be easily expanded to allow more users to connect and to increase the coverage area
- ***Flexibility***- provides anytime, anywhere connectivity

- ***Cost savings***- equipment costs continue to fall as the technology matures
- ***Reduce installation time***- installation of a single piece of equipment can provide connectivity for a number of people
- ***Reliability in harsh environments***- easy to install in emergency and hostile environments

Despite the flexibility and benefits of wireless, there are some *limitations and risks*.

First, Wireless LAN (WLAN) technologies make use of the unlicensed regions of the RF spectrum. Since these regions are unregulated, many different devices make use of them. As a result, these regions are congested and signals from different devices often interfere with each other. In addition, many devices such as microwave ovens and cordless phones use these frequencies and can interfere with WLAN communications.

Second, a major concern with wireless is security. Wireless provides ease of access. It does this by broadcasting data in a manner that allows anyone the ability to access it. However, this same feature also limits the amount of protection wireless can provide for the data. It allows anyone to intercept the communication stream, even unintended recipients. To address these security concerns, techniques have been developed to help secure wireless transmissions including encryption and authentication.

The limitations of wireless technology are given as follows:

- ***Interference –***wireless technology is susceptible to interference from other devices that produce electromagnetic energies. This includes: cordless phones, microwaves, televisions and other wireless LAN implementations.
- ***Network and Data Security-*** Wireless LAN technology is designed to provide access to the data being transmitted, not security of the data. Additionally, it can provide an unprotected entrance into the wired network.

➢ ***Technology-*** technology continues to evolve. Wireless LAN technology does not currently provide the speed or reliability of wired LANs.

1.4. Types of Wireless Network

Wireless networks are grouped into four major categories: Wireless Personal Area networks (WPAN), Wireless Local Area networks (WLAN), Wireless Metropolitan Area Network (WMAN), and Wireless Wide Area networks (WWAN).

Despite these distinct categories, it is difficult to place boundary limitations on a wireless implementation. This is because, unlike a wired network, wireless networks do not have precisely defined boundaries. The range of wireless transmissions can vary due to many factors. Wireless networks are susceptible to outside sources of interference, both natural and man-made. Fluctuations in temperature and humidity can greatly alter the coverage of wireless networks. Obstacles within the wireless environment can also affect the range.

Wireless Personal Area Network *(WPAN)*

This is the smallest wireless network used to connect various peripheral devices such as mice, keyboards and PDAs to a computer. All of these devices are dedicated to a single host with usually use IR or Bluetooth technology.

Wireless Local Area Network (*WLAN)*

WLAN is typically used to extend the boundaries of the local wired network (LAN). WLANs use RF technology and conform to the IEEE 802.11 standards. They allow many users to connect to a wired network through a device known as an Access Point (AP). An Access Point provides a connection between wireless hosts and hosts on an Ethernet wired network.

Wireless Metropolitan Area Network (WMAN)
WMAN provides RF coverage to a metropolitan area such as a city and the surrounding suburbs. WMANs have been created for some time by matching different wireless technologies, and recent advancements have made this more practical.

One wireless technology that is often associated with a WMAN is defined by the 802.16 standard. This standard defines broadband

wireless access and is sometimes referred to as Worldwide Interoperability for Microwave Access (WiMAX). The WiMAX Forum is responsible for compatibility and interoperability testing of wireless broadband equipment such as 802.16.

802.16 technologies are viewed as a direct competitor to other broadband services such as DSL and cable. Although 802.16 wireless networking is typically thought of as a last mile data-delivery solution, the technology might also be used to provide access to users over citywide areas.

Wireless Wide Area Network (*WWAN)*

WWAN networks provide coverage over extremely large areas. A good example of a WWAN is the cell phone network. These networks use technologies such as Code Division Multiple Access (CDMA) or Global System for Mobile Communication (GSM) and are often regulated by government agencies.

	WPAN	WLAN	WWAN
Standards	Bluetooth v2.0+ EDR**	IEEE802.11 a/b/g/n, HiperLAN, HiperLAN2	GSM, GPRS, CDMA
Speed	< 3 Mbps	1-540 Mbps	10-384 Kbps
Range	Short	Medium	Long
Applications	Peer-to-Peer device to device	Home, small business and enterprise networks	PDAs, mobile phones, cellular access

** EDR is Enhanced Data Rate

Speed and ranges are constantly increasing with newer technologies.

1.5. Standard Organizations

Each of the standards organizations discussed below help to guide a different aspect of the wireless networking industry.

The International Telecommunication Union Radio communication Sector (ITU-R) and local entities such as the Federal Communications Commission (FCC) set the rules for what the user can do with a radio transmitter. These organizations manage and regulate frequencies, power levels, and transmission methods. They also work together to

help guide the growth and expansion that is being demanded by wireless users.

Federal Communications Commission *(FCC)*

FCC regulates communications within the United States as well as communications to and from the United States. It is responsible for regulating interstate and international communications by radio, television, wire, satellite, and cable. The task of the FCC in wireless networking is to regulate the radio signals that are used for wireless networking. The FCC has jurisdiction over the 50 states, the District of Columbia, and U.S. possessions. Most countries have governing bodies that function similarly to the FCC.

The FCC and the respective controlling agencies in the other countries typically regulate two categories of wireless communications: licensed spectrum and unlicensed spectrum.

Licensed Spectrum

✓ Frequencies require an approved license application, and the financial costs are very high.

Unlicensed Spectrum

✓ Users do not have to go through the license application procedures before they can install a wireless system

✓ There are no financial costs, you still must abide by transmission regulations and other restrictions. In other words, transmitting in an unlicensed frequency may be free, but there still are rules.

✓ The main disadvantage to transmitting in an unlicensed frequency band is that anyone else can also transmit in that same frequency space.

✓ Unlicensed frequency bands are often very crowded; therefore, transmissions from other individuals can cause interference with your transmissions.

International Telecommunication Union Radio communication Sector *(ITU-R)*

A global hierarchy exists for management of the RF spectrum worldwide. The United Nations has tasked the *International*

Telecommunication Union Radio communication Sector (ITU-R) with global spectrum management. The ITU-R maintains a database of worldwide frequency assignments and coordinates spectrum management through five administrative regions.

The five regions are broken down as follows:

- ✓ **Region A: North and South America** Inter-American Telecommunication Commission (CITEL)
- ✓ **Region B: Western Europe** European Conference of Postal and Telecommunications Administrations (CEPT)
- ✓ **Region C: Eastern Europe and Northern Asia** Regional Commonwealth in the field of Communications (RCC)
- ✓ **Region D: Africa** African Telecommunications Union (ATU)
- ✓ **Region E: Asia and Australasia** Asia-Pacific Telecommunity (APT)

It is important to understand that communications are regulated differently in many regions and countries. For example, European RF regulations are very different from the regulations used in North America. When deploying a WLAN, please take the time to learn about rules and policies of the local *regulatory domain authority*.

Institute of Electrical and Electronics Engineers (*IEEE)*

l, commonly known as the *IEEE*, is a global professional society with more than 400,000 members. The IEEE's mission is to "foster technological innovation and excellence for the benefit of humanity." To networking professionals, that means creating the standards that we use to communicate. The IEEE standards must adhere to the rules of the communications organizations, such as the FCC.

The IEEE is probably best known for its LAN standards, the IEEE 802 project.

IEEE projects are subdivided into working groups to develop standards that address specific problems or needs. For instance, the IEEE 802.3 working group was responsible for the creation of a standard for Ethernet, and the IEEE 802.11 working group was responsible for creating the WLAN standard. The numbers are assigned as the groups are formed, so the 11 assigned to the wireless group indicates that it was the 11th working group formed under the IEEE 802 project.

As the need arises to revise existing standards created by the working groups, task groups are formed. These task groups are assigned a

sequential single letter (multiple letters are assigned if all single letters have been used) that is added to the end of the standard number (for example, 802.11a, 802.11g, and 802.3af). Some letters are not assigned.

For example *o* and *l* are not assigned to prevent confusion with the numbers 0 and 1. Other letters may not be assigned to task groups to prevent confusion with other standards. For example, 802.11x has not been assigned because it can be easily confused with the 802.1X standard and because 802.11*x* has become a common casual reference to the 802.11 family of standards.

It is important to remember that the IEEE standards, like many other standards, are written documents describing how technical processes and equipment should function. Unfortunately, this often allows for different interpretations when the standard is being implemented, so it is common for early products to be incompatible between vendors, as was the case with the early 802.11 products.

Internet Engineering Task Force (*IETF)*

It is an international community of people in the networking industry whose goal is to make the Internet work Better. The mission of the IETF, as defined by the organization in a document known as RFC3935, is "to produce high quality, relevant technical and engineering documents that influence the way people design, use, and manage the Internet in such a way as to make the Internet work better. IETF is responsible for creating Internet standards. Many of these standards are integrated into the wireless networking and security protocols and standards.

The IETF is one of five main groups that are part of the Internet Society (ISOC). The ISOC groups include the following:

- ✓ Internet Architecture Board (IAB)
- ✓ Internet Corporation for Assigned Names and Numbers (ICANN)
- ✓ Internet Engineering Steering Group (IESG)
- ✓ Internet Research Task Force (IRTF)

The IETF is broken into eight subject matter areas: Applications, General, Internet, Operations and Management, Real-Time Applications and Infrastructure, Routing, Security, and Transport.

The IETF is made up of a large number of groups, each addressing specific topics. An IETF working group is created by the IESG and is given a specific charter or specific topic to address.

The results of a working group are usually the creation of a document known as a *Request for Comments (RFC)*. Contrary to its name, an RFC is not actually a request for comments, but a statement or definition. Most RFCs describe network protocols, services, or policies and may evolve into an Internet standard. RFCs are numbered sequentially, and once a number is assigned it is never reused. RFCs may be updated or supplemented by higher-numbered RFCs. As an example, Mobile IPv4 is described in RFC 3344 and updated in RFC 4721. When RFC 3344 was created, it made RFC 3220 obsolete. At the top of the RFC document, it states whether it is updated by another RFC and also if it makes any other RFCs obsolete.

Not all RFCs are standards. Each RFC is given a status, relative to its relationship with the Internet standardization process: Informational, Experimental, Standards Track, or Historic. If it is a Standards Track RFC, it could be a Proposed Standard, Draft Standard, or Internet Standard. When an RFC becomes a standard, it still keeps its RFC number, but it is also given an "STD xxxx" label. The relationship between the STD numbers and the RFC numbers is not one to one. STD numbers identify protocols whereas RFC numbers identify documents. Many of the protocol standards, best current practices, and informational documents produced by the IETF affect WLAN security.

Wi-Fi Alliance

The *Wi-Fi Alliance* is a global, nonprofit industry association of more than 350 member companies devoted to promoting the growth of WLANs. One of the primary tasks of the Wi-Fi Alliance is to market the Wi-Fi brand and raise consumer awareness of new 802.11 technologies as they become available. Because of the Wi-Fi Alliance's overwhelming marketing success, the majority of the worldwide 450 million Wi-Fi users are likely to recognize the Wi-Fi logo seen in the figure below

Wi-Fi logo



The Wi-Fi Alliance's main task is to ensure the interoperability of WLAN products by providing certification testing. During the early days of the 802.11 standard, the Wi-Fi Alliance further defined some of the ambiguous standards requirements and provided a set of guidelines to ensure compatibility between different vendors.

The Wi-Fi Alliance, originally named the Wireless Ethernet Compatibility Alliance (WECA), was founded in August 1999. The name was changed to the Wi-Fi Alliance in October 2002.

Multiple Wi-Fi CERTIFIED programs exist that cover basic connectivity, security, quality of service (QoS), and more. Testing of vendor Wi-Fi products is performed in 14 independent authorized test laboratories in eight countries. The guidelines for interoperability for each Wi-Fi CERTIFIED program are usually based on key components and functions that are defined in the IEEE 802.11-2007 standard and various 802.11 amendments. In fact, many of the same engineers who belong to 802.11 task groups are also contributing members of the Wi-Fi Alliance. However, it is important to understand that the IEEE and the Wi-Fi Alliance are two separate organizations. The IEEE 802.11 task group defines the WLAN standards, and the Wi-Fi Alliance defines interoperability certification programs.

The Wi-Fi CERTIFIED programs include the following:

Core Technology & Security The core technology and security program certifies 802.11a, b, g, and/or n interoperability to ensure that the essential wireless data transmission works as expected. Each device is tested according to its capabilities.

In addition to testing the device's transmission capabilities, each device must support *robust security network (RSN) c*apabilities, security mechanisms that were originally defined in the IEEE 802.11i amendment. Devices must support Wi-Fi Protected Access (WPA) and Wi-Fi Protected Access 2 (WPA2) security mechanisms for personal (WPA2-Personal) or enterprise (WPA2-Enterprise) environments. Additionally, enterprise devices must support *Extensible Authentication Protocol (EAP)*, which is used to validate the identity of the wireless device or user.

Wi-Fi Multimedia (WMM)—Multimedia WMM is based on the QoS mechanisms that were originally defined in the IEEE 802.11e

amendment. WMM enables Wi-Fi networks to prioritize traffic generated by different applications. In a network where WMM is supported by both the access point and the client device, traffic generated by time-sensitive applications such as voice or video can be prioritized for transmission on the half-duplex RF medium. WMM certification is mandatory for all core certified devices that support 802.11n. WMM certification is optional for core certified devices that support 802.11 a, b, or g.

WMM Power Save (WMM-PS)—Multimedia WMM-PS helps conserve battery power for devices using Wi-Fi radios by managing the time the client device spends in sleep mode. Conserving battery life is critical for handheld devices such as barcode scanners and VoWiFi phones. To take advantage of power-saving capabilities, both the device and the access point must support WMM Power Save.

Wi-Fi Protected Setup—Security Wi-Fi Protected Setup defines simplified and automatic WPA and WPA2 security configurations for home and small-business users. Users can easily configure a network with security protection by using a personal identification number (PIN) or a button located on the access point and the client device.

Wi-Fi Direct Wi-Fi Direct enables Wi-Fi devices to connect directly without the use of an access point, making it easier to print, share, synch, and display. Wi-Fi Direct is ideal for mobile phones, cameras, printers, PCs, and gaming devices needing to establish a one to-one connection, or even connecting a small group of devices. Wi-Fi Direct is simple to configure (in some cases as easy as pressing a button), provides the same performance and range as other Wi-Fi certified devices, and is secured using WPA2 security.

CWG-RF—Multimedia Converged Wireless Group-RF Profile (CWG-RF) was developed jointly by the Wi-Fi Alliance and the Cellular Telecommunications and Internet Association (CTIA), now known as The Wireless Association. CWG-RF defines performance metrics for Wi-Fi and cellular radios in a converged handset to help ensure that both technologies perform well in the presence of the other. All CTIA-certified handsets now include this certification.

Voice Personal—Application Voice Personal offers enhanced support for voice applications in residential and small-business Wi-Fi networks. These networks include one access point, mixed voice and data traffic

from multiple devices (such as phones, PCs, printers, and other consumer electronic devices), and support for up to four concurrent phone calls.

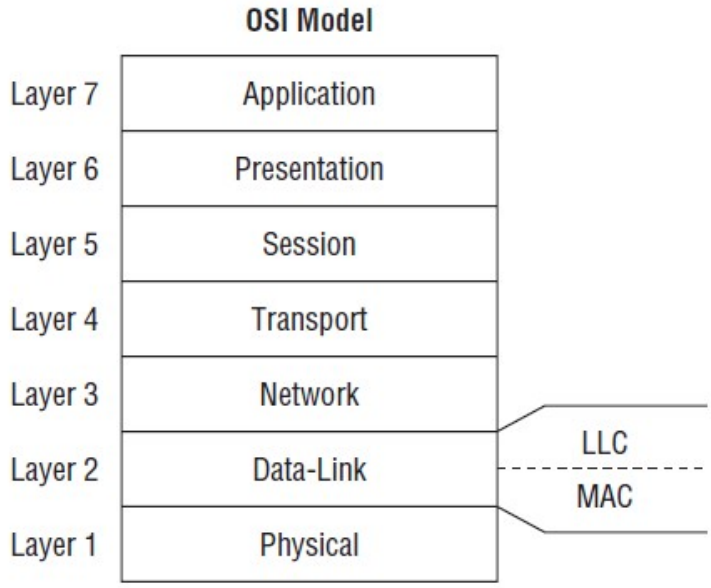
Both the access point and the client device must be certified to achieve performance matching the certification metrics.

As 802.11 technologies evolve, new Wi-Fi CERTIFIED programs will be detailed by the Wi-Fi Alliance. The next certification will probably be Voice Enterprise, which will expand on the elements of the Voice Personal certification program and define enhanced support for voice applications in the enterprise environment. Some aspects of the 802.11r (secure roaming) and 802.11k (resource management) amendments will probably be tested in Voice Enterprise. Remember that the core technologies and security is mandatory for all Wi-Fi certified products, meaning that Voice Enterprise certified devices will be required to support WPA2 and EAP.

International Organization for Standardization (*ISO)*

The *International Organization for Standardization*, commonly known as the *ISO*, is a global, nongovernmental organization that identifies business, government, and society needs and develops standards in partnership with the sectors that will put them to use. The ISO is responsible for the creation of the Open Systems Interconnection (OSI) model, which has been a standard reference for data communications between computers since the late 1970s.

The OSI model is the cornerstone of data communications, and learning to understand it is one of the most important and fundamental tasks a person in the networking industry can undertake. The following graphic shows the seven layers of the OSI.



The IEEE 802.11-2007 standard defines communication mechanisms only at the Physical layer and MAC sub layer of the Data-Link layer of the OSI model.

A communication device can exhibit any one of the following characteristics:

- **Fixed and wired**: This configuration describes the typical desktop computer in an office. Neither weight nor power consumption of the devices allow for mobile usage. The devices use fixed networks for performance reasons.

- **Mobile and wired**: Many of today's laptops fall into this category; users carry the laptop from one hotel to the next, reconnecting to the company's network via the telephone network and a modem.

- **Fixed and wireless**: This mode is used for installing networks, e.g., in historical buildings to avoid damage by installing wires, or at trade shows to ensure fast network setup.

- **Mobile and wireless**: This is the most interesting case. No cable restricts the user, who can roam between different wireless networks. Most technologies discussed in this book deal with this type of device and the networks supporting them. Today's most successful example for this category is GSM with more than 800 million users.

APPLICATIONS OF MOBILE COMPUTING

In many fields of work, the ability to keep on the move is vital in order to utilize time efficiently. The importance of Mobile Computers has been highlighted in many fields of which a few are described below:

a. **Vehicles:** Music, news, road conditions, weather reports, and other broadcast information are received via digital audio broadcasting (DAB) with 1.5 Mbit/s. For personal communication, a universal mobile telecommunications system (UMTS) phone might be available offering voice and data connectivity with 384 kbit/s. The current position of the car is determined via the global positioning system (GPS). Cars driving in the same area build a local ad-hoc network for the fast exchange of information in emergency situations or to help each other keep a safe distance. In case of an accident, not only will the airbag be triggered, but the police and ambulance service will be informed via an emergency call to a service provider. Buses, trucks, and trains are already transmitting maintenance and logistic information to their home base, which helps to improve organization (fleet management), and saves time and money.

Emergencies: An ambulance with a high-quality wireless connection to a hospital can carry vital information about injured persons to the hospital from the scene of the accident. All the necessary steps for this particular type of accident can be prepared and specialists can be consulted for an early diagnosis. Wireless networks are the only means of communication in the case of natural disasters such as hurricanes or earthquakes. In the worst cases, only decentralized, wireless ad-hoc networks survive.

