FUNDAMENTALS OF MOBILE COMPUTING

Chandra Shekhar Rajora
Sindhu Madhuri G

BOOKS ARCADE

Fundamentals of Mobile Computing

Fundamentals of Mobile Computing

Chandra Shekhar Rajora
Sindhu Madhuri G

BOOKS ARCADE
KRISHNA NAGAR, DELHI

Fundamentals of Mobile Computing

Chandra Shekhar Rajora
Sindhu Madhuri G

BOOKS ARCADE

Regd. Office:
F-10/24, East Krishna Nagar, Near Vijay Chowk, Delhi-110051
Ph. No: +91-11-79669196, +91-9899073222
E-mail: info@booksarcade.co.in, booksarcade.pub@gmail.com
Website: www.booksarcade.co.in

Year of Publication 2023

International Standard Book Number-13: 978-81-19199-19-8

CONTENTS

CHAPTER 1

INTRODUCTION TO MOBILE COMPUTING

Chandra Shekhar Rajora, Assistant Professor,
Department of Electronics and Communication, School of Engineering & Technology, Jaipur National University, Jaipur, India,
Email Id- chandra.shekhar@jnujaipur.ac.in

Mobile computing is a form of human-computer interaction where a computer is supposed to be carried around while being used, enabling the communication of data, speech, and video. Mobile software, hardware, and communication all play a part in mobile computing. Ad hoc and infrastructure networks, as well as communication features, protocols, data formats, and specific technologies, all have communication-related challenges. Smart phones or device parts are examples of hardware. Mobile application requirements and characteristics are dealt with by mobile software.

Working Principle of Mobile Computing:

A mobile computing system typically consists of a mobile device, like a laptop, tablet, or smartphone, and a wireless network connection powered by Wi-Fi or cellular wireless technologies, like 5G. Typically, mobile devices have local storage that may be accessed without a network connection. Usually, mobile computers allow users to connect to both wired and wireless technology. Given the collaborative nature of today's workplace, access to shared network resources, especially mobile cloud-based services, is crucial. Mobile devices are powered by integrated, rechargeable batteries, and the majority of them can utilize an alternating current (AC) power source when used from a permanent place.

For specialized and vertical applications, there are numerous mobile computing devices available in addition to laptops, tablets, and smartphones. These are gadgets used for telemetry and control, security, surveillance, and medicinal purposes. The application determines which device to use. For instance, tablets are frequently used for content consumption whereas laptops are better suited for content development. Smartphones have small screens and screen-based keyboards, but they serve as portable computers and communication tools.

Evaluation of Cloud Computing:

Different technologies have arisen in the world of computing today. These have expanded to support the global computer networks that are already in place. The necessity to be confined to a single physical location has been eliminated thanks to mobile computing. We hear phrases like "telecommuting," which refers to working from home or the field while still having access to the same resources as if one were in the office. Mobile computing has become more practical as a result of the development of portable computers and laptops, Personal Digital Assistants (PDA), PC tablets, and smartphones. These devices' mobility guarantees and makes it possible for users to access all services as if they were on their company's internal network. Consider the utilization of iPads and Tablet PCs. Users using this new technology can edit documents, browse the internet, send and receive email, stream live video files, take pictures, and support for phone and video conferencing is also included. Market share has been boosted by the persistent and growing need for improved and durable smart gadgets. Every producer aspires to have a distinct position in the market. Modern applications and services are made possible by these devices thanks to their invention and innovation. For instance, various cell phone makers have developed distinctive smartphones that can carry out the same function as

computers and at the same processing speed. Market share for various companies is a perennial point of contention. For instance, the creators of the operating systems used by Apple's iPhone, Google's Android, Microsoft's Windows Mobile, and Research in Motion's Blackberry are continuously vying with one another to create better products. These vendors have been increasingly innovative as a result of the demand for better, more mobile, resilient, and less expensive technology. Market data and statistics demonstrate an ever-increasing demand to buy and use such gadgets, whether for business or personal use. . Services that are designed for long-term implementation are created or reinvented in this context. It has also compelled other business partners in the sector to embrace improved services. For instance, in order to attract more customers, cellular service providers must innovate and make improvements. This could be in the form of better services, such voice and video service, high-speed internet and data access, etc. Thus, different network generations such as 2G, 2.5G, 3G, and 4G network services have been adopted.

Being able to work from anywhere is the core of mobile computing. The demand for these gadgets has increased due to the widespread usage of iPads, tablets, smartphones, and notebooks. These tools are available to modern workers, allowing them to complete their task from the comfort of their own home. Large volumes of important data are accessible from and stored on these devices. Without having to visit the workplace, executive and upper management can make decisions based on available facts. These gadgets, for instance, can be used to examine sales information and market forecasts or to host a meeting using video or audio conferencing. Due to the increased demand for these features, developers are continually creating new mobile computing applications that support various services.

Utilization of Mobile Computing:

Most aspects of life employ mobile computing, both in business and among consumers. It enables users to spend extended periods of time independent of a power source. Traveling employees who wish to keep connected to their jobs while on the go can benefit from this. It's also helpful for remote workers who might not have access to all the power and connectivity alternatives they would in an office. Mobile computing is used by consumers in a variety of ways, such as the following:

1. Web Browsing
2. Mobile Applications
3. Mobile Communications
4. Streaming Media
5. Internet Access

User data can be gathered by mobile devices and apps in a variety of settings and circumstances. Examples of wearable technology that gathers user data in innovative situations, such as fitness and health settings, include Fitbits and smart watches. Additionally, Internet of Things (IoT) is made possible by mobile computing. IoT is made up of non-traditional computers, sensors, and other devices that can connect and communicate with one another without direct human involvement.

Applications of Mobile Computing

The capacity to stay active is essential in numerous professions in order to utilize time effectively. Numerous fields have emphasized the value of mobile computers; a few are included here. **Vehicles:** Digital audio broadcast (DAB) with a 1.5 Mbit/s bandwidth is used to receive music, news, traffic updates, and other broadcast information. A universal mobile

telecommunications system (UMTS) phone with 384 kbit/s voice and data connectivity might be available for personal communication. The global positioning system (GPS) is used to determine the car's current location (GPS). For the quick exchange of information in emergencies or to assist one another in maintaining a safe distance from one another, cars moving in the same area create a local ad hoc network. In the event of an accident, not only will the airbag deploy, but a provider will also receive an emergency call alerting the police and ambulance service. Already, trains, trucks, and buses send maintenance and logistical data to their base of operations, improving fleet management and saving time and money.

Emergencies

When an accident occurs, an ambulance with a good wireless connection to a hospital can transport crucial data about injured people there. For this specific accident type, the essential preparations can be made, and professionals can be consulted for an early diagnosis. In the event of a natural disaster, such as a hurricane or earthquake, wireless networks serve as the sole means of communication. Only wireless, decentralized ad hoc networks survive in the worst scenarios.

Business

Managers can utilize portable laptops to make important presentations to important clients. They have access to the most recent market share data. They can use this knowledge to update the presentation during a brief break. They have the option of informing the office of potential new offers and setting up meetings to discuss responses to the new bids. Mobile computers can therefore take use of competitive advantages. Today's travelling salesperson requires immediate access to the company's database to make sure that the files on his or her laptop match the current situation, to allow the business to monitor all of its travelling employees' activities, to maintain consistent databases, etc. The laptop can become a truly mobile office with wireless connection, but effective and strong synchronization techniques are required to guarantee data consistency.

Credit Card Verification

When clients use their credit cards at Point of Sale (POS) terminals in stores and supermarkets, the necessary intercommunication between the bank's central computer and the POS terminal can happen swiftly and securely over cellular channels utilizing a mobile computer unit. By doing so, the transaction process can be sped up and the POS terminals can experience less congestion.

Tourism

The largest industry in every nation is tourism. The majority of tourist destinations are located far from populated areas. In this situation, wireless communication is essential for maintaining connectivity for tourists. They seek out travel services, hotel services, meal services, etc. and stay in constant communication with friends and family.

E-Governance

Governments are updating their rural areas using a variety of communication options. Government are connecting rural areas with head quarter offices for monitoring in order to give health, education, safety, farming, weather forecasts, and many other related information to governance.

Education

Pandemic of 2020 has taught us an important lesson: everyone who cannot attend college or schools should be able to receive distance education. Digital equipment and wireless connectivity form the basis of this distance learning idea. With the aid of strong wireless connectivity, all sectors of the economy public and private are now pushing toward online education.

Changing from Wired Networks

Additionally, wireless networks can take the place of wired networks in places like trade exhibitions, old buildings, or with distant sensors. It is frequently impractical to link remote sensors for weather forecasts, earthquake detection, or to give environmental data due to financial considerations. In this case, wireless connections, like those provided by satellite, can be useful. Computers, sensors, or information displays in ancient structures are further instances of wireless networks, as unnecessary wiring there could damage priceless walls or floors.

Infotainment

Wireless networks have the ability to deliver current information at any suitable location. By determining your location via GPS, a local base station, or triangulation, the tour guide may provide you with information on a building's past while simultaneously downloading details about a concert that will be taking place there that same night over a local wireless network. In order to enable, for instance, ad-hoc gaming networks as soon as players meet to play together, entertainment and games are a rising area of wireless network applications.

Mobile Computing's Limitation

A. **Resources limited:** Battery
B. **Interference:** Radio transmission cannot be shielded from interference, which leads to greater rates of transmitted data loss or bit errors, as appropriate.
C. **Bandwidth:** Despite constant growth, wireless devices' transmission speeds are still quite modest when compared to desktop systems. Researchers work towards low-overhead communication protocols that are more effective.
D. **Dynamic changes in the communication environment:** It include fluctuations in the signal strength within a given area, which causes link delays and connection losses.
E. **Network Issues:** Locating the destination connection-service and connection stability
F. **Issues with interoperability:** Caused by inconsistent protocol standards
G. **Security restrictions:** Radio interfaces are vulnerable to eavesdropping risks, and portable equipment are easier to steal. Encryption, authentication, and other security measures must always be present in wireless access, and they must be effective and easy to use.

A simplified reference model

In accordance with the reference model, the protocol stack implemented in the system is shown in the picture. End-systems, like the PDA and computer in the example, need a whole protocol stack, which includes the application layer, transport layer, network layer, data connection layer, and physical layer. The bottom layer services are used by end-system applications to interact with one another. All of the levels are not required necessary for intermediate systems like the interworking unit (Figure 1.1).

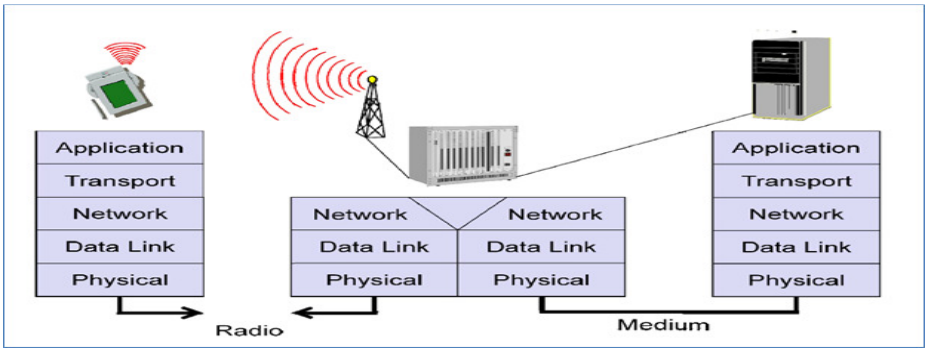


Figure 1.1: A simplified reference model

A personal digital assistant (PDA) is shown in the picture above as an example of a wireless and portable gadget. This PDA connects to the base station in the centre of the image to communicate. A radio transceiver (receiver and transmitter) and an interworking equipment that joins the fixed connection and wireless link together make up the base station. A traditional computer, the Personal Digital Assistant's communication partner, is seen on the right. The diagram displays the protocol stack that has been implemented in the system in accordance with the reference model under each network element (such as a PDA, interworking unit, or computer).

A whole protocol stack, including the application layer, transport layer, network layer, data link layer, and physical layer, is required by end-systems like the PDA and PC in the example. Applications on end systems use the bottom layer's services to interact with one another.

Not all levels are required for intermediate systems like interworking units. The network, data connection, and physical layers are shown in the above diagram. Since only entities at the same level may interact with one another (according to the reference model) (i.e. transport with transport, network with network).

Mobile communication's impact on the layer model

Physical layer

The conversion of a stream of bits into signals that are communicated on the sender side takes place at this layer, which is the lowest in a communication system. The signals are changed back into a bit stream by the physical layer of the receiver. The physical layer controls the creation of the carrier frequency, frequency choice, signal detection (although strong interference may obstruct the signal), modulation of the data into a carrier frequency, and encryption for wireless communication.

Data link layer

Accessing the media, multiplexing several data streams, correcting transmission problems, and synchronizing are the core duties of the data link layer (i.e. detection of a data frame). In a nutshell, the data link layer is in charge of establishing a trustworthy point-to-point connection between two devices or a point-to-multipoint connection between a single sender and a number of recipients.

Network layer

The third layer, referred to as the network layer, is in charge of carrying out network packet routing and creating connections between two entities across several other intermediary systems. Addressing, routing, device location, and handover across other networks are a few

of the subjects. Several options for the internet's network layer protocol (the Internet Protocol IP).

Transport layer

In the reference model, a link from end to end is established via the transport layer. Quality of service, flow, and congestion management are significant issues, particularly if the internet's TCP and UDP transport protocols are employed over a wireless network.

Application layer

The applications are positioned on top of all transmission orientated layers, supplemented by extra layers that may support applications. Service location, support for multimedia applications, adaptive programmes that can cope with changes in transmission characteristics, and wireless access to the World Wide Web via a portable device are some of the context on this layer. The most demanding applications are interactive games and video (high data rate) (low jitter, low latency).

GSM Services

The most popular digital mobile telecommunications network in use today is GSM. In more than 190 countries, it is utilised by more than 800 million individuals. GSM makes it possible to combine various voice and data services and communicate with current networks. Customers find a network appealing because of its services. Bearer, tele, and supplemental services are the three main kinds of services that GSM has established.

Provider services Different data transmission protocols are specified by GSM, with the original GSM allowing for non-voice service data speeds of up to 9600 bit/s. Transparent and opaque, synchronous and asynchronous data transfer is possible with bearer services.

Only the physical layer's (layer 1) capabilities are used by transparent bearer services for data transmission. If there are no transmission defects, data transmission has a constant latency and throughput.

Forward error correction (FEC), which incorporates redundancy into the data stream and aids in the reconstruction of the original data in the event of transmission mistakes, may enhance the quality of transmission. Transparent bearer services don't attempt to recover lost data in the event of, say, shadowing or handover-related pauses. Error correction and flow control are implemented via protocols at layers two and three in non-transparent bearer services. These services include a radio link protocol while using transparent bearer services (RLP). This protocol includes high-level data link control (HDLC) mechanisms and unique selective-reject techniques to force the retransmission of inaccurate data.

GSM defines a number of bearer services enabling interoperability with PSTN, ISDN, and packet switched public data networks (PSPDNs), such as the globally accessible X.25, using transparent and non-transparent services. Full-duplex, synchronous data transmission is possible at data rates of 1.2, 2.4, 4.8, and 9.6 kbit/s, or full-duplex, asynchronous data transfer at rates ranging from 300 to 9,600 bit/s.

GSM mostly concentrates on voice-oriented tele communications. These include message services, basic data connection with terminals that are familiar from the PSTN or ISDN, and encrypted voice transfer (e.g., fax). The provision of high-quality digital voice transmission was the main objective of GSM. While ordinary codecs are used to transmit analogue data for use with conventional computer modems found in, for example, fax machines, special codecs (coder/decoder) are utilised for voice transmission. The emergency number is another another

feature provided by GSM (eg 911, 999). All providers are required to use this service, and it is free. This connection will be established automatically with the nearest emergency centre and has the greatest priority, sometimes taking precedence over other connections. The short messaging service (SMS), which allows for the delivery of messages of up to 160 characters, is a helpful tool for extremely basic message transfer. SMS may be sent and received while being sent via voice or data. It can transmit logos, ring tones, horoscopes, and love notes in addition to "serious" uses like showing road conditions, email headers, or stock quotations.

The extended messaging service (EMS), which succeeded SMS, provides a bigger message size, structured text, and standardised delivery of animated graphics, tiny images, and ring tones. But EMS was seldom ever utilised with MMS. Larger images (GIF, JPG, WBMP), brief video clips, etc. may be sent through MMS, which is included with mobile phones that have tiny cameras. Group 3 fax is another non-voice tele service that is accessible globally. In this service, fax data is transferred as digital data using modems in accordance with ITU-T standards T.4 and T.30 across the analogue telephone network.

Additional services: GSM service providers may provide other services in addition to tele and bearer services. These services, which might differ from provider to provider, provide a variety of improvements for the basic telephonic service. User identification, call redirection or forwarding of current calls, blocking of incoming and outgoing calls, Advice of Charge (AoC), and other services are examples of typical services. There may be access to standard ISDN capabilities like multiparty communication and locked user groups.

GSM Architecture

The radio subsystem (RSS), the network and switching subsystem (NSS), and the operation subsystem make up a GSM system (OSS) (Figure 1.2).

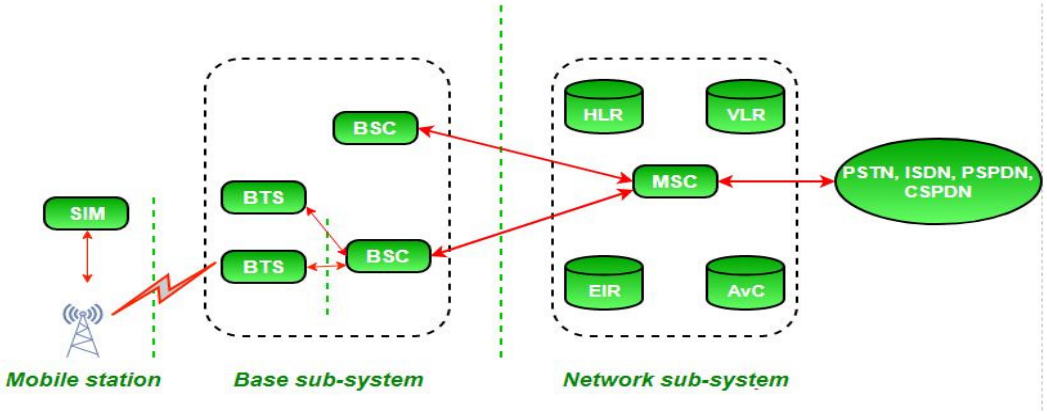


Figure 1.2: GSM Architecture

GSM System's Functional Architecture

Subsystem for switching networks: The NSS is in charge of carrying out subscriber-related tasks including call processing. The following functional units are part of the switching system:

Home Location Register (HLR) is a database for managing and storing subscriptions. HLR maintains a permanent record of information on subscribers, including service profiles, location data, and activity status. A person becomes registered in the HLR of that operator when they purchase a subscription from the PCS provider.

The MSC uses the visitor location register (VLR), a database with temporary subscriber information, to provide services to visiting subscribers. The MSC and VLR are always integrated. The VLR attached to that MSC will ask the HLR for information on the mobile station whenever an MS roams into a new MSC area. In the future, if the mobile station has to make a call, VLR will have all the necessary call setup information.

Authentication centre (AUC): The AUC is a component that offers authentication and encryption settings that confirm the callers' identities and guarantee call secrecy.

Equipment identification record (EIR): An EIR is a database that holds details about the identity of mobile equipment to stop calls from being placed from stations that are stolen, unlicensed, or broken.

The system's telephone switching operations are carried out by the mobile switching centre, or MSC. Calls from and to other telephone and data networks are controlled by it.

Mobile Stations (MS) and Base Station Subsystem: Together, these two radio-specific entities make up the Radio Subsystem (RSS) (BSS). In the illustration, the RSS and NSS are connected through the A interface (solid lines), while the OSS is connected via the O interface (dashed lines).

Base station subsystem (BSS): Each BSS in a GSM network is managed by a base station controller (BSC). The BSS carries out all tasks required to maintain radio connections with an MS, voice coding and decoding, and rate adaption to and from the wireless network component. The BSS also includes many BTSs in addition to a BSC.

BSCs, or base station controllers: All of the control mechanisms and physical connections between the MSC and BTS are provided by the BSC. This high capacity switch performs tasks including handover, cell configuration information, and radio frequency (RF) power level management in BTS. A MSC provides service to a number of BSCs.

Base transceiver station (BTS): The BTS manages the mobile station's radio interface.

Using sectorized antennas, a BTS may create many radio cells. It is linked to the MS through the Um interface and the BSC via the Abis interface. Each of the techniques required for wireless transmission is included in the Um interface (TDMA, FDMA etc.) The radio hardware (transceivers and antennas) required to serve each network cell is known as the BTS.

System for operation and support: All of the switching system's hardware and the BSC are linked to the operations and maintenance centre (OMC). Operation and support system is the term for OMC implementation (OSS). The OSS is the functional component that the network administrator uses to monitor and manage the system. The goal of OSS is to provide the client with efficient assistance for the local, regional, and centralised operational and maintenance tasks necessary for a GSM network. Engineers can monitor, identify, and debug any component of the GSM network using OSS, which offers a network overview.

The mobile station (MS) is made up of a smart card known as the Subscriber Identity Module and portable technology (the terminal) (SIM). Personal mobility is a feature of the SIM that enables access to subscription services for the user, regardless of the terminal being used. The user may make calls from that GSM terminal, receive calls at that terminal, and utilise other subscription services by putting the SIM card into another GSM terminal.

The International Mobile Equipment Identity uniquely identifies the mobile equipment (IMEI). The International Mobile Subscriber Identity (IMSI), which is needed to identify the system

subscriber, as well as other data are stored on the SIM card. Personal mobility is made possible by the independence between the IMEI and the IMSI.

A password or personal identification number may be used to secure the SIM card from illegal usage.

Wireless Interface

The radio interface, which includes numerous multiplexing and media access techniques, is the most intriguing interface in a GSM system. Using cells with BTS and allocating an MS to a BTS, GSM achieves SDMA.

Protocols for GSM

According to the interface, the GSM signalling protocol is divided into three general tiers, as seen below. The physical layer, Layer 1, is in charge of all radio-specific operations. This involves producing bursts in accordance with the five various formats, multiplexing those bursts into TDMA frames, synchronising with the BTS, identifying unused channels, and assessing the channel quality on the downlink. Data encryption and decryption are carried out at the physical layer at Um using GMSK for digital modulation, meaning that encryption is not done end-to-end but rather only between the MS and BSS via the air interface (Figure 1.3).

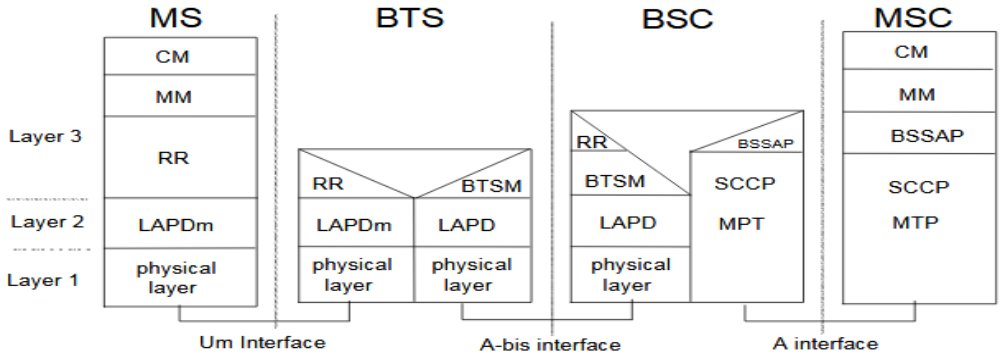


Figure 1.3: Protocols for GSM

Signaling protocol architecture

Channel coding and error detection and correction, which are intimately related to the coding techniques, make up the primary duties of the physical layer. Forward error correction (FEC) systems of many different types are widely used in channel coding.

Higher levels are needed in a GSM network for signalling between entities. At the layer two Um interface, the LAPDm protocol has been developed for this purpose. LAPDm is a variant of HDLC that was evolved from the link access procedure for the D-channel (LAPD) in ISDN systems. Because it doesn't need synchronisation flags or checksumming for error detection, LAPDm is a lightweight version of LAPD. Reordering of data frames, flow management, and reliable data transport across connections are all features of LAPDm.

In GSM, layer three, the network layer, is divided into multiple sublayers. The radio resource management layer is the bottom sublayer (RR). The functions of RR' are provided by the BSC through the BTS management, and only a portion of this layer, RR', is implemented in the BTS (BTSM). Setting up, maintaining, and releasing radio channels are the primary responsibilities

of RR. Mobility management (MM) functions include those for registering, authenticating, identifying, updating locations, and providing a temporary mobile subscriber identity (TMSI).

Last but not least, there are three components in the call management (CM) layer: call control (CC), short message service (SMS), and supplemental service (SS). SMS gives the ability to transport messages through the control channels SDCCH and SACCH, while SS provides features like user identification, call forwarding, and call redirection. Higher levels employ CC for call formation, call clearing, and call parameter changes since it offers a point-to-point connection between two terminals.

Dual tone multiple frequency (DTMF) tones, which are in-band tones that may be sent over the GSM network, are also made available by this layer. These tones are used for dialing in conventional analogue telephone systems as well as for other purposes, such as remote control of answering machines and PIN input in electronic banking.

The ABIS and an interfaces use additional protocols. Systems that employ pulse code modulation (PCM) are frequently used for data transfer at the physical layer. At Abis, layer two is managed by LAPD, and BTS management by BTSM. An MSC and a BSC communicate via Signaling System (SS7). All management data is also sent through this protocol between MSCs, HLRs, VLRs, AuCs, EIRs, and OMCs. A BSS application portion may also be used by an MSC to control a BSS (BSSAP).

Calling and localization

The automated, global localisation of users, for whom the system periodically updates location, is the core function of the GSM system. The HLR always has information about the current location, and the VLR in charge of the MS at the time notifies the HLR of any changes in position. Roaming is switching VLRs while maintaining availability. Roaming may happen inside a single provider's network, between two carriers in the same nation, or even between separate providers in different nations. Several numbers are required in order to find and contact an MS:

International Mobile Station ISDN Number (MSISDN): For a GSM subscriber, the phone number is the sole significant number. This number is made up of the subscriber number, the national destination code, and the country code (SN). IMSI stands for international mobile subscriber identity, and it is used by GSM to uniquely identify subscribers internally. The IMSI is made up of the mobile network code (MNC), the mobile subscriber identity number (MSIN), and the mobile country code (MCC) (MSIN).

Temporary mobile subscriber identity (TMSI): GSM employs the 4 byte TMSI for local subscriber identification in order to conceal the IMSI, which would reveal the precise identity of the user signalling over the air interface.

Mobile station roaming number (MSRN): This short-term address also conceals a subscriber's identity and location. This address is generated by the VLR in response to a request from the MSC, and it is also saved in the HLR.

The current visitor country code (VCC), visitor national destination code (VNDC), identity of the current MSC, and subscriber number are all included in the MSRN. In order to locate a subscriber for an incoming call, the HLR uses the MSRN. The stages involved in a mobile terminated call (MTC) are shown in the following figure 1.4:

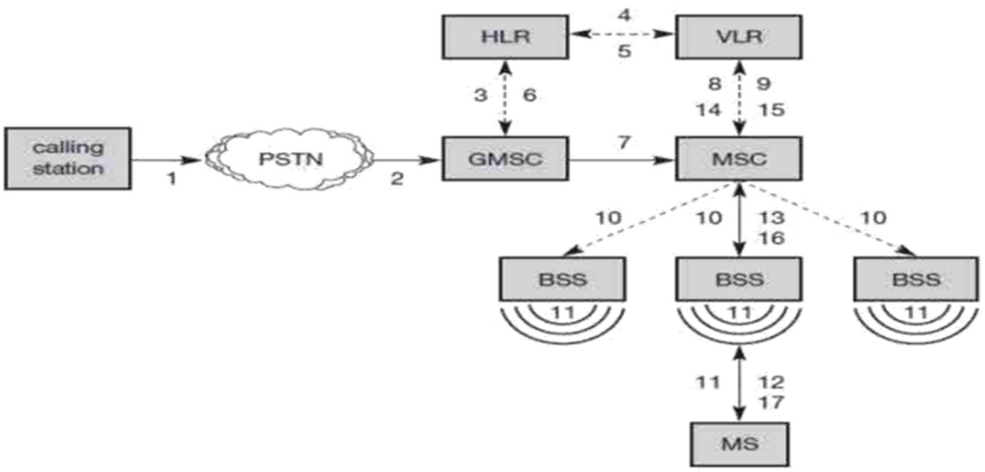


Figure 1.4: Calling and localization

Terminated mobile call (MTC)

Step 1: The user contacts a GSM subscriber's phone number.

Step 2: The fixed network (PSTN) determines that a given number belongs to a GSM network user and sends the call setup to the Gateway MSC (GMSC).

Step 3: The GMSC locates the subscriber's HLR and notifies it of the call setup.

Step 4: The HLR requests an MSRN from the current VLR after checking the number's existence and its subscribed services.

Step 5: VLR transmits HLR the MSRN.

Step 6: The HLR identifies the MSC responsible for MS after receiving the MSRN and sends the information to the GMSC.

Step 7: At this point, the GMSC may send the call setup request to the designated MSC.

Step 8: The MSC asks the VLR for the MS's most recent status.

Step 9: The VLR provides the needed information. Step 10: In the event that an MS is available, the MSC starts paging all of the cells under its control.

Step 11: The paging signal is sent to the MS by all BSS BTSs.

Step 12: If MS responds, step 13: VLR does security checks.

Steps 15 through 17: At that point, the VLR instructs the MSC to establish a connection with the MS.

The stages that occur for a mobile-originated call (MOC) are as follows:

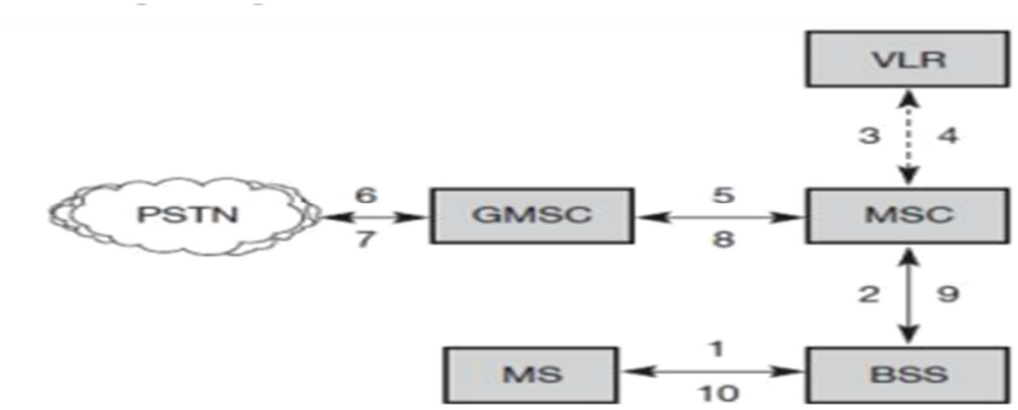


Figure 1.5: Mobile-Originated Call

Step 1: the MS sends a request for a new connection.

Step 2: The BSS sends the MSC this request.

Step 3: The MSC then determines if this user is authorized to initiate a call with the required party and determines whether resources are accessible through the GSM network and the PSTN. The MSC establishes a connection between the MS and the fixed network if all resources are available (Figure1.5).

Handover

Since a single cell cannot cover the whole service area, cellular systems need handover processes. But a handover shouldn't result in a cut-off, commonly known as a call drop. The maximum handover time that GSM aspires for is 60 ms. there are basically two motives for a handover:

1. When a mobile station leaves a BTS's coverage area, the received signal level drops, the error rate goes up, and the radio link's quality declines.
2. Load balancing may result in handover when an MSC/BSC determines that the volume of traffic in one cell is too high and moves certain MS to other cells with a lower load.