Business: Managers can use mobile computers say, critical presentations to major customers. They can access the latest market share information. At a small recess, they can revise the presentation to take advantage of this information. They can communicate with the office about possible new offers and call meetings for discussing responds to the new proposals. Therefore, mobile computers can leverage competitive advantages. A travelling salesman today needs instant access to the company’s database: to ensure that files on his or her laptop reflect the current situation, to enable the company to keep track of all activities of their travelling employees, to keep databases consistent etc. With wireless access, the laptop can be turned into a true mobile office, but efficient and powerful synchronization mechanisms are needed to ensure data consistency.

d.**Credit Card Verification**: At Point of Sale (POS) terminals in shops and supermarkets, when customers use credit cards for transactions, the intercommunication required between the bank central computer and the POS terminal, in order to effect verification of the card usage, can take place quickly and securely over cellular channels using a

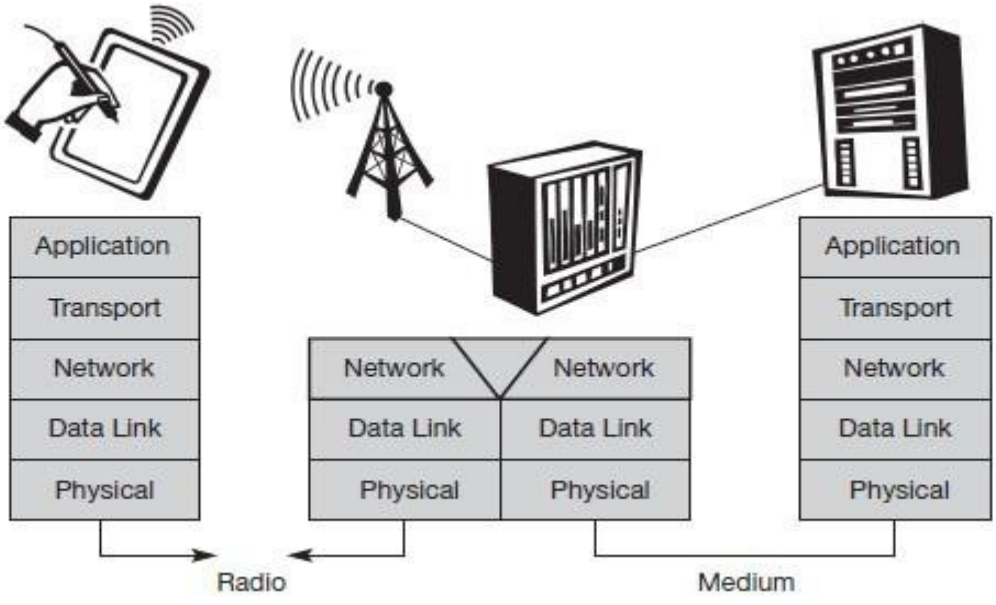
mobile computer unit. This can speed up the transaction process and relieve congestion at the POS terminals.

e.**Replacement of Wired Networks**: wireless networks can also be used to replace wired networks, e.g., remote sensors, for tradeshows, or in historic buildings. Due to economic reasons, it is often impossible to wire remote sensors for weather forecasts, earthquake detection, or to provide environmental information. Wireless connections, e.g., via satellite, can help in this situation. Other examples for wireless networks are computers, sensors, or information displays in historical buildings, where excess cabling may destroy valuable walls or floors.

f.**Entertainment**: wireless networks can provide up-to-date information at any appropriate location. The travel guide might tell you something about the history of abuilding (knowing via GPS, contact to a local base station, or triangulation where you are) downloading information about a concert in the building at the same evening via a local wireless network. Another growing field of wireless network applications lies in entertainment and games to enable, e.g., ad-hoc gaming networks as soon as people meet to play together.

A simplified reference model

The figure shows the **protocol stack** implemented in the system according to the reference model. **End-systems**, such as the PDA and computer in the example, need a full protocol stack comprising the application layer, transport layer, network layer, data link layer, and physical layer. Applications on the end-systems communicate with each other using the lower layer services. **Intermediate systems**, such as the interworking unit, do not necessarily need all of the layers.



A Simplified Reference Model

●**Physical layer**: This is the lowest layer in a communication system and is responsible for the conversion of a stream of bits into signals that can be transmitted on the sender side. The physical layer of the receiver then transforms the signals back into a bit stream. For wireless communication, the physical layer is responsible for frequency selection, generation of the carrier frequency, signal detection (although heavy interference may disturb the signal), modulation of data onto a carrier frequency and (depending on the transmission scheme) encryption.

●**Data link layer**: The main tasks of this layer include accessing the medium, multiplexing of different data streams, correction of transmission errors, and synchronization (i.e., detection of a data frame). Altogether, the data link layer is responsible for a reliable point-to-point connection between two devices or a point-to-multipoint connection between one sender and several receivers.

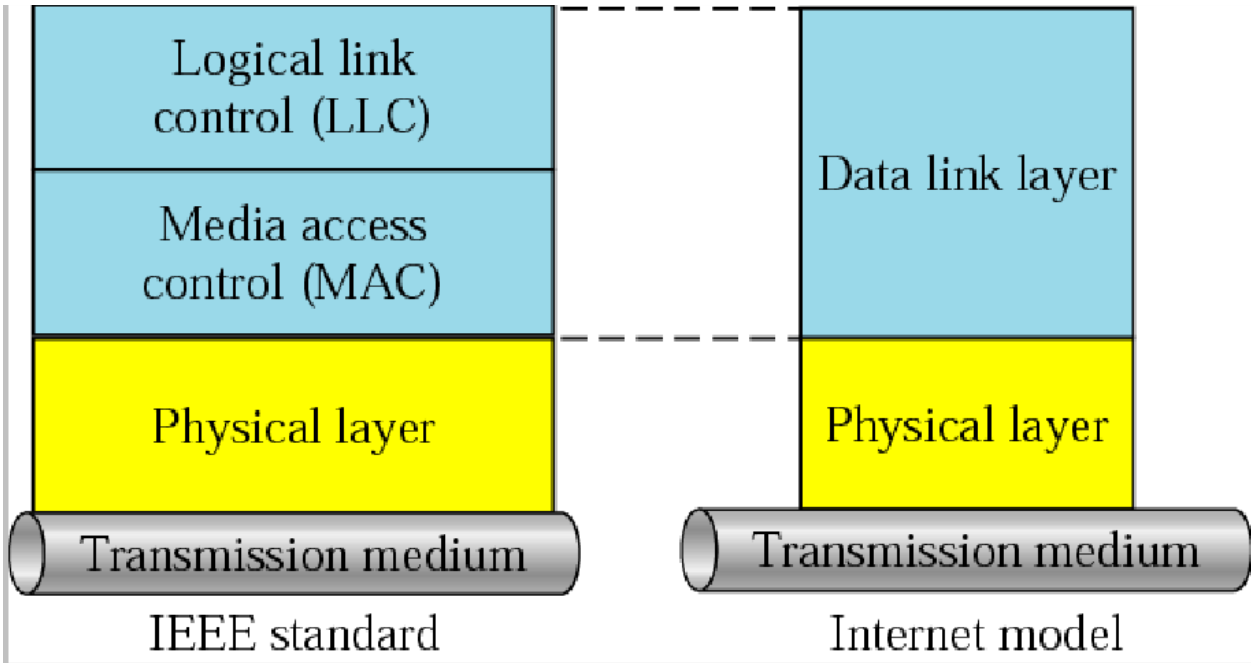
● **Network layer**: This third layer is responsible for routing packets through a network or establishing a connection between two entities over many other intermediate systems. Important functions are addressing, routing, device location, and handover between different networks.

● **Transport layer**: This layer is used in the reference model to establish an end-to-end connection

● **Application layer**: Finally, the applications (complemented by additional layers that can support applications) are situated on top of all transmission oriented layers. Functions are service location, support for multimedia applications, adaptive applications that can handle the large variations in transmission characteristics, and wireless access to the world-wide web using a portable device.

Chapter-2

The **Media Access Control** (**MAC**) data communication protocol sub-layer, also known as the Medium Access Control, is a sublayer of the Data Link Layer specified in the seven-layer OSI model (layer 2). The hardware that implements the MAC is referred to as a **Medium Access Controller**. The MAC sub-layer acts as an interface between the Logical Link Control (LLC) sublayer and the network's physical layer. The MAC layer emulates a full-duplex logical communication channel in a multi-point network. This channel may provide unicast, multicast or broadcast communication service.



LLC and MAC Sub Layers

Motivation for a specialized MAC

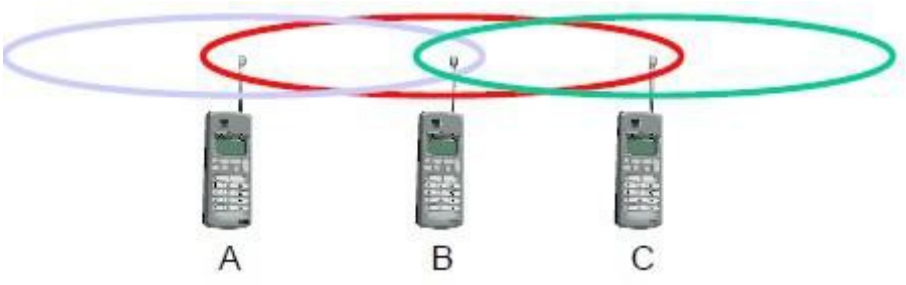
One of the most commonly used MAC schemes for wired networks is carrier sense multiple access with collision detection (CSMA/CD). In this scheme, a sender senses the medium (a wire or coaxial cable) to see if it is free. If the medium is busy, the sender waits until it is free. If the medium is free, the sender starts transmitting data and continues to listen into the medium. If the sender detects a collision while sending, it stops at once and sends a jamming signal. But this scheme doesn't work well with wireless networks. The problems are:

- Signal strength decreases proportional to the square of the distance
- The sender would apply CS and CD, but the collisions happen at the receiver
- It might be a case that a sender cannot "hear" the collision, i.e., CD does not work
- Furthermore, CS might not work, if for e.g., a terminal is "hidden"

Hidden and Exposed Terminals

Consider the scenario with three mobile phones as shown below. The transmission range of A reaches B, but not C (the detection range does not

reach C either). The transmission range of C reaches B, but not A. Finally, the transmission range of B reaches A and C, i.e., A cannot detect C and vice versa



Hidden terminals

 A sends to B, C cannot hear A

 C wants to send to B, C senses a “free” medium (CS fails) and starts transmitting

 Collision at B occurs, A cannot detect this collision (CD fails) and continues with its transmission to B

 A is “hidden” from C and vice versa

Exposed terminals

 B sends to A, C wants to send to another terminal (not A or B) outside the range

 C senses the carrier and detects that the carrier is busy.

 C postpones its transmission until it detects the medium as being idle again

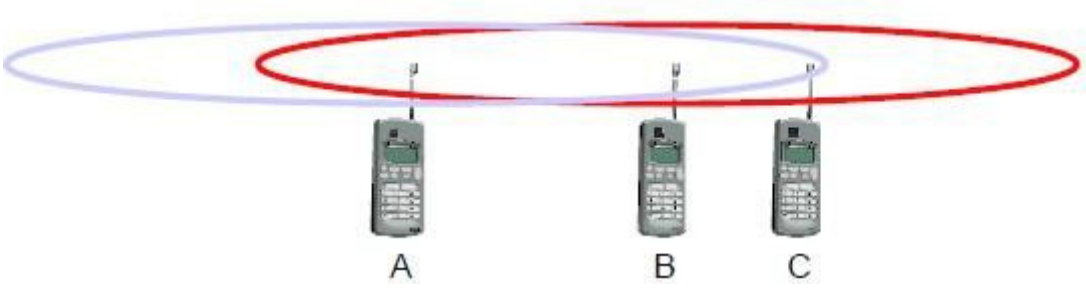
 but A is outside radio range of C, waiting is **not** necessary

 C is “exposed” to B

Hidden terminals cause collisions, where as Exposed terminals causes unnecessary delay.

Near and far terminals

Consider the situation shown below. A and B are both sending with the same transmission power.



Signal strength decreases proportional to the square of the distance

 So, B's signal drowns out A's signal making C unable to receive A's transmission

 If C is an arbiter for sending rights, B drown out A's signal on the physical layer making C unable to hear out A.

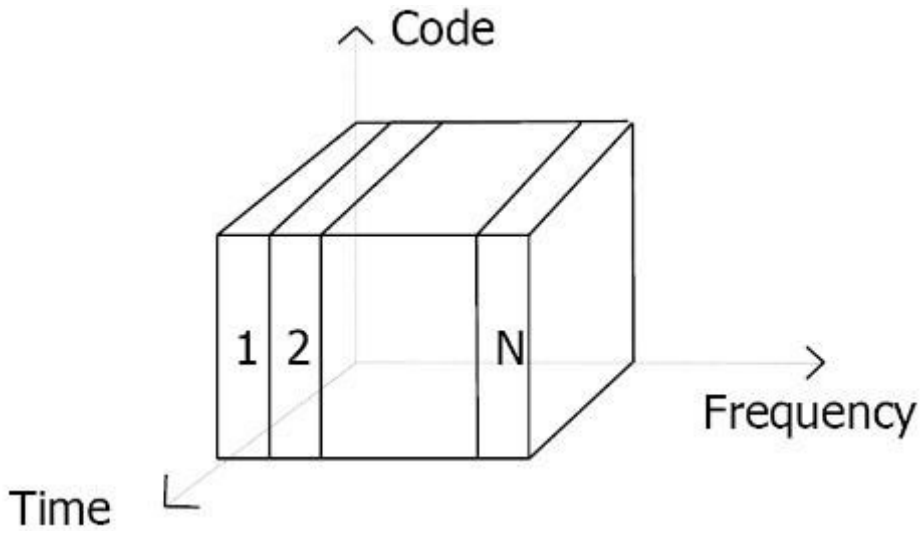
The **near/far effect** is a severe problem of wireless networks using CDM. All signals should arrive at the receiver with more or less the same strength for which Precise power control is to be implemented.

SDMA

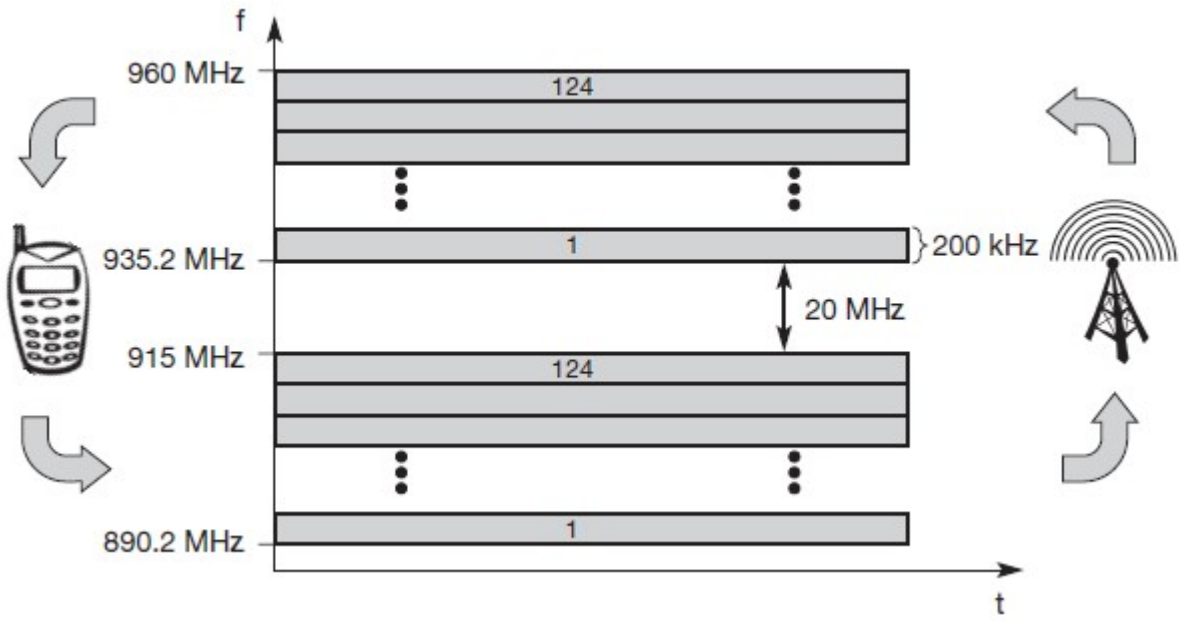
Space Division Multiple Access (SDMA) is used for allocating a separated space to users in wireless networks. A typical application involves assigning an optimal base station to a mobile phone user. The mobile phone may receive several base stations with different quality. A MAC algorithm could now decide which base station is best, taking into account which frequencies (FDM), time slots (TDM) or code (CDM) are still available. The basis for the SDMA algorithm is formed by cells and sectorized antennas which constitute the infrastructure implementing **space division multiplexing (SDM).** SDM has the unique advantage of not requiring any multiplexing equipment. It is usually combined with other multiplexing techniques to better utilize the individual physical channels.

FDMA

Frequency division multiplexing (FDM) describes schemes to subdivide the frequency dimension into several non-overlapping frequency bands.



Frequency Division Multiple Access is a method employed to permit several users to transmit simultaneously on one satellite transponder by assigning a specific frequency within the channel to each user. Each conversation gets its own, unique, radio channel. The channels are relatively narrow, usually 30 KHz or less and are defined as either transmit or receive channels. A full duplex conversation requires a transmit & receive channel pair. FDM is often used for simultaneous access to the medium by base station and mobile station in cellular networks establishing a duplex channel. A scheme called **frequency division duplexing (FDD)** in which the two directions, mobile station to base station and vice versa are now separated using different frequencies.



FDM for multiple access and duplex

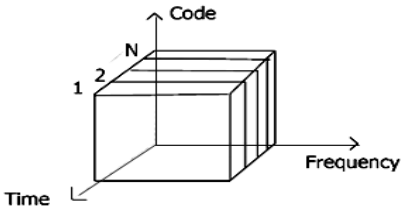
The two frequencies are also known as **uplink**, i.e., from mobile station to base station or from ground control to satellite, and as **downlink**, i.e., from base station to mobile station or from satellite to ground control. The basic frequency allocation scheme for GSM is fixed and regulated by national authorities. All uplinks use the band between 890.2 and 915 MHz, all downlinks use 935.2 to 960 MHz. According to FDMA, the base station, shown on the right side, allocates a certain frequency for up- and downlink to establish a duplex channel with a mobile phone. Up- and downlink have a fixed relation. If the uplink frequency is fu = 890 MHz + n·0.2 MHz, the downlink frequency is fd = fu + 45 MHz,
 ., fd = 935 MHz + n·0.2 MHz for a certain channel n. The base station selects the channel. Each channel (uplink and downlink) has a bandwidth of 200 kHz.

This scheme also has disadvantages. While radio stations broadcast 24 hours a day, mobile communication typically takes place for only a few minutes at a time. Assigning a separate frequency for each possible communication scenario would be a tremendous waste of (scarce) frequency resources. Additionally, the fixed assignment of a frequency to a sender makes the scheme very inflexible and limits the number of senders.

TDMA

A more flexible multiplexing scheme for typical mobile communications is time division multiplexing (TDM). Compared to FDMA, time division

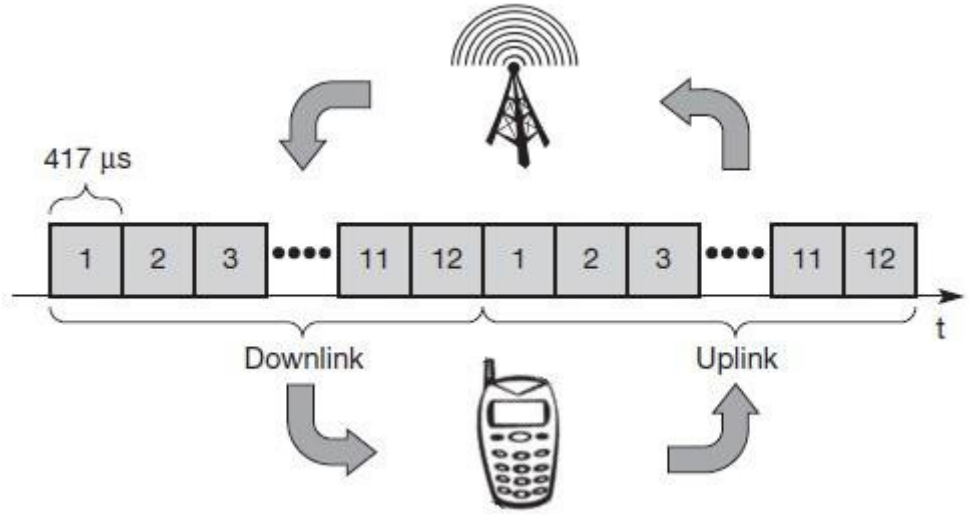
multiple access (TDMA) offers a much more flexible scheme, which comprises all technologies that allocate certain time slots for communication. Now synchronization between sender and receiver has to be achieved in the time domain. Again this can be done by using a fixed pattern similar to FDMA techniques, i.e., allocating a certain time slot for a channel, or by using a dynamic allocation scheme.



Listening to different frequencies at the same time is quite difficult, but listening to many channels separated in time at the same frequency is simple. Fixed schemes do not need identification, but are not as flexible considering varying bandwidth requirements.