The following four GSM handover situations are illustrated (Figure 1.6):

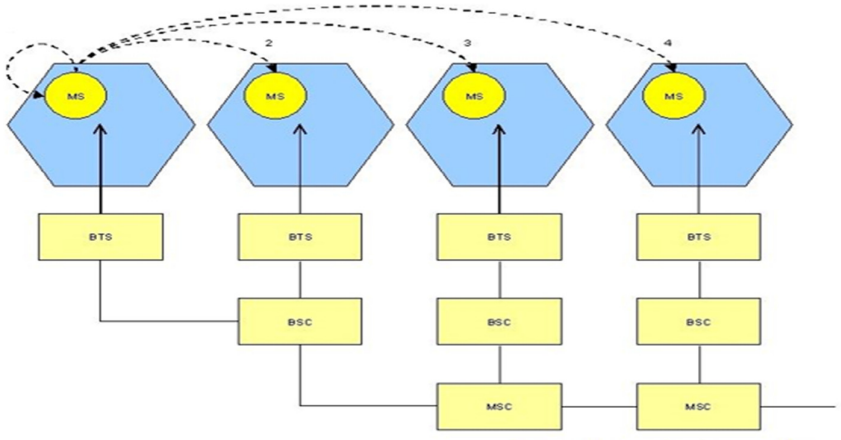


Figure 1.6: GSM handover situations

Handover inside a cell: Narrow-band interference may prevent transmission at a particular frequency within a cell. After then, the BSC could opt to alter the carrier frequency (scenario)

This is a common handover scenario: inter-cell, intra-BSC. While moving between cells, the mobile station remains under the same BSC's management. Following a handover, the BSC releases the old cell and provides a new radio channel to the new cell (scenario 2).

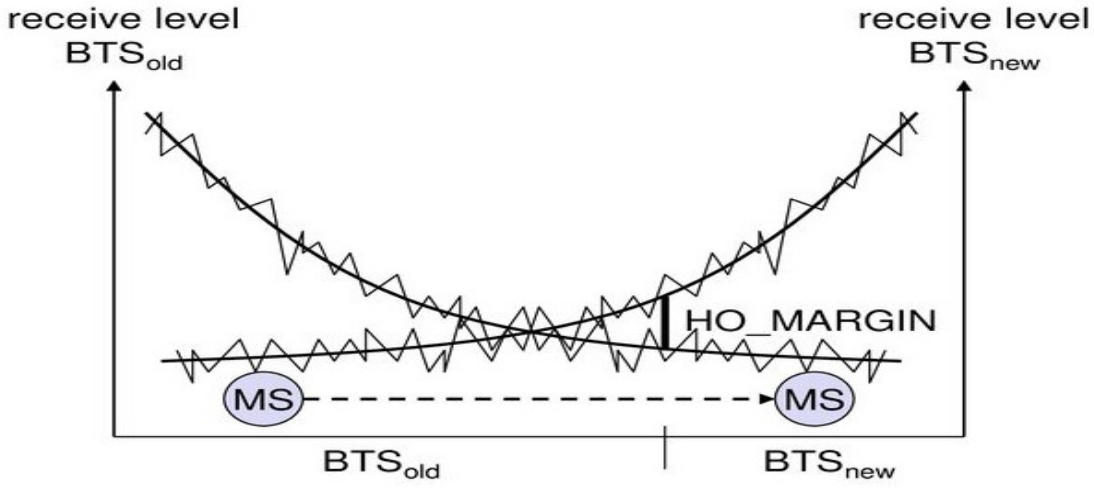


Figure 1.8: Conventional Received Signal Strength Based Handover

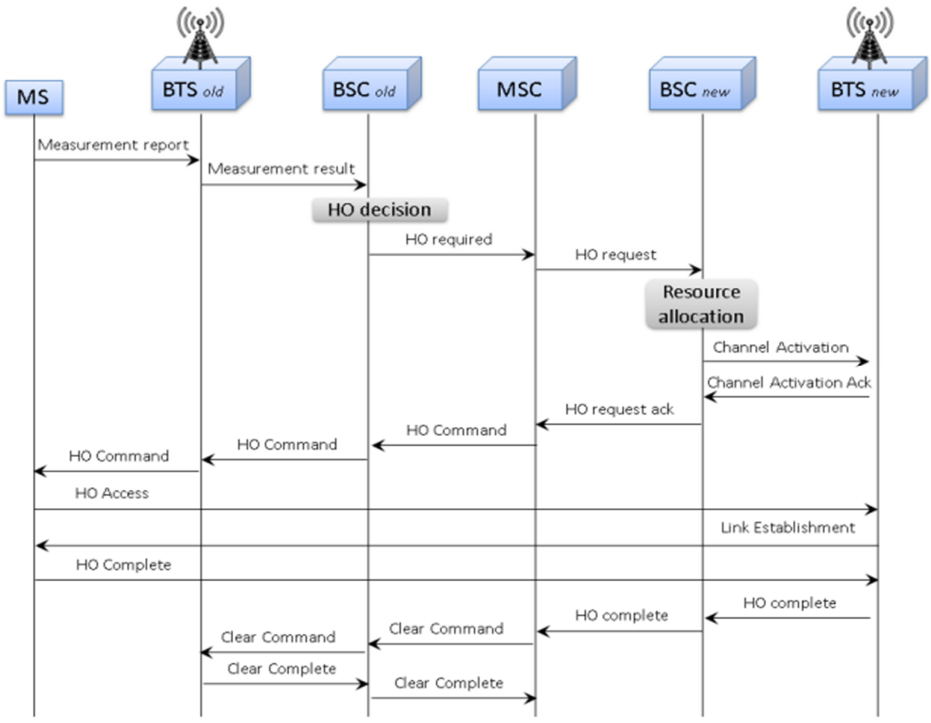


Figure 1.9: depicts the typical end-to-end call flow between the entities during the Handover procedure.

Handover between cells managed by different BSCs: Because a BSC can only handle a certain number of cells, GSM also has to execute handovers between cells controlled by other BSCs. The MSC must then supervise this transfer (scenario 3).

Inter-MSC handover: A handover between two cells from distinct MSCs may be necessary. Together, the two MSCs now complete the handover (scenario 4).

The uplink and downlink quality are periodically measured by MS and BTS in order to give all the information required for a handover due to a poor connection. Every half-second or so, the MS sends measurement reports that include information on the quality of the current connection being utilized for transmission as well as the quality of specific channels in nearby cells (the BCCHs). For smooth handovers across various systems, more advanced handover methods are required. Standard Received Signal Strength-Based Handover (Figure 1.8). The usual end-to-end call flow between the entities throughout the handover process is shown in the following diagram 1.9:

Security

GSM uses private data kept in the AuC and the individual SIM to provide a number of security services. The SIM is password-protected to prevent unauthorized usage and saves private, confidential information. To offer security services in GSM, three specific algorithms have been defined. A3 is used for authentication, A5 is used for encryption, and A8 is used for creating cypher keys. GSM provides a number of security services, including: Authentication and access control Authenticating a legitimate user for the SIM is the first step. To access the SIM, the user requires a private PIN. The subscriber authentication process comes next. This process is based on the challenge-response diagram 1.10 below:

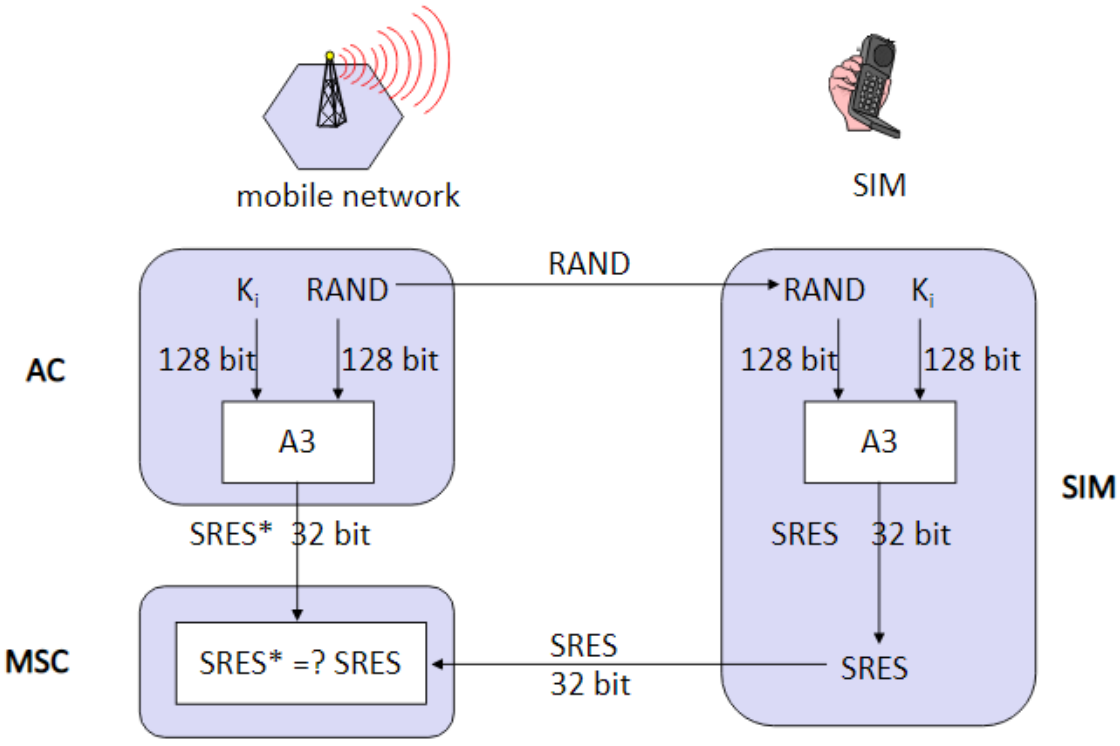


Figure 1.10: GSM Authentication

Subscription Verification

The SIM, which holds the unique authentication key Ki, the user identity IMSI, and the authentication algorithm A3, is the foundation for authentication. For each IMSI, the AuC generates the fundamental random values RAND, signed answers SRES, and cypher keys Kc before sending the data to the HLR. The present VLR asks the HLR for the necessary RAND, SRES, and Kc values. The VLR transmits to the SIM the random value RAND for authentication. The identical action is carried out by the network and subscriber module on both sides using RAND and the key Ki, also known as A3. The SRES produced by the SIM is sent back by the MS, allowing the VLR to compare the two numbers. The subscriber is accepted by the VLR if they match; if not, they are refused.

Confidentiality: All user-related data is encrypted to ensure confidentiality. Following authentication, speech, data, and signaling are encrypted by BTS and MS as indicated below.

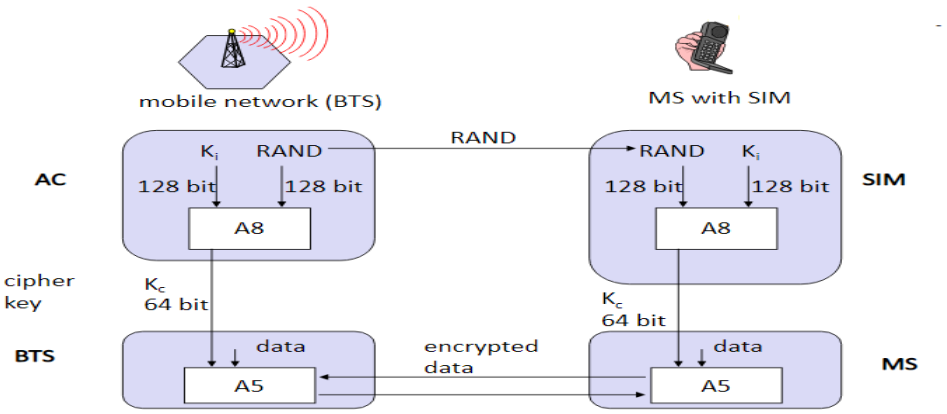


Figure 1.11: GSM key generation and encryption.

All communications sent through the GSM over the air interface are encrypted to protect user privacy. After authentication, MS and BSS may begin encrypting data by using the cypher key Kc, which is created by applying the algorithm A8 to the individual key Ki and a random number. Be aware that the network and the SIM in the MS base their Kc calculations on the same random number, RAND. The actual key Kc is not sent through the air interface. With the help of the cypher key Kc and the algorithm A5, MS and BTS may now encrypt and decode data. User anonymity is ensured through the encryption of all data before to transmission and the avoidance of over-the-air usage of user identification. Instead, after each position update, the VLR generates a new temporary identification (TMSI), which GSM uses to transfer data. Additionally, the TMSI may be modified at any moment by the VLR (Figure 1.11).

Data Services New

There are two fundamental strategies that may be used to improve GSM's data transfer capability. Since the foundation of GSM is connection-oriented traffic channels, such as those with 9.6 kbit/s apiece, it is possible to combine many channels to boost bandwidth. HSCSD, or "high speed circuit switched data," is the name of this technology. The introduction of packet-oriented traffic in GSM, or the paradigm change from connections/telephone thinking to packets/internet thinking, is a more advanced move. The general packet radio service (GPRS) is the name of the system.

High Speed Circuit Switched Data (HSCD): This simple enhancement to GSM's data transmission capabilities allows for greater data rates to be reached by bundling several TCHs.

An MS allots many TDMA slots inside a TDMA frame by requesting one or more TCHs from the GSM network. As more slots may be assigned on the downlink than the uplink, this allocation may be asymmetrical, which is in line with the normal user behaviour of downloading more data than uploading. The fact that HSCD continues to employ GSM's connection-oriented procedures, which are ineffective for computer data transfer, is a significant drawback.

GPRS: Being totally packet-oriented, this next step in flexible and powerful data transfer overcomes the issues of HSCSD. According to the requirement specification, the general packet radio service (GPRS) offers packet mode transfer for applications that exhibit traffic patterns like frequent transmission of small volumes (for example, typical web requests) or infrequent transmissions of small or medium volumes (for example, typical web responses). The GSM system may allot one to eight time slots inside a TDMA frame for the new GPRS radio channels.

The distribution of time slots is demand-based rather than set and predetermined. The active users may share each time slot; uplink and downlink are given distinct allotments of time. The slots are assigned depending on the operator's preferences and the current load. The GPRS concept does not impose a maximum data rate cap and is independent of channel characteristics and channel type (conventional GSM traffic or control channel) (only the GSM transport system limits the rate). The usage of all GPRS services is possible concurrently with other services. Authentication, access control, user identity secrecy, and user information confidentiality are just a few of the security features offered by GPRS.

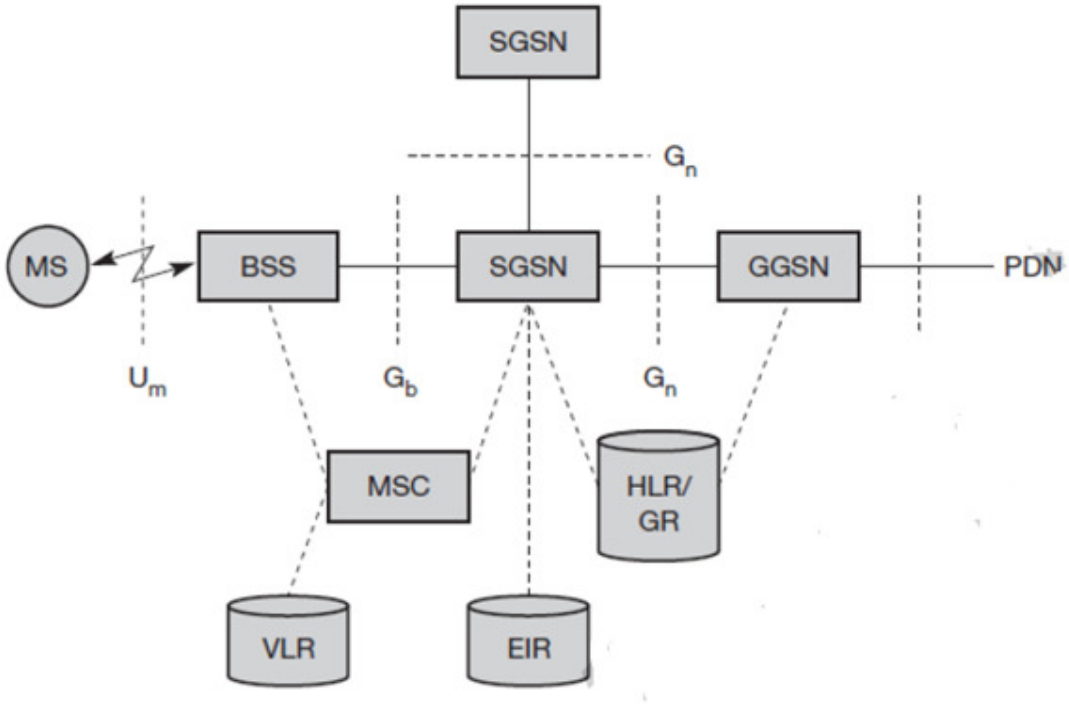


Figure 1.12: GPRS Architecture Reference Model

Two additional network components, referred to as GPRS support nodes (GSN), which are really routers, are introduced by the GPRS design. The standard GSM architecture has been integrated with all GSNs, and several additional interfaces have been established. The gateway GPRS support node (GGSN) serves as an interface between the external packet data networks

and the GPRS network (PDN). This node converts addresses, stores routing information for GPRS users, and encapsulates data before tunnelling it to a user. Through the Gi interface, the GGSN connects to external networks (such as IP or X.25) and sends packets to the SGSN across an IP-based GPRS backbone network (Gn interface). The serving GPRS support node (SGSN), which supports the MS through the Gb interface, is the second new component. For instance, the SGSN obtains user addresses from the GPRS register (GR), maintains track of the specific MSs' locations, is in charge of gathering billing data (such as bytes counted) and carries out many security tasks, such as access control. The SGSN is essentially on the same hierarchy level as an MSC and is linked to a BSC through frame relay. All GPRS-related data is kept in the GR, which is often a component of the HLR.

Architecture Reference Model for GPRS

As previously mentioned, packet data is sent from a PDN straight to the BSS and then to the MS after passing via the GGSN and SGSN. In the GPRS scenario, the MSC, which handles data transfer in the conventional circuit-switched GSM, is simply used for signaling. An MS must connect to the GPRS network by following the mobility management's instructions before transferring any data over it. A temporary logical link identification (TLLI) and a ciphering key sequence number (CKSN) are issued as part of the attachment operation to enable data encryption. A GPRS context is created for each MS and stored both in the MS and the associated SGSN. Mobility management includes features for ciphering, location management, and authentication in addition to attaching and detaching (Figure 1.12).

The protocol architecture of the GPRS transmission plane is shown in the following diagram. The GPRS tunnelling protocol is used to transport all data inside the GPRS backbone, or between the GSNs (GTP). GTP may employ either the dependable TCP (required for the dependable transmission of X.25 packets) or the unreliable UDP (used for IP packets). The GPRS backbone's network protocol is IP (using any lower layers). Between an SGSN and the MS, the subnetwork dependent convergence protocol (SNDCP) is utilised to adjust to the various properties of the underlying networks. User packet data is tunnelled from the MS to the GGSN and vice versa on top of SNDCP and GTP. A specific LLC is employed, which includes ARQ and FEC procedures for PTP (and subsequently PTM) services, to provide high reliability of packet transmission between SGSN and MS.

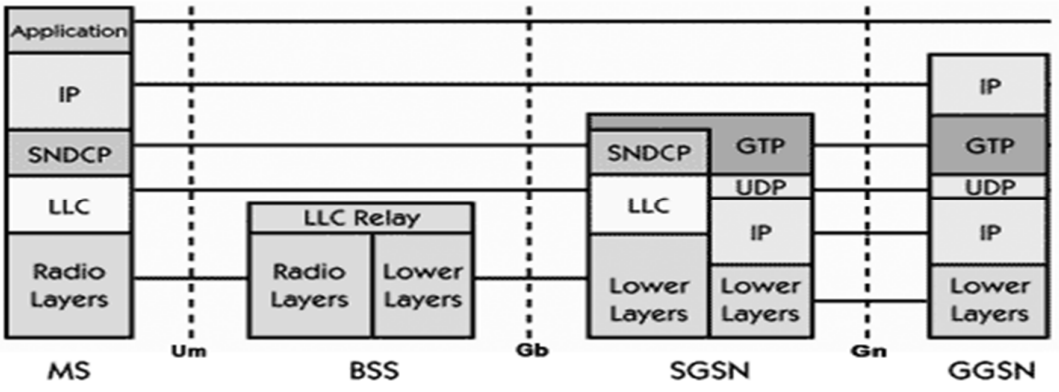


Figure 1.13: GPRS transmission Plane Protocol References Model

Information about routing and QoS is sent between the BSS and SGSN via the base station subsystem GPRS protocol (BSSGP). BSSGP operates on top of a frame relay (FR) network but does not do error correction. Finally, data transmission via the Um interface requires radio link dependent methods. A reliable connection is provided via the radio link protocol (RLC),

and access is controlled by the MAC through signalling protocols for the radio channel and the mapping of LLC frames into the GSM physical channels. In comparison to regular GSM, the radio interface at Um required for GPRS does not need significant modifications (Figure 1.13).

CHAPTER 2

ARCHITECTURE OF THE MOBILE COMPUTING

Lokesh Lodha, Associate Professor,
Department of Electronics and Communication, School of Engineering & Technology, Jaipur National University, Jaipur, India,
Email Id- lokesh.lodha@jnujaipur.ac.in

An architectural strategy that combines cloud-based resources with the processing power of devices such as smartphones or tablets is referred to as mobile cloud computing architecture. Mobile cloud computing (MCC) devices may add resources from various cloud-based accounts remotely rather than locally due to computational advancements. Three key levels make up the three-tier architecture that enables mobile computing, so let's investigate each layer and its functions in turn. Layers of the mobile computing architecture, as shown in Figure 2.1.

- **Presentation Layer (UI):** Users can interact with device handling and rendering through this layer.
- **Application Layer (AL):** This layer enables the execution of rules and business logic.
- **Data Access Layer (DM):** It enables administration and access.

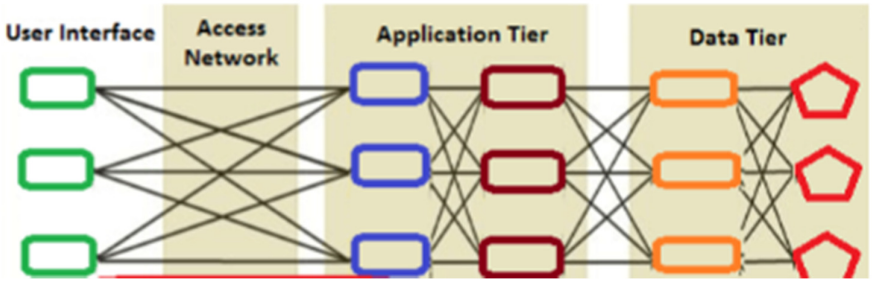


Figure 2.1: Representation of Architecture of the Mobile Computing [2]**.**

1-Tier Presentation Layer:

1. This presentation layer enables full user interfaces and application execution on client devices.
2. Information presentation to edge users is its primary duty.
3. Users have access to all information via screens, speakers, vibration, etc.
4. With the use of input devices like pen drives, mice, keyboards, touch displays, and so forth, users can send information.
5. WAP browsers, specialized client programs, web browsers, and other tools are available for this layer.
6. By using a client-side data source, dynamic HTML, and data cursors, the presentation layer enables accomplishment.
7. The presentation layer must be both device- and context-aware.

2-Tier Application Layer:

Business logic completes all tasks as a server for client requests from workstations at the application layer. The Data Layer is used to fetch or enter data according to business requirements.

A few technologies, including PHP, .Net services, JSP, Java, and others, are used to enable it.

Independent of presentation and databases.

This layer determines the types of data that are necessary and acts as the client in relation to third-tier programming that may be located locally or on a mainframe computer.

Decisions must be made regarding rendering, network administration, security, data store access, and the necessity of various middleware's.

This layer is used as a repository for both temporary and permanent data and to retain data that applications need in storage.

3-Tier Data Access Layer:

The DBMS that provides all of the data for the previous two layers makes up the data access layer.

The DBMS Access Layer is another name for this layer.

All data is saved in a variety of formats, including relational databases and text files.

In this layer, it is possible to update or change without the application tier customers being aware of the change by ignoring the dependencies on the storage mechanism.

Mobile Computing Architecture Benefits:

Data is stored more securely as a result of the installation of the application layer between the presentation and data layers.

The process of updating to new geographic contexts is simpler and faster.

The presentation layer can be used to hide extraneous application layer methods.

Mobile Computing Architecture Drawbacks:

The structure is complicated.

It takes a lot of effort and time to construct.

Aspects of Mobile Computing:

1. The greatest platform for delivering the numerous services and applications found on the internet is software as a service.
2. Utilizing the cloud allows you to scale up while keeping costs under control and maximize your resources.
3. Provide recommendations for the top mobile tools and programs that you can use for utilities and programs.
4. With the aid of this notion, you can connect to any wireless network at any time and access data anywhere without location or time restrictions.
5. Provide the most recent and appropriate operating system and other software for mobile terminals.
6. Check how long it takes for specific program to load.
7. Configure the wireless LAN and WAN networks.

Concept of Mobile Computing:

A wide communication coverage diameter is made possible by mobile computer technology. It is among the quickest and most dependable subfields of computing technology. Three components make up the idea of mobile computing:

Mobile Communication:

A framework for the operation of mobile computer technology is defined by mobile communication. In this context, "mobile communication" refers to a system that enables dependable and smooth wireless device communication. The consistency and dependability of wireless device connection is ensured by this architecture. The mobile communication framework consists of communication tools including protocols, services, bandwidth, and portals required to enable and facilitate the aforementioned services. These gadgets are in charge of facilitating easy communication. The following four categories can be used to categorize mobile communication:

Fixed and Wired: In a Fixed and Wired arrangement, the devices are fixed in place and linked together physically so that they may communicate with one another.

Use a desktop computer as an example.

Fixed and Wireless: The devices in a Fixed and Wireless setup are fixed in place and connected to other devices via a wireless link to enable communication.

For instance, a WiFi router or a communication tower

Wired and mobile: Some devices in a mobile and wired arrangement are mobile and some are wired. Together, they enable connectivity with other gadgets.

Take laptops, as an example.

Mobile and Wireless: The devices can communicate with one another in a mobile and wireless setup regardless of where they are located. Without using any connected connections, they can connect to any network. Example. Wi Fi Dongle.

Mobile Hardware:

Mobile hardware consists of mobiles or mobile device parts. It refers to portable electronic devices, such as laptops, tablet PCs, smartphones, smartwatches, and other PDAs, that have access to mobile services (Personal Digital Assistants). Receptors are fitted in such gadgets to perceive and receive impulses. These devices have full-duplex capability, which allows them to send and receive signals simultaneously.

Mobile Software:

A program called mobile software can be used with mobile hardware and can also make calls to a mobile operating system. It addresses the needs of mobile applications and continues to be in charge of the device engine's entire range of operations.

Types of Mobile Computing:

Mobile computing categories can be separated into two groups:

Portable Computing: Wired communication is required. Although all users are free to move their devices whenever and wherever they want, they must have access to a network line in order to join.

Mobility Computing: It also goes by the name "Mobile Computing," and it refers to wireless communication. With the help of these mobile computing devices, it offers a better environment for users to communicate data from one spot to another at any time without any physical connections.

Advantages of Mobile Computing:

Compared to traditional computers, mobile devices are more compact and portable, making them convenient to use in a variety of settings. They function without electricity, without a direct network connection, and while they are not linked to the network.

Mobile devices have decreased in cost and ease of access throughout time. People are increasingly choosing smartphones and tablets as their main online connectivity devices. A smartphone is frequently less expensive to purchase than a desktop computer.

Wireless transmissions, Users of mobile devices can use applications for text, instant messaging, phone, and video communication.

Companies can now gather more consumer data than was previously feasible through mobile devices and applications. Mobile devices, for instance, can record the user's and the device's geolocation. A marketing firm can target adverts to a user at a specific time by matching the location of a device with the user's behavior on the device using a behavioral analytics program. This is a typical neuro-marketing technique. Additionally, some mobile applications gather biometric information from users, such heart rate measurements, which can be used to track health.

Disadvantages of Mobile Computing:

Due to the vulnerability of many mobile computing devices, users may also experience security difficulties. Despite having a better battery, these devices consume more power. Cost also plays a significant role in consumers' disadvantages, since new technologies' costs rise along with them, making it more challenging for users to purchase.

Mobile Computing Devices

Devices for Mobile Computing:

A mobile computing device typically consists of a metal or plastic body, a RAM, a CPU, a hard drive, a motherboard, a keyboard and mouse that may be distinct body parts or touch-based, a screen, a video card, an operating system, software programs, and finally a network connection. This is comparable to the parts of a personal computer, a non-mobile device. However, mobile devices may also contain additional parts that help make them portable and unique features.

Size: Smaller sizes are necessary for portable electronics. It has always been difficult to reduce size without compromising functionality while creating mobile devices.

Power Source: Rechargeable batteries are typically used to power mobile gadgets. Another important field of research is extending the battery life of mobile devices.

Operating system: Laptops and PCs both use roughly the same OS, but cellphones and other devices have a very distinct OS. They are robust but sized down and created especially for specific devices.

Connectivity: Mobile computing devices have features that enable internet access. The ability to make and receive phone calls is also available on mobile devices like smartphones thanks to mobile broadband networks.

Applications: Programs intended for mobile devices are created specifically to run on a certain OS. The ability of devices to do more than just connect to the internet or make calls is extended by these applications. In addition, mobile computer devices typically have GPS, accelerometer, compass, microphone, camera, and other functionalities.

Mobile computing technology has advanced significantly. Numerous older gadgets, such as the Personal Digital Assistant (PDA), have been phased away. In addition to smartphones, there are a number of different types of mobile computing devices

Laptop

A laptop is a personal computer that is transportable. The OS, programs, and files can all operate on it because it is designed to provide the same functionality as a PC. An alphanumeric keyboard and screen make up a compact, portable personal computer (PC) known as a laptop, laptop computer, or notebook computer. Although 2-in-1 PCs with a detachable keyboard are frequently advertised as laptops or as having a "laptop mode," laptops traditionally feature a clamshell form factor with the screen mounted on the inside of the higher lid and the keyboard mounted on the inside of the lower lid. Laptops are appropriate for mobile use because they can be folded shut for transporting. They are referred to as "lap" computers because when in use, they may essentially be placed on a person's lap. Today, laptops are utilized in many different contexts, including work, school, and gaming, browsing the web, personal multimedia, and basic home computing.

Smartphone

A smartphone is a cell phone that has advanced features. They frequently incorporate features like a camera and GPS, have a touchscreen interface, internet access, the ability to run different applications, and more. An enhanced version of a standard cell phone is a smartphone. Smartphones can access the Internet via Wi-Fi or a cellular network in addition to the same fundamental features calls, voicemail, and text messaging, which requires purchasing a monthly data plan. This implies that you may perform the same tasks on a smartphone that you would typically perform on a computer, such as checking your email, browsing the web, or making purchases online.

Tablet computer

Often viewed as a bridge between a laptop and a smartphone, tablets have touchscreens and virtual keyboards. They outperform smartphones in terms of processing speed, functionality, and screen resolution. Some models could also come with a stylus for easier touch screen use. Handwriting is digitized and can be turned into conventional text by handwriting recognition or it is typed in text form in notebook computers, which additionally allow for the use of a stylus. On a pen-based key layout, where letter keys are placed differently than on a typical QWERTY keyboard, the stylus can also be used to type.

Wearable

The term "wearable computer" can be used in a narrow or broad sense, including smartphones and even typical wristwatches. Wearables might be intended for everyday use in which case they are merely a very specific type of mobile computing. Alternately, they might be used for niche functions like fitness trackers. They could include unique sensors like accelerometers, heart rate monitors, or, for more sophisticated models, electrocardiogram (ECG) and blood oxygen saturation (SpO_2) monitors. We also consider cutting-edge user interfaces like Google Glass, an optical head-mounted display controlled by gestures, within the umbrella of wearable computers. It's possible that specialized wearables will converge with generic all-in-one technology to become smartphones, just as PDAs and mobile phones did with smartphones. More recently introduced wearable computers, such as smartwatches, offer a small number of functionalities that are similar to those of a smartphone in a watch.

E-reader

Although e-readers resemble tablets, their primary use is reading digital documents. Similar to tablet computers, e-book readers, sometimes known as e-readers, are primarily intended for reading e-books. The Kobo, Nook, and Amazon Kindle are illustrative instances. A standard computer display is more difficult to read than the e-ink screens used by the majority of e-readers. Even in direct sunshine, you can read as if it were a conventional book.

CHAPTER 3

MEDIUM ACCESS CONTROL

Dr.Sudhir Kumar Sharma, Professor,
Department of Electronics and Communication, School of Engineering & Technology, Jaipur National University, Jaipur, India,
Email Id- hodece_sadtm@jnujaipur.ac.in

Each network node is individually identified by a hardware identifier called a Medium Access Control (MAC) address. It offers addressing and channel access control features that allow communication between several terminals or network nodes in a given network. Media Access Control is another name for the data transmission protocol's Medium Access Control. The Data Link Control (DLC) layer in the IEEE 802 OSI Reference model of computer networking is split into two sub-layers:

The Medium Access Control (MAC) layer and the Logical Link Control (LLC) layer

The logical link control (LLC) Ethernet sublayer and the physical layer of the reference model are directly connected via the MAC sublayer. As a result, a distinct MAC layer is needed for each kind of network media. The node address on networks that participate in the OSI Reference Model but do not adhere to IEEE 802 standards is known as the Data Link Control (DLC) address. In a multipoint network system, the MAC sublayer simulates a full-duplex logical communication channel. These routes of communication could provide broadcast, multicast, or unicast communication services.

Control of Medium Access When several devices are connected to the same physical connection, MAC sublayer LLC and MAC Sublayer MAC addresses are appropriate to avoid collisions by allowing the system to uniquely identify each device at the data link layer. MAC addresses are issued to all ports on a switch. In order to avoid collisions, the MAC sublayer employs MAC protocols. MAC protocols make use of the MAC algorithm, which takes as inputs a secret key and an arbitrary-length message that has to be verified and outputs a MAC address.

The MAC sublayer carries out the following tasks:

According to IEEE Std 802-2001, the MAC layer's main duties are as follows:

Delimiting and recognizing frames: This function is in charge of setting up and identifying frames.

Addressing: The MAC sublayer handles the transmission of source-station addressing information as well as the addressing of destination stations (both as individual stations and as groups of stations).

Transparent data transmission: It handles the Ethernet sublayer's data transparency for the transfer of LLCs, PDUs, or analogous information.

Protection: The purpose of the MAC sublayer is to safeguard data from mistakes, often by creating and examining frame check sequences.

Access control refers to the prevention of illegal access to the physical transmission medium.

For wired networks, one of the most often used MAC sublayers is Carrier Sense Multiple Access with Collision Detection (CSMA/CD). Before transmitting data, a sender detects the

medium a wire or coaxial cable through the MAC scheme to determine if the medium is free to do so. The sender waits until the medium is free if MAC detects that it is busy. When the medium is open, the sender begins transferring data while continuing to listen into it. If the transmitter detects any form of collision while transferring data, it immediately pauses and transmits a jamming signal. However, wireless networks do not operate well with this plan. The following are a few of the issues that might arise while transferring data through wireless networks;

Signal strength drops in inverse proportion to distance squared. Carrier Sense (CS) and Collision Detection (CD) would be used by the transmitter, but collisions take place at the receiver. It's possible that a sender won't be able to "hear" the collision since the CD won't operate. Additionally, CS may not function, for instance, if the terminals are "hidden".

Hidden and Exposed Terminal

The hidden terminal issue occurs in wireless LANs (wireless local area networks) when two or more stations that are out of each other's range concurrently broadcast to a single receiver. In decentralized networks, where no one body is in charge of regulating transmissions, this is common. This happens when a station is concealed from other stations that are communicating with a wireless access point (AP), but is visible from the AP itself.

Situational Illustration

Assume there are three stations with the call letters STA, STB, and STC, with STA and STC broadcasting and STB receiving. The two transmitters, STA and STC, are not within radio range of one another due to the arrangement of the stations. This is seen in the image 3.1 below.

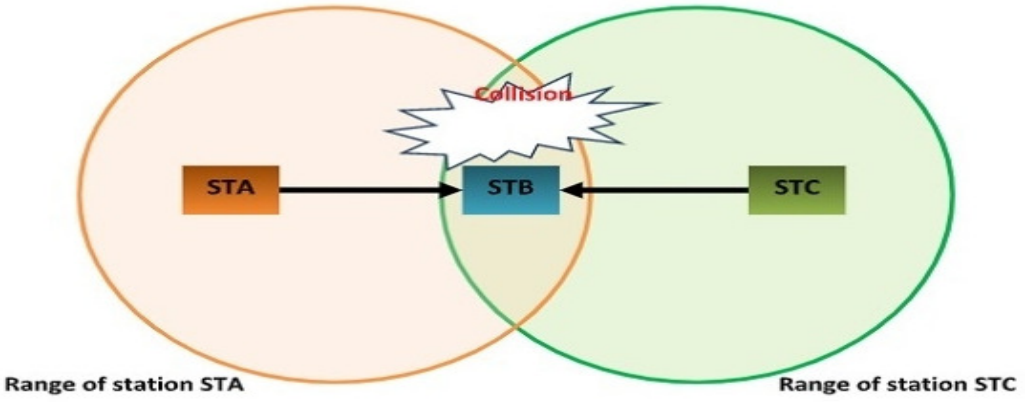


Figure 3.1: Situational Illustration

The figure up above illustrates how station STA begins broadcasting to station STB. Station STC senses that the channel is open since it is outside of STA's radio range and begins broadcasting to STB. STC receives jumbled frames that sometimes collide. The buried terminal issue is the name given to this circumstance.

Solution

The MAC (medium access control) layer protocol IEEE 802.11 RTS/CTS, under the condition that the stations are synchronised and frame sizes and data speed are the same, solves the exposed terminal issue. RTS and CTS stand for request to send and clear to send, respectively.

A RTS frame is sent from the transmitting station to the receiving station. The responding station sends a CTS frame in response. The transmitting station starts broadcasting as soon as it receives a CTS frame. Any station that hears the RTS is near the transmitter and keeps quiet long enough for the CTS to arrive. During the data transfer, any station that hears the CTS is near to the receiving station and keeps quiet.

In the aforementioned illustration, station STC hears CTS frame from station STB but not RTS from station STA. As a result, it delays transmission since it recognises that the STB is busy, preventing a collision.

Spatial Division Multiple Access (SDMA): In mobile communication systems, mean spatial division multiple access (SDMA) is a channel access technique that makes use of the same set of cell phone frequencies over a particular service region. If two cells or tiny areas are separated by an acceptable distance, they may utilise the same set of frequencies (called the reuse distance). By concentrating the signal into precise transmission beams, SDMA improves the system's capacity and transmission quality. SDMA provides service to several users in the same area by using intelligent antennas with beams directed in the general direction of the mobile station. Mobile stations operating outside the range of these directed beams encounter almost no interference from other mobile stations using the same radio frequency as them but under a different base station. The radio energy frequency may have a greater base station range due to the focussed nature of the beams. Due to this feature of SDMA, base stations may provide a wider radio coverage while emitting less energy. Greater gain and clarity are also made possible by this small beam width. In conventional mobile phone network architectures, the base station transmits radio signals without being aware of the position of the mobile station across the cell. The position of the mobile station determines how radio signals are channelled using SDMA technology. The SDMA design avoids duplicate signal transmission in locations where mobile devices are not in use at the moment and conserves important network resources in this way. Frequency reuse is SDMA's key benefit. Even if mobile stations utilise the same allotted frequencies, interference may be close to nil if the reuse distance is kept in the network design.

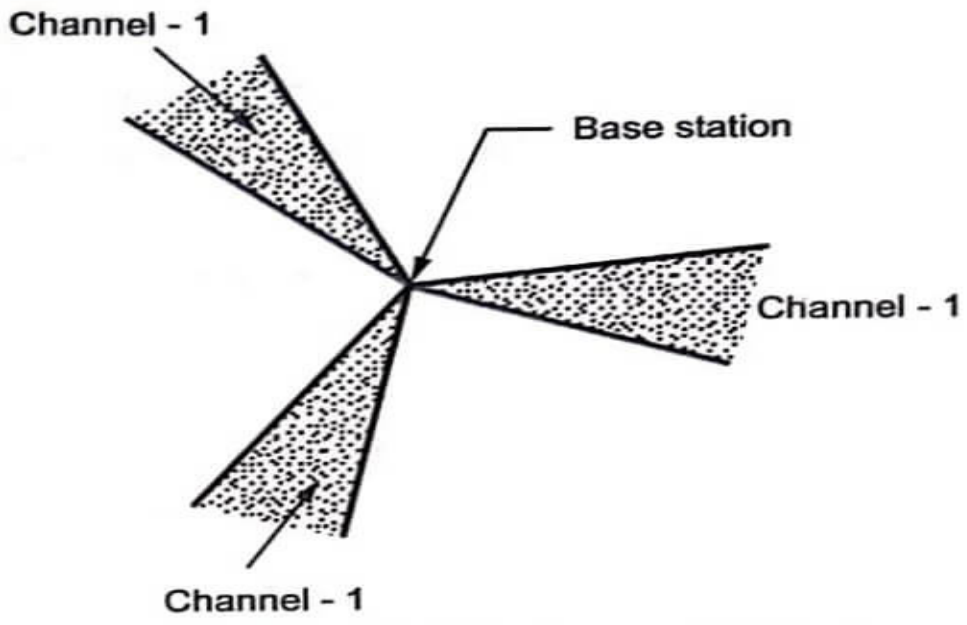


Figure 3.2: Principle of SDMA

Principle of SDMA: A specific area of space is targeted by a narrow radio wave beam. Another tiny beam pointed towards a different area of the space is covered by the same channel once again. Space Partition Multiple Access refers to this division of space from the base station in many directions by highly directed beams (SDMA), as shown in Figure 3.2.

The area is partitioned as shown above, and three channels are broadcast on the same frequency.

Advantages

- The channel bandwidth is preserved.
- Increases the bandwidth's usefulness.
- Wired and wireless communications and the function of SDMA
- Both mobile and satellite communication are possible with SDMA. Signals from the satellite dish antennas are sent to different parts of the earth's surface. They have excellent directionality. Therefore, as illustrated in Fig., the same frequency may be utilised for numerous surface zones.

Area A1 and area A3 are physically separate, as indicated in Fig. Therefore, using highly directional attennas, the same channel-1 is utilized to deliver signals to antennas A1 and A3. The signals of regions A1 and A3 will not interact in any way.

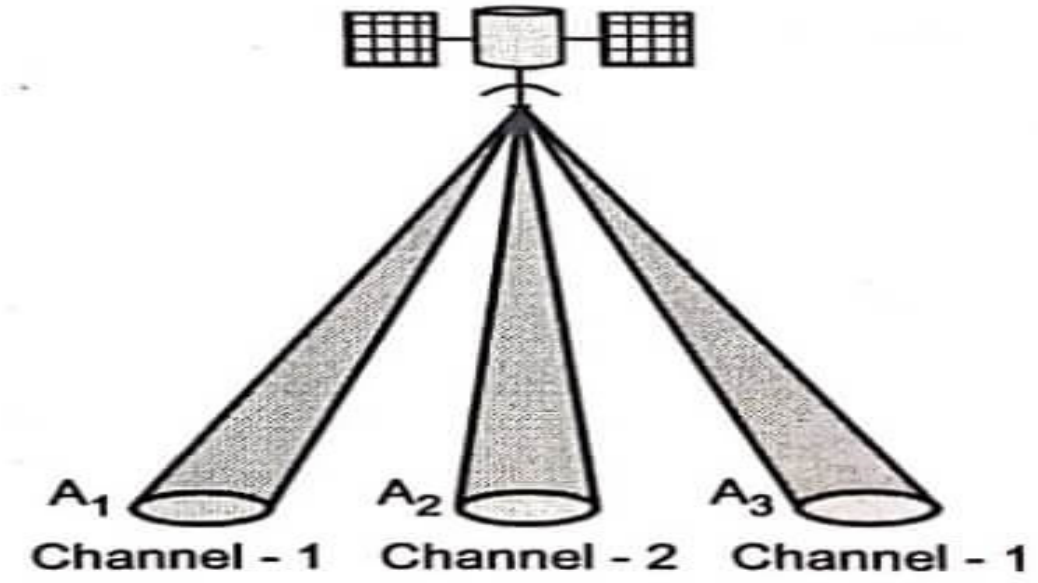


Figure 3.3: Frequency reuse by SDMA inter channel interference.

To prevent co-channel and SDMA inter channel interference, satellite-based SDMA needed accurate antenna alignment and careful zone (area) selection for each transmitter (Figure 3.3). Controlling the transmitting antennas' strength is necessary in cellular (mobile) communication to prevent co-channel and inter channel interference. At the base station of cellular communication, multidirectional horn antennas are used for Space Division Multiple Access (SDMA) (BS). Mobile users are recognized by the base station using their spatial signatures. All of the broadcast signals on the forward connection are completely within the base station's control in terms of power. To prevent cross-channel interference, the broadcast power from each mobile user is dynamically adjusted. Each mobile user's power level is detected by the base station, which then connects it. There are other adaptive antennas in use.

Multiple access protocol- ALOHA, CSMA, CSMA/CA and CSMA/CD

Data Link Layer

In a computer network, the data connection layer is used to convey data between two devices or nodes. The multiple access resolution/protocol and data connection control are two examples of how it splits the layer into sections. The higher layer is referred to as logical data link control because it is responsible for flow control and error control in the data link layer. While the bottom sub-layer is used to manage and minimise channel collisions or multiple access. The multiple access resolutions or media access control are the terms used to describe it.

Data Link Control

A data link control is a dependable method of directing the flow of data packets in a computer network utilising different methods including framing, error control, and flow control.

Multiple access protocol

The data link control may manage the channel when a sender and receiver have a dedicated connection to deliver data packets. Assume there is no specific route for the data to be sent or communicated between the two devices. In such situation, a number of stations concurrently access the channel and send data across it. Collisions and crosstalk might result. As a result, the multiple access protocol is necessary to decrease collision and prevent crosstalk between channels. Consider a classroom full of kids as an example. When a teacher poses a question, all of the pupils (little channels) in the class immediately begin to respond (transferring the data simultaneously). Due to the simultaneous responses from all pupils, data overlaps or is lost. Therefore, it is the teacher's duty (multiple access protocol) to control the pupils and force them to provide a single response. The sorts of multiple access protocols that are separated into several processes are as follows:

Random Access Protocol

All stations in this protocol have an equal priority to transfer data across a channel. One or more stations cannot control or be dependent upon another station while using the random access protocol. Each station sends the data frame in accordance with the active or inactive status of the channel. However, there might be a collision or data conflict if more than one station provides the data over the same channel. The data frame packets may be altered or lost as a result of the collision. As a result, the receiver end does not receive. The various random-access protocols for broadcasting frames on the channel are listed below in the Figure 3.3.

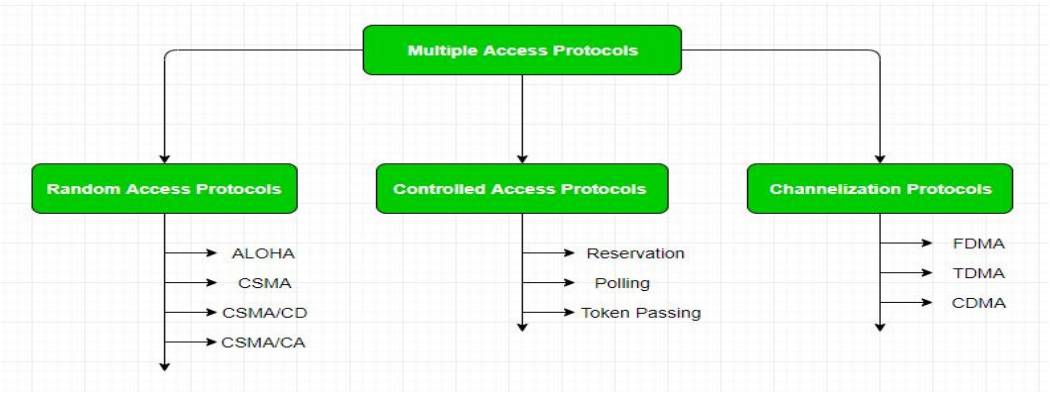


Figure 3.3: Types of Random Access Protocol

Types of Random Access Protocol

- Aloha
- CSMA
- CSMA/CD
- CSMA/CA
- ALOHA

Although it is intended for wireless LAN (Local Area Network), it may also be used to send data across a shared media. When a data frameset is available for transmission, any station may use this approach to concurrently broadcast data across a network.

Aloha Rules

1. Data transmission to a channel is always possible from any station.
2. No carrier sensing is necessary.
3. Data frames may be lost or collide when being sent across several stations.
4. Aloha acknowledges the existence of the frames. As a result, collision detection is absent.
5. Data must be retransmitted after an arbitrary period of time.

Types of Aloha

There are two types of aloha as shown in the Figure 3.4.

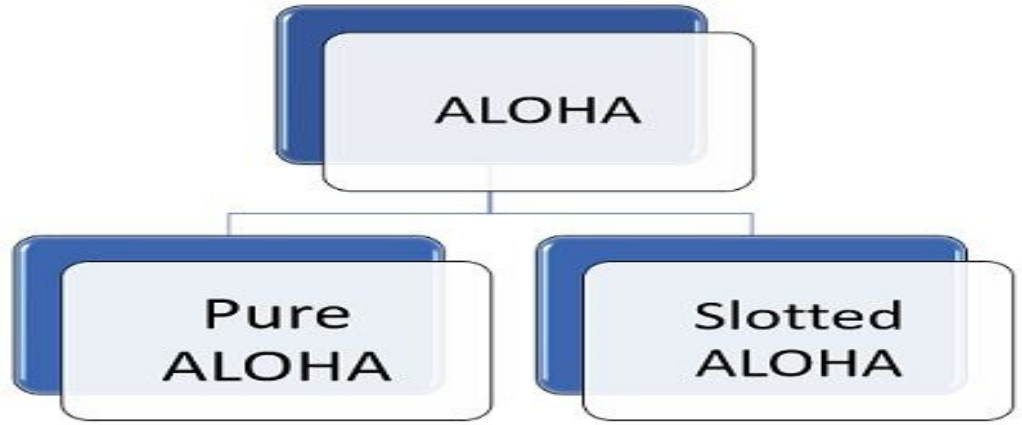


Figure 3.4: Types of Aloha

Pure Aloha

We employ Pure Aloha whenever data can be sent across a channel at stations. When each station broadcasts data to a channel in pure Aloha without first confirming if the channel is free or not, there is a potential that a collision may take place and the data frame would be lost. The pure Aloha waits for the receiver to acknowledge transmission of a data frame to a channel by any station. The station waits for a random period of time, known as the back off time, if it does not recognise the receiving end within the allotted amount of time (Tb). Furthermore, the station can presume that the frame has been lost or destroyed. As a result, it sends the frame again until the recipient successfully receives all of the data.

1. Pure Aloha has a total vulnerability of 2 * Tfr.
2. The throughput is at its highest when G = 1/2, or 18.4%.
3. The formula for a successful data frame transfer is S = G * e - 2 G.

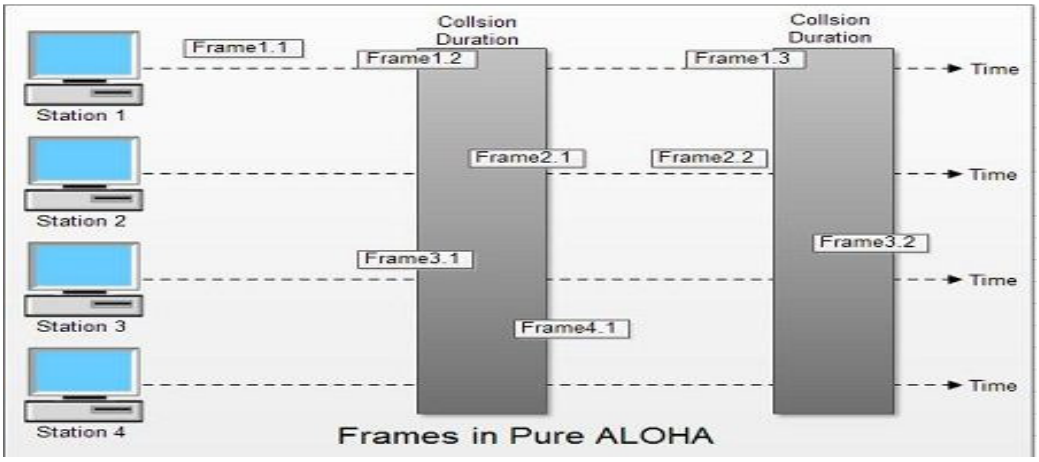


Figure 3.5: Pure Aloha.

There are four stations for accessing a common channel and sending data frames, as shown in the above diagram 3.5. Due to the fact that most stations deliver their frames at the same time, some frames clash. Frames 1.1 and 2.2 are the only ones that are successfully transferred to the receiving end. Other frames are also lost or damaged at the same time. Collisions may happen whenever two frames land on a common channel at the same time, and both will suffer damage. If the initial bit of the new frame enters the channel before the final bit of the previous frame has finished. The data frame must be sent again since both frames have completed entirely.

Slotted Aloha

Pure Aloha has a very high chance of striking a frame, therefore the slotted Aloha is made to outperform its efficiency. Slotted Aloha divides the shared channel into set time intervals known as slots. As a result, only the first frame may be transmitted to each slot by a station if it wishes to send a frame to a shared channel, and only one frame may be sent to each slot overall (Figure 3.6). The station will have to wait until the beginning of the slot for the following time if it is unable to transmit data to the beginning of the slot. When attempting to transmit a frame at the start of two or more station time slots, the risk of a collision still exists.

1. The slotted Aloha's maximum throughput occurs when G = 1, or 37%.

2. S = G * e - 2 G is the chance of sending the data frame correctly in the slotted Aloha.

3. Tfr is the overall amount of vulnerable time needed in scheduled Aloha.

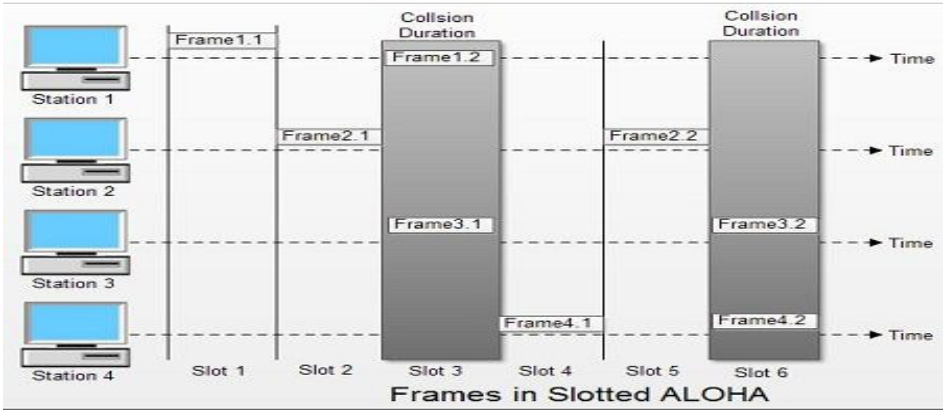


Figure 3.6: Slotted Aloha.

CSMA (Carrier Sense Multiple Access)

Before transferring the data, a carrier sensing multiple access system that uses the media access protocol detects if a channel is busy or idle. It indicates that the station may transmit data to the channel even if it is not in use. If not, it must wait until the channel is empty. The likelihood of a collision on a transmission medium is thereby decreased.

Access Modes for CSMA

1-Persistent: Each node in the CSMA 1-Persistent mode, which specifies it, senses the shared channel first, and if the channel is empty, delivers the data right away. Otherwise, it must wait for the channel to be idle, monitor its state, and broadcast the frame without condition as soon as it is.

Non-Persistent: This CSMA access mode stipulates that each node must first feel the channel in order to send data; if the channel is inactive, the data is sent right away. If not, the station must wait for an arbitrary amount of time (not constantly), and when it discovers the channel is empty, it sends the frames.

1-Persistent and Non-persistent modes are combined to form P-Persistent. Each node observes the channel in the P-Persistent mode, and if the channel is idle, it transmits a frame with a P probability. If the data is not transferred, the frame restarts with the next time slot after waiting for a (q = 1-p probability) random period.

Prior to the transmission of the frame on the shared channel, the station's dominance is determined using an O-persistent approach. Each station waits for its time to broadcast the data again if it is discovered that the channel is dormant (Figure 3.7).

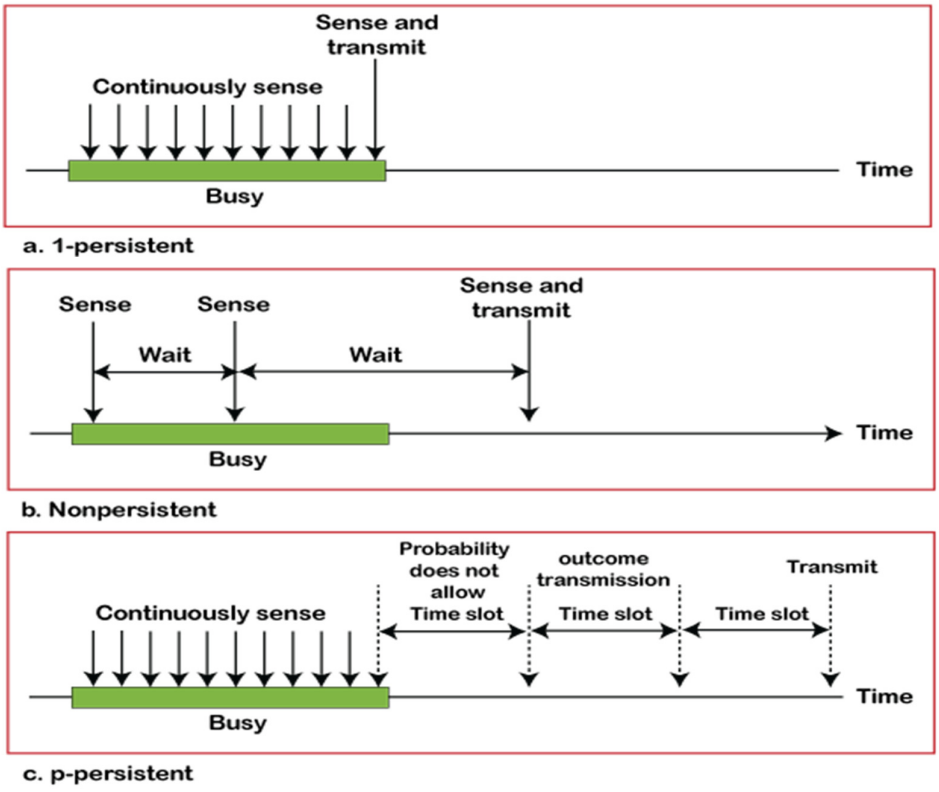


Figure 3.7: Access Modes for CSMA

CSMA/ CD

To transfer data frames, a carrier sensing multiple access/collision detection network protocol is used. The medium access control layer is used by the CSMA/CD protocol. Therefore, before broadcasting the frames, it detects the shared channel. If the channel is empty, it sends a frame to verify that the transmission was successful. The station transmits another frame if the first one was successfully received. The station sends a jam/stop signal to the shared channel to cease data transmission if any collision is found in the CSMA/CD. Following that, it holds off transmitting a frame to a channel for an unknown amount of time.

CSMA/ CA

For carrier transmission of data frames, it is a carrier sense multiple access/collision avoidance network protocol. It uses a medium access control layer and is a protocol. Data frames are acknowledged after being transmitted to a channel in order to determine if the channel is clear. The data frame has successfully been delivered to the receiver if the station only gets one (own) acknowledgement. However, if it receives two signals—its own and another in which frames collide—a frame collision happens in the common channel. Detects the frame collision when a sender gets a signal of acknowledgement.

The strategies used by the CSMA/ CA to prevent collisions are as follows:

Space between frames: In this strategy, the station waits for the channel to become empty before sending the data. If the channel is empty, the station does not transmit the data right away. Instead, it waits for a while; this time frame is referred to as the Interframe space, or IFS. The priority of the station is often determined by the IFS time, however.

Window for contention: The complete amount of time is split into several slots in the Contention window. The station/sender selects a random slot number of slots as a wait period when it is ready to transmit the data frame. It merely refreshes the timer to deliver data packets when the channel is dormant and does not resume the full procedure if the channel is still active.

Acknowledgment: If the acknowledgement is not received beforehand, the sender station uses the acknowledgment technique to deliver the data frame to the shared channel.

Protocol for Controlled Access

On a shared channel, it is a technique for reducing data frame collision. In the controlled access technique, every station engages in communication and chooses whether to provide a data frame that has been accepted by every other station. This implies that until all other stations are rejected, a single station cannot provide the data frames. Reservation, polling, and token passing are the three different forms of regulated access available.

Protocols for channelization

It is a channelization technique that enables many stations to split the entire useable bandwidth on a shared channel according to their time, location, and codes. To transfer the data frames to the channel, it may simultaneously access all of the stations.

According to their time, location, and codes, the following are the numerous ways to view the channel:

1. FDMA (Frequency Division Multiple Access)
2. TDMA (Time Division Multiple Access)
3. CDMA (Code Division Multiple Access)

FDMA (Frequency Division Multiple Access)

It uses the frequency division multiple access (FDMA) technique to split the available bandwidth into equal bands, allowing numerous users to transmit data to the subchannel using various frequencies. To avoid interference from other stations and channel crosstalk, a specific band is set aside for each station (Figure 3.8).

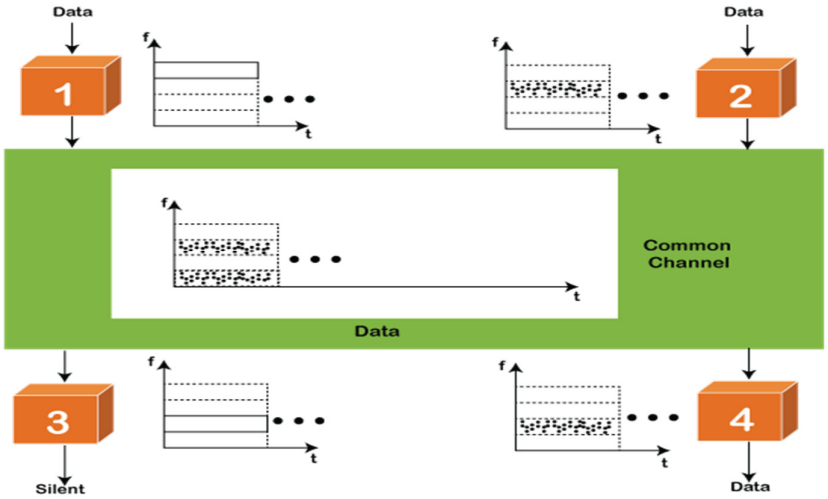


Figure 3.8: Frequency Division Multiple Access

Time Division Multiple Access (TDMA)

A channel access technique is called Time Division Multiple Access (TDMA). It enables numerous stations to share the same frequency spectrum. Additionally, it separates the channel into several frequency slots that assign stations to transmit the data frames in order to prevent collisions in the common channel. By separating the signal into several time slots, it may be sent across the same frequency bandwidth. However, TDMA includes a synchronisation overhead that adds synchronisation bits to each time slot to specify each station's time slot.