FixedTDM

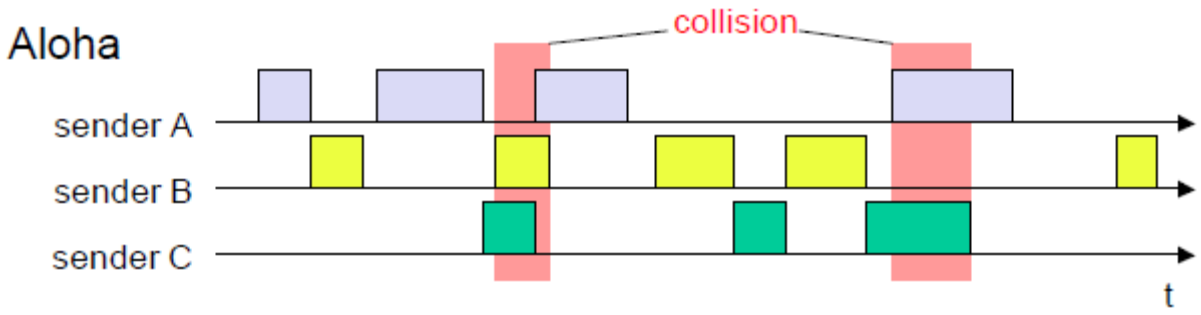
The simplest algorithm for using TDM is allocating time slots for channels in a fixed pattern. This results in a fixed bandwidth and is the typical solution for wireless phone systems. MAC is quite simple, as the only crucial factor is accessing the reserved time slot at the right moment. If this synchronization is assured, each mobile station knows its turn and no interference will happen. The fixed pattern can be assigned by the base station, where competition between different mobile stations that want to access the medium is solved



The above figure shows how these fixed TDM patterns are used to implement multiple access and a duplex channel between a base station and mobile station. Assigning different slots for uplink and downlink using the same frequency is called **time division duplex (TDD)**. As shown in the figure, the base station uses one out of 12 slots for the downlink, whereas the mobile station uses one out of 12 different slots for the uplink. Uplink and downlink are separated in time. Up to 12 different mobile stations can use the same frequency without interference using this scheme. Each connection is allotted its own up- and downlink pair. This general scheme still wastes a lot of bandwidth. It is too static, too inflexible for data communication. In this case, connectionless, demand-oriented TDMA schemes can be used

Classical Aloha

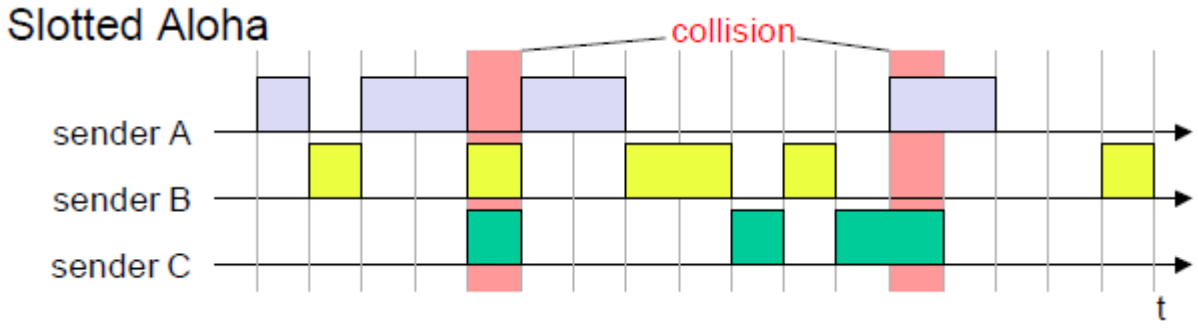
In this scheme, TDM is applied without controlling medium access. Here each station can access the medium at any time as shown below



This is a random access scheme, without a central arbiter controlling access and without coordination among the stations. If two or more stations access the medium at the same time, a **collision** occurs and the transmitted data is destroyed. Resolving this problem is left to higher layers (e.g., retransmission of data). The simple Aloha works fine for a light load and does not require any complicated access mechanisms.

Slotted Aloha

The first refinement of the classical Aloha scheme is provided by the introduction of time slots (**slotted Aloha**). In this case, all senders have to be **synchronized**, transmission can only start at the beginning of a **time slot** as shown below.



The introduction of slots raises the throughput from 18 per cent to 36 per cent, i.e., slotting doubles the throughput. Both basic Aloha principles occur in many systems that implement distributed access to a medium. Aloha systems work perfectly well under a light load, but they cannot give any hard transmission guarantees, such as maximum delay before accessing the medium or minimum throughput.

Carrier sense multiple access

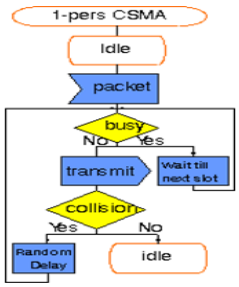
One improvement to the basic Aloha is sensing the carrier before accessing the medium. Sensing the carrier and accessing the medium only if the carrier is idle decreases the probability of a collision. But, as already mentioned in the introduction, hidden terminals cannot be detected, so, if a hidden terminal transmits at the same time as another sender, a collision might occur at the receiver. This basic scheme is still used in most wireless LANs. The different versions of CSMA are:

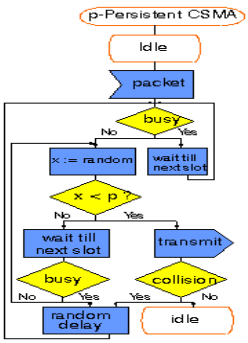
 1-persistent CSMA: Stations sense the channel and listens if its busy and transmit immediately, when the channel becomes idle. It's called 1-persistent CSMA because the host transmits with a probability of 1 whenever it finds the channel idle.

 non-persistent CSMA: stations sense the carrier and start sending immediately if the medium is idle. If the medium is busy, the station pauses a random amount of time before sensing the medium again and repeating this pattern.

 p-persistent CSMA: systems nodes also sense the medium, but only transmit with a probability of p, with the station deferring to the next slot with the probability 1-p, i.e., access is slotted in addition

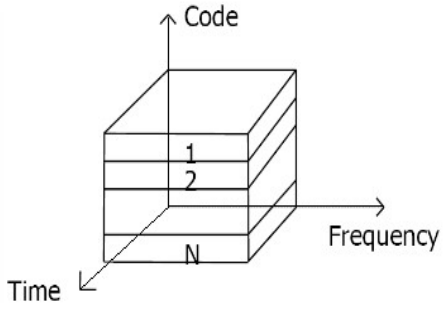
CSMA with collision avoidance (**CSMA/CA**) is one of the access schemes used in wireless LANs following the standard IEEE 802.11. Here sensing the carrier is combined with a back-off scheme in case of a busy medium to achieve some fairness among competing stations.





CDMA

Code division multiple access systems apply codes with certain characteristics to the transmission to separate different users in code space and to enable access to a shared medium without interference



All terminals send on the same frequency probably at the same time and can use the whole bandwidth of the transmission channel. Each sender has a unique random number, the sender XORs the signal with this random number. The receiver can “tune” into this signal if it knows the pseudo random number, tuning is done via a correlation function

Disadvantages:

 higher complexity of a receiver (receiver cannot just listen into the medium and start receiving if there is a signal)

 all signals should have the same strength at a receiver Advantages:

 all terminals can use the same frequency, no planning needed

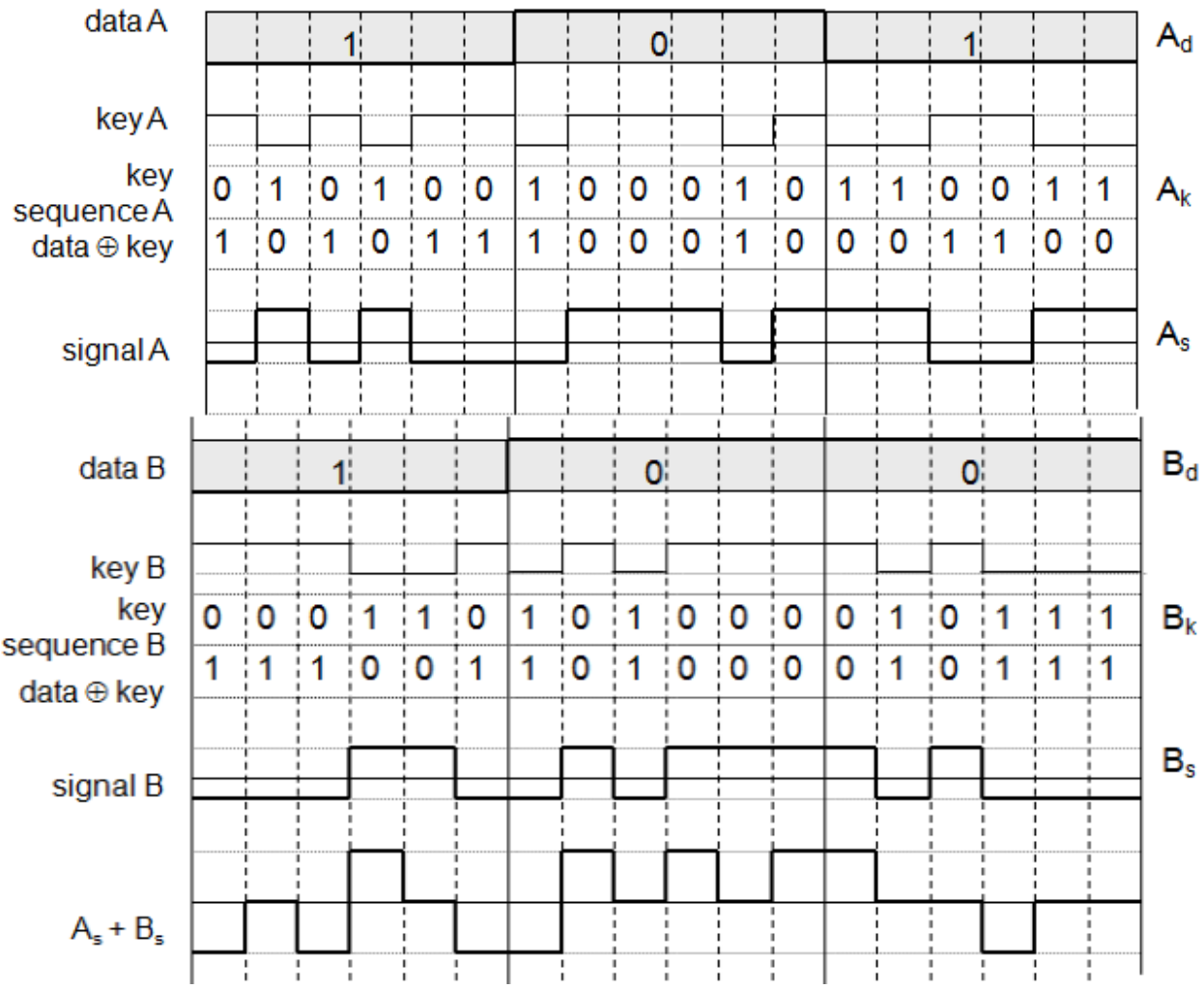
 huge code space (e.g. 232) compared to frequency space

 interferences (e.g. white noise) is not coded

 forward error correction and encryption can be easily integrated

- Sender A
  - sends A_d = 1, key A_k = 010011 (assign: "0"= -1, "1"= +1)
  - sending signal $A_s = A_d * A_k$ = (-1, +1, -1, -1, +1, +1)
- Sender B
  - sends B_d = 0, key B_k = 110101 (assign: "0"= -1, "1"= +1)
  - sending signal $B_s = B_d * B_k$ = (-1, -1, +1, -1, +1, -1)
- Both signals superimpose in space
  - interference neglected (noise etc.)
  - $A_s + B_s$ = (-2, 0, 0, -2, +2, 0)
- Receiver wants to receive signal from sender A
  - apply key A_k bitwise (inner product)
    - A_e = (-2, 0, 0, -2, +2, 0) • A_k = 2 + 0 + 0 + 2 + 2 + 0 = 6
    - result greater than 0, therefore, original bit was "1"
  - receiving B
    - B_e = (-2, 0, 0, -2, +2, 0) • B_k = -2 + 0 + 0 - 2 - 2 + 0 = -6, i.e. "0"

The following figure shows a sender A that wants to transmit the bits 101. The key of A is shown as signal and binary sequence Ak. The binary "0" is assigned a positive signal value, the binary "1" a negative signal value. After spreading, i.e., XORing Ad and Ak, the resulting signal is As.



Coding and spreading of data from sender A and sender B

The same happens with data from sender B with bits 100. The result is Bs. As and Bs now superimpose during transmission. The resulting signal is simply the sum As + Bs as shown above. A now tries to reconstruct the original data from Ad. The receiver applies A's key, Ak, to the received signal and feeds the result into an integrator. The integrator adds the products, a comparator then has to decide if the result is a 0 or a 1 as shown below. As clearly seen, although the original signal form is distorted by B's signal, the result is quite clear. The same happens if a receiver wants to receive B's data.

DHCP

Need for Mobile IP

The IP addresses are designed to work with stationary hosts because part of the address defines the network to which the host is attached**.** A host cannot change its IP address without terminating on-going sessions and restarting them after it acquires a new address. Other link layer mobility solutions exist but are not sufficient enough for the global Internet.

Mobility is the ability of a node to change its point-of-attachment while maintaining all existing communications and using the same IP address.

Nomadicity allows a node to move but it must terminate all existing communications and then can initiate new connections with a new address.

Mobile IP is a network layer solution for homogenous and heterogeneous mobility on the global Internet which is scalable, robust, secure and which allows nodes to maintain all ongoing communications while moving.

Design Goals: Mobile IP was developed as a means for transparently dealing with problems of mobile users. Mobile IP was designed to make the size and the frequency of required routing updates as small as possible. It was designed to make it simple to implement mobile node software. It was designed to avoid solutions that require mobile nodes to use multiple addresses.

Requirements: There are several requirements for Mobile IP to make it as a standard. Some of them are:

1. *Compatibility*: The whole architecture of internet is very huge and a new standard cannot introduce changes to the applications or network protocols already in use. Mobile IP is to be integrated into the existing operating systems. Also, for routers also it may be possible to enhance its capabilities to support mobility instead of changing the routers which is highly impossible. Mobile IP must not require special media or MAC/LLC protocols, so it must use the same interfaces and mechanisms to access the lower layers as IP does. Finally, end-systems enhanced with a mobile IP

implementation should still be able to communicate with fixed systems without mobile IP.

2. *Transparency*: Mobility remains invisible for many higher layer protocols and applications. Higher layers continue to work even if the mobile computer has changed its point of attachment to the network and even notice a lower bandwidth and some interruption in the service. As many of today's applications have not been designed to use in mobile environments, the effects of mobility will be higher delay and lower bandwidth.

Scalability and efficiency: The efficiency of the network should not be affected even if a new mechanism is introduced into the internet. Enhancing IP for mobility must not generate many new messages flooding the whole network. Special care is necessary to be taken considering the lower bandwidth of wireless links. Many mobile systems have a wireless link to an attachment point. Therefore, only some additional packets must be necessary between a mobile system and a node in the network. It is indispensable for a mobile IP to be scalable over a large number of participants in the whole internet, throughout the world.

4. *Security*: Mobility possesses many security problems. A minimum requirement is the authentication of all messages related to the management of mobile IP. It must be sure for the IP layer if it forwards a packet to a mobile host that this host really is the receiver of the packet. The IP layer can only guarantee that the IP address of the receiver is correct. There is no way to prevent faked IP addresses and other attacks
The goal of a mobile IP can be summarized as: 'supporting end-system mobility while maintaining scalability, efficiency, and compatibility in all respects with existing applications and Internet protocols'.

Entities and terminology

The following defines several entities and terms needed to understand mobile IP as defined in RFC 3344.

Mobile Node (MN): A mobile node is an end-system or router that can change its point of attachment to the internet using mobile IP. The MN keeps its IP address and can continuously communicate with any other system in the internet as long as link-layer connectivity is given. Examples are laptop, mobile phone, router on an aircraft etc.

Correspondent node (CN): At least one partner is needed for communication. In the following the CN represents this partner for the MN. The CN can be a fixed or mobile node.

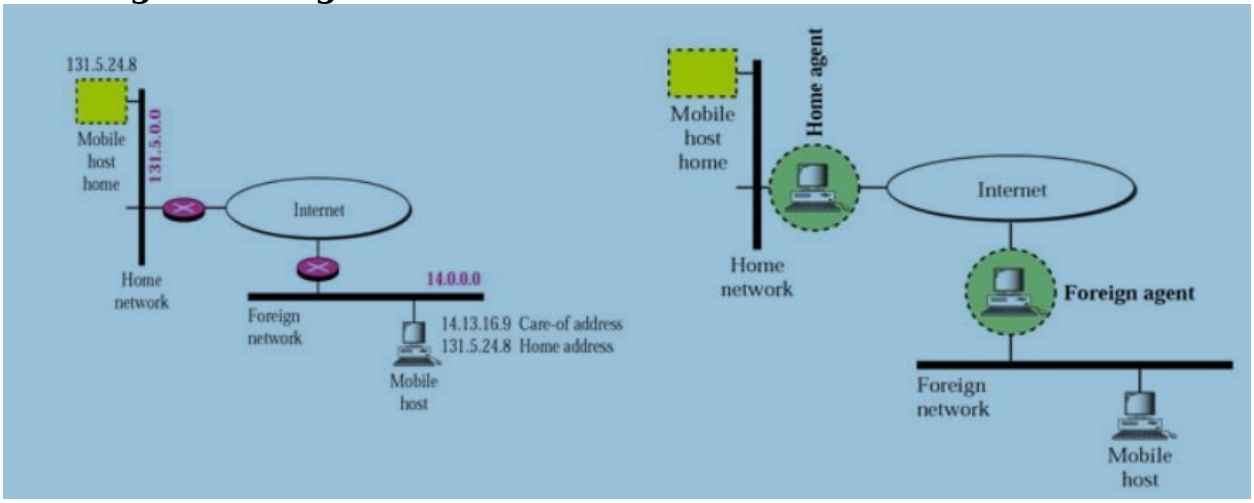
Home network: The home network is the subnet the MN belongs to with respect to its IP address. No mobile IP support is needed within the home network.

Foreign network: The foreign network is the current subnet the MN visits and which is not the home network.

Foreign agent (FA): The FA can provide several services to the MN during its visit to the foreign network. The FA can have the COA, acting as tunnel endpoint and forwarding packets to the MN. The FA can be the default router for the MN. FAs can also provide security services because they belong to the foreign network as opposed to the MN which is only visiting. FA is implemented on a router for the subnet the MN attaches to.

Care-of address (COA): The COA defines the current location of the MN from an IP point of view. All IP packets sent to the MN are delivered to the COA, not directly to the IP address of the MN. Packet delivery toward the MN is done using a tunnel, i.e., the COA marks the tunnel endpoint, i.e., the address where packets exit the tunnel. There are two different possibilities for the location of the COA:

1.Foreign agent COA 2.Co-located COA



Home agent (HA): The HA provides several services for the MN and is located in the home network. The tunnel for packets toward the MN starts at the HA. The HA maintains a location registry, i.e., it is informed of the MN's location by the current COA. Three alternatives for the implementation of an HA exist.

1. The HA can be implemented on a router that is responsible for the home network. This is obviously the best position, because without optimizations to mobile IP, all packets for the MN have to go through the router anyway.

2. If changing the router's software is not possible, the HA could also be implemented on an arbitrary node in the subnet. One disadvantage of this solution is the double

crossing of the router by the packet if the MN is in a foreign network. A packet for the MN comes in via the router; the HA sends it through the tunnel which again crosses the router.
3. Finally, a home network is not necessary at all. The HA could be again on the ‘router’ but this time only acting as a manager for MNs belonging to a virtual home network. All MNs are always in a foreign network with this solution.

Chapter-3

Traditional TCP

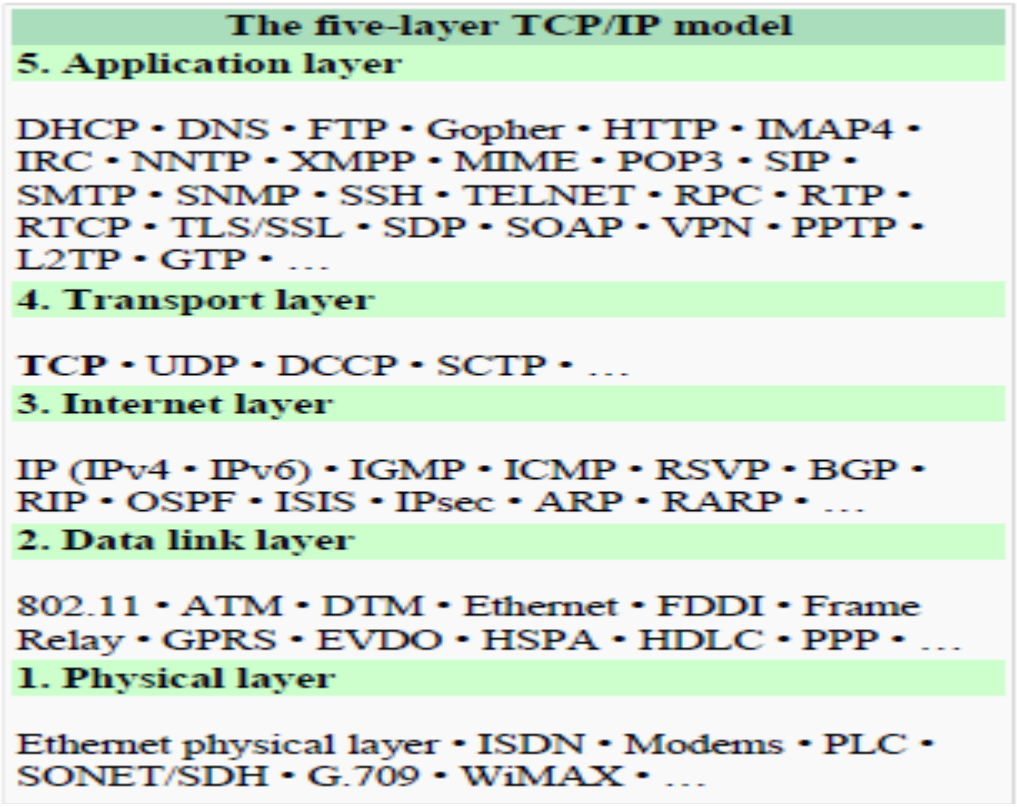
The **Transmission Control Protocol** (**TCP**) is one of the core protocols of the Internet protocol suite, often simply referred to as TCP/IP. TCP is reliable, guarantees in-order delivery of data and incorporates congestion control and flow control mechanisms.
TCP supports many of the Internet's most popular application protocols and resulting applications, including the World Wide Web, e-mail, File Transfer Protocol and Secure Shell. In the Internet protocol suite, TCP is the intermediate layer between the Internet layer and application layer.
The major responsibilities of TCP in an active session are to:
• **Provide reliable in-order transport of data**: to not allow losses of data.