CDMA

A channel access technique is code division multiple access (CDMA). In CDMA, all stations may transfer data over the same channel at once. In other words, it always permits each station to send data frames on the shared channel at full frequency. It is not necessary to divide the available bandwidth on a shared channel into time periods. When many stations submit data to a channel at once, each station's data frames are separated by a different coding pattern. For data transmission across a shared channel, each station uses a unique code that is distinct from the others. For instance, there are several people conversing constantly in a room. If just two people engage with each other using the same language, data is received by the users. Similar situations arise in a network when many stations concurrently connect with one another using various coding languages.

CHAPTER 4

MOBILE NETWORK LAYER: MOBILE IP

Chandra Shekhar Rajora, Assistant Professor,
Department of Electronics and Communication, School of Engineering & Technology, Jaipur National University, Jaipur, India,
Email Id- chandra.shekhar@jnujaipur.ac.in

his is a standard communication protocol developed by the IETF (Internet Engineering Task Force) that enables users of mobile devices (such as laptops, PDAs, mobile phones, etc.) to switch between networks while keeping their permanent IP (Internet Protocol) address. Mobile IP is an improvement to the internet protocol (IP) that provides methods for forwarding internet traffic to mobile devices (also known as mobile nodes) while they are connected over a network other than their home network, as described in RFC (Request for Comments) 2002. The instance that follows illustrates how a datagram transfers inside the Mobile IP framework from one location to another. The internet host first sends a datagram to the mobile node using its residential address (normal IP routing process). The datagram is sent to the mobile node (MN) via the standard IP (Internet Protocol) procedure if it is connected to its home network. If not, the datagram is picked up by the home agent. The home agent (HA) sent the datagram to the foreign agent if the mobile node (MN) was on a different network. The datagram is sent to the mobile node by the foreign agent (FA). Using standard IP routing techniques, datagrams are transferred from the MN to the Internet host. The packets are sent to the foreign agent if the mobile node is connected to a foreign network. The FA transmits the datagram to the host over the Internet. The aforementioned examples of wireless communications show how to send datagrams to a mobile node using wireless transceivers. Additionally, regardless of whether the mobile node is on a domestic or international network, all datagrams between the Internet host and the MN utilise the mobile node's home address. The Internet host never sees the care-of address (COA), which is exclusively used for communication with mobility agents (Figure 4.1).

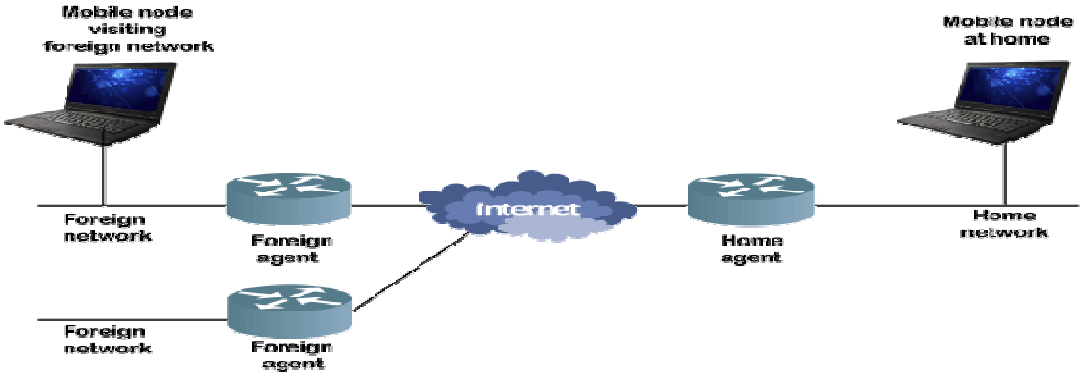


Figure 4.1: Mobile IP

Mobile IP components

The following three elements make up the mobile IP:

1. Mobile Node (MN)

The mobile node is a system or device that has network roaming capabilities, such as a cell phone, PDA (Personal Digital Assistant), or laptop.

2. Home Agent (HA)

The mobile node receives a number of services from the home agent, which is situated in the home network. At the home agent, the packet tunnel leading to the mobile node begins. The current COA informs the home agent of the mobile node's position in order to maintain a location register (care of address). There are the following options for implementing a HA. A router in charge of the home network may be equipped with a home agent. Since all packets for the MN must pass via the router even with mobile IP optimization, this is undoubtedly the ideal place. The home agent might potentially be installed on any unspecified node in the subset if updating the router's firmware is not an option. If the MN is on a different network, the packet will double-cross the router, which is one of the greatest drawbacks of this system. The router receives a packet for the mobile node; the HA passes it over the tunnel, which once again crosses the router.

3. Foreign Worker (FA)

The mobile node may get a variety of services from the foreign agent while it is connected to the foreign network. The COA (care or address) may be used by the FA to send packets to the MN and function as a tunnel endpoint. The foreign agent may serve as the MN's default router. Because they are a part of a foreign network as opposed to an MN who is merely visiting, foreign agents are also qualified to provide security services. In essence, FA is a router that transmits packets from the home agent to the mobile node and may serve as the point of attachment for the mobile node when it roams to a foreign network.

4. Addressing Care (COA)

From an IP perspective, the Care- of address identifies the mobile node's present location. Instead than going straight to the MN's IP address, all IP packets directed to the MN are sent to the COA. A tunnel is used to transport packets to the mobile node. To be more exact, the COA identifies the address at which packets leave the tunnel as the endpoint of the tunnel.

The location of the care of address might be any of two places:

Foreign Agent COA: The COA may be found at the foreign agent; in other words, the COA is the foreign agent's IP address. The tunnel endpoint, which is the foreign agent, passes packets to the MN. This COA may be shared as a standard COA by several MN who use the FA.

Co-located COA: If the MN has temporarily obtained a second IP address that serves as a COA, the COA is said to be co-located. The tunnel endpoint is located at the mobile node, and this address is now topologically accurate. DHCP and other services may be used to get a co-located address. If MNs desire a COA, one issue with this strategy is the need for additional addresses. Given the limited supply of IPv4 addresses, this is not always a wise decision.

5. Correspondent Node (CN) Communication requires a minimum of one partner. This partner is represented by the corresponding node for the MN. A permanent or moving node might serve as the corresponding node.

6. Home Network

The subset of networks to which the MN belongs in terms of its IP address is the home network. Within this network, mobile IP capability is not required.

7. Global network

The MN currently browses a subset of the international network, which is distinct from the domestic network.

IP process for mobile

The three primary stages of the mobile IP process are as follows:

1. Agent Research

HA and FA use the ICMP router discovery protocol to promote their services on the network during the agent discovery phase (IROP).

Agent solicitation and agent marketing, which are really extensions of router discovery techniques, are the two methods that Mobile IP describes.

Agent promotion: In the first technique, FA and HA occasionally use unique agent promotion messages to promote their existence. You might think of these messages as beacons emitted into the network. Internet control message protocol (ICMP) messages in accordance with RFC 1256 are used together with certain mobility enhancements for this advertising.

Solicitation of agents: The mobile node must send agent solicitations if there are no agent advertising, the inter arrival time is too long, and an MN has not yet received a COA. These requests are based once again on RFC 1256 for router requests.

2. Registration The primary goal of registration is to provide the home agent with updated location information so that packets may be forwarded correctly.

Depending on where the COA is located, there are two procedures to register a mobile IP address.

The MN submits its registration request to the FA, which is sending the request to the HA, if the COA is at the FA. The HA has now established a mobility binding using the home IP address and current COA of the mobile node.

The lifespan of the registration, which is negotiated upon registration, is also included in the mobility bidding. A mobile node should register before it expires since registration expires automatically after the lifespan and is erased. The HA sends a reply message back to the FA after configuring the mobility binding, who then sends it to the MN.

Registration might be a lot easier if the COA was also housed there. Direct requests from the mobile node to the HA are possible, and vice versa. By the way, MNs returning to their home network must also complete this registration process.

Mobile IP is required

Because a portion of an IP address specifies the network to which a host is connected, IP addresses are created to function with stationary hosts. Without ending ongoing connections and resuming them again after acquiring a new address, a host cannot change its IP address. Although there are other link layer mobility options, they are insufficient for the global Internet. The capacity of a node to switch points of attachment while retaining all ongoing connections and utilizing the same IP address is known as mobility.

A node with nomadic status is able to relocate, but only after terminating all ongoing conversations and opening up fresh connections using a new address. On the global Internet, mobile IP is a scalable, reliable, and secure network layer solution for homogeneous and heterogeneous mobility that enables nodes to keep all active connections while moving. Design Objectives to address mobile user issues in a transparent manner, mobile IP was created. The goal of mobile IP was to minimize the size and frequency of necessary routing updates. It was

created to make the implementation of mobile node software easier. It was created to steer clear of solutions that need the usage of multiple addresses by mobile nodes.

Requirements: To become a standard, Mobile IP must meet a number of criteria. Among them are:

1. Compatibility: The internet's whole infrastructure is exceedingly complex, and a new standard cannot make modifications to the network protocols or applications now in use. The old operating systems will include mobile IP. Additionally, rather than altering the routers, which is very difficult, it would be able to increase their capacity to accommodate mobility. Finally, end-systems upgraded with a mobile IP implementation should still be able to connect with stationary systems without mobile IP, and mobile IP must not need specific media or MAC/LLC protocols in order to reach the lower levels.

2. Transparency: For many upper layer protocols and applications, mobility is opaque. Even though the mobile computer's point of connection to the network has changed due to higher layers continue to function even if there is a reduced bandwidth and occasional service interruptions. Mobility will result in increased latency and lesser bandwidth since many of today's apps were not designed to be used in mobile situations.

3. Scalability and efficiency: Even if a new mechanism is added to the internet, the network's effectiveness shouldn't be impacted. The network must not be overrun with additional communications as a result of IP enhancement for mobility. Because wireless networks have a smaller bandwidth, more caution must be exercised. A wireless connection to an attachment point is included in many mobile systems. Therefore, just a small number of extra packets should be required to communicate between a mobile system and a network node. A mobile IP must be scalable over a huge number of users on the whole internet, anywhere in the globe.

4. Security: There are various security issues with mobility. All communications pertaining to the control of mobile IP must at the very least be authenticated. When sending a packet to a mobile host, the IP layer must be certain that the host is indeed the intended recipient. Only the receiver's IP address may be guaranteed by the IP layer. Attacks using false IP addresses and other methods cannot be stopped. 'Supporting end-system mobility while retaining scalability, efficiency, and compatibility in all aspects with current applications and Internet protocols' may be summed up as the objective of a mobile IP (Figure 4.2).

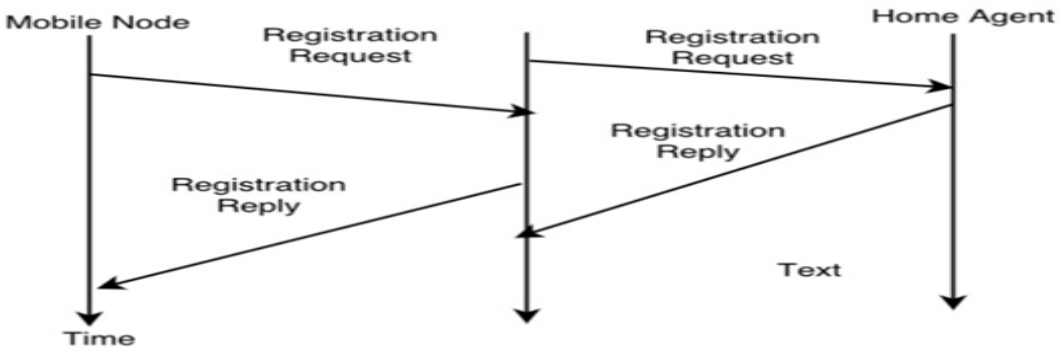


Figure 4.2: Registration

3. Tunneling

Between a tunnel entrance and a tunnel endpoint, a tunnel creates a virtual conduit for data packets. Packets entering a tunnel are sent within the tunnel and remain unaltered when they

exit the tunnel. With the use of encapsulation, tunneling, or transmitting a packet via a tunnel, is accomplished. The transfer of data across a public network that is solely intended for use inside a private network typically a business network is referred to as tunneling, sometimes known as "port forwarding."

IP packet transmission

Think about the scenario in which a correspondent node (CN) wishes to transmit an IP packet to the MN in the example above. Support for disguising the MN's movement was one of the prerequisites for mobile IP. Without needing to know MN's precise location, CN delivers the packet as normal to its IP address, which is shown in the below figure 4.3.

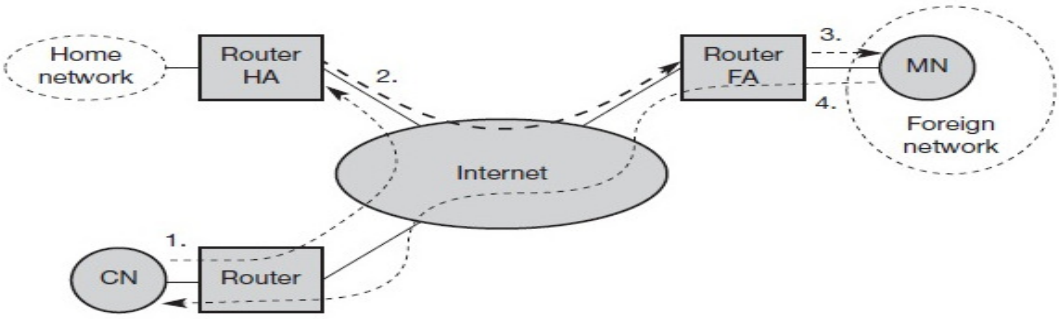


Figure 4.3: IP packet transmission

MN and CN are the source and destination addresses of an IP packet sent by CN. The packet is sent to the router in charge of MN's home network via the internet since it doesn't know MN's current location. The internet's standard routing 4 techniques are used for this. Knowing that MN is not now connected to its home network, the HA now intercepts the packet. Instead of being sent as normal into the subnet, the packet is encapsulated and tunneled to the COA. The previous IP header is replaced with a new one that displays COA as the new destination and HA as the source of the encapsulated packet (step 2). Encapsulating the packet now, the foreign agent transmits the original packet with CN as the source and MN as the destination to the MN (step 3). Once again, MN movement is not evident. As it would have done on the home network, it gets the packet with the identical sender and recipient addresses Figure 4.4.

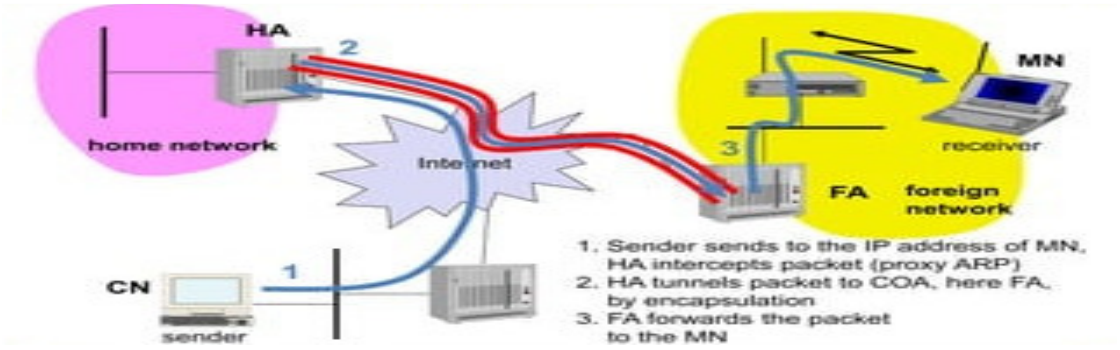


Figure 4.4: Data Transfer to the Mobile System.

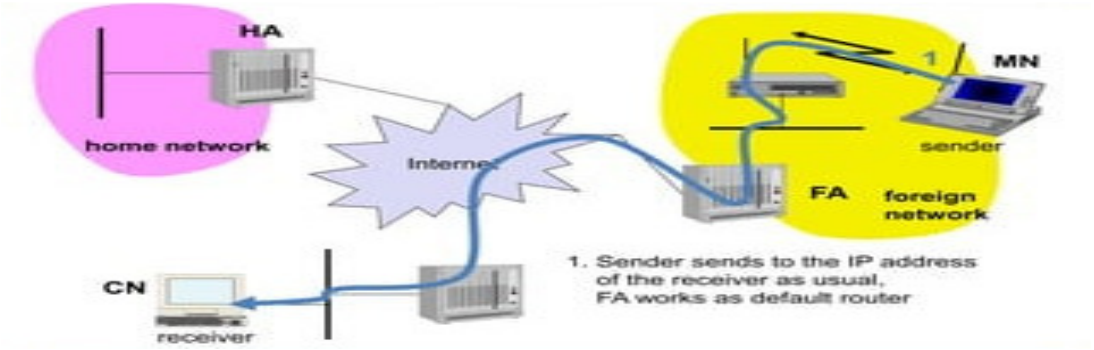


Figure 4.5: Data Transfer from the Mobile System.

It is quite easy to send packets from the mobile node (MN) to the CN. The MN transmits the packet as normal, using the CN's address as the destination and its own fixed IP address as the source (step 4). As the default router, the router with the FA forwards the packet just as it would for any other node in the foreign network. The rest is in the fixed internet as normal as long as CN is a fixed node. The same procedures as explained in steps 1 through 3 would apply now in the reverse way if CN were likewise a mobile node located in a foreign network Figure 4.5.

How Mobile IP Works: For a mobile host, Mobile IP has two addresses: a home address and a care-of address. The care-of address varies when the mobile host switches between networks; the home address is constant. A home agent and a foreign agent are necessary to make the change of address obvious to the rest of the Internet. The application layer is where an agent's specialised role is carried out. The care-of address is referred to as a co-located care-of address when the mobile host and the foreign agent are one and the same. A mobile host goes through three steps in order to connect with a distant host: agent detection, registration, and data transmission.

Agent Detection

When a mobile node leaves its home network, it must locate a foreign agent. Mobile IP outlines two approaches to address this issue: agent solicitation and agent marketing.

Advertising for agents

With this technique, home agents and foreign agents occasionally broadcast unique agent advertising messages into the subnet to announce their existence. Mobile IP employs the ICMP router advertising packet and an agent advertisement message, not a new packet type for agent advertisement. To prevent forwarding of adverts, the TTL field of the IP packet is set to 1 for all of them. The type is set to 9, and the code may either be 0 or 16, depending on whether the agent also routes traffic from non-mobile nodes or just routes mobile traffic. While the addresses themselves are listed as follows, the total number of addresses broadcast with this packet is in #addresses. The lifetime of this advertising indicates how long it will be in effect. A node may choose the router that is most ready to receive a new node by using the preference levels for each address.

The following fields are specified in the extension for mobility: Type is set to 16, and length is equal to 6 + 4*(number of addresses) depending on the number of COAs included with the message. The sequence number displays the overall number of ads delivered by the agent since activation. The maximum lifespan in seconds a node may ask for during registration is specified by the agent using the registration lifetime. The features of an agent are thoroughly described in the sections that follow.

When employing a collocated COA at the MN, the R bit (registration) indicates if a registration with this agent is necessary. The B bit may be set if the agent is presently too busy to accept new registrations. The next two bits indicate whether the agent provides services on the connection where the advertising has been delivered as a home agent (H) or a foreign agency (F).

The tunnel's mode of encapsulation is defined by bits M and G. Although IP-in-IP encapsulation is the required standard, M and G may define minimum and generic routing encapsulation, respectively. The V bit defined the usage of header compression in accordance with RFC 1144 in the first iteration of mobile IP (RFC 2002). Now that the field r is set to zero at the same bit location, it must be disregarded. The new field T shows that the FA supports reverse tunnelling. The COAs advertised are in the fields below. A foreign agent must promote at least one COA when setting the F bit. Now, either the home agent or a foreign agent may send agent ads to a mobile node in a subnet. The MN may locate itself in this manner, among others.

Solicitation of agents

The mobile node must send agent solicitations if there are no agent advertising, the inter-arrival duration is too long, and an MN has not already received a COA. Although caution must be used to prevent a network overflow from these solicitation messages, an MN may essentially do an unending search for an FA who is sending out solicitation messages. To prevent overwhelming the network, if a node receives no response to its solicitations, it must exponentially reduce the pace of solicitations until it hits a maximum gap in between solicitations (typically one minute). Anytime, not only when the MN is not tied to one, is a good time to find a new agent.

Following these processes of solicitations or ads, the MN is now eligible to get a COA, either one for an FA or a co-located COA.

Registration of Agents

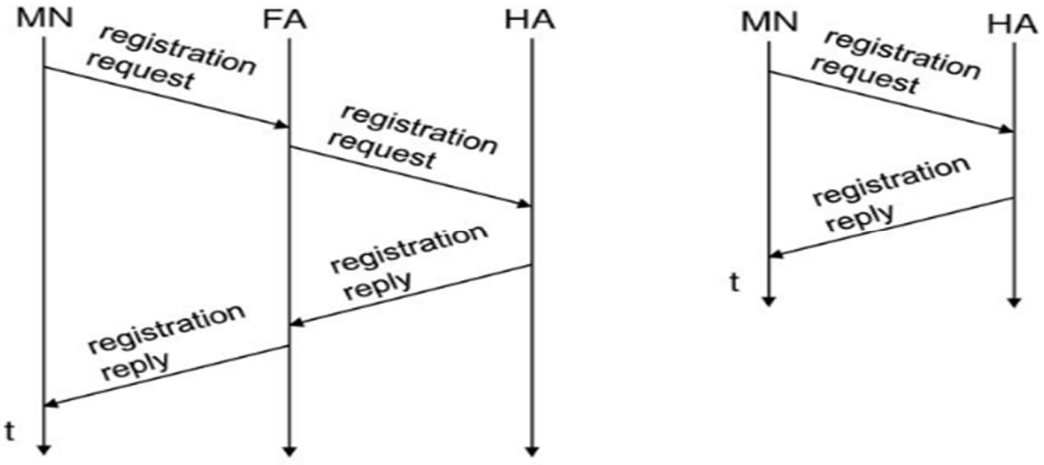


Figure 4.6: Registration of Agents

The MN must register with the HA after receiving a COA. The basic goal of registration is to provide the HA with updated location information so that packets may be forwarded correctly. Depending on where the COA is, there are two distinct methods to register: If the COA is at the FA, the MN submits its registration request with the COA to the FA, who then sends the

request to the HA. The HA now creates a mobility binding using the home IP address and current COA of the mobile node. The lifespan of the registration, which is negotiated upon registration, is also included. The lifetime registration automatically expires and is erased, therefore an MN should reregister before it does. To prevent no longer-used mobility bindings, this method is required. The HA sends a reply message to the FA after configuring the mobility binding, who then passes it on to the MN (Figure 4.6).

Mobile node registration via the FA or directly with the HA

Registration may be easier if the COA is co-located since the MN may submit requests to the HA directly and vice versa. In order to register directly with the HA, MNs must follow the same registration process when they return to their home network.

The registration requests employ port number 434 and UDP packets. The MN's interface address is set as the packet's IP source address, and the FA or HA's or HA's IP destination address.

Request for Registration

The first field type for a registration request is set to 1. An MN may tell the HA whether or not it should keep previous mobility bindings by using the S bit. This permits several binds at once. Setting the B bit often means that an MN additionally wishes to receive the broadcast packets that the HA in the home network has already received. An MN handles the decapsulation at the tunnel endpoint if a co-located COA is used. This behavior is indicated by the D bit. The bits M and G, as previously specified for agent advertisements (Figure 4.7).

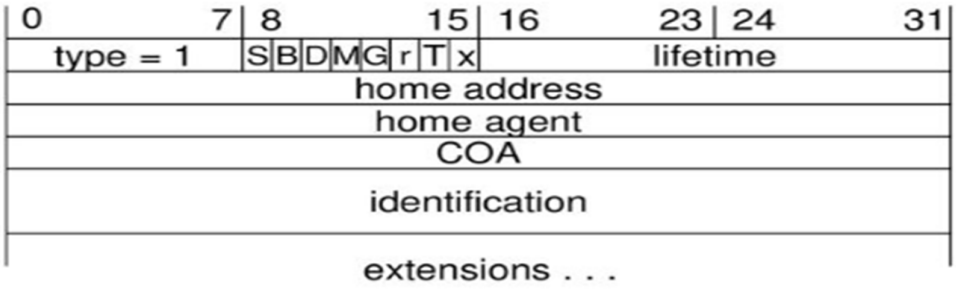


Figure 4.7: Mobile IP Registration Request

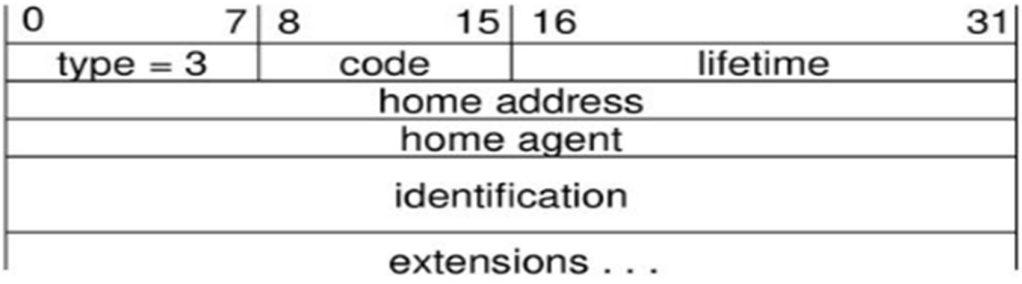


Figure 4.8: Mobile IP Registration Reply

The first field type for a registration request is set to 1. An MN may tell the HA whether or not it should keep previous mobility bindings by using the S bit. This permits several binds at once. Setting the B bit often means that an MN additionally wishes to receive the broadcast packets that the HA in the home network has already received. An MN handles the decapsulation at the tunnel endpoint if a co-located COA is used. This behavior is indicated by the D bit. The bits M and G indicate the usage of minimum encapsulation or generic routing encapsulation, respectively, as has previously been established for agent ads. T stands for reverse tunneling, while r and x are both set to 0 figure 4.8.

Lifetime indicates the number of seconds that the registration is valid. Deregistration is indicated by a value of zero; infinity is indicated by all bits being set. Home address is the MN's fixed IP address, home agent is the HA's fixed IP address, and COA stands for the tunnel endpoint. To identify a request and match it with registration answers, the MN generates a 64 bit identity. This field is used to defend against registration replay attacks.

The extensions must at the very least provide authentication parameters.

An UDP packet with a registration reply has a type field set to 3 and a code reflecting the outcome of the registration request. The lifetime field shows the number of seconds that, assuming the registration was successful, it will be active. The addresses of the MN and the HA, respectively, are home address and home agent. The registration requests and responses are matched using the 64-bit identity. The value is depending on the registration's identifying field and authentication method. Once again, the extensions must at least include authentication parameters.

Tunneling and encapsulation

A tunnel creates a virtual channel for data packets to travel through between its input point and its terminus. When packets enter a tunnel, they are sent there and remain unaltered when they exit. Utilizing encapsulation, or transmitting a packet via a tunnel, is known as tunneling (Figure 4.9).

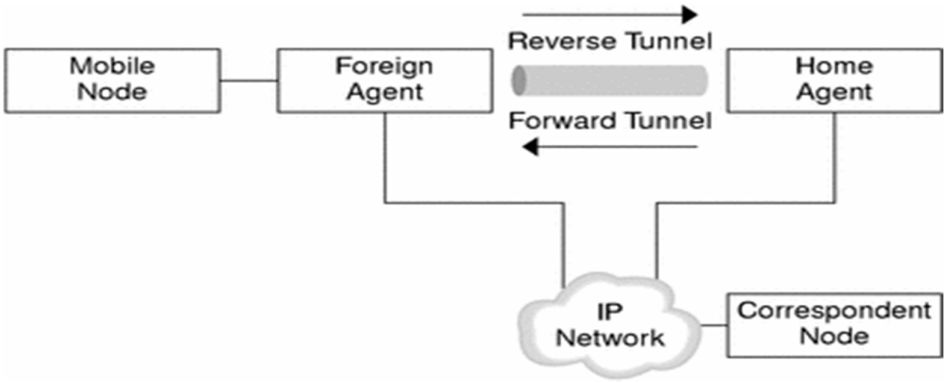


Figure 4.9: Mobile IP with a Reverse Tunnel

Encapsulation

Encapsulation is a technique for transferring a packet's data portion into a new packet's data portion. A packet consists of a packet header and data. Decapsulation refers to the opposite procedure, which involves removing a packet from the data portion of another packet. When a packet is transmitted from a higher protocol layer to a lower layer or from a lower to a higher layer, encapsulation and decapsulation are commonly used. In order to route the packet to the COA, the HA takes the original packet with the MN as destination, inserts it into the data

portion of a new packet, and sets the new IP header. Outer header is the name of the new heading.

Encapsulation of IP in IP

The encapsulation required for the tube between HA and COA may be done in several ways. According to RFC 2003, IP-in-IP encapsulation is required for mobile IP. A packet within the tunnel is seen in the next picture (Figure 4.10).

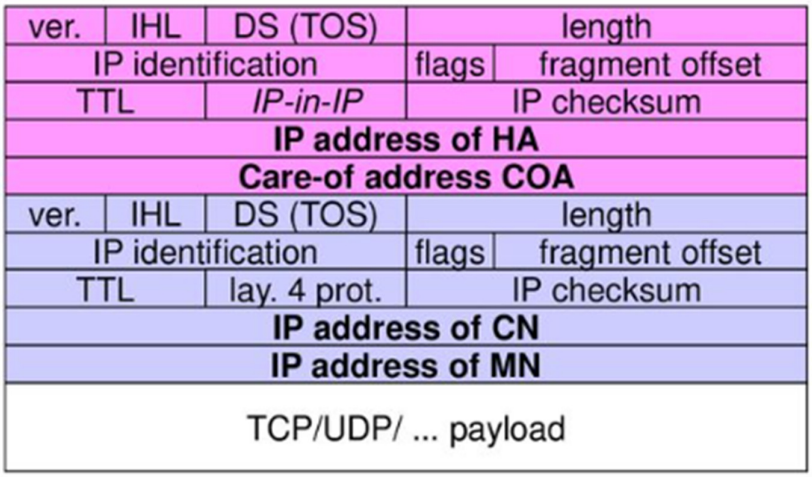


Figure 4.10: Encapsulation of IP in IP

Internet header length (IHL), which represents the length of the outer header in 32 bit words, is indicated by the version field, which is 4 for IP version 4. The length field contains the whole encapsulated packet, whereas DS (TOS) is simply copied from the inner header. The fields up to TTL are set in accordance with RFC 791 and have no particular significance for mobile IP. For the packet to get to the tunnel destination, the TTL has to be high enough. The kind of protocol used in the IP payload is indicated in the following field by the symbol IPin-IP. An IPv4 packet follows this outer header, hence this field is set to 4, the IPv4 protocol type. As usual, the IP checksum is computed. The following parameters are the tunnel exit point's destination address and the tunnel entry's source address (the HA's IP address) (the COA).

The inner header begins with the same fields as the outer header if the outer header has no choices after it. This header virtually stays the same during encapsulation, displaying the packet's original sender CN and recipient MN. The sole modification is a 1 reduction in TTL. This indicates that from the perspective of the initial packet, the whole tunnel is seen as a single hop. This is a crucial aspect of tunneling because it enables the MN to act as if it were connected to the household network. No matter how many actual hops a packet must go through in the tunnel, the MN is only one (logical) hop away. The two headers are followed by the payload.

Minimal encapsulation

In order to eliminate repeats of similar fields in IP-in-IP encapsulation, minimal encapsulation (RFC 2004), as seen below, is an optional encapsulation approach for mobile IP. Both the tunnel's beginning and ending points are identified (Figure 4.11).

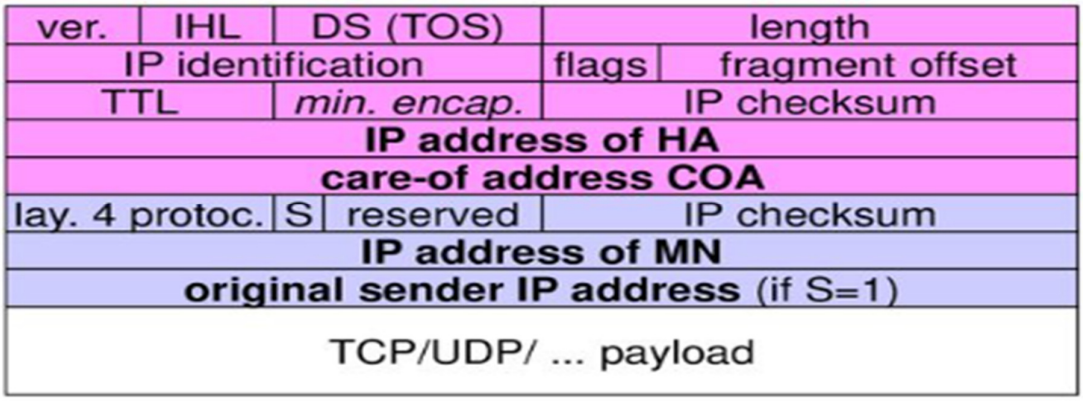
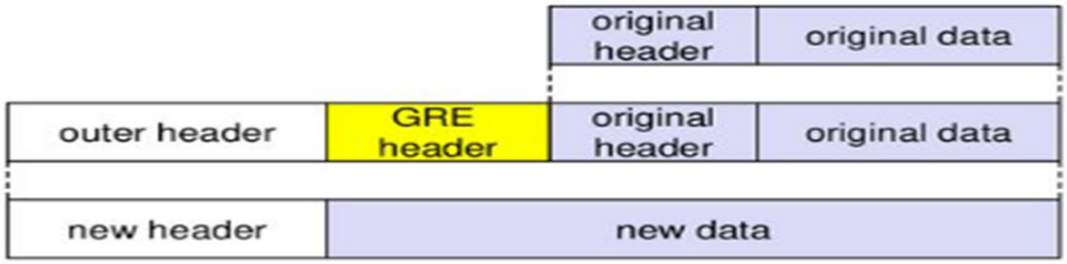


Figure 4.11: Minimal encapsulation

The value 55 for the minimum encapsulation protocol is included in the field for the type of the following header. For limited encapsulation, a separate inner header is used. It is necessary to know the kind of the subsequent protocol as well as the MN's address. The original sender address of the CN is provided if the S bit is set since it is often impossible to exclude the source. The inner header no longer has a field for fragmentation offset, hence minimum encapsulation cannot be used with fragmented packets.

Routing Encapsulation Generally

Generic routing encapsulation (GRE) enables the encapsulation of packets of one protocol suite inside the payload section of a packet of another protocol suite, as seen below, in contrast to IP-in-IP and Minimal encapsulation which only function for IP packets (Figure 4.12).



4.12: Routing Encapsulation Generally

One protocol suite's packet is extracted, along with its original packet header and payload, and a new GRE header is appended. Together, they make up the new packet's data portion. The second protocol suite's header is then placed in front.

The following image displays the fields of a packet within the tunnel between HA and COA using RFC 1701's GRE encapsulation method (Figure 4.13). With HA as the source address and COA as the destination address, the outer header is a typical IP header. This outer IP header's protocol type is 47 for GRE.

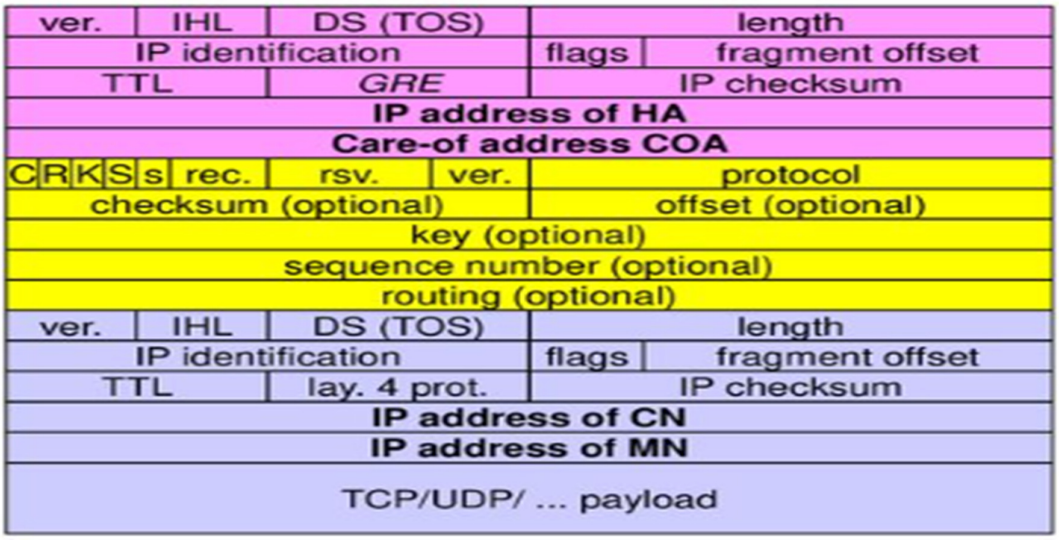


Figure 4.13: displays the fields of a packet within the tunnel between HA and COA using RFC 1701's GRE encapsulation method

The GRE header begins with a number of flags indicating the presence or absence of certain fields. The smallest GRE header is simply 4 bytes in size. The checksum field's presence is indicated by the C bit and provides accurate facts. The IP checksum of the GRE header and payload is present in the checksum field if C is specified. The presence of the offset and routing fields and the validity of their contents are indicated by the R bit. The offset reflects the initial source routing entry's offset in bytes. If present, the routing field has a configurable length and source routing fields. Additionally, the GRE provides a key field that may be used for authentication. The K bit is set if this field is present. If the s bit is set, tight source routing is utilised, and the sequence number bit S specifies whether the sequence number field is present.

Further feature that sets GRE apart from IP-in-IP and minimum encapsulation is the recursion control field (rec.). This parameter is a counter that indicates how many recursive encapsulations are permitted. This field's default value should be 0, which would limit encapsulation to one level. The following reserved fields are not used during receiving and must be set to zero. For the GRE version, the version field has a value of 0. The protocol of the packet that follows the GRE header is represented by the subsequent 2 byte protocol field. Following the source address of the correspondent node and the destination address of the mobile node is the standard header of the first packet. A reduced version of the GRE header as per RFC 2784. Again, the presence of a checksum is indicated by the field C. Seven reserved bits come after the following 5 bits that are set to zero. The number 0 is present in the version field. Once again, the protocol type specifies the payload protocol in accordance with RFC 3232. If the flag C is set, the reserved1 field and the checksum field come next. The final field is set to zero after being constant zero.

Optimizations

If a situation where the HA is on the other side of the planet and the MN is on the same subnetwork as the node to which it is communicating happens. It is known as a triangle routing issue because it causes the network between the CN and the HA to incur needless overheads.

If a situation where the HA is on the other side of the planet and the MN is on the same subnetwork as the node to which it is communicating happens. It is known as a triangle routing issue because it causes the network between the CN and the HA to incur needless overheads.

The CN should be informed of the MN's present location as a remedy to this issue. By storing the location in a binding cache, which is a component of the CN's routing table, the CN may learn the location. The CN is informed of the location via HA. It requires the following four messages:

The node that wishes to know the location of an MN right now sends a binding request to the HA. After determining if it is permitted to share the location, HA sends back a binding update.

Binding update: This tells the CN where an MN is right now, according to the HA. It includes the COA and the MN's fixed IP address. An acknowledgment may be requested in this message. When a node receives a binding update message, it responds with this acknowledgment if desired.

Binding warning: If a node decapsulates a packet for an MN but does not record the current FA of this MN, it emits a binding warning. It includes the home address of MN and the address of the target node. The HA now sends a binding update to the node that plainly has the incorrect COA for the MN since the receiver may be the HA.

The old FA would not learn anything about the new location of MN without the information given by the new FA. In this instance, CN still tunnels its packets for MN to the old FA, FAold, since it is unaware of the changed location. Now when packets containing the destination MN are detected, this FA is aware that it is not the MN FA at this time. These packets may now be sent by FAold to FAnew, the new COA of MN in this case. This packet forwarding is one more way to improve the fundamental Mobile IP, which facilitates seamless handovers. Without this improvement, as the MN switches from one FA to another, all packets in transit would be lost.

FAold sends a binding warning message to CN in order to inform it that its binding cache is out of date. Then CN asks for a binding update. (The alert may also be sent directly to the HA, causing an upgrade.) The HA notifies the CN of the changed location by sending an update, which is acknowledged. Now that triangle routing is no longer necessary, CN may deliver its packets straight to FAnew. Unfortunately, there are a number of security issues brought on by this mobile IP optimization for triangular routing.

Reverse Tunnelling

The reverse route from MS to CN seems to be relatively straightforward since the MN may transmit its packets directly to the CN as in any other typical IP scenario. The packets' final destination address is CN. But as can be shown below, it has several issues: - Quite frequently, firewalls are built to only permit the passage of packets with topologically accurate addresses in order to provide straightforward protection against incorrectly configured systems with unknown addresses. While on a foreign network, MN continues to transmit packets with its fixed IP address as source, which is not topologically valid. Firewalls often filter packets from outside that come from inside network machines using a source address. This suggests that an MN cannot communicate with a computer that is a part of its home network (Figure 4.14).

An MN in a foreign network cannot send multi-cast packets in a fashion that they originate from its home network without a reverse tunnel, even if the nodes in the home network may participate in a multi-cast group. It's possible that the foreign network won't even offer the necessary technological framework for multi-cast communication (multi-cast backbone, Mbone).

The previous TTL may be too low for the packets to reach the same target nodes as before if the MN switches to a different foreign network. If the TTL has to be changed while the user is moving, mobile IP is no longer transparent. No matter how many hops are really required to connect the foreign network to the domestic network, a reverse tunnel representing only one hop is required. Reverse tunnelling is a definition of an extension of mobile IP based on the aforementioned principles (per RFC 2344). To address the aforementioned issues, it specifies topologically valid reverse tunnelling and was meant to be backward compatible with mobile IP.

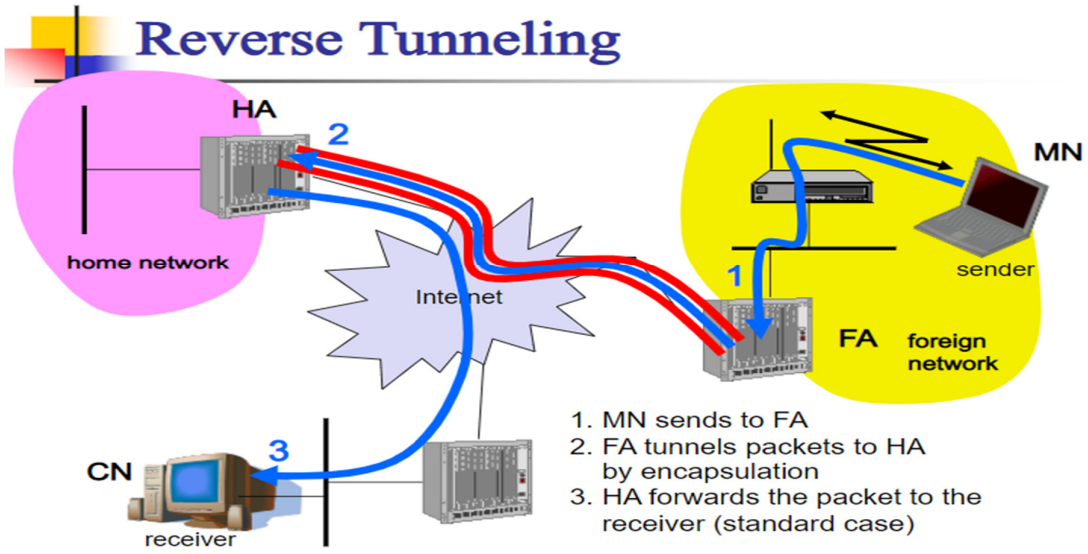


Figure 4.14: Reverse tunneling

Firewall issues are not resolved by reverse tunneling, which may also be exploited to get around security measures (tunnel hijacking). Data path optimization, wherein packets are sent to a sender through the HA and via the tunnel (double triangle routing);

The four extra messages are combined in the following figure in the event that an MN modifies its FA. The CN may ask the HA for the current location. If permitted by the MN, the HA sends an update message that contains the MN's COA. This update message is acknowledged by the CN, which also saves the mobility binding. As of right now, the CN may transfer data straight to the foreign agent FAold. The packets are sent to the MN via FAold. In this case, an FA is where the COA is situated. The CN, not the HA, is now responsible for data encapsulation for tunnelling to the COA. The MN could now relocate and sign up with FAnew, a new foreign agency. Additionally, the HA receives this registration in order to update its location information. Additionally, FAnew updates FAold on MN's new registration. The address of FAold is included in MN's registration message for this reason. This information is sent through an update message, which FAold acknowledges.

IPv6

Designing Mobile IP support for IPv6 (Mobile IPv6) takes use of IPv6's potential as well as the knowledge learned from developing Mobile IP support for IPv4. Thus, Mobile IPv6 and Mobile IPv4 have a lot in common but Mobile IPv6 is incorporated into IPv6 and provides many more benefits. The key distinctions between mobile IPv4 and mobile IPv6 are outlined in this section:

As with Mobile IPv4, there is no need to instal specialised routers as "foreign agents". Anywhere that supports mobile IPv6 may use it without the local network providing any specialised assistance.

The protocol includes support for route optimization as a core feature rather than as a series of unconventional extensions.

Mobile IPv6 route optimization is safe even in the absence of pre-established security relationships. The widespread deployment of route optimization between all mobile nodes and correspondent nodes is anticipated.

Mobile IPv6 has support for enabling route optimization to effectively coexist with routers that carry out "ingress filtering."

The IPv6 Neighbor Unreachability Detection ensures that the mobile node and its default router are symmetrically reachable at the present location.

When compared to Mobile IPv4, less overhead is created since the majority of packets delivered to a mobile node when away from home use an IPv6 routing header rather than IP encapsulation.

Since mobile IPv6 utilises IPv6 Neighbor Discovery rather than ARP, it is independent of any specific connection layer. This enhances the protocol's resilience as well.

In Mobile IPv6, managing "tunnel soft state" is no longer necessary thanks to the usage of IPv6 encapsulation (and the routing header).

The Mobile IPv6 dynamic home agent address discovery technique only provides the mobile node with a single response. The IPv4 directed broadcast method generates distinct responses from each home agent.

Protocol for Dynamic Host Configuration (DHCP)

On IP networks, DHCP is a mechanism for automated setup. A machine may join an IP-based network via DHCP even without a pre-configured IP address. In accordance with the DHCP protocol, devices are given distinct IP addresses, which are then released and renewed when they enter and depart the network.

When a new computer joins a network, DHCP may provide it access to all the data it needs for complete network integration, such as the IP address, subnet mask, domain name, and addresses for a DNS server and the default router. DHCP is a particularly desirable source of care-of-addresses for mobile IP since it offers an IP address. The client/server concept of DHCP is shown here. A server (DHCPDISCOVER in the example) receives a request from DHCP clients and answers. To reach every device in the LAN, a client uses MAC broadcasts to deliver requests. To transmit requests to a DHCP server across inter-working units, a DHCP relay may be required.

Think about a situation where there is only one client and two servers. Below is an example of a DHCP client initialization as shown in the Figure 4.15:

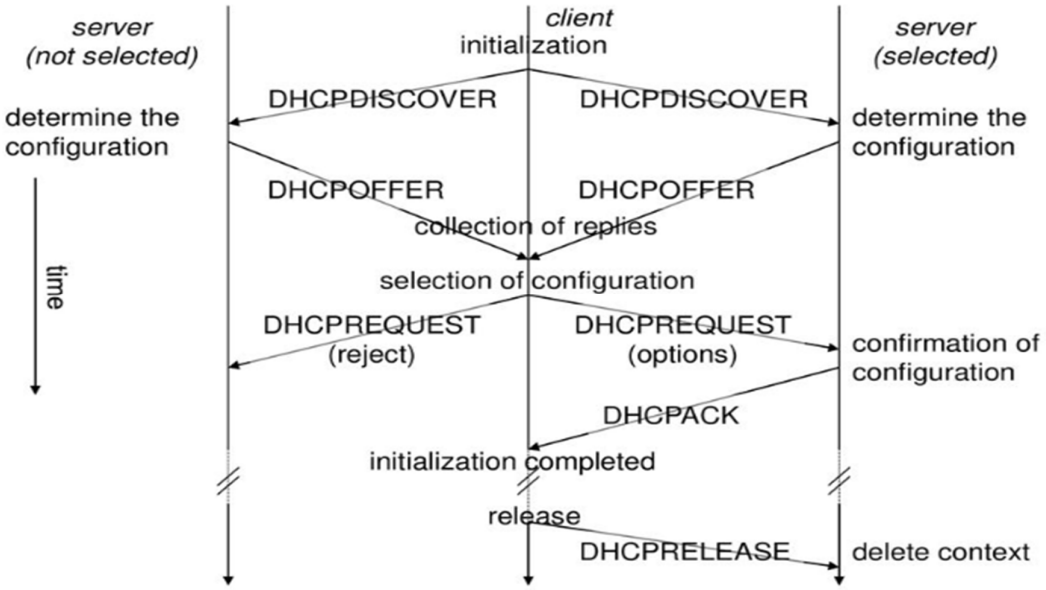


Figure 4.15: Client Initialization via DHCP.

A DHCPDISCOVER is broadcast by the client into the subnet. This broadcast could be sent through a relay. Two servers in the scenario are shown to receive this broadcast and choose which configuration to provide to the client. When a client makes a request, servers respond with DHCPOFFER and provide a list of configuration options. The customer may now choose from the available setups. Using DHCPREQUEST, the client responds to the servers, accepting one configuration and rejecting the others. A server may release the reserved configuration for further potential clients if it gets a DHCPREQUEST with a refusal. The server that has the configuration that the client approved now uses DHCPACK to verify the configuration. With this, the startup stage is over. When a client departs a subnet, it should use DHCPRELEASE to relinquish the configuration it obtained from the server. Now that the client's context has been saved, the server may release it and provide the settings once again. The configuration a client receives from a server is only leased for a certain period of time; it must sometimes be reconfirmed. If not, the settings will be released by the server. This configurable delay is beneficial in the event that a node crashes or is relocated away without relinquishing the context.

The purchase of care-of addresses for mobile nodes may be supported through DHCP. The same is true for all other required settings, such as default router addresses, DNS server addresses, timeserver addresses, etc. The mobile node's access point subnet should have a DHCP server, or at the very least, a DHCP relay should transmit the messages. To guard against rogue DHCP servers, RFC 3118 provides authentication for DHCP communications. A DHCP server cannot trust a mobile node without authentication, and vice versa.

CHAPTER 5

MEMORY MANAGEMENT

Anil Agarwal, Associate Professor,
Department of Electronics and Communication, School of Engineering & Technology, Jaipur National University, Jaipur, India,
Email Id- anil.agarwal@jnujaipur.ac.in

For every computer system, memory is a crucial resource. Additionally, when it comes to programming mobile devices, memory is a crucial resource since manufacturers only put a little amount of it in devices in an effort to keep the cost of the device cheap, despite the fact that all running applications compete for it. Memory chips also require some power, the quantity of which is dependent on the amount of memory present in the device, in addition to being a significant cost issue. While the heap and stack are only memory spaces in theory. Additional thought should be given to the simplicity and naturalness with which sharing data occurs while utilising the heap, as opposed to stack-based variables, where references should only be used sparingly or not at all since the stack may overwrite referenced data as the execution progresses.

Stack

Generally speaking, ephemeral items, or those that have a short lifespan, should be kept on the stack. According to the theory, allocating an object to a stack is simpler if it will only be used briefly in any case since it will be automatically deallocated as the execution moves forward. Additionally, utilising stack memory is often already reserved, therefore looking for an appropriate memory space does not need to be done during allocation. The usage of stack for transitory objects is further encouraged by this problem. The drawback of automated variable allocation and deallocation is that variables on the stack may be wiped and replaced by other variables, which might cause issues when utilising references to such variables. This can never be an issue if references are always created from subsequent activations. However, it might be challenging to do this in reality, particularly during the maintenance period.

Heap

Allocate space to the heap for any data structures whose size or structure might change dynamically. The argument is that it is difficult to reserve enough space from the stack since it is impossible to predict their size in advance. The need that the object continue to exist regardless of the programme phase is another factor that sometimes justifies utilising the heap rather than the stack. In other words, memory for an object should be allocated from the heap if its function is global.

As was previously said, these variables may be declared static inside of methods, thereby making them global. Additionally, if big objects need to be allocated, they may go on the heap rather than the stack, whose size may be constrained by the application. The caller of this function is responsible for making sure that memory is freed when the data structure is no longer required if the run-time system does not automate garbage collection. In order to prevent unintentional usage of the region, it is a good idea to reset the reference to it at the same time. It's also crucial to keep in mind that utilising the heap might be slower than using the stack since the active application might need to look for an appropriate memory space.

Design Patterns for Memory Limited

The most crucial rule to remember when creating designs for gadgets with a limited amount of memory is to not squander it. Here, we concentrate on their use in the creation of apps for mobile devices.

Structures for linear data

Preference should be given to linear data structures. There are a number of reasons why linear data structures are often preferable to non-linear ones for memory management, including the following:

Less dispersion In contrast to non-linear data structures, which may be placed everywhere in memory, linear data structures consume memory space from a single position. Of course, the former reduces the likelihood of fragmentation.

Less time spent looking. In contrast to non-linear structures, where one memory request is necessary for each allocated element, reserving a linear block of memory for many things just requires one search for an appropriate memory element in the run-time environment. This may potentially result in a significant performance issue when combined with a design where one object allocates a lot of child objects.

Controlling the design time. Since fewer reservations are issued for linear blocks, they are simpler to handle at design time. Typically, this results in cleaner designs.

Watching. Addressing may be done under supervision since it is easy to verify that the index being used relates to a legitimate object.

A better cache. Since cache functions internally with blocks of memory, it is more probable when utilising linear data structures that the next data piece is already in cache. Another problem is that the majority of caches assume that data structures be consumed in ascending order of memory regions used. Therefore, it is advantageous to include this into designs when appropriate.

Less memory is used by Index. When assigning items to a vector of 256 objects, supposing that this is the maximum number of objects, an index of just 8 bits may be used, as opposed to an absolute reference to an object, which typically requires 32 bits. Additionally, it is possible to confirm that there won't be any erroneous indexing.

Fundamental Design Choices

Following, we discuss several fundamental ideas that facilitate the use of linear data structures. The goal is to provide some instances of how linear data structures might be used to assist design composition rather than to provide a comprehensive checklist.

At the start of a programme, all the memory should be allocated. The application will always have the memory it needs thanks to this, and memory allocation will only ever go wrong at the start of the programme. When the most crucial or necessary characteristics, like emergency calls, are taken into account, where resources must always be accessible, reservation of all the resources is very alluring. This kind of technique is often best suited for singularly purpose-optimized devices, and it cannot usually be applied to smartphones, with the exception of a small number of specific unique instances. Even if you only need one thing, allocate memory for many. Then, a policy where a number of objects are reserved with a single allocation request may be created. Then, as required, these items may be utilised. As a result, there are fewer

requests for allocation, which results in a simpler structure for the memory. The method also boosts speed since there will be fewer memory allocations and better cache utilisation.

Utilize typical allotment sizes. When the next reservation is created, it is simple to reuse a deallocated portion of memory with a standard allocation size. As a consequence, memory fragmentation may be avoided, at least in part. Reuse things. Utilizing a pool of free things may be necessary for recycling outdated items. For handling free and utilised data structures, we need a data structure. This suggests that the programmer deliberately chooses the object creation and destruction strategy throughout the design phase.

Early release and late allocation. The programmer may provide extra memory management choices by always deallocating as soon as feasible since new objects can be allocated to the space that was just freed up. In contrast, the developer may make sure that all potential deallocations have been completed before the allocation by allocating RAM as late as feasible. Before allocating new objects, it is very important to make sure that memory-intensive objects are deallocated. Because heap often provides the first acceptable memory space, or, in a stack-like implementation, on one end, this is the cause. Then, fragmentation may be avoided or at the very least, its effects may be mitigated when big objects are deallocated before allocating others. When appropriate, use ROM or persistent storage. Due to physical limitations, it is sometimes not even necessary to maintain all of the data structures in programme memory. For instance, all unsaved data would be destroyed if the battery is taken out of the device. Introduce the practise of saving all data to permanent storage as soon as feasible in such circumstances. A user interface that requires the user to commit to finishing an input to the calendar or contacts, for example, may help with this. Static data, such as dynamic library and application IDs or strings used in programmes, may be obtained in a similar way. Additionally, even if there is no chance of data loss, writing big, seldom used items to permanent storage may be advantageous from the standpoint of memory usage, protecting the device's memory for more crucial data.

Differential Link Libraries

In programming, the words EXE and DLL are often used. When programming, you may export your finished product as either an EXE or a DLL. The abbreviation EXE designates the file as a programme and is an abbreviated form of the word executable. On the other hand, DLL, which stands for Dynamic Link Library, is often used to refer to a collection of processes and functions that may be utilised by other applications.

The simplest programme package would have at least one EXE file, maybe supplemented by one or more DLL files. The entry point, or section of the code, where the operating system is expected to start the execution of the programme, is included in an EXE file. DLL files are unable to run independently because they lack this entry point.

DLL files' greatest benefit is their reusability. As long as the programmer is aware of the names and parameters of the functions and procedures in the DLL file, the DLL file may be utilised in other programmes. DLL files are perfect for sharing device drivers due of this functionality. The DLL would make it easier for the hardware and the programme that wants to utilise it to communicate. As long as the programme is able to call the DLL's functions, it is not necessary for it to understand how to access the hardware.A process and memory space must be created in order to launch an EXE.

In order for the software to function correctly, this is required. A DLL does not have its own memory space or process since it is called by another programme rather than being started independently. It only makes use of the application's calling process and memory. A DLL may

therefore only have restricted access to resources since they may already be used by the programme or by other DLLs.

Dynamic Link Library

DLL files, are a particular kind of file that hold instructions that other applications may use to carry out certain tasks. In this approach, a single file's capabilities may be shared by numerous applications, perhaps even at the same time.

Dynamic and Static DLLs

Typically, a static DLL is created at application startup and remains in memory until the program is closed. A dynamic DLL (plugin) is loaded and unloaded as required, for example, a distinct plugin for each kind of communications (email, SMS, and MMS).

Plug-in

A plug-in (also known as a plugin, add-in, add-on, addon, or extension) is a software component that enhances an existing computer program with a particular function. Using Dynamically Loaded Libraries: Some General Guidelines It is best to construct reusable or shareable components using dynamically loaded libraries rather than individually storing them in each application that uses them. Memory is then used up as a result. Additionally, sharing may occur when a single model is built as a dynamically loaded library and utilised in a variety of devices, each of which has a unique user interface. In terms of dynamic libraries, variation or management point may be better to implement. As a result, variation or management may be directly linked to a software component, increasing control. In addition, the library may be quickly updated or modified if necessary.

The input for all software development processes, including automated testing, may need to be in a format that can be handled immediately. The delivery of binary components could then be necessary in order to run the tests. A single library that is in charge of a certain set of tasks may be assembled by an organisational unit. A distinct deliverable that may be quickly included into the finished system is produced as a consequence of requesting a library.

Constitutes an Application

The simplest definition of an application is a piece of software that can be started and stopped on its own and completes a specific goal. In addition, it is often required to link a user interface to a program since doing so makes it easier to see how the application behaves. Under a technical sense, an application may be thought of as a segment of executable code that, in certain circumstances, the user or the system can cause to run.

Application Development Workflow

Consistency in user experience may be the most crucial design consideration when creating an application for a mobile device. The actions must be straightforward, single-minded, and easy to do with the fewest possible keystrokes. It follows that this affects how apps must be created. A typical process for the development of applications for the mobile environment, with an emphasis on usability and user behaviors, and it includes:

1. Scope
2. Performance-related factors
3. Designing user interfaces
4. Communications and I/O,
5. Communications and I/O,

This procedure in the paragraphs that follow.

Scoping

The essential objective of the application, as well as what it can and cannot achieve, must be known before beginning the design of an application for the mobile environment. In particular, a subset of functionalities that will be included in the implementation must be chosen when creating a mobile version of a desktop application. Additionally, if the device's physical attributes suggest limitations, these must be taken into consideration. By visualising (seeing) the application using images, mockups (models), and prototypes, scoping may be aided. This will be useful for explaining the application's scope and goals to other developers. The relative relevance of the features to users should also be taken into account. For example, if clock times are only seldom input, it could be sufficient to utilise a somewhat cumbersome user interface; even if the action may be bothersome, it is only occasionally required that the user can still perform it. However, a thoughtful user interface should be used for entries that are made often.

Factors Considering Performance

After scoping is finished, performance should be taken into account. Applications require baseline measurements for general responsiveness. For instance, you may specify how quickly the application's menus should appear. The level of responsiveness overall is crucial to the user experience. For the most crucial cases, particular measurements should be developed in addition to general responsiveness. Because of this, the application designer is compelled to take into account the sequences of actions that let the user to do certain tasks. Utilizing more outdated (or just less competent) gear for preliminary testing is one method of performance design. While this paints a bleak picture of the implementation options for the programme, the design may begin even before the target device is really ready, and with fewer assumptions, there is a higher likelihood that the consumers will be happy with the performance.

A genuine implementation should be used to test each assumption. Start with a few critical features and their performance, and only go on to less crucial aspects if the core features have an acceptable degree of performance, is a popular strategy. A decent general rule of thumb is to assume that the application will be held to higher standards in the future. In example, the concept that the code should be written in its entirety first in order to identify the greatest bottlenecks is sometimes wrong since the overall performance is frequently the most crucial factor. Then, rather than individual lines of code, concerns relating to data structures, their organization in memory, used algorithms, and the design of the user interface should be taken into account first. In other words, rather of focusing on the symptoms of performance issues, one should consider their underlying causes.

Additionally, it should be remembered that portability may suffer if performance is overemphasized. As a result, even while it's crucial to take into account whether the chosen implementation principles can meet performance requirements, one shouldn't feel compelled to optimize the development only for performance. Instead, a reality check on what can really be done is required.

Designing user interfaces

As was previously said, it is crucial to understand the main use cases and elements that make up a mobile application before moving on to the technical design. It's time to concentrate on the proper user interface if the performance offered by the prototype implementation in studies is satisfactory.

The end-user productivity and responsiveness might be regarded as the most crucial aspects in user interface design, along with scoping and innovations. According to the first, tasks that are common and natural for the end-user may be completed quickly and effortlessly. The latter means that the user feels in control when engaging in the activities, which often entails limiting the amount of time the user must wait for activities to finish. More significant, however, is that the user is never left in the dark about what the device is really doing. The inclination of users to repeat activities if a response is not seen right away further fortifies the design.

This promotes designs that provide feedback on user-initiated activities even while those operations are still being carried out behind the scenes. For example, some applications can be always active even if the user has never started them, so this may call for a strategy where the user is deceived into thinking that an already completed task takes place only on her command in a proactive manner, or where the device allows the user to believe the task is completed when it is actually not (for instance, the phone claims to be ready after a reboot even if it has not yet loaded contacts from SIM).

A forced flow of control may need to be designed in specific circumstances, but caution must be used to prevent user annoyance. Keeping the user informed of what has really been stored to disc is an additional problem if the user wants to switch off the device. The amenities that are readily accessible are very significant while creating the user interface. It is unrealistic to assume that the usability and user experience would be retained by copying the user interface from one kind of device to another type of device. Instead, one should base their design of user interfaces on what the user perceives as natural when a certain sort of technology is accessible.