• **Control congestions in the networks**: to not allow degradation of the network performance,

• **Control a packet flow between the transmitter and the receiver**: to not exceed the receiver's capacity.

TCP uses a number of mechanisms to achieve high performance and avoid 'congestion collapse', where network performance can fall by several orders

of magnitude. These mechanisms control the rate of data entering the network, keeping the data flow below a rate that would trigger collapse. There are several mechanisms of TCP that influence the efficiency of TCP in a mobile environment. Acknowledgments for data sent, or lack of acknowledgments, are used by senders to implicitly interpret network conditions between the TCP sender and receiver.



Congestion Control

A transport layer protocol such as TCP has been designed for fixed networks with fixed end- systems. Congestion may appear from time to time even in carefully designed networks. The packet buffers of a router are filled and the router cannot forward the packets fast enough because the sum of the input rates of packets destined for one output link is higher than the capacity of the output link. The only thing a router can do in this situation is to drop packets. A dropped packet is lost for the transmission, and the receiver notices a gap in the packet stream. Now the receiver does not directly tell the sender which packet is missing, but continues to acknowledge all in-sequence packets up to the missing one.

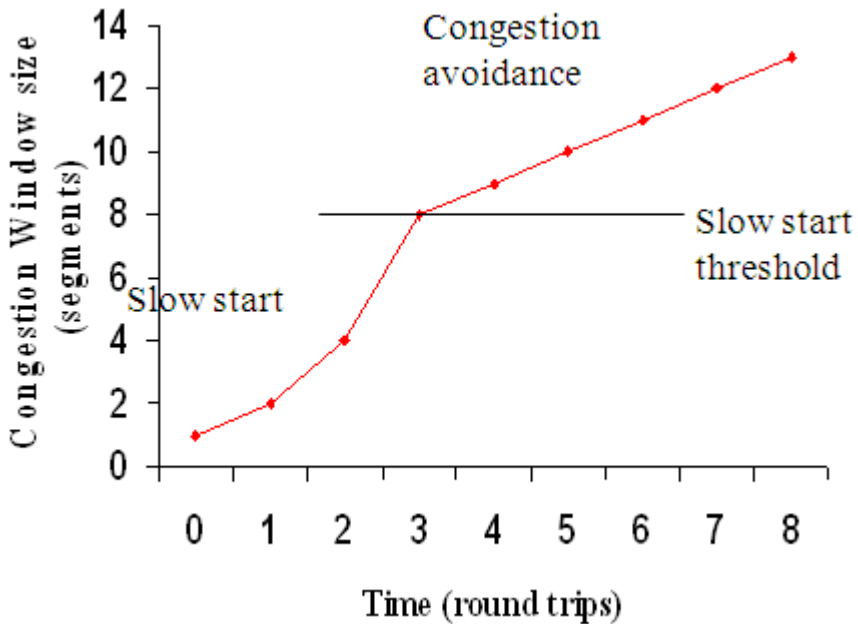
The sender notices the missing acknowledgement for the lost packet and assumes a packet loss due to congestion. Retransmitting the missing packet and continuing at full sending rate would now be unwise, as this might only increase the congestion. To mitigate congestion, TCP slows down the

transmission rate dramatically. All other TCP connections experiencing the same congestion do exactly the same so the congestion is soon resolved.

Slow start

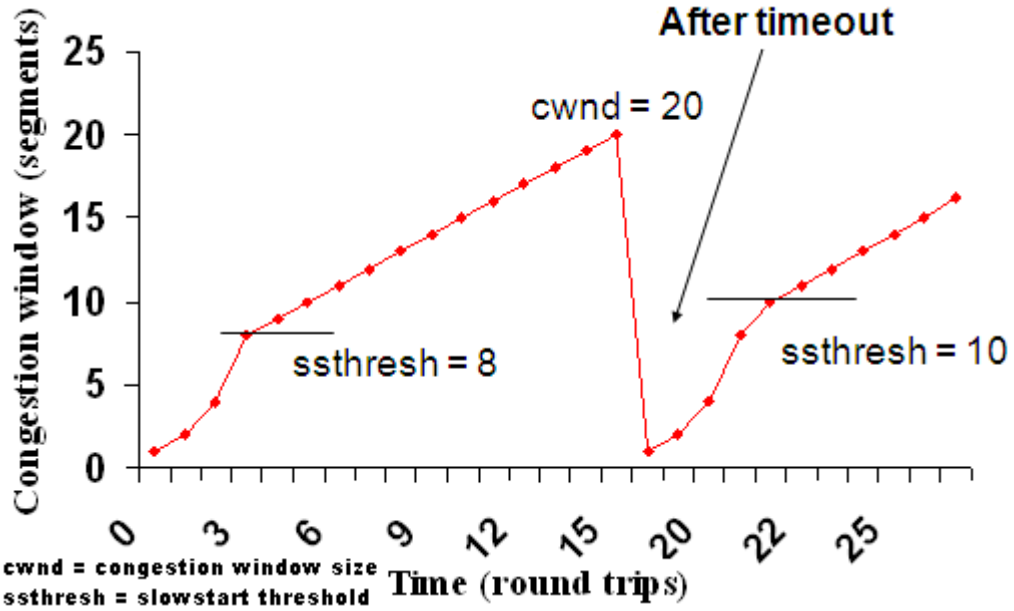
TCP's reaction to a missing acknowledgement is quite drastic, but it is necessary to get rid of congestion quickly. The behavior TCP shows after the detection of congestion is called **slow start.** The sender always calculates a **congestion window** for a receiver. The start size of the congestion window is one segment (TCP packet). The sender sends one packet and waits for acknowledgement. If this acknowledgement arrives, the sender increases the congestion window by one, now sending two packets (congestion window = 2). This scheme doubles the congestion window every time the acknowledgements come back, which takes one round trip time (RTT). This is called the exponential growth of the congestion window in the slow start
mechanism.

But doubling the congestion window is too dangerous. The exponential growth stops at the **congestion threshold**. As soon as the congestion window reaches the congestion threshold, further increase of the transmission rate is only linear by adding 1 to the congestion window each time the acknowledgements come back.



Linear increase continues until a time-out at the sender occurs due to a missing acknowledgement, or until the sender detects a gap in transmitted data because of continuous acknowledgements for the same packet. In

either case the sender sets the congestion threshold to half of the current congestion
window. The congestion window itself is set to one segment and the sender starts sending a single segment. The exponential growth starts once more up to the new congestion threshold, then the window grows in linear fashion



Fast retransmit/fast recovery

The congestion threshold can be reduced because of two reasons. First one is if the sender receives continuous acknowledgements for the same packet. It informs the sender that the receiver has got all the packets upto the acknowledged packet in the sequence and also the receiver is receiving something continuously from the sender. The gap in the packet stream is not due to congestion, but a simple packet loss due to a transmission error. The sender can now retransmit the missing packet(s) before the timer expires. This behavior is called **fast retransmit**. It is an early enhancement for preventing slow-start to trigger on losses not caused by congestion. The receipt of acknowledgements shows that there is no congestion to justify a slow start. The sender can continue with the current congestion window. The sender performs a **fast recovery** from the packet loss. This mechanism can improve the efficiency of TCP dramatically. The other reason for activating slow start is a time-out due to a missing acknowledgement. TCP using fast retransmit/fast recovery interprets this congestion in the network and activates the slow start mechanism.

The advantage of this method is its simplicity. Minor changes in the MH's software results in performance increase. No changes are required in FA or CH.

The disadvantage of this scheme is insufficient isolation of packet losses. It mainly focuses on problems regarding Handover. Also it effects the efficiency when a CH transmits already delivered packets.

Problems with Traditional TCP in wireless environments

Slow Start mechanism in fixed networks decreases the efficiency of TCP if used with mobile receivers or senders.

Error rates on wireless links are orders of magnitude higher compared to fixed fiber or copper links. This makes compensation for packet loss by TCP quite difficult.

Mobility itself can cause packet loss. There are many situations where a soft handover from one access point to another is not possible for a mobile end-system.

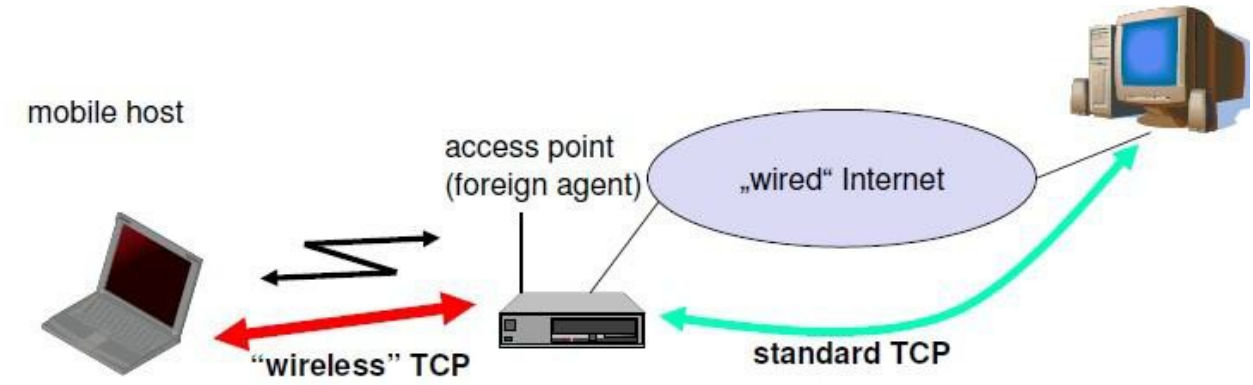
Standard TCP reacts with slow start if acknowledgements are missing, which does not help in the case of transmission errors over wireless links and which does not really help during handover. This behavior results in a severe performance degradation of an unchanged TCP if used together with wireless links or mobile nodes

Classical TCP Improvements

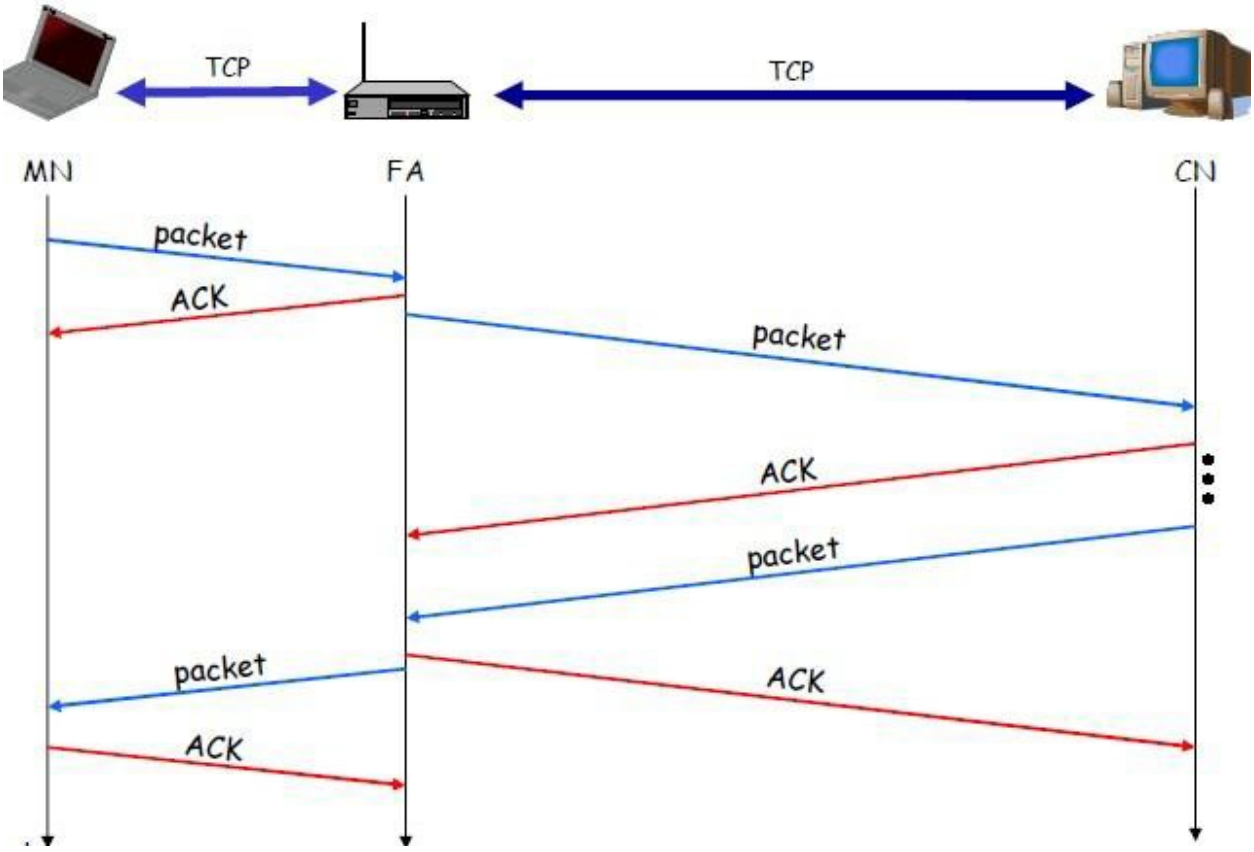
Indirect TCP (I-TCP)

Indirect TCP segments a TCP connection into a fixed part and a wireless part. The following figure shows an example with a mobile host connected via a wireless link and an access point to the ‘wired’ internet where the correspondent host resides.

Standard TCP is used between the fixed computer and the access point. No computer in the internet recognizes any changes to TCP. Instead of the mobile host, the access point now terminates the standard TCP connection, acting as a proxy. This means that the access point is now seen as the mobile host for the fixed host and as the fixed host for the mobile host. Between the access point and the mobile host, a special TCP, adapted to wireless links, is used. However, changing TCP for the wireless link is not a requirement. A suitable place for segmenting the connection is at the foreign agent as it not only controls the mobility of the mobile host anyway and can also hand over the connection to the next foreign agent when the mobile host moves on.



During handover, the buffered packets, as well as the system state (packet sequence number, acknowledgements, ports, etc), must migrate to the new agent. No new connection may be established for the mobile host, and the correspondent host must not see any changes in connection state. Packet delivery in I-TCP is shown below



Advantages of I-TCP

No changes in the fixed network necessary, no changes for the hosts (TCP protocol) necessary, all current optimizations to TCP still work

Simple to control, mobile TCP is used only for one hop between, e.g., a foreign agent and mobile host

1. transmission errors on the wireless link do not propagate into the fixed network

2. therefore, a very fast retransmission of packets is possible, the short delay on the mobile hop s known

It is always dangerous to introduce new mechanisms in a huge network without knowing exactly how they behave.

New optimizations can be tested at the last hop, without jeopardizing the stability of the Internet.

It is easy to use different protocols for wired and wireless networks.

Disadvantages of I-TCP

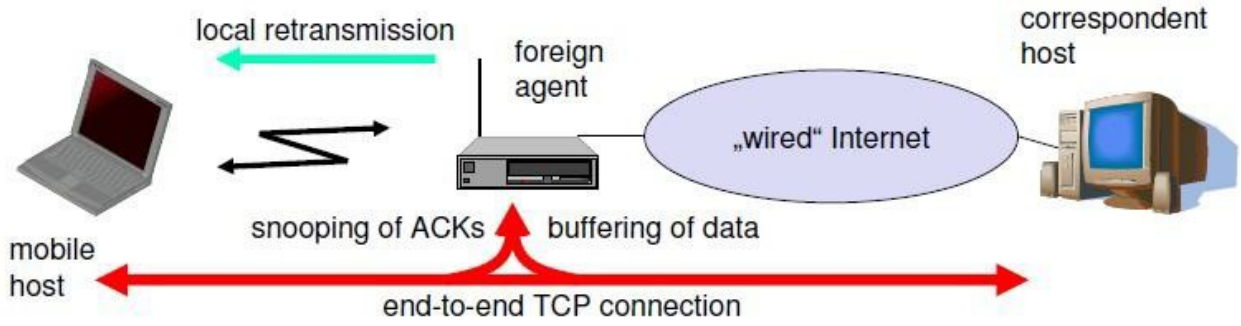
Loss of end-to-end semantics:- an acknowledgement to a sender no longer means that a receiver really has received a packet, foreign agents might crash.

Higher latency possible:- due to buffering of data within the foreign agent and forwarding to a new foreign agent

Security issue:- The foreign agent must be a trusted entity

Snooping TCP

The main drawback of I-TCP is the segmentation of the single TCP connection into two TCP connections, which loses the original end-to-end TCP semantic. A new enhancement, which leaves the TCP connection intact and is completely transparent, is Snooping TCP. The main function is to buffer data close to the mobile host to perform fast local retransmission in case of packet loss.



Snooping TCP as a transparent TCP extension

Here, the foreign agent buffers all packets with **destination mobile host** and additionally ‘snoops’ the packet flow in both directions to recognize

acknowledgements. The foreign agent buffers every packet until it receives an acknowledgement from the mobile host. If the FA does not receive an acknowledgement from the mobile host within a certain amount of time, either the packet or the acknowledgement has been lost. Alternatively, the foreign agent could receive a duplicate ACK which also shows the loss of a packet. Now, the FA retransmits the packet directly from the buffer thus performing a faster retransmission compared to the CH. For transparency, the FA does not acknowledge data to the CH, which would violate end-to-end semantic in case of a FA failure. The foreign agent can filter the duplicate acknowledgements to avoid unnecessary retransmissions of data from the correspondent host. If the foreign agent now crashes, the time-out of the correspondent host still works and triggers a retransmission. The foreign agent may discard duplicates of packets already retransmitted locally and acknowledged by the mobile host. This avoids unnecessary traffic on the wireless link.
For data transfer from the mobile host with **destination correspondent host**, the FA snoops into the packet stream to detect gaps in the sequence numbers of TCP. As soon as the foreign agent detects a missing packet, it returns a negative acknowledgement (NACK) to the mobile host. The mobile host can now retransmit the missing packet immediately. Reordering of packets is done automatically at the correspondent host by TCP.

Advantages of snooping TCP:

- The end-to-end TCP semantic is preserved.
- Most of the enhancements are done in the foreign agent itself which keeps correspondent host unchanged.
- Handover of state is not required as soon as the mobile host moves to another foreign agent. Even though packets are present in the buffer, time out at the CH occurs and the packets are transmitted to the new COA.
- No problem arises if the new foreign agent uses the enhancement or not. If not, the approach automatically falls back to the standard solution.

Disadvantages of snooping TCP

- Snooping TCP does not isolate the behavior of the wireless link as well as I-TCP. Transmission errors may propagate till CH.
- Using negative acknowledgements between the foreign agent and the mobile host assumes additional mechanisms on the mobile host. This approach is no longer transparent for arbitrary mobile hosts.

Mobile TCP

Both I-TCP and Snooping TCP does not help much, if a mobile host gets disconnected. The **M-TCP (mobile TCP)** approach has the same goals as I-TCP and snooping TCP: to prevent the sender window from shrinking if bit errors or disconnection but not congestion cause current problems. M-TCP

wants to improve overall throughput, to lower the delay, to maintain end-to-end semantics of TCP, and to provide a more efficient handover. Additionally, M-TCP is especially adapted to the problems arising from lengthy or frequent disconnections. M-TCP splits the TCP connection into two parts as I-TCP does. An unmodified TCP is used on the standard host-**supervisory host (SH)** connection, while an optimized TCP is used on the SH-MH connection.
The SH monitors all packets sent to the MH and ACKs returned from the MH. If the SH does not receive an ACK for some time, it assumes that the MH is disconnected. It then chokes the sender by setting the sender's window size to 0. Setting the window size to 0 forces the sender to go into **persistent mode**, i.e., the state of the sender will not change no matter how long the receiver is disconnected. This means that the sender will not try to retransmit data. As soon as the SH (either the old SH or a new SH) detects connectivity again, it reopens the window of the sender to the old value. The sender can continue sending at full speed. This mechanism does not require changes to the sender's TCP. The wireless side uses an adapted

TCP that can recover from packet loss much faster. This modified TCP does not use slow start, thus, M-TCP needs a **bandwidth manager** to implement fair sharing over the wireless link.
Advantages of M-TCP:
- It maintains the TCP end-to-end semantics. The SH does not send any ACK itself but forwards the ACKs from the MH.
- If the MH is disconnected, it avoids useless retransmissions, slow starts or breaking connections by simply shrinking the sender's window to 0.
- As no buffering is done as in I-TCP, there is no need to forward buffers to a new SH. Lost packets will be automatically retransmitted to the SH.

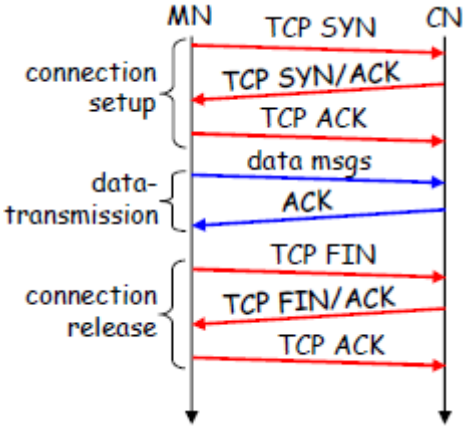
Disadvantages of M-TCP:
- As the SH does not act as proxy as in I-TCP, packet loss on the wireless link due to bit errors is propagated to the sender. M-TCP assumes low bit error rates, which is not always a valid assumption.
- A modified TCP on the wireless link not only requires modifications to the MH protocol software but also new network elements like the bandwidth manager.

Transmission/time-out freezing
Often, MAC layer notices connection problems even before the connection is actually interrupted from a TCP point of view and also knows the real reason for the interruption. The MAC layer can inform the TCP layer of an upcoming loss of connection or that the current interruption is not caused by congestion. TCP can now stop sending and 'freezes' the current state of its congestion window and further timers. If the MAC layer notices the upcoming interruption early enough, both the mobile and correspondent

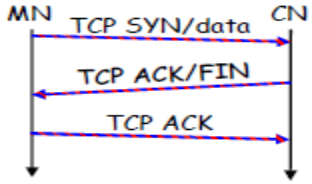
host can be informed. With a fast interruption of the wireless link, additional mechanisms in the access point are needed to inform the correspondent host of the reason for interruption. Otherwise, the correspondent host goes into slow start assuming congestion and finally breaks the connection.
As soon as the MAC layer detects connectivity again, it signals TCP that it can resume operation at exactly the same point where it had been forced to stop. For TCP time simply does not advance, so no timers expire.

Selective retransmission
A very useful extension of TCP is the use of selective retransmission. TCP acknowledgements are cumulative, i.e., they acknowledge in-order receipt of packets up to a certain packet. A single acknowledgement confirms reception of all packets upto a certain packet. If a single packet is lost, the sender has to retransmit everything starting from the lost packet (go-back-n retransmission). This obviously wastes bandwidth, not just in the case of a mobile network, but for any network.
Using selective retransmission, TCP can indirectly request a selective retransmission of packets. The receiver can acknowledge single packets, not only trains of in-sequence packets. The sender can now determine precisely which packet is needed and can retransmit it. The **advantage** of this approach is obvious: a sender retransmits only the lost packets. This lowers bandwidth requirements and is extremely helpful in slow wireless links. The disadvantage is that a more complex software on the receiver side is needed. Also more buffer space is needed to resequence data and to wait for gaps to be filled.

Transaction-oriented TCP
Assume an application running on the mobile host that sends a short request to a server from time to time, which responds with a short message and it requires reliable TCP transport of the packets. For it to use normal TCP, it is inefficient because of the overhead involved. Standard TCP is made up of three phases: setup, data transfer and release.



First, TCP uses a three-way handshake to establish the connection. At least one additional packet is usually needed for transmission of the request, and requires three more packets to close the connection via a three-way handshake. So, for sending one data packet, TCP may need seven packets altogether. This kind of overhead is acceptable for long sessions in fixed networks, but is quite inefficient for short messages or sessions in wireless networks. This led to the development of transaction-oriented TCP (T/TCP).
T/TCP can combine packets for connection establishment and connection release with user data packets. This can reduce the number of packets down to two instead of seven. The obvious **advantage** for certain applications is the reduction in the overhead which standard TCP has for connection setup and connection release. Disadvantage is that it requires changes in the software in mobile host.



Chapter-4

Data Dissemination

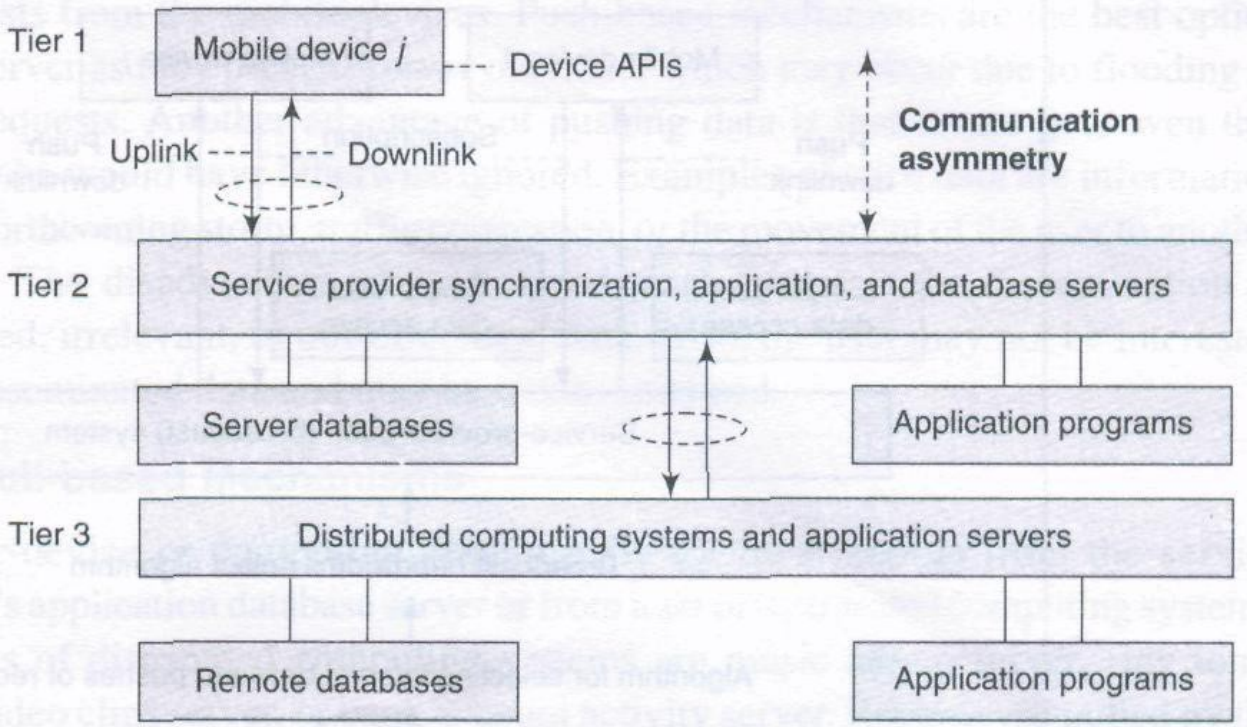
Ongoing advances in communications including the proliferation of internet, development of mobile and wireless networks, high bandwidth availability to homes have led to development of a wide range of new-information centered applications. Many of these applications involve **data**

dissemination, i.e. *delivery of data from a set of producers to a larger set of consumers*.
Data dissemination entails distributing and pushing data generated by a set of computing systems or broadcasting data from audio, video, and data services. The output data is sent to the mobile devices. A mobile device can select, tune and cache the required data items, which can be used for application programs.
Efficient utilization of wireless bandwidth and battery power are two of the most important problems facing software designed for mobile computing. Broadcast channels are attractive in tackling these two problems in wireless data dissemination. Data disseminated through broadcast channels can be simultaneously accessed by an arbitrary number of mobile users, thus increasing the efficiency of bandwidth usage.

Communications Asymmetry

One key aspect of dissemination-based applications is their inherent communications asymmetry. That is, the communication capacity or data volume in the downstream direction (from servers-to-clients) is much greater than that in the upstream direction (from clients-to- servers). Content delivery is an asymmetric process regardless of whether it is performed over a symmetric channel such as the internet or over an asymmetric one, such as cable television (CATV) network. Techniques and system architectures that can efficiently support asymmetric applications will therefore be a requirement for future use.
Mobile communication between a mobile device and a static computer system is intrinsically asymmetric. A device is allocated a limited bandwidth. This is because a large number of devices access the network. Bandwidth in the downstream from the server to the device is much larger than the one in the upstream from the device to the server. This is because mobile devices have limited power resources and also due to the fact that faster data transmission rates for long intervals of time need greater power dissipation from the devices. In GSM networks data transmission rates go up to a maximum of 14.4 kbps for both uplink and downlink. The communication is symmetric and this symmetry can be maintained because GSM is only used for voice communication.

The below figure shows communication asymmetry in uplink and downlink in a mobile network. The participation of device APIs and distributed computing systems in the running of an application is also shown



Communication asymmetry in uplink and downlink and participation of device APIs and distributed computing systems when an application runs

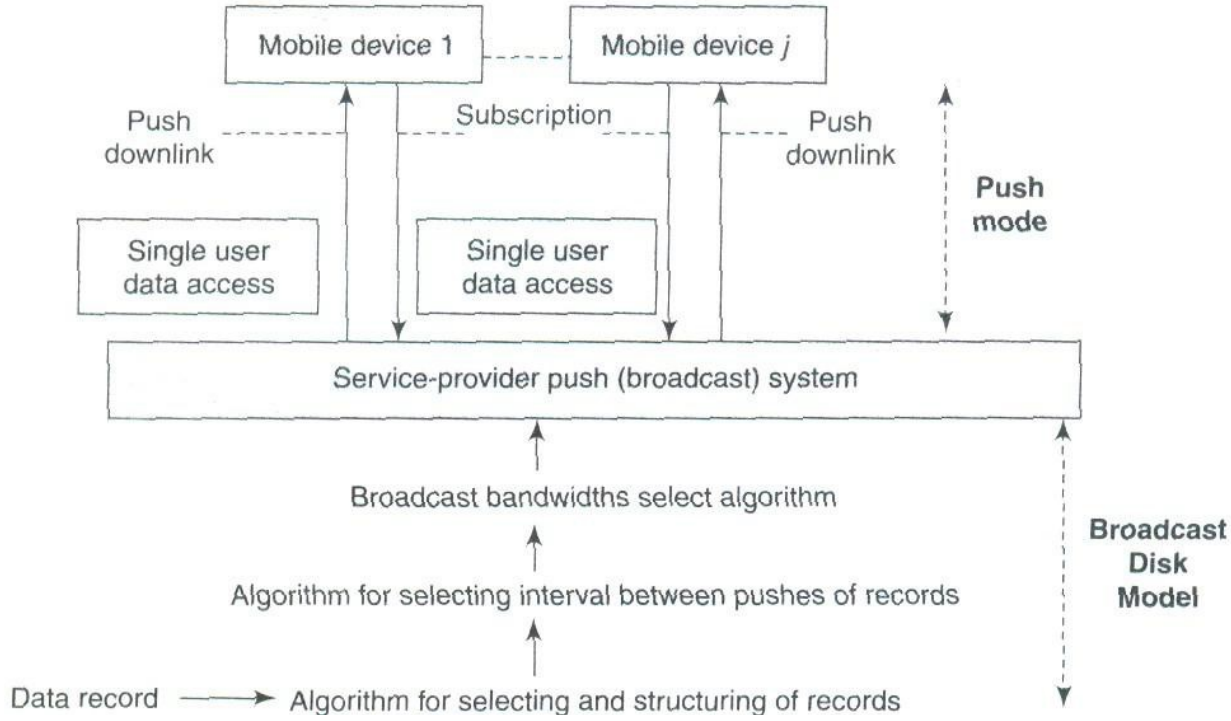
Classification of Data-Delivery Mechanisms

There are two fundamental information delivery methods for wireless data applications: Point-to-Point access and Broadcast. Compared with Point-to-Point access, broadcast is a more attractive method. A single broadcast of a data item can satisfy all the outstanding requests for that item simultaneously. As such, broadcast can scale up to an arbitrary number of users. There are three kinds of broadcast models, namely *push-based* broadcast, *On-demand* (or *pull- based*) broadcast, and *hybrid* broadcast. In push based broadcast, the server disseminates information using a periodic/aperiodic broadcast program (generally without any intervention of clients). In on demand broadcast, the server disseminates information based on the outstanding requests submitted by clients; In hybrid broadcast, push based broadcast and on demand data deliveries are combined to complement each other. In addition, mobile computers consume less battery power on monitoring broadcast channels to receive data than accessing data through point-to-point communications.

Data-delivery mechanisms can be classified into three categories, namely, push-based mechanisms (publish-subscribe mode), pull-based mechanisms (on-demand mode), and hybrid mechanisms (hybrid mode).

Push-based Mechanisms

The server pushes data records from a set of distributed computing systems. Examples are advertisers or generators of traffic congestion, weather reports, stock quotes, and news reports. The following figure shows a push-based data-delivery mechanism in which a server or computing system pushes the data records from a set of distributed computing systems. The data records are pushed to mobile devices by broadcasting without any demand. The push mode is also known as ***publish-subscribe*** mode in which the data is pushed as per the subscription for a push service by a user. The subscribed query for a data record is taken as perpetual query till the user unsubscribe to that service. Data can also be pushed without user subscription.



Push-based data-delivery mechanism

Push-based mechanisms function in the following manner:

1. A structure of data records to be pushed is selected. An algorithm provides an adaptable multi-level mechanism that permits data items to be pushed uniformly or non-uniformly after structuring them according to their relative importance.

2. Data is pushed at selected time intervals using an adaptive algorithm. Pushing only once saves bandwidth. However, pushing at periodic intervals is important because it provides the devices that were disconnected at the

time of previous push with a chance to cache the data when it is pushed again.

3. Bandwidths are adapted for downlink (for pushes) using an algorithm. Usually higher bandwidth is allocated to records having higher number of subscribers or to those with higher access probabilities.

4. A mechanism is also adopted to stop pushes when a device is handed over to another cell.
The application-distribution system of the service provider uses these algorithms and adopts bandwidths as per the number of subscribers for the published data records. On the device handoff, the subscription cancels or may be passed on to new service provider system.

Advantages of Push based mechanisms:
Push-based mechanisms enable broadcast of data services to multiple devices.

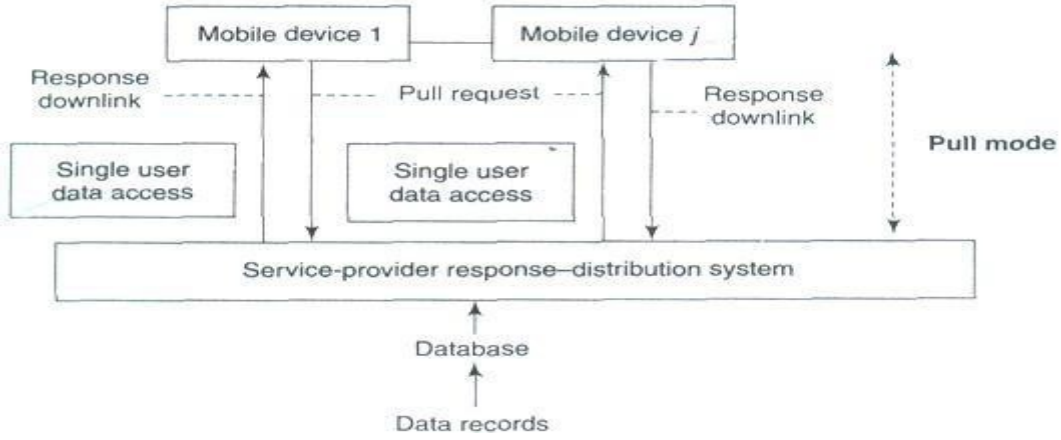
The server is not interrupted frequently by requests from mobile devices.

These mechanisms also prevent server overload, which might be caused by flooding of device requests

Also, the user even gets the data he would have otherwise ignored such as traffic congestion, forthcoming weather reports etc

Disadvantages:
Push-based mechanisms disseminate of unsolicited, irrelevant, or out-of-context data, which may cause inconvenience to the user.

Pull based Mechanisms
The user-device or computing system pulls the data records from the service provider's application database server or from a set of distributed computing systems. Examples are music album server, ring tones server, video clips server, or bank account activity server. Records are pulled by the mobile devices on demand followed by the selective response from the server. Selective response means that server transmits data packets as response selectively, for example, after client-authentication, verification, or subscription account check. The pull mode is also known as the on-demand mode. The following figure shows a pull-based data-delivery mechanism in which a device pulls (demands) from a server or computing system, the data records generated by a set of distributed computing systems.



Pull based Delivery Mechanism

Pull-based mechanisms function in the following manner:
1. The bandwidth used for the uplink channel depends upon the number of pull requests.

2. A pull threshold is selected. This threshold limits the number of pull requests in a given period of time. This controls the number of server interruptions.

3. A mechanism is adopted to prevent the device from pulling from a cell, which has handed over the concerned device to another cell. On device handoff, the subscription is cancelled or passed on to the new service provider cell

In pull-based mechanisms the user-device receives data records sent by server on demand only.
Advantages of Pull based mechanisms:
With pull-based mechanisms, no unsolicited or irrelevant data arrives at the device and the relevant data is disseminated only when the user asks for it.

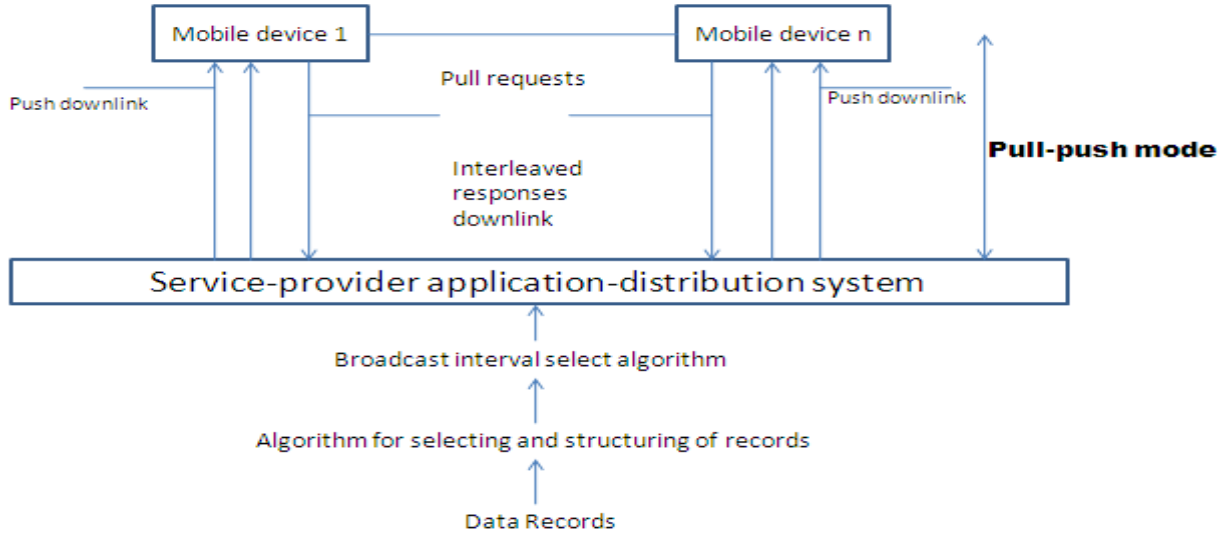
Pull-based mechanisms are the best option when the server has very little contention and is able to respond to many device requests within expected time intervals.

Disadvantages:
The server faces frequent interruptions and queues of requests at the server may cause congestion in cases of sudden rise in demand for certain data record.

In on-demand mode, another disadvantage is the energy and bandwidth required for sending the requests for hot items and temporal records

Hybrid Mechanisms

A hybrid data-delivery mechanism integrates pushes and pulls. The hybrid mechanism is also known as interleaved-push-and-pull (IPP) mechanism. The devices use the back channel to send pull requests for records, which are not regularly pushed by the front channel. The front channel uses algorithms modeled as broadcast disks and sends the generated interleaved responses to the pull requests. The user device or computing system pulls as well receives the pushes of the data records from the service provider's application server or database server or from a set of distributed computing systems. Best example would be a system for advertising and selling music albums. The advertisements are pushed and the mobile devices pull for buying the album.



Hybrid interleaved push-pull-based data-delivery mechanism

The above figure shows a hybrid interleaved, push-pull-based data-delivery mechanism in which a device pulls (demands) from a server and the server interleaves the responses along with the pushes of the data records generated by a set of distributed computing systems. Hybrid mechanisms function in the following manner:

1. There are two channels, one for pushes by front channel and the other for pulls by back channel.

2. Bandwidth is shared and adapted between the two channels depending upon the number of active devices receiving data from the server and the number of devices requesting data pulls from the server.

3. An algorithm can adaptively chop the slowest level of the scheduled pushes successively The data records at lower level where the records are

assigned lower priorities can have long push intervals in a broadcasting model.

Advantages of Hybrid mechanisms:

- The number of server interruptions and queued requests are significantly reduced.

Disadvantages:

- IPP does not eliminate the typical server problems of too many interruptions and queued requests.
- Another disadvantage is that adaptive chopping of the slowest level of scheduled pushes

Selective Tuning and Indexing Techniques

The purpose of pushing and adapting to a broadcast model is to push records of greater interest with greater frequency in order to reduce access time or average access latency. A mobile device does not have sufficient energy to continuously cache the broadcast records and hoard them in its memory. A device has to dissipate more power if it gets each pushed item and caches it. Therefore, it should be activated for listening and caching only when it is going to receive the selected data records or buckets of interest. During remaining time intervals, that is, when the broadcast data buckets or records are not of its interest, it switches to idle or power down mode.

Selective tuning is a process by which client device selects only the required pushed buckets or records, tunes to them, and caches them. *Tuning means getting ready for caching at those instants and intervals when a selected record of interest broadcasts*. Broadcast data has a structure and overhead. Data broadcast from server, which is organized into buckets, is interleaved. The server prefixes a directory, hash parameter (from which the device finds the key), or index to the buckets. These prefixes form the basis of different methods of selective tuning. Access time **(taccess)** is the time interval between pull request from device and reception of response from broadcasting or data pushing or responding server. Two important factors affect taccess – (i) number and size of the records to be broadcast and (ii) directory- or cache-miss factor (if there is a miss then the response from the server can be received only in subsequent broadcast cycle or subsequent repeat broadcast in the cycle).