The fact that various activities are natural with various technologies makes the matter worse. For instance, editing Excel macros on a Communicator-type device appears quite feasible, yet on a standard mobile phone with its more constrained capacity, being able to read the numbers could be sufficient. Naturally, the size of the screen and the limited input options have an impact on the design overall. This problem may be partially resolved by utilising PCs for specific work and just moving the results to mobile devices. Additionally, one may decide whether to focus on a single tool that handles everything or on devices and programmes with specialised functions.

Norman (1998) offers one perspective on this issue, contending that a versatile technique is more difficult to utilise than one that is application- and purpose-specific. In reality, however, it seems that the latter strategy is also continually gaining popularity, at least when taking into account the gadgets that are now in use. The price of manufacturing is one of the causes of this. When new hardware features are included in a mobile phone rather than a separate device, it might be less expensive. Additionally, certain software functions are practically free.

Memory and Data Model Issues

Mobile devices, as was previously said, have very limited possibilities for application creation. This is connected to the price per unit of a device, where more advanced technology results in an increase in price per unit, but power consumption and device size also indicate certain limitations. The result might be a gadget that has a number of drawbacks but can still be used for the envisioned use cases.

The representation of data affects where it can be found in memory, how the system performs under load, and how the application disposes of data. This indicates that data structures and memory use in general must be properly taken into account for an application developer. Since their technical implementation may depend on DLLs, dynamically loaded libraries may also be seen as a problem that is strongly tied to data model and memory issues.

Information and output

How the application interacts with resources that are outside of its control depends on how communications and I/O are specified. This comprises both internal resources on devices, such as files and internal systems, as well as external resources that need a communications method in order to be accessed.

To provide only a few examples, the latter offers socket-based communication, files on servers, Web Services, and distant databases. Usability is significantly impacted by how the programme manages local and distant resources. When using the present implementation approaches, connecting with distant resources is often slower than accessing local resources.

In certain circumstances, choosing to load some data from a distant place in advance of the user's activities might significantly enhance the user experience. This, however, is often not practical and should only be used in unique situations when users' intents can be precisely predicted in advance.

The degree of abstraction of data that is transferred and stored is a crucial factor to take into account when it comes to communications and I/O. For instance, while utilising files, one may take into account the following layers of abstraction:

Text streams, where data becomes more intelligible but may still be relatively unstructured and unreadable for a human reader, binary streams, where data is stored in a manner that is unreadable without additional software.

XML forward-only readers and writers, which include additional meta-data

The XML Document Object Model, which often enables complicated automated processing of supplied data

Different tools for altering data are available at various degrees of abstraction. The data processing is simpler and the files are more self-contained as the degree of abstraction increases. Because programming, debugging, and maintenance will be simpler and it is more probable that standard components will be usable or that possibilities for reuse exist inside the organization, this suggests that developer productivity increases. The strategy may not be appropriate in situations when a big quantity of data has to be processed quickly since at the same time, the amount of overhead in sending, processing, and storing grows. This may result in needs that conflict with one another and make the design more challenging. The fact that it is seldom feasible to include many implementations of the same function in the device, even though their features would be different, makes the design more challenging. If a cellular data connection is envisaged, connection charges may also play a significant role in the decision-making process. For instance, when a wireless LAN connection is available, one would want to download as much data as possible, yet only accept bare-bones connectivity while utilizing GPRS.

CHAPTER 6

TRADITIONAL TCP

Puneet Kalia, Associate Professor,
Department of Electronics and Communication, School of Engineering & Technology, Jaipur National University, Jaipur, India,
Email id- puneet.kalia@jnujaipur.ac.in

One of the fundamental protocols of the Internet protocol suite, sometimes known as TCP/IP, is the Transmission Control Protocol (TCP). TCP is dependable, ensures that data is delivered in the correct sequence, and includes congestion and flow management techniques. The World Wide Web, email, the File Transfer Protocol, and Secure Shell are just a few of the most well-known application protocols and applications that TCP enables. TCP is the layer between the Internet layer and the application layer in the Internet protocol suite.

In an active session, TCP's primary duties are to:

1. Offer dependable data transit in order to prevent data loss.
2. Manage network congestion to prevent deterioration of network performance,
3. Maintain proper packet flow between the transmitter and the receiver so as not to overload it.

To achieve high performance and prevent "congestion collapse," where network performance might drop by many orders of magnitude, TCP employs a variety of methods. These processes regulate the data flow into the network, preventing it from reaching a level that would cause collapse. The effectiveness of TCP in a mobile context is influenced by a number of TCP methods. Senders utilize the receipt of acknowledgments for data delivered, or the absence thereof, to infer implicitly the state of the network between the TCP sender and receiver.

Congestion mitigation

TCP and other transport layer protocols were created for fixed networks with fixed end systems. Even networks that have been well planned sometimes experience congestion. Because the total of the input rates of packets intended for one output link is larger than the capacity of the output link, a router's packet buffers are full and the router is unable to forward the packets at a fast enough rate. A router's sole option in this case is to discard packets. The receiver detects a pause in the packet stream and loses a dropped packet for the transmission.

Now, the receiver acknowledges every packet in sequence up to the missing one without immediately informing the sender which packet is missing. The sender determines a packet loss was caused by congestion after seeing the absence of an acknowledgment for the lost packet. It wouldn't be a good idea to retransmit the lost packet and keep sending at your maximum pace right now since this might make the congestion worse. TCP significantly reduces the transmission rate in order to reduce congestion. The exact same behaviour is followed by all other TCP connections facing the same congestion, which causes it to disappear quickly.

Although TCP's response to a missed acknowledgment is fairly severe, it is vital to swiftly clear congestion. Slow start refers to the behaviour that TCP exhibits when congestion is detected. For a receiver, the sender will always determine the congestion window. The congestion window's initial size is one segment (TCP packet). One packet is sent, and the sender then waits for a response. If the acknowledgment is received, the sender sends one more packet, increasing the congestion window to two (congestion window = 2). Every time the

acknowledgements return, this approach doubles the congestion window, which requires a round trip (RTT). This is referred to as the sluggish start mechanism's congestion window's exponential development. However, increasing the congested window is too risky. At the congestion barrier, exponential growth comes to an end. Once the congestion window hits the congestion threshold, there is no other way to raise the transmission rate other than by adding 1 each time an acknowledgment is received. The linear rise keeps on until the sender experiences a time-out as a result of a missed acknowledgment or until the sender notices a gap in the data being transferred as a result of repeated acknowledgements for the same packet. The sender adjusts the congestion threshold to half of the current congestion window in either scenario. The sender begins transmitting a single segment when the congestion window is already set to one segment. Up to the new congestion level, the exponential increase resumes, after which the window expands linearly.

Fast retransmit/fast recovery

There are two factors that may lower the congestion threshold. The first is if the sender consistently gets acknowledgments for the same packet. It notifies the sender that the recipient has received every packet up to the accepted packet in the sequence and is also continuing to receive packets from the sender. Not congestion, but a straightforward packet loss caused by a transmission mistake, is to blame for the pause in the packet stream. The missing packet(s) may now be sent again before the timer runs out. Fast retransmit is the term for this activity. It is an early improvement to stop sluggish starts from happening on losses that aren't brought on by congestion. The fact that acknowledgements have been received proves there isn't enough traffic to warrant a sluggish start. With the current congestion window, the sender may proceed. The sender quickly recovers from the lost packet. This method has the potential to significantly boost TCP's effectiveness. The second trigger for sluggish start is a time-out brought on by an undeceived acknowledgment. The slow start technique is activated by TCP when it detects network congestion utilizing rapid retransmit/fast recovery. This approach has the benefit of being straightforward. Performance is improved by small software modifications to the MH. FA and CH don't need any modifications. This scheme's drawback is that packet losses aren't sufficiently isolated. It mostly focuses on issues with handover. Additionally, it reduces efficiency when a CH sends packets that have already been delivered.

Traditional TCP issues in wireless contexts

If TCP is used in stationary networks with mobile receivers or senders, the Slow Start method reduces TCP's efficiency. Wireless connections have error rates that are orders of magnitude greater than stationary fibre or copper lines. This makes it difficult for TCP to compensate for packet loss. Packet loss might be caused by mobility itself. A gentle handover from one access point to another is often not achievable for a mobile end-system. Standard TCP responds with a delayed start if acknowledgements are absent, which is ineffective in the event of wireless connection transmission problems and ineffective during handover. If this behaviour is combined with wireless networks or mobile nodes, an unaltered TCP suffers a significant performance reduction.

Classical TCP Improvements

Indirect TCP (I-TCP): TCP connections are divided into a fixed portion and a wireless portion using indirect TCP (I-TCP). The example with a mobile host linked to the correspondent host's "wired" internet through a wireless connection and an access point is shown in the accompanying Figure 6.1.

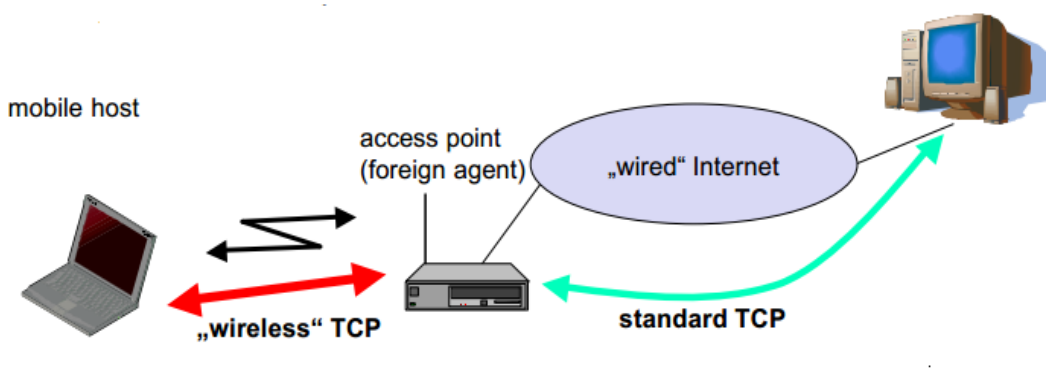


Figure 6.1: Classical TCP Improvements

The access point and the fixed PC communicate using standard TCP. No internet-connected machine can detect any modifications to TCP. Now, the access point, acting as a proxy, closes the regular TCP connection in place of the mobile host. As a result, the access point is now regarded as both the fixed host for the mobile host and the mobile host for the fixed host.

Specifically designed for wireless communications, a specific TCP is used between the access point and the mobile host. It is not necessary to alter TCP for the wireless connection, however. The foreign agent is an appropriate location for segmenting the connection because it not only already controls the mobility of the mobile host but also has the ability to transfer the connection to the next foreign agent when the mobile host changes. The foreign agent transmits all information in both ways and serves as a proxy. The FA acknowledges and transmits a packet to the MH if CH (correspondent host) sends one. On a successful reception, MH acknowledges, but only the FA uses this. To preserve dependable data delivery, CH doesn't notice when a packet is lost over the wireless connection and FA attempts to retransmit it locally. The FA acknowledges and transmits a packet to CH if the MH sends one. Due to the shorter round trip time, mobile hosts may immediately retransmit the packet if a packet is lost over the wireless channel. The foreign agent is now in charge of managing packet loss in the wired network.

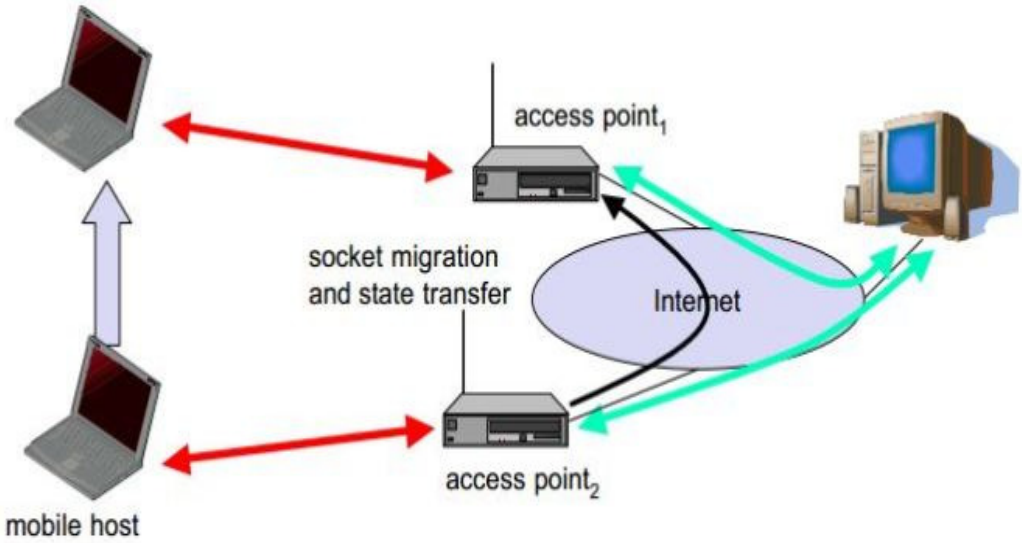


Figure 6.2: Socket migration after handover of a mobile host (I-TCP)

Migration of sockets and states after the transfer of a mobile host

The system state (packet sequence number, acknowledgements, ports, etc.), as well as the buffered packets, must be sent to the new agent during handover. For the mobile host, no new connections may be made, and there must be no changes to the connection status on the correspondent host. The following illustrates packet delivery in I-TCP.

Benefits of I-TCP

All present TCP protocol optimizations continue to function; no modifications to the fixed network are required; no changes are required for the hosts.

Mobile TCP is utilized just for one hop between, for example, a foreign agent and mobile host 1, making it easy to manage. The fixed network is not affected by transmission faults on the wireless connection

Consequently, a very quick packet retransmission is feasible due to the known small mobile hop latency.

It is never a good idea to add new processes into a complex network without fully understanding their behavior.

It is simple to employ various protocols for wired and wireless networks, allowing for the testing of new improvements at the last hop without compromising the reliability of the Internet.

Disadvantages of I-TCP

The loss of end-to-end semantics, which implies that a packet acknowledgment to a sender no longer guarantees that a recipient really got it, and the possibility of foreign agents crashing.

Data buffering inside the foreign agent and forwarding to a new foreign agent may result in higher latency.

Security concern: The foreign agent must represent a reliable organisation

Snooping TCP

The segmentation of the single TCP connection into two TCP connections, which results in the loss of the original end-to-end TCP meaning, is the fundamental downside of I-TCP. Snooping TCP is a brand-new improvement that entirely maintains the TCP connection while being completely transparent. In case of packet loss, the primary purpose is to buffer data around the mobile host for quick local retransmission (Figure 6.3).

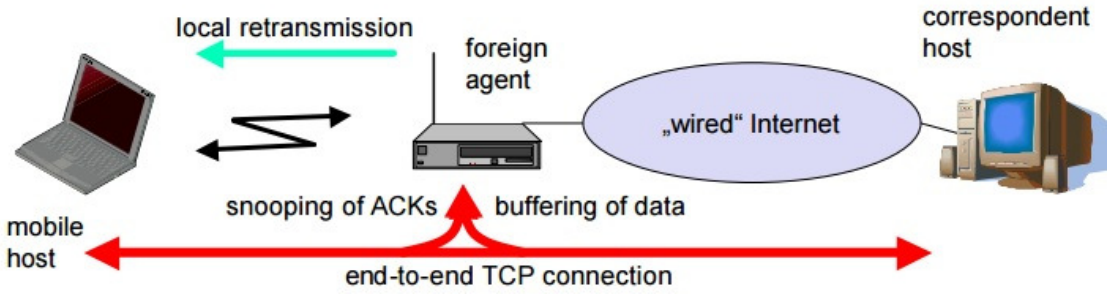


Figure 6.3: Snooping TCP

The foreign agent in this case buffers all packets with mobile hosts as the destinations and also ‘snoops' packet flow in both directions to identify acknowledgements. Every packet is buffered by the foreign agent until the mobile host acknowledges it. The packet or the acknowledgment has been lost if the FA doesn't get a response from the mobile host within a certain length of time. As an alternative, the foreign agent can get a duplicate ACK that also indicates a packet loss.

In contrast to the CH, the FA now performs a quicker retransmission by sending the packet straight from the buffer. Since acknowledging data to the CH would violate end-to-end semantic in the event of an FA failure, the FA does not do so for the sake of transparency. To prevent pointless retransmissions of data from the correspondent host, the foreign agent may filter the duplicate acknowledgements. The correspondent host's time-out still functions and forces a retransmission if the foreign agent now fails. Duplicate packets that have been locally retransmitted and acknowledged by the mobile host may be discarded by the foreign agent. By doing this, extra traffic on the wireless network is prevented.

The FA peeks into the packet stream during data transmission from the mobile host to the destination correspondent host to look for gaps in the TCP sequence numbers. The foreign agent sends a negative acknowledgement (NACK) back to the mobile host as soon as it notices a missing packet. The lost packet may now be sent again right away by the mobile host. TCP automatically rearranges packets at the correspondent host.

Snooping TCP has the following benefits:

The end-to-end TCP semantic is kept. The majority of improvements are made inside the foreign agent, keeping correspondent hosts unaltered. As soon as the mobile host switches to another foreign agent, there is no need for a handover of state. Despite the fact that there are packets in the buffer, the packets are transferred to the new COA when the CH times out. Whether the new foreign agent employs the enhancement or not, there is no issue. If not, the strategy immediately reverts to the default resolution.

The drawbacks of intercepting TCP

Snooping TCP does not isolate the wireless link's activity as effectively as I-TCP does. Transmission faults can spread all the way to CH. The mobile host is assumed to have extra mechanisms when negative acknowledgements are used between the foreign agent and the mobile host. Any mobile hosts can no longer use this method invisibly. If certain encryption techniques are used end-to-end between the correspondent host and mobile host, snooping and buffering data may be worthless. Snooping TCP may be utilised if encryption is applied above the transport layer (for example, SSL/TLS).

3TCP on mobile

In the event that a mobile host loses connection, neither I-TCP nor Snooping TCP are very helpful. Similar to I-TCP and snooping TCP, the M-TCP (mobile TCP) strategy aims to keep the sender window from becoming smaller if bit errors or disconnections but not congestion cause the present issues. M-TCP seeks to increase overall throughput, decrease latency, preserve end-to-end TCP semantics, and provide a more effective handover. Additionally, M-TCP is specifically designed to address issues brought on by prolonged or frequent disconnections. Like I-TCP, M-TCP divides the TCP connection into two halves. The SH-MH connection uses an optimized TCP whereas the normal host-supervisory host (SH) connection uses an unmodified TCP (Figure 6.4).

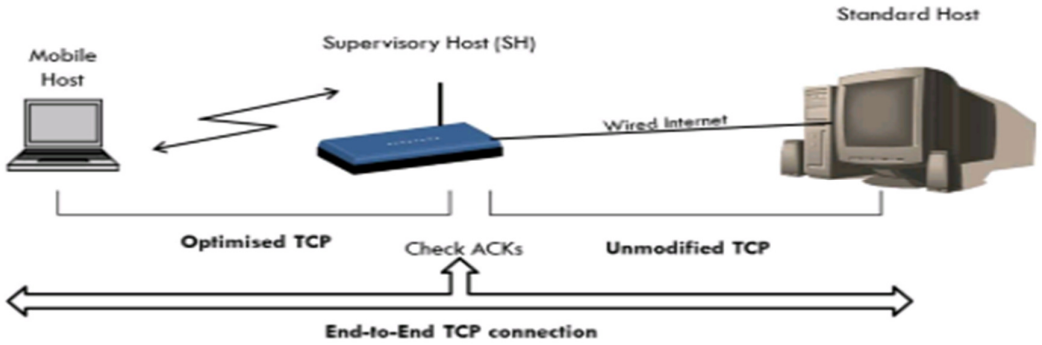


Figure 6.4: Mobile-TCP

Similar to the proxy in the I-TCP, the supervisory host (SH) is in charge of sharing data between the two components. The M-TCP method counts on a wireless network with a minimal bit error rate. As a result, it doesn't cache or retransmit data over the SH. The original sender must retransmit a packet if it is lost on the wireless connection. This preserves the end-to-end semantics of TCP.

Time-out and transmission freezing

Often, the MAC layer detects connection issues long before the connection is officially broken from a TCP perspective and also understands the true cause of the break. The MAC layer may alert the TCP layer to an impending loss of connection or the fact that congestion is not the root cause of the present outage. The congestion window and other timers are now "frozen" and TCP may cease transmitting. Both the mobile and correspondent host may be alerted if the MAC layer detects the impending disruption in time. Additional procedures in the access point are required in the event of a sudden break in the wireless connection in order to notify the correspondent host of the cause for the break. If not, the connecting host assumes congestion and starts to start slowly until disconnecting completely.

Advantages

1. The moment the MAC layer recognizes connection once again, it informs TCP that it may pick up where it left off, precisely where it had to stop. Timer expirations are not an issue since TCP simply does not progress.
2. It provides a technique to restart TCP connections even after lengthy breaks in the connection, and since it is independent of other TCP protocols like sequence number and acknowledgment, it may be used with encrypted data.

Disadvantages

1. The software of the MH, CH, and FA needs to undergo several adjustments.
2. Selective retransmission

The use of selective retransmission is a highly beneficial addition of TCP. TCP acknowledgements are cumulative, i.e., they confirm receipt of packets in the order they were sent up until a specific packet. A single acknowledgment verifies that all packets up to a particular packet have been received. Senders must go back and retransmit everything beginning with the lost packet if even one packet is lost (go-back-n retransmission). This blatantly wastes bandwidth across all networks, not just those connected to mobile devices.

TCP may obliquely request a selective resend of packets via selective retransmission. The receiver is not limited to acknowledging trains of packets that are sent in order.

The sender may now accurately identify which packet is required and resend it. This method has the clear benefit of resending just the missing packets. This reduces bandwidth needs and is quite beneficial for sluggish wireless networks. The drawback is that the receiver side software must be more complicated. Additionally, extra buffer space is required to wait for gaps to be filled and to reorder data.

Transact-Centric TCP

Consider a mobile application that periodically makes a brief request to a server, which answers with a brief message. This application needs dependable TCP transmission of the packets. It would be inefficient to utilise standard TCP because of the added overhead. Setup, data transport, and release are the three stages of a typical TCP connection. First, the connection is established through TCP using a three-way handshake. The request typically needs at least one more packet, and it takes three more packets to complete a three-way handshake to end the connection. Therefore, TCP may need a total of seven packets to convey one data packet. In fixed networks, this sort of cost is tolerable for extended sessions, but in wireless networks, it is highly ineffective for brief messages or sessions. As a result, transaction-oriented TCP (T/TCP) was created.

T/TCP has the ability to mix user data packets with packets for connection formation and release. As a result, there will only be two packets instead of seven. The apparent benefit for certain applications is the decrease in overhead associated with connection establishment and release in regular TCP.

The drawback is that it necessitates software adjustments for the mobile host chooser retransmission. The use of selective retransmission is a highly beneficial addition of TCP. TCP acknowledgements are cumulative, i.e., they confirm receipt of packets in the order they were sent up until a specific packet. A single acknowledgment verifies that all packets up to a particular packet have been received. Senders must go back and retransmit everything beginning with the lost packet if even one packet is lost (go-back-n retransmission). This blatantly wastes bandwidth across all networks, not just those connected to mobile devices.

TCP may obliquely request a selective resend of packets via selective retransmission. The receiver is not limited to acknowledging trains of packets that are sent in order. The sender may now accurately identify which packet is required and resend it. This method has the clear benefit of resending just the missing packets. This reduces bandwidth needs and is quite beneficial for sluggish wireless networks. The drawback is that the receiver side software must be more complicated. Additionally, extra buffer space is required to wait for gaps to be filled and to reorder data.

Transact-Centric TCP

Consider a mobile application that periodically makes a brief request to a server, which answers with a brief message. This application needs dependable TCP transmission of the packets. It would be inefficient to utilise standard TCP because of the added overhead. Setup, data transport, and release are the three stages of a typical TCP connection. First, the connection is established through TCP using a three-way handshake. The request typically needs at least one more packet, and it takes three more packets to complete a three-way handshake to end the connection. Therefore, TCP may need a total of seven packets to convey one data packet. In fixed networks, this sort of cost is tolerable for extended sessions, but in wireless networks, it is highly ineffective for brief messages or sessions. As a result, transaction-oriented TCP (T/TCP) was created.

T/TCP has the ability to mix user data packets with packets for connection formation and release. As a result, there will only be two packets instead of seven. The apparent benefit for certain applications is the decrease in overhead associated with connection establishment and release in regular TCP.

The disadvantage is that it requires software modifications on the mobile host and any related hosts. With this method, mobility is no longer concealed. T/TCP also has a number of security flaws.

Wireless Networking

Data management, processing, and communication all depend on computing technologies. Wireless simply refers to a connection between devices that is made without the use of a physical cable. Wireless computing is the transport of data or information between computers or devices that are "wirelessly network connected" but not physically attached to each other. Mobile devices, Wi-Fi, wireless printers and scanners, for instance, are some examples. Mobile devices are not physically connected, yet we can nevertheless transfer data via them. A computing device that is mobile does not need a network connection or any other kind of link to move data or information between devices. Laptops, tablets, smartphones, etc. are a few examples. Without a link to the base or a central network, mobile computing enables the transfer of data/information, audio, video, or any other document. These computers are currently the most popular technology. The following list of wireless and mobile computing technologies includes:

Global System for Mobile Communications (GSM): GSM is a modern wireless data communication technology that uses circuit switching. Throughout the middle of the 1980s, ETSI (European Telecommunications Standards Institute) established it in Europe. The GSM network has four distinct components, each with a distinctive set of functions: the Mobile Station, the BSS (Base Station Subsystem), the NSS (Network Switching Subsystem), and the OSS (Operation and Support Subsystem). GSM is a mobile communication system that is widely utilized, as the name suggests. It works in the 900, 1800, and 1900 MHz range of frequencies. To improve mobile communication, TDMA (Time Division Multiple Access) was used in the development of GSM. It is the most popular and necessary mobile communication system nowadays. It can transmit data at a maximum speed or rate of 9.6 kilobits per second (kbps) (Kilobits per second).

Code-Division Multiple Access (CDMA): This wireless computing technique is a kind of CDMA. During World War II, it is developed. This technique is mostly employed because it improves network quality, has more storage space for data communication than TDMA, uses power regulation to reduce system noise and interference, and offers more security by encrypting user transmission data into a special code. CDMA uses the entire spectrum of accessible frequencies for transmission rather than assigning any user a specific frequency. It uses frequencies between 800 MHz and 1.9 GHz to function. It employs Soft Handoff, which minimizes signal hiccups.

Wireless in Local Loop: A popular technique for wireless communication networks is WLL. Another name for it is a Fixed Wireless Loop. As wireless systems are less expensive because the expense of cable installation is not added, WLL is relatively simple to create and takes less time to implement. WLL offers cutting-edge customer service services and allows users to connect wirelessly to the local phone station. It offers fast data rates and high-quality data transfer. Local Multipoint Distribution Service (LMDS) and Multichannel Multipoint Distribution Service are the two main WLL approaches that are available (MMDS).

General Packet Radio Service: Packet-based wireless communication technology includes GPRS. It was created by ETSI (European Telecommunications Standards Institute). A data transfer rate of up to 114Kbps is possible using GPRS. It has a very low cost, is quite steady, and has a maximum data rate of 114 kbps (Kilobits per second). It supports the following protocols: Internet Protocol (IP), X.25 (a standard protocol for packet-switched data transfer), Point-to-Point Protocol (PPP), and is based on the modulation technique known as Gaussian Minimum-Shift Keying (GMSK). The two essential modules needed to allow GPRS on a GSM or TDMA network are the Gateway GPRS Service Node (GGSN) and the Serving GPRS Service Node (SGSN).

Short Message Service: SMS was initially developed for GSM (Global System for Mobile)-enabled phones and mobile devices. Using this service, text messages can be sent between two or more mobile devices even when there is no Internet connection. This method is the best for wireless communication because it is simple, comfy, and straightforward to use. Less time is needed for communication in this service. Text message sending does not require an internet connection. It permits the delivery of brief communications, i.e., those that are no longer than 160 characters. Standardized communication protocols are used by SMS. Short Message Service Center receives SMS (SMSC).

There were problems with mobile computing.

The use of mobile computer technology has many benefits. It offers a wide range of capabilities, including portability, cloud, and productivity. But along with these benefits, employing mobile computing technologies also provide some distinctively eye-catching problems. The problems we encounter when using fixed and wireless networks for mobile computing are listed below.

Expensive because of Wireless Medium: The implementation cost for mobile computing technology is usually high because it primarily concentrates on wireless infrastructure. It also has problems with efficiency, delays, and security that we must take into account when planning the project.

Problem caused by Device Mobility: One of the most important benefits of mobile computing technology is the portability of the device. But it is also a significant problem for it. We must install equipment that meets the highest standards in order to benefit from mobile computing technology's device mobility feature. Therefore, we must reorganize the setup environment of the mobile device whenever its environment changes. The device mobility feature needs to be frequently configured in accordance with the location, environment, and surrounds of a mobile device.

Problems with Security in Mobile Computing: Without a doubt, this is the most significant and frequently discussed problem relating to mobile computing. It results from mobile computing's capacity to share a medium.

CHAPTER 7

DATA MANAGEMENT ISSUES IN MOBILE COMPUTING

Dr. Sudhir Kumar Sharma, Professor,
Department of Electronics and Communication, School of Engineering & Technology, Jaipur National University, Jaipur, India,
Email Id- hodece_sadtm@jnujaipur.ac.in

One of the primary issues with mobile information systems is finding data management technologies that can provide simple data access from and to mobile devices. A variant of distributed computing can be referred to as mobile computing. Mobile databases are delivered in two different situations: The whole database is distributed among the wired components, with full or partial replication possible. A base station or fixed host manages its own database with DBMS-like capabilities, as well as additional capability for identifying mobile units and additional query and transaction management functions to match the needs of mobile environments. The database is dispersed among the wired and wireless parts. The task of managing the data is divided between the mobile units and base stations, or fixed hosts. The following are some of the problems that can occur when managing data for mobile databases:

Security

1. Data left at a fixed location is safer than data sent via a mobile device. Mobile data is less secure because of this.
2. Data are also getting more brittle, thus strategies need to be able to make up for their loss.
3. Authorizing access to crucial data and appropriate methods is the most crucial requirement in this mobile environment.

Replication and spread of data

1. Data is distributed unevenly among mobile devices and base stations.
2. In data replication and distribution, there is greater data availability and less expensive remote access.
3. Consistency restrictions make managing caches more difficult.
4. The Caches give the mobile units access to the most recent and frequently used data.
5. It handles all transactions on its own. Data can be accessed most effectively, and higher security levels are accessible.

Replication problems

The cost of updates and signaling has increased as a result of the growth in the number of copies. Anywhere and at any moment, mobile hosts can move.

Labor division

The division of work in query processing has changed somewhat as a result of some aspects of the mobile environment. In some circumstances, the client must work independently of the server.

Transaction models

The problems of transactional accuracy and fault tolerance are much worse in a mobile context. The ACID properties atomicity, consistency, isolation, and durability must all be met by

transactions. A mobile transaction is carried out sequentially according to the mobility of the mobile unit, maybe on different data sets and through multiple base stations.

It becomes difficult to enforce ACID characteristics when the mobile computers are unconnected. A mobile transaction is expected to last a lengthy time because of the disconnect in mobile units.

Recovery and tolerance for faults

A system's capacity to function correctly even in the face of internal errors is known as fault tolerance. There are two categories of faults: transitory and permanent. A temporary defect will gradually disappear without any visible intervention, while a permanent flaw will persist unless it is fixed by an outside agent. The mobile database environment has to handle failures in communication, media, transactions, and sites. There is a site failure at MU as a result of low battery power. It is not appropriate to consider a voluntary shutdown in MU to be a failure. Transaction failures during handoff most typically occur when Mu crosses the cells. Network segmentation and the affection of the routing algorithms are both greatly influenced by the failure of MU.