Directory Method

One of the methods for selective tuning involves broadcasting a directory as overhead at the beginning of each broadcast cycle. If the interval between

the start of the broadcast cycles is T, then directory is broadcast at each successive intervals of T. A directory can be provided which specifies when a specific record or data item appears in data being broadcasted. For example, a directory (at header of the cycle) consists of directory start sign, 10, 20, 52, directory end sign. It means that after the directory end sign, the 10th, 20th and 52nd buckets contain the data items in response to the device request. The device selectively tunes to these buckets from the broadcast data.

A device has to wait for directory consisting of start sign, pointers for locating buckets or records, and end sign. Then it has to wait for the required bucket or record before it can get tuned to it and, start caching it. Tuning time ttune is the time taken by the device for selection of records. This includes the time lapse before the device starts receiving data from the server. In other words, it is the sum of three periods—time spent in listening to the directory signs and pointers for the record in order to select a bucket or record required by the device, waiting for the buckets of interest while actively listening (getting the incoming record wirelessly), and caching the broadcast data record or bucket.

The device selectively tunes to the broadcast data to download the records of interest. When a directory is broadcast along with the data records, it minimizes ttune and taccess. The device saves energy by remaining active just for the periods of caching the directory and the data buckets. For rest of the period (between directory end sign and start of the required bucket), it remains idle or performs application tasks. Without the use of directory for tuning, ttune = taccess and the device is not idle during any time interval.

Hash-Based Method

Hash is a result of operations on a pair of key and record. Advantage of broadcasting a hash is that it contains a fewer bits compared to key and record separately. The operations are done by a hashing function. From the server end the hash is broadcasted and from the device end a key is extracted by computations from the data in the record by operating the data with a function called hash function (algorithm). This key is called hash key.

Hash-based method entails that the hash for the hashing parameter (hash key) is broadcasted. Each device receives it and tunes to the record as per the extracted key. In this method, the records that are of interest to a device or those required by it are cached from the broadcast cycle by first extracting and identifying the hash key which provides the location of the record. This helps in tuning of the device. Hash-based method can be described as follows:

1. A separate directory is not broadcast as overhead with each broadcast cycle.

2. Each broadcast cycle has hash bits for the hash function H, a shift function S, and the data that it holds. The function S specifies the location of the

record or remaining part of the record relative to the location of hash and, thus, the time interval for wait before the record can be tuned and cached.
3. Assume that a broadcast cycle pushes the hashing parameters **H*(Rí)*** [H and S] and record ***Rí***. The functions H and S help in tuning to the H(***Rí***) and hence to ***Rí*** as follows—H gives a key which in turn gives the location of H*(**Rí**)* in the broadcast data. In case H generates a key that does not provide the location of H(***Rí***) by itself, then the device computes the location from S after the location of H(***Rí***). That location has the sequential records ***Rí*** and the devices tunes to the records from these locations.

4. In case the device misses the record in first cycle, it tunes and caches that in next or some other cycle.

Index-Based Method

Indexing is another method for selective tuning. Indexes temporarily map the location of the buckets. At each location, besides the bits for the bucket in record of interest data, an offset value may also be specified there. While an index maps to the absolute location from the beginning of a broadcast cycle, an offset index is a number which maps to the relative location after the end of present bucket of interest. Offset means a value to be used by the device along with the present location and calculate the wait period for tuning to the next bucket. All buckets have an offset to the beginning of the next indexed bucket or item.
Indexing is a technique in which each data bucket, record, or record block of interest is assigned an index at the previous data bucket, record, or record block of interest to enable the device to tune and cache the bucket after the wait as per the offset value. The server transmits this index at the beginning of a broadcast cycle as well as with each bucket corresponding to data of interest to the device. A disadvantage of using index is that it extends the broadcast cycle and hence increases taccess.
The index I has several offsets and the bucket type and flag information. A typical index may consist of the following:
1. Ioffset(1) which defines the offset to first bucket of nearest index.

2. Additional information about *Tb,* which is the time required for caching the bucket bits in full after the device tunes to and starts caching the bucket. This enables transmission of buckets of variable lengths.

3. Ioffset (next) which is the index offset of next bucket record of interest.
4. Ioffset(end) which is the index offset for the end of broadcast cycle and the start of next cycle. This enables the device to look for next index I after the time interval as per Ioffset(end). This also permits a broadcast cycle to consist of variable number of buckets.

5. Itype, which provides the specification of the type of contents of next bucket to be tuned, that is, whether it has an index value or data.

6. A flag called dirty flag which contains the information whether the indexed buckets defined by Ioffset(1) and Ioffset(next) are dirty or not. An indexed bucket being dirty means that it has been rewritten at the server with new values. Therefore, the device should invalidate the previous caches of these buckets and update them by tuning to and caching them.

The advantage of having an index is that a device just reads it and selectively tunes to the data buckets or records of interest instead of reading all the data records and then discarding those which are not required by it. During the time intervals in which data which is not of interest is being broadcast, the device remains in idle or power down mode.
Transmission of an index I only once with every broadcast cycle increases access latency of a record as follows: This is so because if an index is lost during a push due to transmission loss, then the device must wait for the next push of the same index-record pair. The data tuning time now increases by an interval equal to the time required for one broadcast cycle.
An index assignment strategy ***(I, m)*** is now described. *(I, m)* indexing means an index I is transmitted *m* times during each push of a record. An algorithm is used to adapt a value of **m** such that it minimizes access (caching) latency in a given wireless environment which may involve frequent or less frequent loss of index or data. Index format is adapted to (I, *m)* with a suitable value of m chosen as per the wireless environment. This decreases the probability of missing I and hence the caching of the record of interest
Indexing reduces the time taken for tuning by the client devices and thus conserves their power resources. Indexing increases access latency because the number of items pushed is more (equals m times index plus *n* records).

Distributed Index Based Method

Distributed index-based method is an improvement on the (I, *m)* method. In this method, there is no need to repeat the complete index again and again. Instead of replicating the whole index *m* times, each index segment in a bucket describes only the offset I' of data items which immediately follow. Each index I is partitioned into two parts—I' and I". I" consists of
unrepeated *k* levels (sub-indexes), which do not repeat and I' consists of top I repeated levels (sub-indexes).
Assume that a device misses I(includes I' and I' once) transmitted at the beginning of the broadcast cycle. As I' is repeated *m* - I times after this, it tunes to the pushes by using I', The access latency is reduced as I' has lesser levels.

Flexible Indexing Method

Assume that a broadcast cycle has number of data segments with each of the segments having a variable set of records. For example, let *n* records, *Ro* to *Rn-1,* be present in four data segments, *R()* to *Ri-1, Ri* to *Rj-1* , *Rj* to *Rj-1* and *Rk* to *Rn-1*. Some possible index parameters are (i) Iseg, having just 2 bits for the offset, to specify the location of a segment in a broadcast cycle, (ii) Irec, having just 6 bits for the offset, to specify the location of a

record of interest within a segment of the broadcast cycle, (iii) Ib, having just 4 bits for the offset, to specify the location of a bucket of interest within a record present in one of the segments of the broadcast cycle. Flexible indexing method provides dual use of the parameters (e.g., use of Iseg or Irec in an index segment to tune to the record or buckets of interest) or multi-parameter indexing (e.g., use of Iseg, Irec, or Ib in an index segment to tune to the bucket of interest).
Assume that broadcast cycle has *m* sets of records (called segments). A set of binary bits defines the index parameter Iseg,. A local index is then assigned to the specific record (or bucket). Only local index (Irec or Ib) is used in (Iloc, m) based data tuning which corresponds to the case of flexible indexing method being discussed. The number of bits in a local index is much smaller than that required when each record is assigned an index. Therefore, the flexible indexing method proves to be beneficial.

AlternativeMethods

Temporal Addressing Temporal addressing is a technique used for pushing in which instead of repeating I several times, a temporal value is repeated before a data record is transmitted. When temporal information contained in this value is used instead of address, there can be effective synchronization of tuning and caching of the record of interest in case of non- uniform time intervals between the successive bits. The device remains idle and starts tuning by synchronizing as per the temporal (time)-information for the pushed record. Temporal information gives the time at which cache is scheduled. Assume that temporal address is 25675 and each address corresponds to wait of 1 ms, the device waits and starts synchronizing the record after 25675 ms.

Broadcast Addressing: Broadcast addressing uses a broadcast address similar to IP or multicast address. Each device or group of devices can be assigned an address. The devices cache the records which have this address as the broadcasting address in a broadcast cycle. This address can be used along with the pushed record. A device uses broadcast address in place of the index I to select the data records or sets. Only the addressed device(s) caches the pushed record and other devices do not select and tune to the record. In place of repeating I several times, the broadcast address can be repeated before a data record is transmitted. The advantage of using this type of addressing is that the server addresses to specific device or specific group of devices.

Use of Headers: A server can broadcast a data in multiple versions or ways. An index or address only specifies where the data is located for the purpose of tuning. It does not specify the details of data at the buckets. An alternative is to place a header or a header with an extension with a data object before broadcasting. Header is used along with the pushed record. The device uses header part in place of the index / and in case device finds from the header that the record is of interest, it selects the object and caches it. The header can be useful, for example it can give information

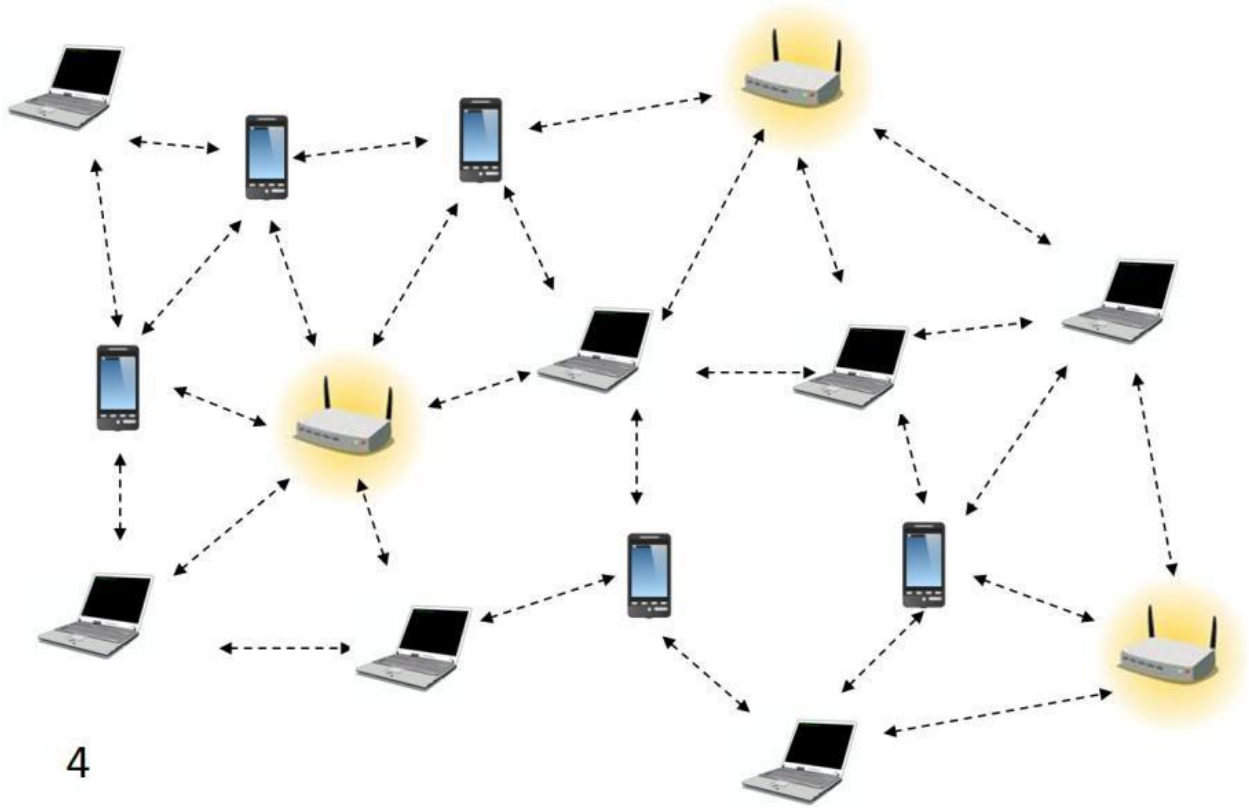
about the type, version, and content modification data or application for which it is targeted.

Chapter-5

Mobile Ad Hoc Networks (MANETs)

Mobile Ad hoc NETworks (MANETs) are wireless networks which are characterized by dynamic topologies and no fixed infrastructure. Each node in a MANET is a computer that may be required to act as both a host and a router and, as much, may be required to forward packets between nodes

which cannot directly communicate with one another. Each MANET node has much smaller frequency spectrum requirements that that for a node in a fixed infrastructure network. A MANET is an autonomous collection of mobile users that communicate over relatively bandwidth constrained wireless links. Since the nodes are mobile, the network topology may change rapidly and unpredictably over time. The network is decentralized, where all network activity including discovering the topology and delivering messages must be executed by the nodes themselves, i.e., routing functionality will be incorporated into mobile nodes.



A mobile ad hoc network is a collection of wireless nodes that can dynamically be set up anywhere and anytime without using any pre-existing fixed network infrastructure.

MANET- Characteristics

- Dynamic network topology
- Bandwidth constraints and variable link capacity
- Energy constrained nodes
- Multi-hop communications
- Limited security
- Autonomous terminal

Distributed operation

Light-weight terminals

Need for Ad Hoc Networks

Setting up of fixed access points and backbone infrastructure is not always viable

– Infrastructure may not be present in a disaster area or war zone

– Infrastructure may not be practical for short-range radios; Bluetooth (range ~ 10m)

Ad hoc networks:

– Do not need backbone infrastructure support

– Are easy to deploy

– Useful when infrastructure is absent, destroyed or impractical

Properties of MANETs

MANET enables fast establishment of networks. When anew network is to be established, the only requirement is to provide a new set of nodes with limited wireless communication range. A node has limited capability, that is, it can connect only to the nodes which are nearby. Hence it consumes limited power.

A MANET node has the ability to discover a neighboring node and service. Using a service discovery protocol, a node discovers the service of a nearby node and communicates to a remote node in the MANET.

MANET nodes have peer-to-peer connectivity among themselves.

MANET nodes have independent computational, switching (or routing), and communication capabilities.

The wireless connectivity range in MANETs includes only nearest node connectivity.

The failure of an intermediate node results in greater latency in communicating with the remote server.

Limited bandwidth available between two intermediate nodes becomes a constraint for the MANET. The node may have limited power and thus computations need to be energy- efficient.

There is no access-point requirement in MANET. Only selected access points are provided for connection to other networks or other MANETs.

MANET nodes can be the iPods, Palm handheld computers, Smartphones, PCs, smart labels, smart sensors, and automobile-embedded systems\

MANET nodes can use different protocols, for example, IrDA, Bluetooth, ZigBee, 802.11, GSM, and TCP/IP.MANET node performs data caching, saving, and aggregation.

MANET mobile device nodes interact seamlessly when they move with the nearby wireless nodes, sensor nodes, and embedded devices in automobiles so that the seamless connectivity is maintained between the devices.

MANET challenges

To design a good wireless ad hoc network, various challenges have to be taken into account:

Dynamic Topology: Nodes are free to move in an arbitrary fashion resulting in the topology changing arbitrarily. This characteristic demands dynamic configuration of the network.

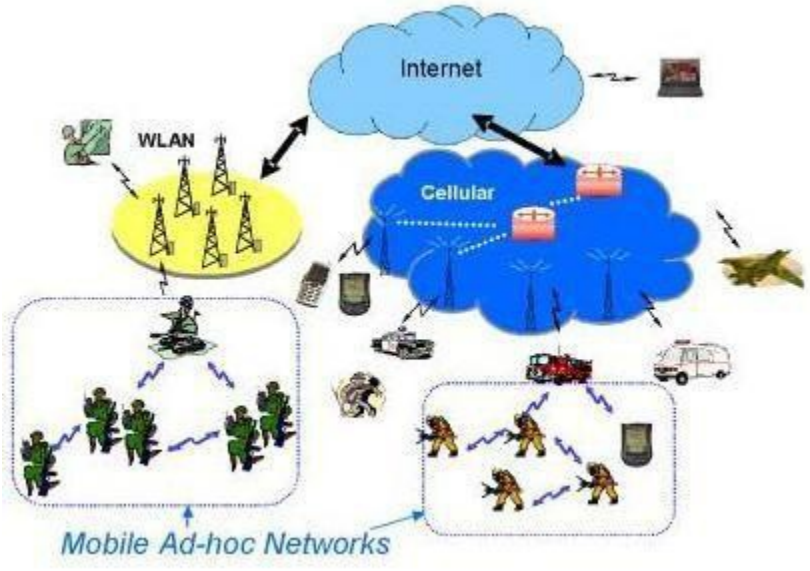
Limited security: Wireless networks are vulnerable to attack. Mobile ad hoc networks are more vulnerable as by design any node should be able to join or leave the network at any time. This requires flexibility and higher openness.

Limited Bandwidth: Wireless networks in general are bandwidth limited. In an ad hoc network, it is all the more so because there is no backbone to handle or multiplex higher bandwidth

Routing: Routing in a mobile ad hoc network is complex. This depends on many factors, including finding the routing path, selection of routers, topology, protocol etc.

Applications of MANETS

The set of applications for MANETs is diverse, ranging from small, static networks that are constrained by power sources, to large-scale, mobile, highly dynamic networks. The design of network protocols for these networks is a complex issue. Regardless of the application, MANETs need efficient distributed algorithms to determine network organization, link scheduling, and routing. Some of the main application areas of MANET's are:



Military battlefield– soldiers, tanks, planes. Ad- hoc networking would allow the military to take advantage of commonplace network technology to maintain an information network between the soldiers, vehicles, and military information headquarters.

Sensor networks – to monitor environmental conditions over a large area

- ***Local level*** – Ad hoc networks can autonomously link an instant and temporary multimedia network using notebook computers or palmtop computers to spread and share information among participants at e.g. conference or classroom. Another appropriate local level application might be in home networks where devices can communicate directly to exchange information.
- ***Personal Area Network (PAN)*** – pervasive computing i.e. to provide flexible connectivity between personal electronic devices or home appliances. Short-range MANET can simplify the intercommunication between various mobile devices (such as a PDA, a laptop, and a cellular phone). Tedious wired cables are replaced with wireless connections. Such an ad hoc network can also extend the access to the Internet or other networks by mechanisms e.g. Wireless LAN (WLAN), GPRS, and UMTS.
- ***Vehicular Ad hoc Networks*** – intelligent transportation i.e. to enable real time vehicle monitoring and adaptive traffic control
- ***Civilian environments*** – taxi cab network, meeting rooms, sports stadiums, boats, small aircraft
- ***Emergency operations*** – search and rescue, policing and fire fighting and to provide connectivity between distant devices where the network infrastructure is unavailable. Ad hoc can be used in emergency/rescue operations for disaster relief efforts, e.g. in fire, flood, or earthquake. Emergency rescue operations must take place where non- existing or

damaged communications infrastructure and rapid deployment of a communication network is needed. Information is relayed from one rescue team member to another over a small hand held.

Routing in MANET's

Routing in Mobile Ad hoc networks is an important issue as these networks do not have fixed infrastructure and routing requires distributed and cooperative actions from all nodes in the network. MANET's provide point to point routing similar to Internet routing. The major difference between routing in MANET and regular internet is the route discovery mechanism. Internet routing protocols such as RIP or OSPF have relatively long converge times, which is acceptable for a wired network that has infrequent topology changes. However, a MANET has a rapid topology changes due to node mobility making the traditional internet routing protocols inappropriate. MANET-specific routing protocols have been proposed, that handle topology changes well, but they have large control overhead and are not scalable for large networks. Another major difference in the routing is the network address. In internet routing, the network address (IP address) is hierarchical containing a network ID and a computer ID on that network. In contrast, for most MANET's the network address is simply an ID of the node in the network and is not hierarchical. The routing protocol must use the entire address to decide the next hop.

Some of the fundamental differences between wired networks & ad-hoc networks are:

 Asymmetric links: - Routing information collected for one direction is of no use for the other direction. Many routing algorithms for wired networks rely on a symmetric scenario.

 Redundant links: - In wired networks, some redundancy is present to survive link failures and this redundancy is controlled by a network administrator. In ad-hoc networks, nobody controls redundancy resulting in many redundant links up to the extreme of a complete meshed topology.

 Interference: - In wired networks, links exist only where a wire exists, and connections are planned by network administrators. But, in ad-hoc networks links come and go depending on transmission characteristics, one transmission might interfere with another and nodes might overhear the transmission of other nodes.

 Dynamic topology: - The mobile nodes might move in an arbitrary manner or medium characteristics might change. This result in frequent changes in topology, so snapshots are valid only for a very short period of time. So, in

ad-hoc networks, routing tables must somehow reflect these frequent changes in topology and routing algorithms have to be adopted.

Summary of the difficulties faced for routing in ad-hoc networks

- Traditional routing algorithms known from wired networks will not work efficiently or fail completely. These algorithms have not been designed with a highly dynamic topology, asymmetric links, or interference in mind.
- Routing in wireless ad-hoc networks cannot rely on layer three knowledge alone. Information from lower layers concerning connectivity or interference can help routing algorithms to find a good path.
- Centralized approaches will not really work, because it takes too long to collect the current status and disseminate it again. Within this time the topology has already changed.
- Many nodes need routing capabilities. While there might be some without, at least one router has to be within the range of each node. Algorithms have to consider the limited battery power of these nodes.
- The notion of a connection with certain characteristics cannot work properly. Ad-hoc networks will be connectionless, because it is not possible to maintain a connection in a fast changing environment and to forward data following this connection. Nodes have to make local decisions for forwarding and send packets roughly toward the final destination.
- A last alternative to forward a packet across an unknown topology is flooding. This approach always works if the load is low, but it is very inefficient. A hop counter is needed in each packet to avoid looping, and the diameter of the ad-hoc network.

Mobile Ad Hoc Networks (MANETs)

Types of MANET Routing Algorithms:

1. Based on the information used to build routing tables :

• *Shortest distance algorithms*: algorithms that use distance information to build routing tables.

• *Link state algorithms*: algorithms that use connectivity information to build a topology graph that is used to build routing tables.

2. Based on when routing tables are built:

• *Proactive algorithms*: maintain routes to destinations even if they are not needed. Some of the examples are Destination Sequenced Distance Vector (DSDV), Wireless Routing Algorithm (WRP), Global State Routing (GSR), Source-tree Adaptive Routing (STAR), Cluster-Head Gateway Switch Routing (CGSR), Topology Broadcast Reverse Path Forwarding (TBRPF), Optimized Link State Routing (OLSR) etc.

- Always maintain routes:- Little or no delay for route determination
- Consume bandwidth to keep routes up-to-date
- Maintain routes which may never be used

Advantages: low route latency, State information, QoS guarantee related to connection set-up or other real-time requirements

Disadvantages: high overhead (periodic updates) and route repair depends on update frequency

• *Reactive algorithms*: maintain routes to destinations only when they are needed. Examples are Dynamic Source Routing (DSR), Ad hoc-On demand distance Vector (AODV), Temporally ordered Routing Algorithm (TORA), Associativity-Based Routing (ABR) etc

only obtain route information when needed

Advantages: no overhead from periodic update, scalability as long as there is only light traffic and low mobility.

Disadvantages: high route latency, route caching can reduce latency

• *Hybrid algorithms*: maintain routes to nearby nodes even if they are not needed and maintain routes to far away nodes only when needed. Example is Zone Routing Protocol (ZRP).

Which approach achieves a better trade-off depends on the traffic and mobility patterns.

Mobile Ad Hoc Networks (MANETs)

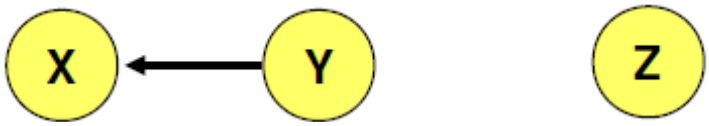
Destination sequence distance vector (DSDV)

Destination sequence distance vector (DSDV) routing is an example of proactive algorithms and an enhancement to distance vector routing for ad-hoc networks. Distance vector routing is used as routing information protocol (RIP) in wired networks. It performs extremely poorly with certain network changes due to the count-to-infinity problem. Each node exchanges its neighbor table periodically with its neighbors. Changes at one node in the network propagate slowly through the network. The strategies to avoid this problem which are used in fixed networks do not help in the case of wireless ad-hoc networks, due to the rapidly changing topology. This might create loops or unreachable regions within the network.

DSDV adds the concept of sequence numbers to the distance vector algorithm. Each routing advertisement comes with a sequence number. Within ad-hoc networks, advertisements may propagate along many paths. Sequence numbers help to apply the advertisements in correct order. This avoids the loops that are likely with the unchanged distance vector algorithm.

Each node maintains a routing table which stores next hop, cost metric towards each destination and a sequence number that is created by the destination itself. Each node periodically forwards routing table to neighbors. Each node increments and appends its sequence number when sending its local routing table. Each route is tagged with a sequence number; routes with greater sequence numbers are preferred. Each node

advertises a monotonically increasing even sequence number for itself. When a node decides that a route is broken, it increments the sequence number of the route and advertises it with infinite metric. Destination advertises new sequence number.
When X receives information from Y about a route to Z,



Let destination sequence number for Z at X be S(X), S(Y) is sent from Y

If S(X) > S(Y), then X ignores the routing information received from Y

If S(X) = S(Y), and cost of going through Y is smaller than the route known to X, then X sets Y as the next hop to Z

If S(X) < S(Y), then X sets Y as the next hop to Z, and S(X) is updated to equal S(Y)

Besides being loop-free at all times, DSDV has low memory requirements and a quick convergence via triggered updates. Disadvantages of DSDV are, large routing overhead, usage of only bidirectional links and suffers from count to infinity problem.

Mobile Ad Hoc Networks (MANETs)

Dynamic Source Routing

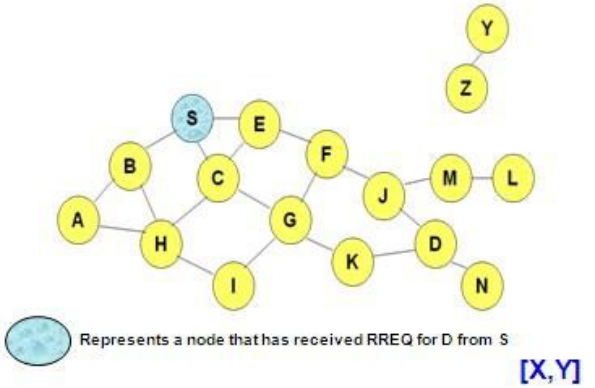
The Dynamic Source Routing protocol (DSR) is a simple and efficient routing protocol designed specifically for use in multi-hop wireless ad hoc networks of mobile nodes. DSR allows the network to be completely self-organizing and self-configuring, without the need for any existing network infrastructure or administration. The protocol is composed of the two main mechanisms of "Route Discovery" and "Route Maintenance", which work together to allow nodes to discover and maintain routes to arbitrary destinations in the ad hoc network. All aspects of the protocol operate entirely on-demand, allowing the routing packet overhead of DSR to scale automatically to only that needed to react to changes in the routes currently in use.

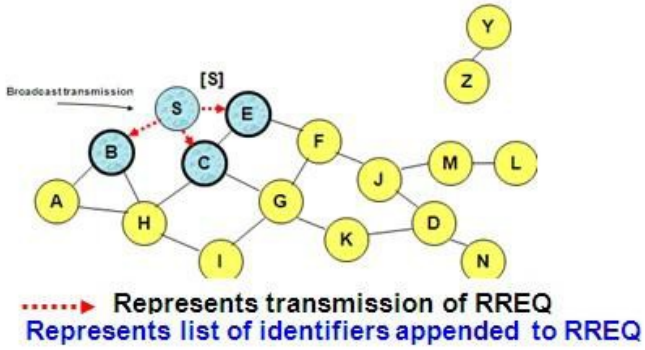
Route discovery. If the source does not have a route to the destination in its route cache, it broadcasts a route request (RREQ) message specifying the destination node for which the route is requested. The RREQ message includes a route record which specifies the sequence of nodes traversed by the message. When an intermediate node receives a RREQ, it checks to see if it is already in the route record. If it is, it drops the message. This is done to prevent routing loops. If the intermediate node had received the RREQ before, then it also drops the message. The intermediate node forwards the RREQ to the next hop according to the route specified in the header. When the destination receives the RREQ, it sends back a route reply message. If the destination has a route to the source in its route cache, then it can send

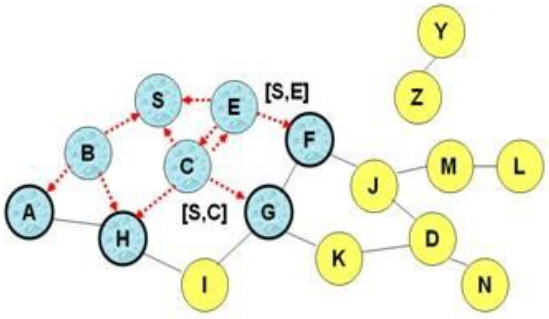
a route response (RREP) message along this route. Otherwise, the RREP message can be sent along the reverse route back to the source. Intermediate nodes may also use their route cache to reply to RREQs. If an intermediate node has a route to the destination in its cache, then it can append the route to the route record in the RREQ, and send an RREP back to the source containing this route. This can help limit flooding of the RREQ. However, if the cached route is out-of-date, it can result in the source receiving stale routes.

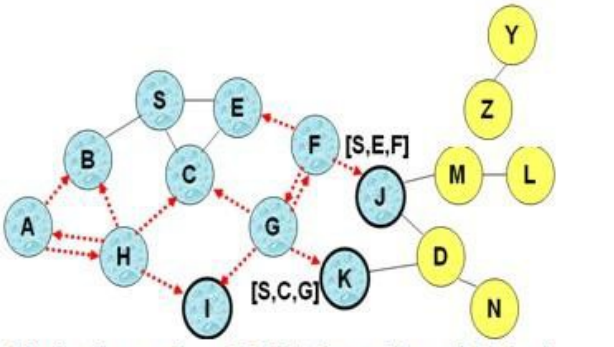
Route maintenance. When a node detects a broken link while trying to forward a packet to the next hop, it sends a route error (RERR) message back to the source containing the link in error. When an RERR message is received, all routes containing the link in error are deleted at that node.

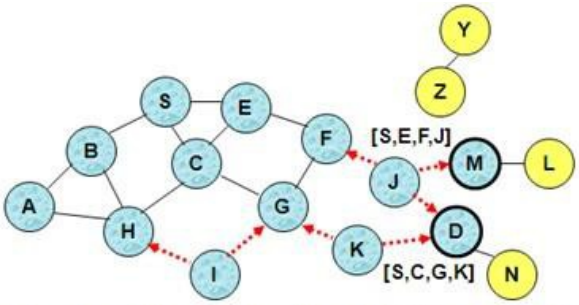
As an example, consider the following MANET, where a node S wants to send a packet to D, but does not know the route to D. So, it initiates a route discovery. Source node S floods Route Request (RREQ). Each node appends its own identifier when forwarding RREQ as shown below.



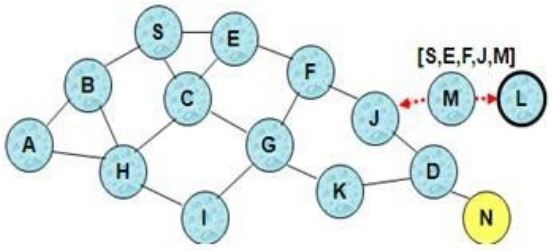






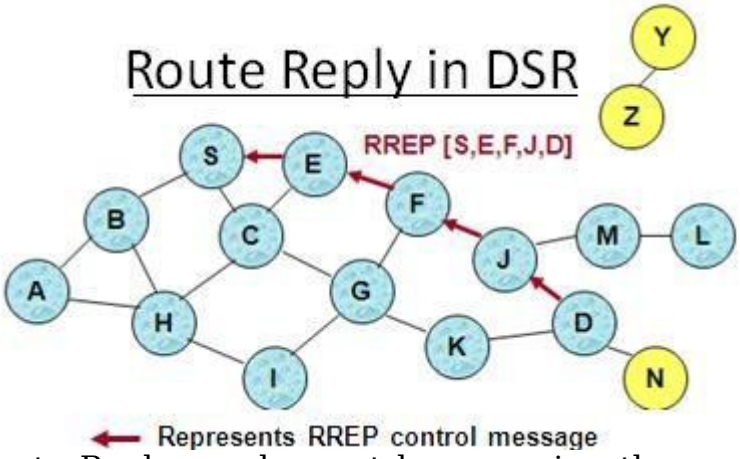


• Nodes J and K both broadcast RREQ to node D
• Since nodes J and K are hidden from each other, their transmissions may collide

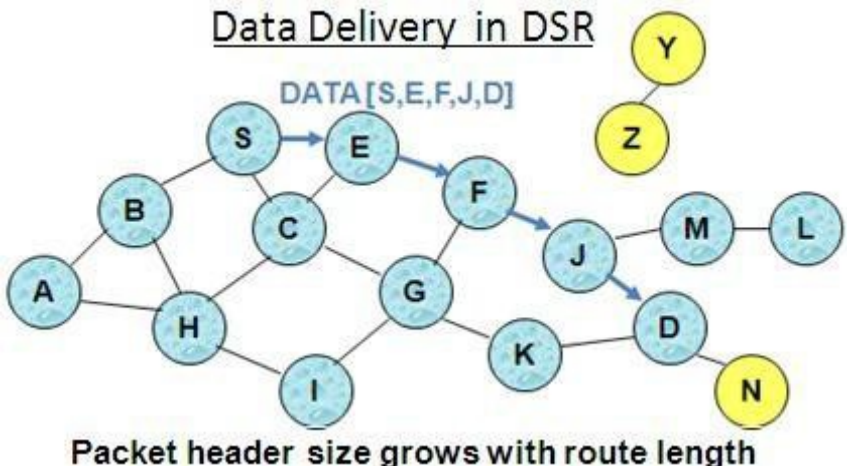


Node D does not forward RREQ, because node D is the intended target of the route discovery

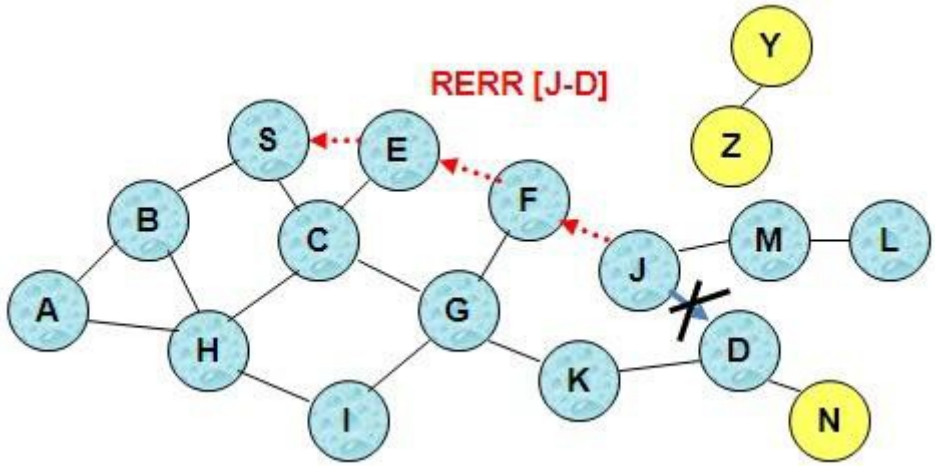
Destination D on receiving the first RREQ, sends a Route Reply (RREP). RREP is sent on a route obtained by reversing the route appended to received RREQ. RREP includes the route from S to D on which RREQ was received by node D.



Route Reply can be sent by reversing the route in Route Request (RREQ) only if links are guaranteed to be bi-directional. If Unidirectional (asymmetric) links are allowed, then RREP may need a route discovery from S to D. Node S on receiving RREP, caches the route included in the RREP. When node S sends a data packet to D, the entire route is included in the packet header
{hence the name source routing}. Intermediate nodes use the source route included in a packet to determine to whom a packet should be forwarded



sends a route error to S along route J-F-E-S when its attempt to forward the data packet S (with route SEFJD) on J-D fails. Nodes hearing RERR update their route cache to remove link J-D



Advantages of DSR:

- Routes maintained only between nodes who need to communicate-- reduces overhead of route maintenance
- Route caching can further reduce route discovery overhead
- A single route discovery may yield many routes to the destination, due to intermediate nodes replying from local caches

Disadvantages of DSR:

- Packet header size grows with route length due to source routing
- Flood of route requests may potentially reach all nodes in the network
- Care must be taken to avoid collisions between route requests propagated by neighboring nodes -- insertion of random delays before forwarding RREQ
- Increased contention if too many route replies come back due to nodes replying using their local cache-- Route Reply *Storm* problem. Reply storm may be eased by preventing a node from sending RREP if it hears another RREP with a shorter route

An intermediate node may send Route Reply using a stale cached route, thus polluting other caches

An optimization for DSR can be done called as Route Caching. Each node caches a new route it learns by *any means.* In the above example, When node S finds route [S,E,F,J,D] to node D, node S also learns route [S,E,F] to node F. When node K receives Route Request [S,C,G] destined for node, node K learns route [K,G,C,S] to node S. When node F forwards Route Reply RREP [S,E,F,J,D], node F learns route [F,J,D] to node D. When node E forwards Data [S,E,F,J,D] it learns route [E,F,J,D] to node D. A node may also learn a route when it overhears Data packets. Usage of Route cache can speed up route discovery and can also reduce propagation of route requests. The disadvantages are, stale caches can adversely affect performance. With passage of time and host mobility, cached routes may become invalid.

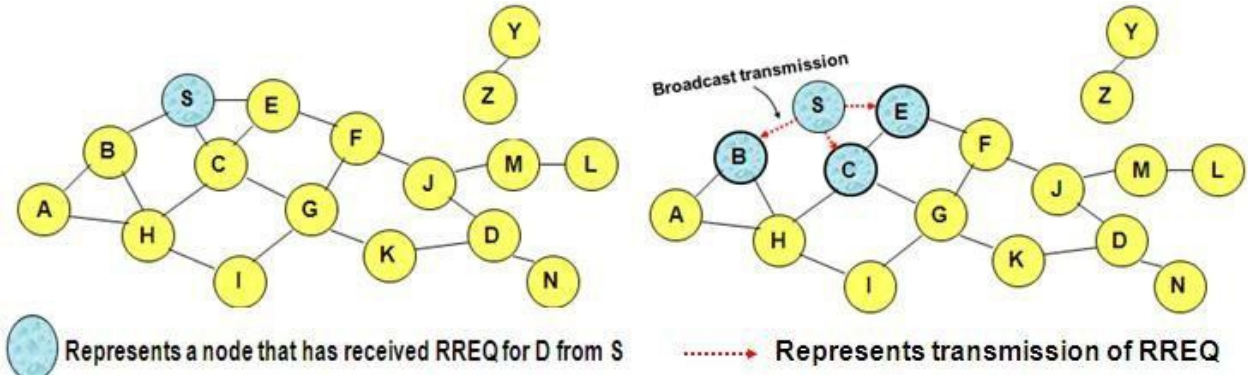
Ad Hoc On-Demand Distance Vector Routing (AODV)

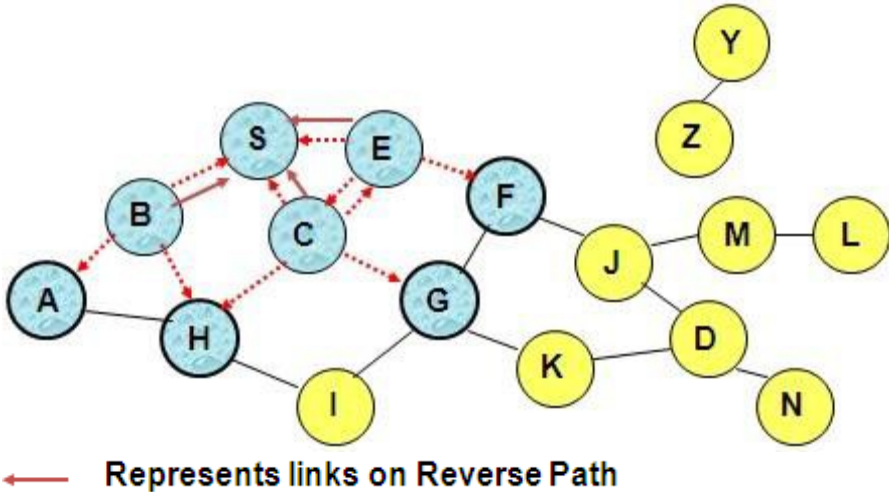
AODV is another reactive protocol as it reacts to changes and maintains only the active routes in the caches or tables for a pre-specified expiration time. Distance vector means a set of distant nodes, which defines the path to destination. AODV can be considered as a descendant of DSR and DSDV algorithms. It uses the same route discovery mechanism used by DSR. DSR includes source routes in packet headers and resulting large headers can sometimes degrade performance, particularly when data contents of a packet are small. AODV attempts to improve on DSR by maintaining routing tables at the nodes, so that data packets do not have to contain routes. AODV retains the desirable feature of DSR that routes are maintained only between nodes which need to communicate. However, as opposed to DSR, which uses source routing, AODV uses hop-by-hop routing by maintaining routing table entries at intermediate nodes.