Interfaces to databases

A. Design of user-friendly query interfaces, like Query by Icons, takes into account the screen size, memory and battery capacity restrictions, and the constrained communication bandwidth.
B. How to use the voice and pen instead of the mouse and keyboard.
C. The use of a computer to implement a drawing-based graphical database interface.

More than only data synchronization is necessary for modern mobile apps. They need a full range of data management services, such as robust data modelling, mobile and systems ” for schema deployment and versioning, rules-based data distribution, quick and secure bidirectional data transactions, mobile device-based database services, and tight transaction-level incorporation with numerous enterprise information sources. The environment for mobile computing is seen as a distributed computing. The entire database might be split into wired parts, much like in mobile switching stations. This is one strategy. However, the following method distributes the complete database among the computer systems' wired and wireless components. The architecture of the database and database replication are two examples of the factors that affect and complicate database administration.

Database: An organized collection of documents or data is known as a database. Data is stored in databases in a certain logical order. The server or network are not always accessible to mobile devices, nor does the device constantly retrieve data from them for use in computations. Instead, when it is linked to the server or network, the device temporarily stores certain particular data that may be needed for calculations in the future. Saving a copy of specific data or a section of a database from a linked system with a big database is called caching. The mobile device database contains the cached data. By storing the cached data in the database, it is made possible for computation to continue even when the device is not connected to the network.

Database Hoarding

Database hoarding might occur right inside the application layer. A straightforward design where a mobile device API directly pulls the data from a database is shown in the following figure 7.1. It also displays a different straightforward design, such as IBM DB2 Everyplace, where a mobile device API directly accesses data from a database using a software (DB2e).

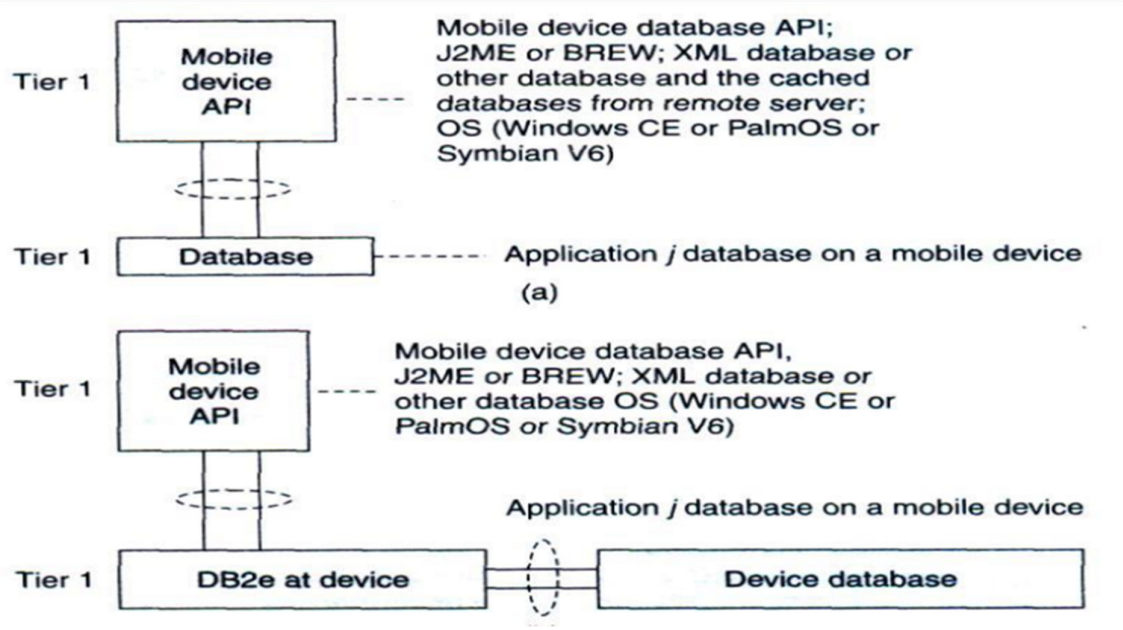


Figure 7.1: Sending queries and obtaining data from a local database via an API on a mobile device (Tier 1) and DB2e is used to get data from a database through an API on a mobile device (Tier 1)

Because the databases are designed specifically for a mobile device, are not intended to be spread to many devices, are not synced with new updates, and are stored on the device itself, both of the two systems fall under the category of one-tier database architecture. Examples include downloadable music, ringtones, and other media. An on-device relational database engine called IBM DB2 Everyplace (DB2e) has been created. It is compatible with J2ME and the majority of mobile operating systems. At the synchronisation, application, or enterprise server, DB2e and DB2 databases are synchronised. The two-tier or multi-tier databases shown below use the database architecture. Here, copies of the databases are cached at the client tiers and are located on distant servers. A cache is a collection of things or records that are kept on the device. Databases are stored at the corporate or application layer, where the database server employs connectivity and business logic to get the data and provide it to the device. Every mobile device linked to the server receives and updates a local copy of the database. The cached local copy is used by the computing API on the mobile device (first tier). The computer architecture described above is used by the API at the first layer (tier 1) to access the cached data records. Using business logic, the server obtains and sends the data records from tier 2 or tier 3 to tier 1 while synchronizing the local copies on the device. These locally stored copies serve as device caches.

The benefit of hoarding is that it eliminates access latency, which is the delay experienced while getting a record from a server via a wireless mobile network. Hoarded or cached data is instantly accessible via the client device API. The data is stored at the device once it has cached the server's dispersed data. The drawback of hoarding is that it requires ongoing maintenance of the consistency of the cached data with the server's database.

Caching of data

Distributed or sent from business servers or application databases to mobile devices are hoarded copies of the databases stored at the servers. In a multiprocessor system with a shared

main memory and copies of the main memory data stored at various places, the copies cached at the devices correspond to the cache memories at the processors.

Protocols for Cache Access: The pushed (disseminated) data records from a server are cached on a client device. In comparison to the pull (on-demand) method of data retrieval, caching of the pushed data results in a shorter access interval. Data records may be cached depending on submitted "hot records" (the most needed database records at the client device). The ratio of the two parameters access probabilities (at the device) to pushing rates (from the server) for each record may also be used as the basis for caching. Cost-based data replacement, often known as data caching, is this technique.

Pre-fetching: Another option to caching distributed data is pre-fetching. Pre-fetching involves locating and retrieving records that may be needed in the future. Pre-fetching from the pushed data on the client device is an alternative to caching while keeping future demands in mind.

Prefetching lowers the burden on the server. In addition, the expense of cache-misses may be decreased. The time required to access a record at the server in the event that it cannot be located in the device database when needed by the device API is referred to as the "cost of cache-misses."

Mechanisms for Caching Invalidation

The client device may invalidate a cached record. This might be as a result of the database server changing or expiring the entry. When a cached data item or record is modified, expires, or is invalidated at another computer system or server, the cache is invalidated, rendering the item or record useless. When a processor writes (modifies) cache-data in a multiprocessor system, cache invalidation methods are employed to synchronise the data at other processors. Cache invalidation mechanisms are also active when mobile devices have disseminated copies from the server.

A cache is made up of several records. Each record is referred to as a cache-line, and duplicates of these records may be kept on other hardware or servers. One of four potential tags denoting the cache's state—modified (after rewriting), exclusive, shared, and invalidated (after expiration or when fresh data becomes available—can be applied to the cache at the mobile devices or server databases at any given moment. The letters M, E, S, and I, respectively, stand in for these four states (MESI). The following statuses are denoted by the various tags:

a) The exclusive status, denoted by the E tag, denotes that the data record is intended only for internal use and cannot be accessed by any other server or device.

b) The S tag designates the shared state, indicating that other people may utilise the data record.

c) The device cache has been changed, as indicated by the M tag.

d) The I tag denotes an invalidated state, indicating that the record that was previously shared and utilised for calculations is no longer present in the server database.

In the accompanying picture, a data record I and its copy on the mobile device j are shown in their four potential states in the server database at any one time (Figure 7.2).

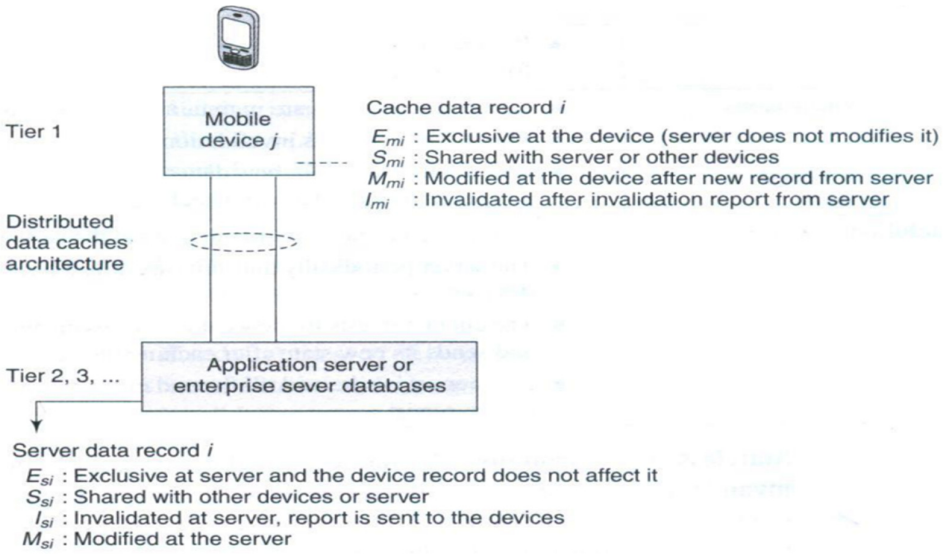


Figure 7.2: A data record may be in any one of four states (M, E, S, or /) in the server database and device j cache.

Cache consistency is a crucial element for cache management in a mobile context (also called cache coherence). In order to do this, a method must be put in place to guarantee that a database entry is same both on the server and in device caches, and that only valid cache entries are utilised for calculations. The server initiates or triggers the mobile device's cache invalidation procedures. Stateless asynchronous, stateless synchronous, stateful asynchronous, and stateful synchronous are the four potential invalidation techniques.

Stateless Asynchronous: The invalidation of the cache is broadcast to all of the server's clients as part of a stateless approach. The records kept in device caches are not tracked by the server. Regardless of whether the device cache contains that specific item or not, it merely uniformly broadcasts invalidation reports to all clients. According to the definition of "asynchronous," an item's invalidation information is provided as soon as its value changes. The server does not save information about a data record's current state (whether Emi, Mmi, Smi, or Imi) in cache for subsequent broadcasting. The server merely promotes the information about invalidation.

When data is pushed from the server, the client has two options: request an updated copy of the record or cache the pertinent record. When the matching data record at the server is invalidated and changed, the server advertises (deleted or replaced).

The benefit of the asynchronous technique is that there are less frequent, pointless data report transfers, which increases the mechanism's bandwidth efficiency. The drawbacks of this strategy are that (a) every client device receives an invalidation report, regardless of whether that client needs the copy or not, and (b) client devices assume that the copy is valid for use in calculations as long as there isn't an invalidation report. As a result, even in the event of a connection loss, the devices may still be utilising erroneous data, and the server may not be notified of client state changes after sending the invalidation report. This mode is similarly stateless, meaning that the server is unaware of the status of the data records it stores and broadcasts to all client devices. In contrast to the asynchronous technique, the server in this case broadcasts invalidation information both on a regular basis and whenever the related data record is updated or deleted. This technique provides synchronisation because even if a connection failure prevents the device from detecting the in-between period report, the device anticipates the periodend report of invalidation and makes a request for the same if it is not

received at the end of the period (deleted or replaced). If a connection problem prevents the client device from receiving the periodic report, it asks the server to deliver it.

The benefit of the synchronous method is that client devices regularly get information about the validity (and therefore invalidity) of the data caches. Since the device-client may send update requests for faulty data to the server in response to the periodic invalidation alerts, cached data is more reliable. Through regular exchanges, this also aids the server and devices in maintaining cache consistency. This mode of cache invalidation has the following drawbacks: (a) unnecessary transfers of data invalidation reports occur; (b) every client device receives an advertised invalidation report on a regular basis, whether or not that client has a copy of the invalidated data; and (c) during the interval between two invalidation reports, the client devices assume that the copy is valid for use in computations as long as there is no invalidation report. Consequently, the devices utilise previously stored data in the event of connection outages. After sending the invalidation report, the server is uninformed of state changes at the clients since the invalidation occurred within the interim time.

Decentralized Asynchronous the AS (asynchronous stateful) scheme is another name for the stateful asynchronous mechanism. The word "stateful" denotes that the cache invalidation reports are disseminated to all client devices, but are instead delivered just to those that are impacted. The server keeps track of each data record's current state in the client device caches (a record I may have its state as Emi, Mmi, Smi, or Imi). The server's home location cache (HLC) is where this state data is kept. A programme called HA (home agent) keeps the HLC up to date. This is comparable to a mobile network's HLR at the MSC. To allow storage of each record at the HLC, the client device notifies the HA of the condition of each record. Only the device-clients who are impacted by the data invalidation get the invalidation information from the server when and as the records are invalidated. These device-clients then ask the server for fresh or changed data to replace the invalidated data based on the invalidation information. After the server's data records have updated the client device's cache, the client device updates the server's cache-state record by sending information about the new state.

The server maintains track of the status of cached data at the client device, which is a benefit of the stateful asynchronous method. This makes it possible for the server to update the HLC and stay in sync with the status of the data in the device cache. The stateful asynchronous mode also has the benefit of preventing a deluge of irrelevant reports on other devices by only sending invalidation reports to the impacted clients. The AS technique has the drawback that client devices assume the copy is valid for use in calculations if there isn't an invalidation report. As a result, the devices utilise invalidated data when there is a connection failure.

Defined Synchronous: The client-caches are where the server stores the information about the current state (Emi, Mmi, Smi, or Imi) of data records. Using the home agent, the server saves the cache record state in the home location cache (HLC) (HA). When a client-relevant data record is invalidated, updated (removed, or replaced), or replaced at the server, the server broadcasts the invalidation information to the clients at regular intervals. Because the device expects to receive the period-end report of invalidation and requests it if it is not received at the end of the period, this technique assures synchronisation even if the in-between period report is not noticed by the device due to a connection failure.

The benefit of the stateful synchronous method is that reports detecting invalidity (and indirectly, validity) of data caches are generated at regular intervals, and the server also updates the client-cache states that are kept in the HLC on a regular basis. By synchronising with the client device when incorrect data is changed and made legitimate, this allows for communication. Additionally, since the invalidation report is delivered on a regular basis, a

device may ask the server to transmit it if it hasn't arrived after a certain amount of time. Thus, each client may get regular updates on any server adjustments. When a connection failure is discovered at the device and the invalidation report is not received within the predetermined time frame, the device does not utilise the invalidated data. Instead, it asks for an invalidation update from the server. The high bandwidth need for frequent transmission of invalidation reports to each device and updating requests from each client device is a drawback of the stateful synchronous method.

Maintenance of Data Caches in Mobile Environments

Assume that during an application, a device requires a data record. The server must receive a request for the data record (this mechanism is called pulling). Access latency is the length of time it takes for the application programme to access a certain record. Access latency is eliminated when the record is cached and stored locally. In order to reduce access latency in a mobile context, data cache maintenance is required.

Data records cached for applications are not invalidated at the device when changed at the server but not when changed at the device, which is referred to as data cache inconsistency. The three techniques listed below may be used to preserve data cache consistency:

I. The cache invalidation mechanism (server-initiated scenario) delivers invalidation reports when records become invalid (asynchronously or periodically) (synchronous).

II. Polling method (client-initiated case): Polling refers to determining a data record's status from the server, such as whether it is valid, invalid, changed, or exclusive. The application programme periodically polls each cached record copy during a calculation. The device requests the changed data and replaces the previous cached record copy if it discovers that the record has been altered or rendered invalid.

Each cached record is given a TTL (timetolive), or time-to-live mechanism (client-initiated example). The TTL assignment is adaptive (configurable) based on the record's prior update intervals. The cached record copy is polled once the TTL expires. If it has been changed, the device asks the server to update the outdated cached record with the new information.

The TTL mechanism is comparable to the polling method when TTL is set to 0.

Maintenance of Web Caches in Mobile Environments

A web server (such as a traffic information server or a train information server) may be linked to mobile devices or their servers. Similar to how server data is maintained in a cache, the web cache at the device stores and manages the web server data. There is access latency if a device-based application requires a data record from the web that is not already cached on the web. In a mobile setting, web cache upkeep is required to reduce access latency while downloading from websites due to disconnections. There are two ways to ensure web cache consistency. These are: Mechanism for time-to-live (TTL) (client-initiated case): The procedure is the same as that described for maintaining the data cache.

II. Client-initiated power-aware computing mechanism: The CRC (cyclic redundancy check) bits may be stored in each web cache that is kept on the device. Suppose there are n CRC bits and n cached bits. N is a huge increase over n. The server stores n CRC bits similarly. The CRC bits at both are the same as long as the server and device records are consistent. The associated CRC bits at the server are likewise changed whenever any of the cached entries are. The cached record CRC is polled and received from the website server once the TTL expires or on-demand for the web cache records via the client API. The client device retrieves the changed portion of

the website hypertext or database for use by the API if the n CRC bits at the server are discovered to have been updated and the change is found to be significantly greater than a predetermined threshold. The API, however, utilises the earlier cache if the change is minimal. Since N > n, the web cache maintenance method's power consumption is significantly higher than the current method's (in which the device polls for a significant change in the CRC bits at the server and transmits records only when there is a significant change in the CRC bits), which only transmits invalidation reports and all invalidated record bits when there is a significant change in the CRC bits).

Computing that is client-server

A distributed computing architecture known as client-server computing uses servers and clients as its two kinds of nodes. A computer system that answers to requests from one or more clients is referred to as a server. A computer system that asks the server to provide a resource or carry out a job is referred to as a client. The client may store these records on the client device or get them directly from the server. The data may be accessible at the client's request, through broadcasts, or by server distribution.

The client and server may be running on the same computer system or may be running on distinct computers. N-tier architecture for client-server computing is possible (N = 1, 2...). The number of tiers, N, is equal to 1 when the client and server are on the same computer system. N = 2 when the client and server are on separate computer systems connected to a network. To get client requests at the server or server answers at the client, a command interchange protocol (like HTTP) is utilized Client-server architecture with two tiers.

The 2, 3, or N-tier client-server architectures are described in the following subsections. A connecting, synchronizing, data, or command exchange protocol is used to link one layer to the others.

Two-tier & Three-tier Architecture

The term "tier" is often used to describe the logical or functional layering of software on several physical locations or hardware. A multi-tier architecture is a method of dividing software into multiple distinct domains that each perform a specific function, such as data administration, logic, or display. Two-tier and three-tier architecture are primarily the topic of this article.

Two-Tier Architecture

The Client-Server concept is the foundation of the two-tier architecture. It is made up of a database layer and a client-application tier. Direct communication exists between the Client-Application server and the Database server. Because there is no middleware, the two components may exchange data or information quickly (Figure 7.3).

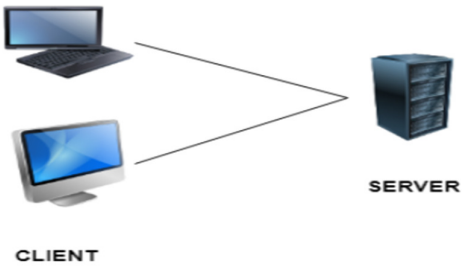


Figure 7.3: Two-Tier Architecture

The programmes for interacting with users and storing data on the database server are both included in the client-application. The client application sends the request to the server, which responds with data after processing it. Because of this, the client application also manages the application layer's presentation layer (application interface) (logical operations). C, C++, Java, Python, PHP, Rails, and.NET are just a few of the languages that may be used to create the client-application layer. The data management layer, on the other hand, is handled by the database server. The data storage (database or file system) and techniques for storing and retrieving data from the data storage make up the data management layer. Databases like MySQL, MongoDB, PostgreSQL, and SQLite are often used. Hosting is available both on-site and in the cloud. Examples of two-tiered applications include desktop programmes, video games, and music players.

Advantages:

A. Implementation is quick and simple, and communication is quicker.
B. It works well in settings with static business rules or logic activities.
C. Disadvantages:
D. Performance suffers as the number of users grows since it is difficult to scale.
E. Data integrity problems might occur as a result of the server handling many requests at once.

Three-Tier Architecture

The Client-Application tier of the three-tier architecture is separated into two, similar to the two-tier architecture. In other words, it is a client-server architecture that is modular and made up of three different tiers: presentation (client), application, and data. Modern online apps are an excellent illustration (Figure 7.4).

The Presentational tier is the application's uppermost layer. This interface layer converts operations and data into terms the user can comprehend. Additionally, it receives user requests, delivers them to the application layer for processing, and then provides the finished product back to the user. The HTML, CSS, VBScript, and Javascript frameworks React, Angular, and Vue.js may be used to build this layer. Additionally, distribution is available using web browsers like Chrome and Firefox.

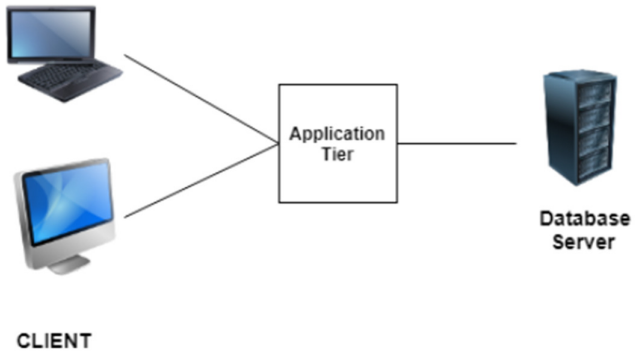


Figure 7.4: Three-Tier Architecture

The intermediate layer is the application tier. Business rules and logical operations are included in this layer and are used for processing instructions, validating logic, making decisions, interacting with databases, and organising web applications. Frameworks such as Django, Rails, Spring, Laravel, and.NET may be used to implement it. Javascript is also usable when

run on Node.js. It is possible to deploy on shared or dedicated internal servers, including Microsoft's Internet Information Server, Nginx, Apache, and Puma.

The back-end layer is the Data tier. It is comparable to the two-tier database server in architecture.

Advantages:

Due to the dispersed deployment of application servers, it grows horizontally effectively. For instance, the presentational layer may load-balance across the current servers when there are additional web requests. Because the concerns are separated, the individual components may be reused more effectively (interface, logic and data management).

The presence of middleware between the client and the server improves data integrity. Each layer is handled as a separate entity, which simplifies maintenance. Because cashing requests at the presentational layer reduces network use, performance is enhanced.

Three-tiered flaw:

- Increased complexity.

N-Tier

Applications that are spread among three or more different machines in a distributed network are referred to be N-Tier applications.

The 3-tier Application, which is divided into three categories, is the most prevalent kind of n-tier.

Computer programming for user interfaces

Business logic is stored in a more centralised computer, while necessary data is stored in a database-managing computer.

With the largest amount of flexibility, this architectural style enables software developers to build reusable applications and systems.

A number of tiers or levels are employed in N-tier, such as 2-tier, 3-tier, or 4-tier, etc. It also goes by the name "Multi-Tier Architecture."

An established paradigm of software architecture is the n-tier architecture. By offering solutions for scalability, security, fault tolerance, reusability, and maintainability, it is appropriate to handle enterprise level client-server applications. It aids in the creation of reusable and adaptable apps by developers.

N-Tier Structure

Here, the display, application, and database levels of an n-tier system are shown diagrammatically.

Diagram of an N-tier architecture

Depending on the needs, these three levels may be further separated into several sub-layers.

Several well-known websites that use this architecture include, MakeMyTrip.com

Amazon.com, IRCTC of Indian Railways, Sales Force corporate application, etc.

There are a few basic words to keep in mind in order to comprehend the idea more fully.

Distributed network: This kind of network design uses passing messages to coordinate and communicate the operations of the components situated at network machines. Although it is really a collection of several systems located at various nodes, to the user it is one system. It offers a single data transmission network that may be independently maintained by many networks. An example of a distributed network is one in which many customers are linked on one side through LAN architecture and on the other side via high-speed switches and a server rack housing service nodes.

Server-Client Architecture: It is a request-response service offered via the internet or through an intranet where the client (one programme) seeks a service from a server (another programme). In this paradigm, the client will behave as a single programme or piece of code that carries out a number of operations via the network. On the other hand, the server is a collection of different programmes that provides the desired result sets to the client system. In this scenario, a client computer acts as an interface for an end user to ask a server for a service or a resource. The server, on the other hand, processes the request and shows the end user the outcome. An ATM is a good example of a client-server model. The client with a user interface and some rudimentary application processing is an ATM machine, whereas the server is a bank for processing applications within the vast customer databases.

Platform: A platform is a system that allows applications programmes to execute in computer science or the software business. It is made up of a mix of hardware and software that includes built-in instructions for processors or microprocessors to carry out certain tasks. In plainer terms, a platform is a basis or system where any apps may operate and carry out an operation to complete a certain job. An illustration of Platform Consider a personal computer running Mac OS X or Windows 2000 as two instances of distinct platforms. A database is a collection of data that has been structured for easy access, management, and updating.

N-Tier Architecture Advantages

A. Using n-tier architecture in your program has several advantages. Scalability, simplicity of administration, adaptability, and security are these.
B. Secure: Various techniques may be used to independently secure each of the three stages.
C. Simple to manage: Each layer may be managed independently, adding or changing information without impacting the other levels.
D. Scalable: You can scale up or down each tier without impacting the other tiers if you need to add additional resources.
E. Flexible: In addition to isolated scalability, you may increase each layer whatever your needs need.

In other words, n-tier architecture allows you to embrace new technologies and add additional components without having to completely rethink your programme or rewrite your whole application, making it simpler to grow and manage. In the meanwhile, you may store private or sensitive data in the logic layer and keep it separate from the display tier to increase security.

Other advantages are:

More efficiency in development: The ability for various teams to operate on each tier of an N-tier architecture makes it ideal for development. You can be certain that the design and presentation experts work on the presentation layer and the database specialists work on the data tier in this fashion.

Adding new features is simple: may add a new feature to the appropriate tier without changing the other tiers if you choose to do so.

Simple to reuse: Each layer of the programme may be readily reused for other software projects since it is separated into distinct tiers. For instance, you can simply duplicate the logic and presentation layers and then construct a new data tier if you wish to utilise the same application with a different set of data.

Computing that is Context-Aware

A mobile device's context refers to the conditions, events, uses, or physical setting in which the device is being used. For illustration, suppose a cell phone is in use in a crowded, bustling region. If the gadget is aware of the background noise, it may automatically increase the speaker level during the discussion and then automatically decrease it when the user departs the area. Additionally, if there is a brief loss of connection during the chat, the device may automatically add background sounds so that the user is not bothered by the brief silences. This is one instance where the computer system is aware of the physical environment in which the communication is occurring.

A context-aware computing system has user, device, and application interfaces so that it can use them to keep track of past and present situations, circumstances, or actions. Examples include the current mobile network, nearby devices or systems, changes in the connectivity network's state, physical parameters like the current time of day, currently available nearest connectivity, and currently remaining memory and battery power.

Database Context Issues

The word "context" describes the interconnected circumstances in which a group of items, records, components, or phenomena exist or take place. Every communication, piece of data, component, or object has a purpose. But when they are taken into account together with the circumstances connecting them to one another and to the environment, they take on a larger significance. Better, more effective computing techniques arise from knowing the environment in which a device is intended to work.

Regulatory Context: Let's look at a few recordings with structural arrangements to illustrate what is meant by structural context. A person's name, location, experience, and accomplishments are each given a unique significance. However, when combined to create a CV, these areas take on implications that go beyond their separate connotations. The fact that they are now grouped in a structure that suggests an interaction between them gives them this meaning. A context for these records is defined by the structure of the resume, which contains the records and how they are related to one another. Thus, the data take on a new significance when seen in the context of a résumé (which is a structure). Structural contexts include contexts like the context of a person's résumé. The manner or format in which entries in a database are arranged provides context in certain situations.

Think of a different illustration, like a line in a phone book. A name, an address, and a 10-digit number are among the entries in a series that it contains. Each entry in a record has a distinct significance. However, a grouping of these information demonstrates an interaction and, as a result, establishes a context, such as a telephone directory.

Contexts, both explicit and implicit the context might be either explicit or implicit. Implicit context allows for omissions by omitting unnecessary facts, adopts different worldviews, and makes changes to messages invisibly in order to deal with incompatible protocols, interfaces, or APIs. Implicit context alters contextual messages by looking at call histories, managing

omissions, identifying receivers, and managing omissions. Take a look at the context "Contacts," which has a list of contacts. In the context of Contacts, the name, email address, and phone number are implied in a contact. The system takes an independent perspective, utilises the phone number implicitly, and deploys CDMA or GSM methods for connecting to the mobile network when a computing device uses a contact to call a number using a name record. The records' definition of "Contact" includes the context CDMA. The usage of the email ID record and SMTP (simple mail transfer protocol) or another mail sending protocol is implicit to the system when a computing system utilises a contact to send an email using a name record. When an email is sent, the name is immediately changed to the email ID. The implicit context also handles interfaces with incompatibilities, such as email sending and receiving programmes that handle data in various formats. Take a look at the context document. Contact or personal information is an extrinsic context in a document. Contact information for a document's authors is extrinsic in the context of processing that document. To create a connection between the document and the contact, the contacts context is imported into the document context.

Computing that is Context-Aware

Application-aware computing follows context-aware computing. This is true since the context includes the APIs (implicit or explicit contexts). For instance, if the context is a contact, the phonetalk program will adjust to utilise the contact's phone number as well as GSM or CDMA connection.

Contextual computing reduces the probability of mistakes. It aids in eliminating action's ambiguity (s). It aids in determining the anticipated reaction of the system to calculations. For instance, if a name is entered in a personal biodata context, calculations also call for the location, experience, and accomplishments that go along with that name. This is so because the context of biodata requires all four, which are connected. When a name is entered in the context of a telephone directory, the address and phone number that go along with that name are also needed for calculations. This is due to the fact that all three have a connection to telephone directories. When doing calculations, the name must be treated differently in two separate contexts (personal biodata and telephone directory).

Types of Context in Context-aware Computing

Physical context, computer context, user context, temporal context, and structural context are the five kinds of contexts that are crucial in context-aware computing.

Actual Context: The physical environment itself may serve as context. Service disconnection, light level, noise level, and signal intensity are the factors that define a physical environment. For instance, if a service outage occurs while a conversation is being held, the mobile device may detect the change in the environment and will intersperse background noise to mask the impacts of the outage.

Additionally, since the mobile device can detect light levels, its display brightness changes depending on the time of day. At night or in low light, it will be less bright. The device display is modified in response to the physical situation.

Computer Environment: A context-aware computer environment's context could also be the computing context. The relationships and circumstances of the network connection protocol being used (Bluetooth, ZigBee, GSM, GPRS, or CDMA), bandwidth, and available resources

determine the computing environment. Resources include things like a keypad, a display, a printer, and a cradle.

The device that a mobile device rests on in order to connect to a nearby computer is known as a cradle. Think of a smartphone resting on a cradle. It finds the computing environment and syncs with and downloads data from the PC via ActiveSync. When a mobile device is close to a computer with a Bluetooth interface, it finds another computing context resource and connects to the computer wirelessly via Bluetooth. It employs a GSM, CDMA, GPRS, or EDGE connection while operating independently and connecting to a mobile network. It also learns about another computer environment. The system reacts in accordance with the network connection protocol and the computing situation.

User Context: The user context is comprised of the user's location, profile, and nearby people. Reza B. Far writes, "Within the domain of user interfaces, we may define context as the total of the connections between the user interface components, the state of the user, the main purpose of the system, and all of the other aspects that enable users and computer systems to interact.