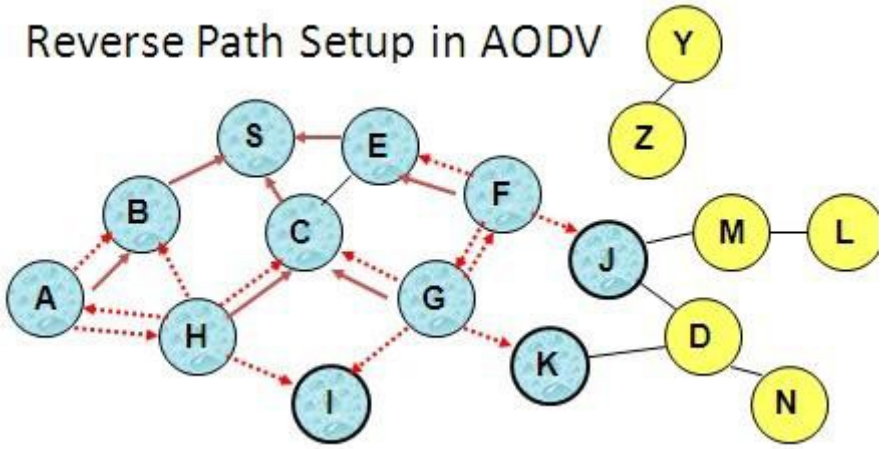
Route Discovery. The route discovery process is initiated when a source needs a route to a destination and it does not have a route in its routing table. To initiate route discovery, the source floods the network with a RREQ packet specifying the destination for which the route is requested. When a node receives an RREQ packet, it checks to see whether it is the destination or whether it has a route to the destination. If either case is true, the node generates an RREP packet, which is sent back to the source along the reverse path. Each node along the reverse path sets up a forward pointer to the node it received the RREP from. This sets up a forward path from the source to the destination. If the node is not the destination and does not have a route to the destination, it rebroadcasts the RREQ packet. At intermediate nodes duplicate RREQ packets are discarded. When the source node receives the first RREP, it can begin sending data to the destination. To determine the relative degree out-of-datedness of routes, each entry in the node routing table and all RREQ and RREP packets are tagged with a destination sequence number. A larger destination sequence

number indicates a more current (or more recent) route. Upon receiving an RREQ or RREP packet, a node updates its routing information to set up the reverse or forward path, respectively, only if the route contained in the RREQ or RREP packet is more current than its own route.

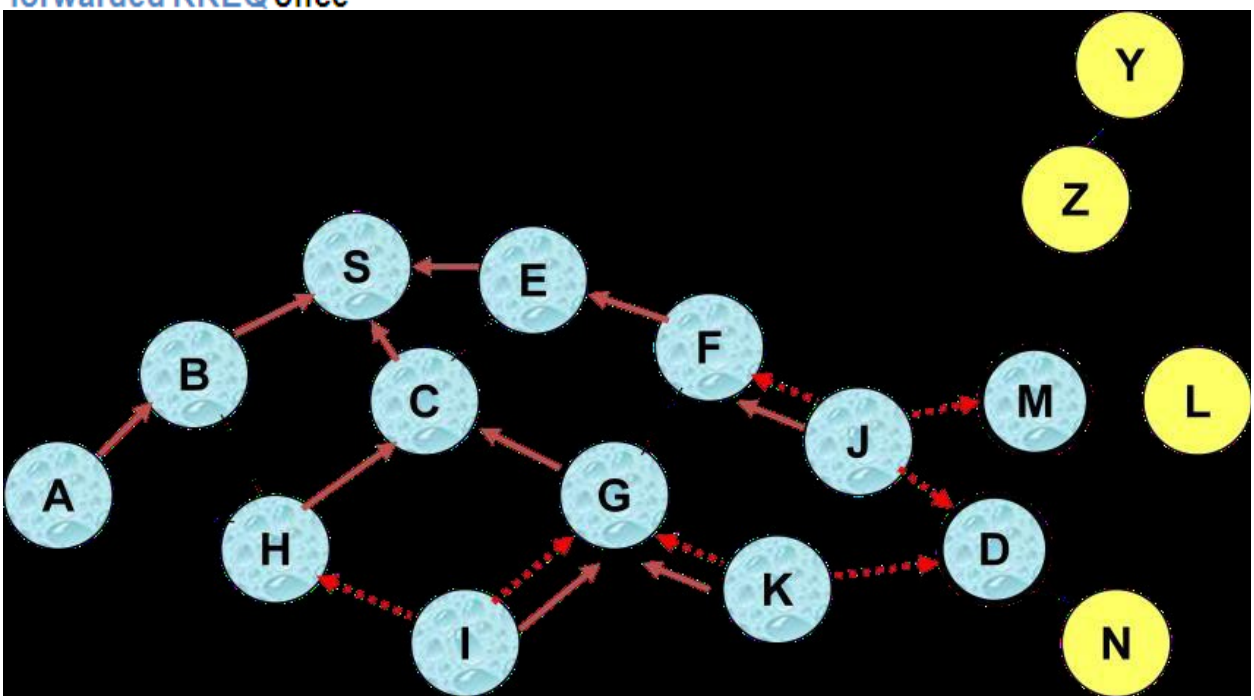
Route Maintenance. When a node detects a broken link while attempting to forward a packet to the next hop, it generates a RERR packet that is sent to all sources using the broken link. The RERR packet erases all routes using the link along the way. If a source receives a RERR packet and a route to the destination is still required, it initiates a new route discovery process. Routes are also deleted from the routing table if they are unused for a certain amount of time.

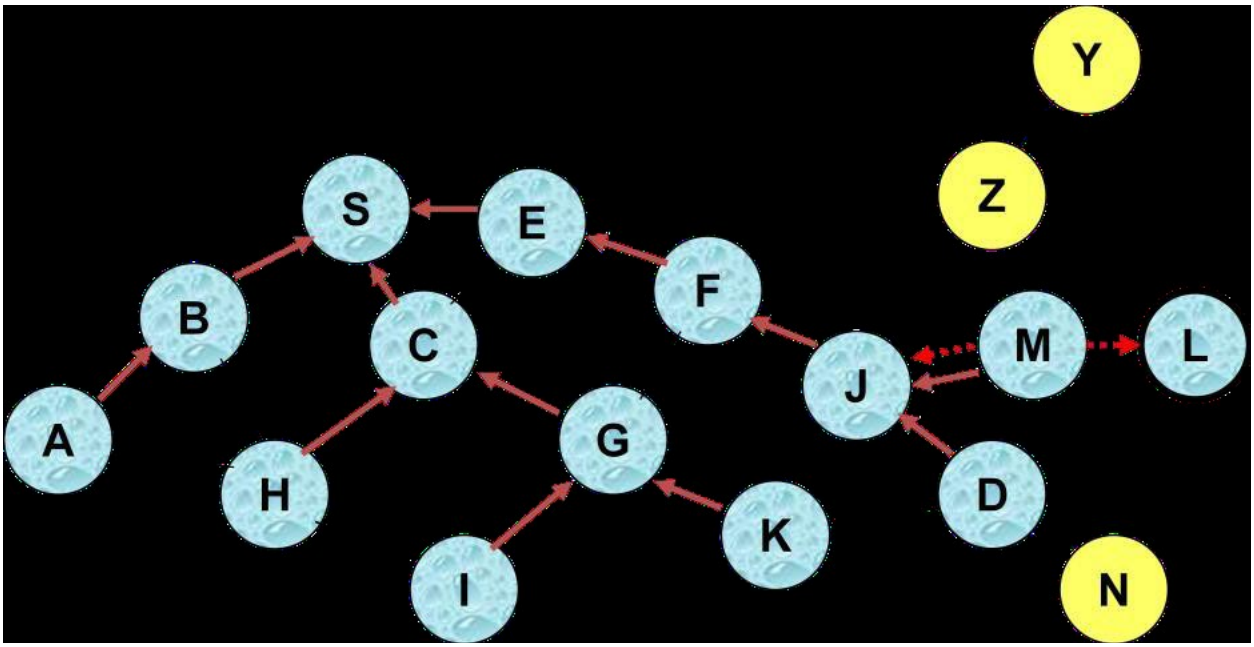




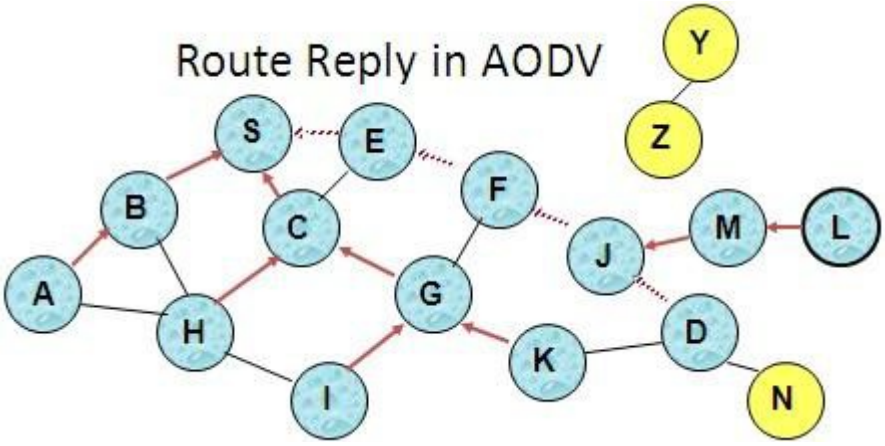


Node C receives RREQ from G and H, but does not forward it again, because node C has already forwarded RREQ once



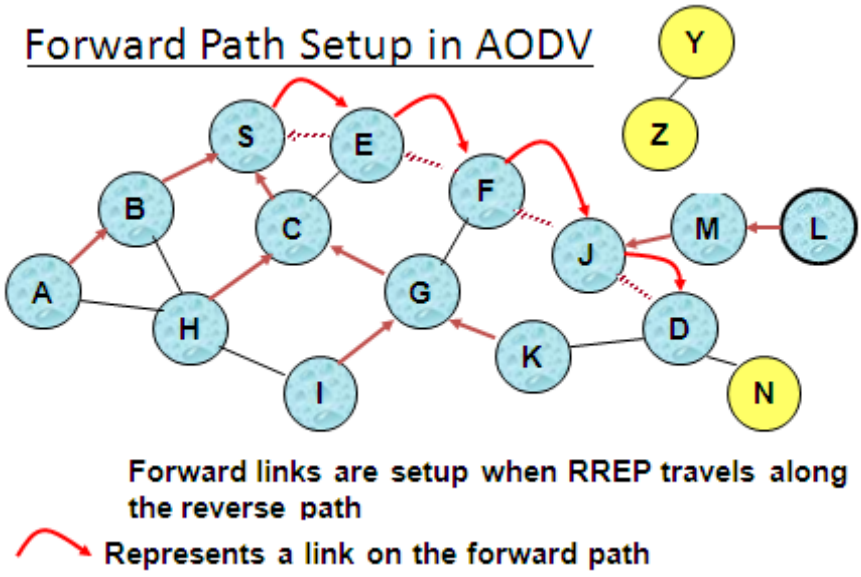


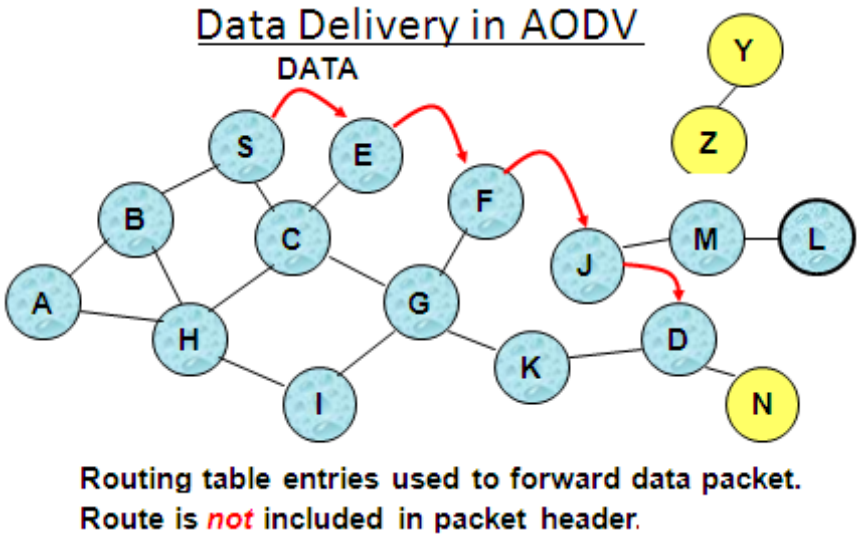
Node D does not forward RREQ, because node D is the intended target of the RREQ



Represents links on path taken by RREP

An intermediate node (not the destination) may also send a Route Reply (RREP) provided that it knows a more recent path than the one previously known to sender S. To determine whether the path known to an intermediate node is more recent, *destination sequence numbers* are used. The likelihood that an intermediate node will send a Route Reply when using AODV is not as high as DSR. A new Route Request by node S for a destination is assigned a higher destination sequence number. An intermediate node which knows a route, but with a smaller sequence number, cannot send Route Reply

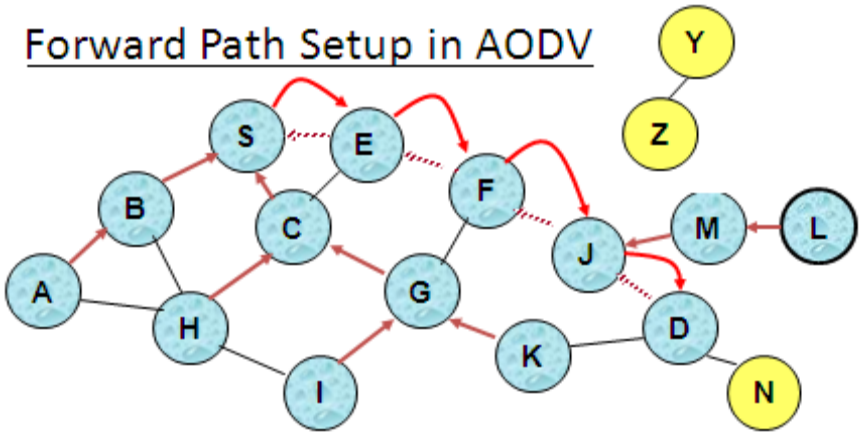




When node X is unable to forward packet P (from node S to node D) on link (X,Y), it generates a RERR message Node X increments the destination sequence number for D cached at node X. The incremented sequence number *N* is included in the RERR. When node S receives the RERR, it initiates a new route discovery for D using destination sequence number at least as large as *N*. When node D receives the route request with destination sequence number N, node D will set its sequence number to N, unless it is already larger than N.

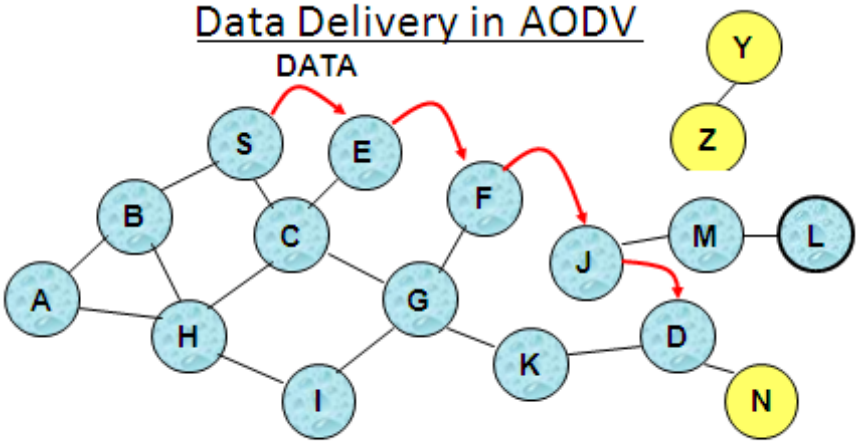
An intermediate node (not the destination) may also send a Route Reply (RREP) provided that it knows a more recent path than the one previously known to sender S. To determine whether the path known to an intermediate node is more recent, *destination sequence numbers* are used. The likelihood that an intermediate node will send a Route Reply when

using AODV is not as high as DSR. A new Route Request by node S for a destination is assigned a higher destination sequence number. An intermediate node which knows a route, but with a smaller sequence number, cannot send Route Reply



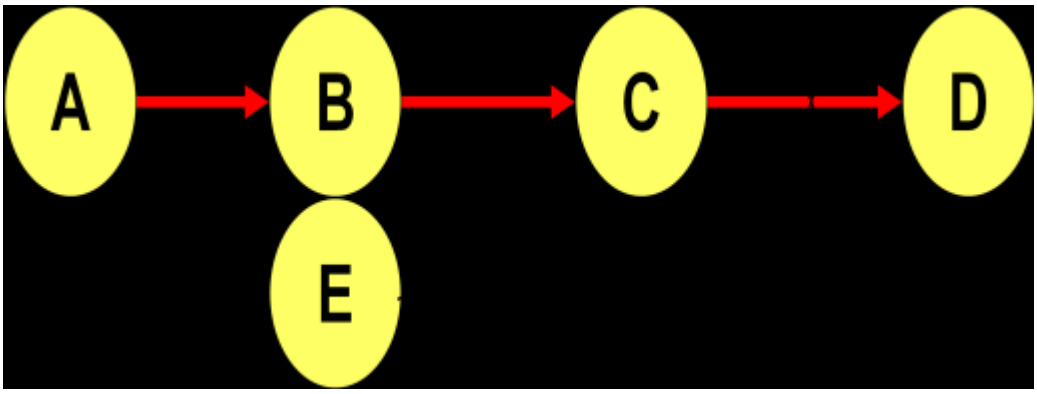
Forward links are setup when RREP travels along the reverse path

Represents a link on the forward path



Routing table entries used to forward data packet.
Route is *not* included in packet header.

When node X is unable to forward packet P (from node S to node D) on link (X,Y), it generates a RERR message Node X increments the destination sequence number for D cached at node X. The incremented sequence number *N* is included in the RERR. When node S receives the RERR, it initiates a new route discovery for D using destination sequence number at least as large as *N*. When node D receives the route request with destination sequence number N, node D will set its sequence number to N, unless it is already larger than N.
Sequence numbers are used in AODV to avoid using old/broken routes and to determine which route is newer. Also, it prevents formation of loops.



Assume that A does not know about failure of link C-D because RERR sent by C is lost.
Now C performs a route discovery for D. Node A receives the RREQ (say, via path C-E-A)
Node A will reply since A knows a route to D via node B resulting in a loop (for instance, C-E-A-B-C)
Neighboring nodes periodically exchange hello message and absence of hello message indicates a link failure. When node X is unable to forward packet P (from node S to node D) on link (X,Y), it generates a **RERR message**. Node X increments the destination sequence number for D cached at node X. The incremented sequence number *N* is included in the RERR. When node S receives the RERR, it initiates a new route discovery for D using destination sequence number at least as large as *N.* When node D receives the route request with destination sequence number N, node D will set its sequence number to N, unless it is already larger than N.

Security in MANET's

Securing wireless ad-hoc networks is a highly challenging issue. Understanding possible form of attacks is always the first step towards developing good security solutions. Security of communication in MANET is important for secure transmission of information. Absence of any central co-ordination mechanism and shared wireless medium makes MANET more vulnerable
to digital/cyber attacks than wired network there are a number of attacks that affect MANET. These attacks can be classified into two types:

1. External Attack: External attacks are carried out by nodes that do not belong to the network. It causes congestion sends false routing information or causes unavailability of services.

2. Internal Attack: Internal attacks are from compromised nodes that are part of the network. In an internal attack the malicious node from the network gains unauthorized access and impersonates as a genuine node. It

can analyze traffic between other nodes and may participate in other network activities.

Denial of Service attack: This attack aims to attack the availability of a node or the entire network. If the attack is successful the services will not be available. The attacker generally uses radio signal jamming and the battery exhaustion method.

Impersonation: If the authentication mechanism is not properly implemented a malicious node can act as a genuine node and monitor the network traffic. It can also send fake routing packets, and gain access to some confidential information.

Eavesdropping: This is a passive attack. The node simply observes the confidential information. This information can be later used by the malicious node. The secret information like location, public key, private key, password etc. can be fetched by eavesdropper.

Routing Attacks: The malicious node makes routing services a target because it's an important service in MANETs. There are two flavors to this routing attack. One is attack on routing protocol and another is attack on packet forwarding or delivery mechanism. The first is aimed at blocking the propagation of routing information to a node. The latter is aimed at disturbing the packet delivery against a predefined path.

Black hole Attack:: In this attack, an attacker advertises a zero metric for all destinations causing all nodes around it to route packets towards it. A malicious node sends fake routing information, claiming that it has an optimum route and causes other good nodes to route data packets through the malicious one. A malicious node drops all packets that it receives instead of normally forwarding those packets. An attacker listen the requests in a flooding based protocol.

Wormhole Attack: In a wormhole attack, an attacker receives packets at one point in the network, —tunnels them to another point in the network, and then replays them into the network from that point. Routing can be disrupted when routing control message are tunnelled. This tunnel between two colluding attacks is known as a wormhole.

Replay Attack: An attacker that performs a replay attack are retransmitted the valid data repeatedly to inject the network routing traffic that has been captured previously. This attack usually targets the freshness of routes, but can also be used to undermine poorly designed security solutions.

Jamming: In jamming, attacker initially keep monitoring wireless medium in order to determine frequency at which destination node is

receiving signal from sender. It then transmit signal on that frequency so that error free receptor is hindered.

Man- in- the- middle attack: An attacker sites between the sender and receiver and sniffs any information being sent between two nodes. In some cases, attacker may impersonate the sender to communicate with receiver or impersonate the receiver to reply to the sender.

Gray-hole attack: This attack is also known as routing misbehavior attack which leads to dropping of messages. Gray-hole attack has two phases. In the first phase the node advertise itself as having a valid route to destination while in second phase, nodes drops intercepted packets with a certain probability.