Temporal Context: The relationship between time and the occurrence of an event or action is defined as temporal context. There is an internal or extrinsic temporal context for a collection of interface elements. Consider the scenario when a user instantly pushes the dial button on a mobile device. The gadget then looks for a number as an input. The user will then take that into account while dialling and enter the number. Let's now imagine that the user adds a contact to the mobile device at a later period. The system asks the user to type a number once again.

The user will take this into account when deciding which number to add to their contacts list and keep on their device for further usage. The gadget then looks for the contact's name as input.

In these situations, the system reacts in accordance with the temporal environment. Depending on the instances and sequences in which the VUI (voice user interface) components occur, the context for those elements also establishes a temporal context.

Regulatory Context: The term "structural context" refers to the order and structure that the components or records form. The structural context of GUI components is present.

Extrinsic structural context is another possibility depending on the kind of context. The structural placements of the GUI components on the display screen determine how they interact. The hour and minute components come into play when time is the context.

Models of transactions

A transaction is the sequential execution of associated instructions for a particular database activity. Data integrity must be preserved, and the ACID criteria must be followed by database transaction models. Here are some guidelines:

Atomicity: A transaction must have all of its activities finished. A transaction must be undone if it cannot be completed (rolled back). The assumption is that an operation is a single, indivisible item (atomic unit).

Consistency: For a certain database, a transaction must be consistent in that it upholds the integrity requirements and adheres to the defined consistency criteria. Consistency indicates that after the transaction, the data is not in a conflicting state.

Isolation: There shouldn't be any interference between two transactions that are running concurrently. A transaction should also be invisible to any intermediate outcomes from any other transactions.

Durability: A transaction must continue once it is finished and cannot be stopped or deleted. For instance, after a transaction involving the transfer of a balance from account A to account B is complete and done, there should be no rolling back of the transaction.

Process of Recovering Data

In the event of physical media destruction, purposeful assault on the database and transaction logging data, or media failure, data cannot be recovered. However, in other circumstances, data recovery is feasible. A recovery management architecture is shown in the figure below. It makes use of a recovery manager to guarantee atomicity and robustness. Atomicity makes sure that a begun but uncommitted transaction fails and that the failed transactions are recorded in the log file. A committed transaction is protected against failure and is recovered through durability. Secondary storage houses stable state databases that are used at the beginning and conclusion of transactions. Fetch commands are delivered to the database manager by the recovery manager after transaction instructions have been issued to it. Using a database buffer, the database manager handles the queries throughout the transaction. The committed transactions and database buffer data are transferred to the secondary storage by the recovery manager using the flush instructions. The recovery manager recognises the outcomes of actions. The backup storage is used to restore lost operations. Data that was lost during the transaction may be found and recovered (Figure 7.5).

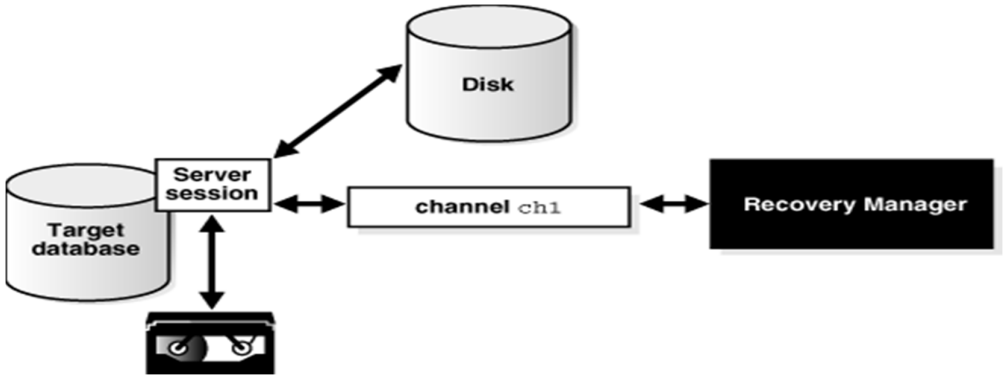


Figure 7.5: Architecture for Recovery Management

The recovery manager makes use of a log file, which records events as follows:

1. Each insertion, deletion, replacement, and addition command for an update transaction must be recorded.
2. Database read commands are not recorded
3. Log files are kept on another kind of storage media.
4. Following the final stable state database's storage, log entries are flushed away.

The following fields are included in every recorded entry.

Pre-operation and post-operation values of the object; transaction type (begin, commit, or rollback transaction); transaction ID; operation-type; object on which the operation is conducted; and

Data recovery also employs a process known as the Aries algorithm. The algorithm's fundamental stages are:

I. Review the data from the previous checkpoint and locate any dirty records in the buffer that were rewritten after the procedure was begun.
II. Complete and create the final page for all buffered actions noted in the update log.
III. Revert to pre-transaction settings and undo all write actions.

The following are the recovery models used to data recovery processes:

I. The comprehensive recovery approach produces incremental backups of the changes as well as database backups. All transactions are recorded from the database's most recent backup.

II. The bulk logged recovery model involves recording and backing up activities for large amounts of data records, but not complete logging and backup. The size of bulk logging is limited to an absolute minimum. This improves efficiency. By recovering the database using the bulk transaction log file backup, we can restore the database up to the point of failure. This contrasts with the complete recovery concept, where every operation is tracked.

III. The incremental changes are not reported even if the basic recovery model creates complete backups. The database may be restored to the most recent backup of the specified database.

Problems with data management in mobile databases

One of the key problems with mobile information systems is the availability of data management technologies that can facilitate simple data access from and to mobile devices. Distributed computing may be thought to take on a mobile form. The following are the two distribution possibilities for mobile databases: The whole database is spread across the connected components, with full or partial replication possible. In order to fulfil the demands of mobile settings, a base station or fixed host controls its own database with DBMS-like capability, along with extra capabilities for identifying mobile units and additional query and transaction management functions. The database is dispersed throughout the wired and wireless parts. The task of managing the data is divided between the mobile devices and base stations, or stationary hosts. The following are some of the problems that might occur while managing data for mobile databases:

1. Mobile database architecture: - The global name resolution issue is exacerbated by the frequent shutdowns and the need to handle queries.

2. Security: - When compared to mobile data, data left in a fixed place is safer. Mobile data is thus less secure. Techniques must be able to make up for data loss as data are getting more volatile. The most crucial requirements in this setting are adequate procedures and permitting access to crucial data.

3. Data replication and distribution - Here, data is distributed unevenly across mobile devices and base stations. In data distribution and replication, there is more data availability and less expensive distant access. Consistency restrictions make managing caches more difficult. The Caches provide the mobile devices access to the most recent and frequently used data. Their own transactions are processed. High security and most effective data access are both offered.

4. Problems with replication - As there are more copies, the cost of updates and signalling has increased. Anywhere and at any moment, mobile hosts may move.

5. Division of labour - Because of certain aspects of the mobile environment, there is a certain shift in the division of labour in query processing. In certain circumstances, the client must operate independently of the server.

6. Transaction models - Issues with transaction accuracy and fault tolerance are exacerbated in a mobile context. The ACID properties—atomicity, consistency, isolation, and durability—must all be met by transactions.A mobile transaction is carried out sequentially according to the movement of the mobile unit, sometimes across numerous data sets and across different base stations. It becomes difficult to enforce ACID characteristics when the mobile computers are unconnected. There is an anticipation that a mobile transaction would take a long time because of disconnect in mobile units.

7. Recovery and fault tolerance - A system's capacity to operate successfully even in the face of internal problems is referred to as fault tolerance. There are two categories of faults: transitory and permanent. A temporary defect will gradually vanish without any visible intervention, while a permanent flaw will persist until it is fixed by an outside agent. The mobile database environment must handle failures in communication, media, transactions, and sites. There is a site failure at MU as a result of low battery power. It is not appropriate to consider a voluntary shutdown in MU to be a failure. The majority of the time when Mu crosses cells, a transaction will fail at handoff. A major factor in network segmentation and the affection of the routing algorithms is MU failure. The definition of mobile computing is as follows:

A. Limiting the availability of resources
B. Recurring disconnects
C. Extreme mobility
D. Limited bandwidth

8. Position-based service - Determining the location of mobile users, which must be done in order to allow a location-based service, is one of the most difficult jobs that must be carried out. Cache information turns into a sale when customers relocate. Techniques for eviction are crucial in this situation. Issues with location and services include: Different mobile mapping standards and user privacy

A. Market potential
B. Interoperability

A issue arises when location-dependent queries are updated and then spatial queries are used to update the cache.

9. Query processing - Query optimization becomes the most challenging due to the mobility and quick resource changes of mobile units. When mobility is taken into account, query processing is impacted. A query answer must be sent to mobile units that could be travelling. In centralised systems, input/output costs are the ones that have the most impacts. The most significant factor in dispersed contexts is communication cost. There are ways to create location-based queries. Because the mobile host may be positioned in many places, it is challenging to estimate the communication costs in dispersed contexts. Dynamic optimization techniques are necessary in the mobile dispersed scenario.

CHAPTER 8

WIRELESS LAN IN MOBILE COMPUTING

Sindhu Madhuri G, Assistant Professor
Department of Computer Science Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Karnataka – 562112
Email Id- g.sindhumadhuri@jainuniversity.ac.in

LAN without cables is known as WLAN. Through LAN, mobile users may access data and network resources. Office cabling should be replaced with WLAN in order to speed up internet access and increase communication flexibility. It is used as an addition to a campus' or buildings wired LAN.

Application for Wireless LAN

Wireless LAN has several uses and application areas. The most dynamic environments are best suited for wireless LAN. These are the applications:

Cross-Building Interconnect: Wireless may be used to link LANs in adjacent structures. A point-to-point wireless network is utilised in this instance to connect two buildings. Bridges and routers are the devices linked.

Nomadic Access: This technology creates a wireless connection between a LAN hub and a mobile data terminal with an antenna, such as a laptop or tablet computer.

In a large setting, such as a campus or a company operating out of a collection of buildings, nomadic access is especially helpful.

Ad hoc networking is the temporary establishment of a peer-to-peer network to serve an urgent requirement. For instance, a group of workers may meet in a conference room to do business, each having a laptop computer.

For the length of the meeting, the staff members temporarily connect their computers into a network.

Requirements for Wireless LAN

The same kinds of criteria that apply to wired LANs, such as high capability, short-range coverage, and broadcast capabilities, must also be met by wireless LANs. Additionally, there are a few needs unique to the wireless LAN context. The most crucial prerequisites are as follows:

Throughput: To optimise capacity, the medium access control protocol should utilise the wireless medium as effectively as feasible.

Nodes: A wireless LAN may need to accommodate hundreds of nodes spread over many cells.

Connection to the backbone LAN: A wireless LAN requires an interconnection framework.

Service area: A wireless LAN's usual service area has a diameter of between 100 and 300 metres. Users would wish to purchase and use wireless LAN equipment without needing to get a licence for the frequency band that the LAN uses. License-free operation

Handoff/roaming: The wireless LAN's MAC protocol should make it possible for mobile stations to switch from one cell to another.

Dynamic configuration: The LAN's MAC addressing and network management features should allow for the automatic, dynamic insertion, deletion, and relocation of end systems without interfering with other users.

Battery power consumption: Workstations used by mobile employees must have a long battery life when connected to wireless adapters. When no one is using the network, wireless LAN implementations offer features such a sleep mode that use less power.

Benefits of Wireless LAN

A. Mobility: When workers have access to data and information from any place, productivity rises.
B. Low Implementation Cost: WLANs are simple to instal, move, modify, and administer.
C. Installation Ease: Installing a WLAN may be simple and quick, and it can save the need to run cables through walls.
D. Wireless technology enables network expansion by allowing it to reach areas where cables are unable to.
E. Reliability: WLAN is immune to various cable malfunctions.
F. Scalability: WLAN may be set up in a number of different topologies to suit the requirements of various installations and applications.
G. ISM band usage: WLAN uses unlicensed, publicly accessible ISM bands for operation.

Technology for Wireless LAN

The transmission mechanism that is utilised to create wireless LANs is often used to classify them. Products for wireless LANs nowadays fit into one of the following groups:

IR (Infrared) LANs Since infrared light cannot pass through opaque walls, each IR LAN cell can only cover one room. For IR data transfer, three transmission methods are used.

I. Point-to-point communications may be built using direct beam infrared technology. The radiated power and level of concentrating determine the range in this mode. A single base station that is in line of sight of every other station on the LAN is used in an omnidirectional configuration

(ii).This station is often suspended from the ceiling.

(iii) In a diffused arrangement, every IR transmitter is concentrated and pointed towards a specific location on a ceiling that is diffusely reflecting light. All of the nearby receivers take up the omnidirectional reradiation of IR energy that hits the ceiling.

LANs that use spread spectrum technology are known as spread spectrum LANs. These LANs typically function in the ISM frequency ranges.

Spread spectrum LAN employs a multiple cell configuration. The architecture inside a cell might either be hub or peer to peer. To link stations that are joined to the wired LAN and stations that are a part of wireless LANs in other cells, a hub that is normally installed on the ceiling and connected to a backbone wired LAN is used in a hub topology. A peer-to-peer technology lacks a central processing unit. Access is restricted using a MAC technique like CSMA.

Narrowband microwave LANs don't employ spread spectrum, but they do operate at microwave frequencies. The use of a microwave radio frequency band for relatively narrow bandwidth signal transmission is referred to as "narrowband microwave."

Variety of WLAN

IEEE 802.11: The IEEE approved the original WLAN definition in June 1997. The 2.4 GHz frequency range and 2Mbps data rate are specified. Using various encoding methods, this standard developed into many distinct versions.

HYPER LAN: The European Telecommunication Standard Institute (ETSI) launched it in 1996.

Network group for broadband radio access at ETSI. The current Hyper LAN/1 version provides up to 24 Mbps of bandwidth and operates in the 5 GHz frequency spectrum.

BIG INDUSTRY LEADERS LIKE IBM, ERICSON, AND NOKIA PROMOTE BLUETOOTH. It was given the name Harold Bluetooth after the Danish monarch. At 2.2 GHz, it provides a data throughput of 1Mbps. PAN is another name for it (Personal area network).

MANET: This working group is responsible for researching and creating the mobile ad hoc network standard (MANET).

The most well-known WLAN family, for which there are several devices, is defined by the IEEE standard 802.11. The standard's number indicates that it is a member of the family of 802.X LAN standards. The main objective of the standard was to provide a straightforward and reliable WLAN that provides time-limited and asynchronous services.

Architecture

A basic service set (BSS), which consists of a few stations running the same MAC protocol, is the simplest component of a WLAN. A BSS may be connected to a backbone distribution system (DS) through an access point or it may be isolated (AP). The AP serves as a relay point and a bridge. Client stations in a BSS don't speak to one another directly. When two BSS stations desire to communicate with one another, the MAC frame is first transmitted from the source station to the AP and then from the AP to the destination station.

The BSS is referred to be an independent BSS if all of its stations are mobile and disconnected from other BSSs (IBSS). An ad hoc network is an IBSS. A distribution system connects two or more basic service sets to form an extended service set (ESS). The distribution system, which may be any kind of communication network, is a wired backbone LAN. The logical link control level sees the ESS as a single logical LAN. A station incorporates an access point into its design. The AP is the station's internal logic that grants access to the DS by functioning as both a station and a provider of DS services. There are two major fundamental system architectures that may be seen in wireless networks. The two kinds of WLAN are:

A. Infrastructure mode
B. Ad-hoc mode

Infrastructure Mode

BS or access point is used to link MSs. Similar to a star network, wireless nodes and access points communicate with one another rather than directly with other wireless devices. In this case, access points serve as a network bridge Figure 8.1.

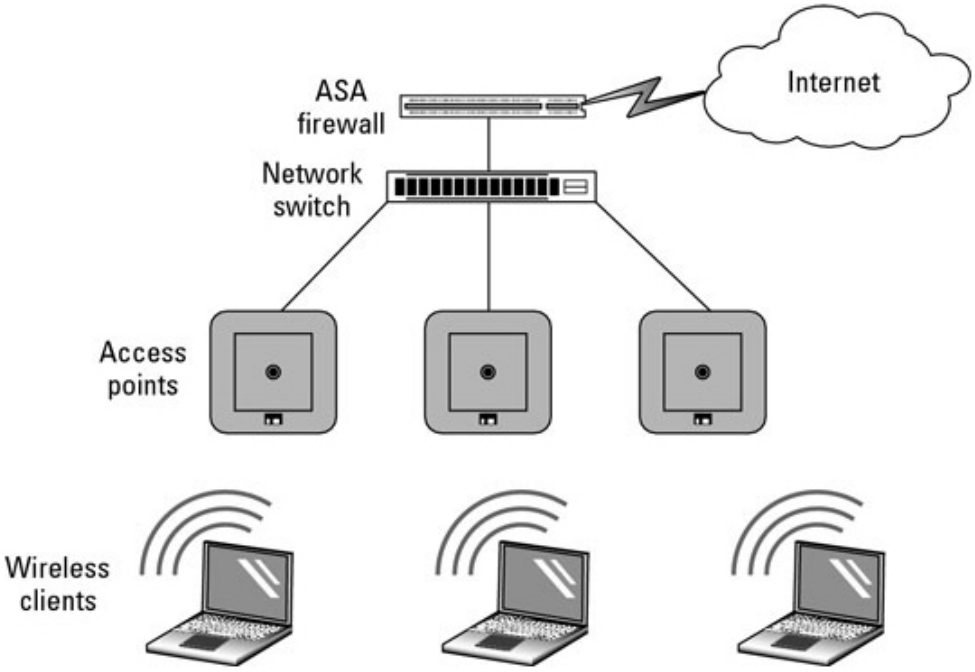


Figure 8.1: Infrastructure Mode

Ad-hoc Mode

There is no access point while in ad-hoc mode. Multiple MS may speak with one another directly. If nodes can physically interact with one another, or if they are within radio range of one another, they can communicate (Figure 8.2).

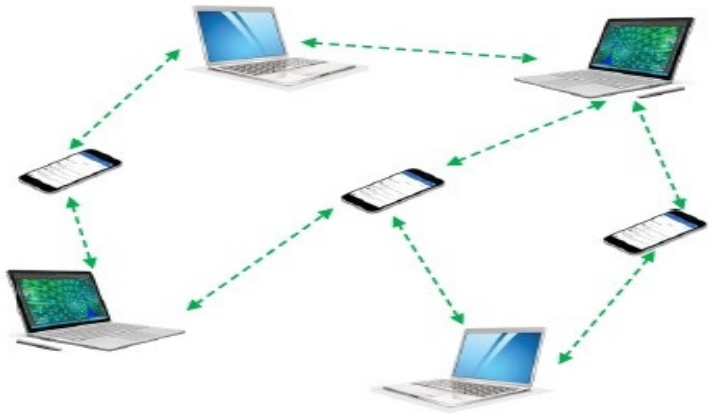


Figure 8.2: Ad-hoc WLAN mode

AD HOC Network in Mobile Computing

When devices link up and communicate with one another, an ad hoc network is created on the spot. Ad hoc is a Latin phrase that literally translates to "for this," meaning that it was created on the spot. Wireless local area networks typically make up ad hoc networks (LANs). Instead of depending on a base station or access points as in wireless LANs for the coordination of data transport, the devices communicate with one another directly. Each device takes part in the

routing process by choosing a path using the routing algorithm and transmitting data to other devices along that route.

Features of ad hoc networks:

Several of the features are listed below.

A. The mobile devices involved talk to each other directly and transfer data on their own without the use of hardware or established infrastructure.
B. Self-repairing.
C. Automatically set.
D. Additional names include Wireless Ad-hoc Network (WANET), spontaneous networks, and on-the-fly networks.
E. Each device serves as its own router, sending data packets to other nodes and devices.

Functioning of ad hoc network:

The way that MANETs function is as follows.

A. Work exclusively as an individual or in a group within a vast network, such as the internet.
B. Hardware and access point absent.
C. Direct and independent interaction between mobile devices
D. The devices begin to look for one another and communicate.
E. When a node is far away, the nodes in the path between it and the target node serve as routers, sending data one by one to the destination node.
F. Devices can dynamically add or remove nodes at any moment, allowing them to join or leave the network.
G. Similar to how they use batteries for power, the devices have their own energy reserves.

Use of mobile ad hoc networks:

Ad hoc network have many applications, including:

Gaming services: These are used for group gaming over local area networks in classrooms, contests, etc.

Military Services: For meetings, to instantly communicate information to all distant battalions or troops, etc.

Commercial Use: Regional Seminars, Events, and Conferences

The industry sector.

Emergency services for situations requiring services without infrastructure, such as earthquakes, disaster relief, firefighting, and natural disasters.

Education Sector: For the purpose of exchanging lectures, etc. in classrooms, labs of schools, colleges, etc.

Ad-Hoc Networks' characteristics:

The qualities are listed after, each with an explanation.

A. No Centralized Control: The operation is entirely dependent on how well the participating devices cooperate.
B. Devices changing at Random: Devices are voluntarily and rapidly entering and exiting the network.
C. Frequent changes in the network's topology, or how the devices are arranged, limited battery life, Limited human interference, and Less Security: These networks face greater security risks than wired networks.
D. Because the gadgets used in these networks are now so small, theft is a possibility.
E. They are easily attacked.
F. Each device serves two roles—one as a router and the other as a device in front of the network.
G. Limited Bandwidth: The capability and range of data transmissions on these networks are quite limited.
H. Limited resources, including battery life, memory capacity, backups, etc.
I. Scalability benefits and low costs go hand in hand with excellent performance.

Cons and Advantages of Ad hoc Network:

Pros:

Divorce from central network management. Each node's ability to act as both a router and a host demonstrates its autonomy. Nodes that can configure themselves and heal themselves do not need human assistance. Extremely expandable and suitable for adding extra network hubs.

Cons:

Due to numerous restrictions like noise, interference situations, etc., resources are limited. A lack of facilities for authorization. Less physical security leaves them more vulnerable to assaults. High latency, which refers to a significant delay in data flow between two nodes that are asleep.

IEEE 802.11 Services

The following services must be offered by the Wireless LAN in accordance with IEEE 802.11 specifications. Distribution: When a frame has to cross the Ds to move from a station in one BSS to a station in another BSS, this service is the one that stations mostly utilise to exchange MAC frames. Data may be sent between a station on an IEEE 802.11 LAN and a station on an integrated IEEE 802.x LAN using the integration service. A wired LAN that is physically linked to the DS and whose stations may logically connect to an IEEE 802.11 LAN through integration services is referred to be integrated. The address translation and media conversion logic necessary for data sharing are handled by the integration service.

Establishes the first connection between a station and an AP via association.

A wireless LAN station's identification and address must be known before it may send or receive frames. A station must connect with an AP inside a certain BSS in order to do this. To ease address frame delivery and routing, the AP may then transmit this information to other APs inside the ESS.

Reassociation: Allows a mobile station to switch between BSSs by transferring an existing association from one AP to another.

Disassociation: A signal from a station or an AP indicating the end of an existing association. This warning is sent by a station before it leaves an ESS or shuts down.

Authentication: Used to prove a station's identification to another station. Stations that want to connect with other stations utilise this authentication service to verify their identities.

Whenever an existing authentication has to be cancelled, this service is called.

Privacy: Used to ensure that only the intended recipient may view the contents of a communication. To ensure privacy, the standard allows for the optional use of encryption.

Architecture for Protocols

The physical layer and MAC layer are the two levels that make up the 802.11 protocol architecture. Physical layer convergence protocol (PLCP) and physical medium dependent sub layer are divisions of the physical layer (PMD). The three primary functions of the MAC layer are media access, user data fragmentation, and encryption. The carrier sense signal known as clear channel assessment is provided by the PLCP layer. For data transmission, it transmits the incoming frame from the wireless media to the MAC protocol Data unit (MPDU). Modulation, encoding, and decoding of signals are handled by the PMD layer. It allows for the wireless medium-based real transmission and receiving of physical layer object between MS.

Wireless ad hoc network (WANET)

In order to link two or more wireless devices to one another without the need of standard network infrastructure equipment, such as a wireless router or access point, a wireless ad hoc network, or WANET, is a sort of local area network (LAN). A PC, laptop, or smartphone with a Wi-Fi interface is often used to create an ad hoc network (Figure 8.3). Devices like wireless sensors are designed to operate largely in an ad hoc way in other circumstances.

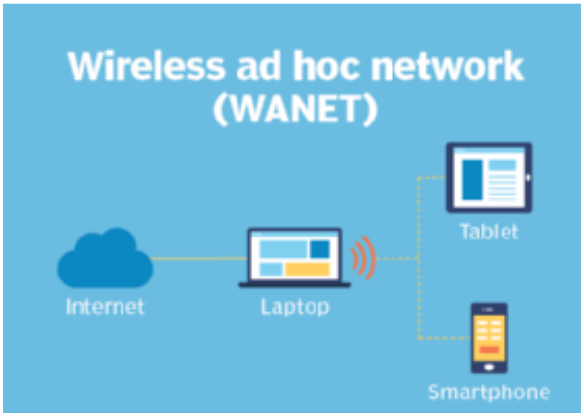


Figure 8.3: shows how to use an ad hoc network to connect devices to the internet.

Central servers are not required for operations like file sharing or printing since the devices in the ad hoc network may access each other's resources directly using simple peer-to-peer (P2P) or point-to-multipoint protocols. A group of devices, or nodes, such as a wirelessly enabled computer or smartphone, are in charge of network functions including routing, security, addressing, and key management in a WANET.

Working of an ad hoc network

Ad hoc devices need to be able to serve as a wireless router when connected, hence they need a wireless network adapter or chip. Each wireless device must be configured for ad hoc mode rather than infrastructure mode when configuring a wireless ad hoc network. The same wireless frequency channel number and service set identification (SSID) must be used by all wireless devices connecting to an ad hoc device.

An ad hoc device assumes this function and organises the flow of messages to each node in the network instead of depending on a wireless base station, such as a wireless access point (WAP) or Wi-Fi router. Packets are forwarded to and from one another by the many wireless endpoints linked to an ad hoc network. Ad hoc wireless networks, which are by their very nature improvised, are most helpful when wireless infrastructure is unavailable, such as when there are no access points or routers nearby and cabling cannot be extended to the site where more wireless communication is required. It's crucial to remember that not all ad hoc networks are created on a computer or mobile device. Wi-Fi access points may really be set up to operate in both ad hoc and infrastructure modes. Wi-Fi routers or a mix of WAPs and wireless controllers that offer the required network intelligence are often used to build and administer Wi-Fi networks that are set for infrastructure mode. A PC or smartphone may also set up an ad hoc network to provide momentary wireless network connectivity. Short-lived ad hoc networks often do not need or are not suited for the usage of more complex network protocols and network services available in infrastructure-based wireless networks, such as IEEE 802.1x authentication.

The usage will determine whether to utilise infrastructure mode or ad hoc mode. For instance, users should choose infrastructure mode with an on-site or cloud-based wireless LAN (WLAN) controller if they want a WAP to function as a permanent access point. However, if a user wants to build up a temporary wireless network for a few devices, ad hoc mode can be a decent choice. Utilizing a smartphone with cellular connectivity and Wi-Fi ad hoc mode enables laptops with Wi-Fi capabilities to connect to the network and access the internet via the smartphone's cellular internet connection. There is no need for a WAP or WLAN controller with this approach.

Ad hoc networks are useful for crises like natural catastrophes, armed conflicts, or while travelling since they need less setup and can be set up rapidly. These networks may be created fast since dynamic and adaptive routing technologies are available. These spontaneous, on-demand networks may be used to quickly and cheaply create a tiny, all-wireless LAN without the need for pricey wireless infrastructure hardware. If wireless access points or routers malfunction, they can function effectively as a temporary internet connection.

Ad hoc wireless network types

Depending on the application and purpose, several WANET types exist. The capabilities of the wireless equipment, the physical setting, and the goal of the communication all play a role in selecting the sort of wireless ad hoc network.

MANET

Mobile devices talk to one another directly in a mobile ad hoc network. A MANET is a wireless mobile device network that is self-organizing and self-configuring and without infrastructure. A "spontaneous network" or "on-the-fly" network are other names for a MANET. Smart home lighting, ad hoc streetlight networks, ad hoc robot networks, disaster relief ad hoc networks,

and hospital ad hoc networks are a few examples of MANETs. These networks often communicate via proprietary or non-TCP/IP networking technologies.

Internet protocols including TCP/IP (Transmission Control Protocol/Internet Protocol) and UDP (User Datagram Protocol) are supported by IMANET Internet-based mobile ad hoc networks (UDP). On each connected device, the iMANET uses a TCP/IP network-layer routing protocol to connect mobile nodes and create dispersed routes automatically. In addition, IMANETs may be utilised to gather sensor data for data mining in a range of applications, including air quality monitoring.

SPAN

Smartphone ad hoc networks are P2P networks that are created without the use of cellular carrier networks, wireless access points, or other conventional network infrastructure hardware by utilising existing hardware like Wi-Fi and Bluetooth and software protocols built into a smartphone operating system (OS). SPANs enable multi-hop relays in contrast to conventional hub-and-spoke networks, such as Wi-Fi Direct. Sending communication from device A to device C while utilizing device B as an intermediate is known as multi-hop relay. Therefore, for traffic to reach its destination, device A and device C do not need the establishment of a direct P2P connection. There is no group leader in this kind of application since SPANs are totally dynamic; as a result, peers may join or depart without disrupting the network.

Automobile ad hoc network

Devices in cars that are used to communicate with equipment on the side of the road make up this sort of network. The in-car safety and security system OnStar is one example.

WMN

Mesh clients, mesh routers, and mesh gateways are typically included in wireless mesh networks, which are composed of radio networks built up in a mesh topology. The devices, or nodes, in mesh networking are interconnected such that at least some, if not all, have several pathways to other nodes. As a result, there are more paths for data to travel between user pairs, making the network more resilient in the event of a node or connection failure. When an infrastructure-based wireless network cannot be built using network cabling, such as when a temporary wireless network is needed, WMNs may be beneficial.

Benefits of ad hoc networks

When just a few devices need to be connected, ad hoc mode, which doesn't need a centralised access point, may be simpler to set up than infrastructure mode. For instance, two laptops may be linked directly in ad hoc mode to establish a temporary Wi-Fi network without a router if the user is in a hotel room without Wi-Fi. Ad hoc mode is further expanded upon by the Wi-Fi Direct standard, a specification that enables Wi-Fi Direct-certified devices to communicate with one another without the need for a wireless router or an internet connection. It makes it possible for devices to directly interact via Wi-Fi signals.

Wireless ad hoc networks also provide the following advantages:

Ad hoc networks offer a low-cost method of direct client-to-client or client-to-internet communication because they don't require infrastructure hardware like access points or wireless routers. Ad hoc networks are simple to set up and provide an efficient method of communicating with nearby devices when time is of the essence and running cabling is not feasible.

Problems with ad hoc networks

Due to its restrictions, certain Wi-Fi-enabled equipment, such as some Android smartphones, wireless printers, and bespoke IoT sensors, do not support ad hoc mode by default and instead connect to networks in infrastructure mode. This is a significant disadvantage of wireless ad hoc networking. Ad hoc communications may sometimes be enabled on endpoint devices by installing third-party software.

Ad hoc mode should not be used to build up a bigger, more permanent network that can serve many more endpoints. Instead, use infrastructure mode. The wireless radios and antennas of wireless routers that act as access points are generally stronger and cover a larger area. Because endpoint antennas were not intended to be as strong as specifically designed WAPs, ad hoc networks often have challenges with low wireless communication range.

Additionally, ad hoc networks may not scale effectively. Ad hoc networks are more difficult to maintain as they grow in size because there is sometimes no one hub through which all traffic passes. For instance, increased wireless interference may happen when several devices are linked through a P2P MANET ad hoc network since each device must establish a direct P2P connection with each of the other devices rather of travelling via a single access point as in a hub-and-spoke design. The data will be sent via other devices en route if a device is too far away from the one it wants to connect to; this is slower than sending it through a single access point that serves as a centralised wireless bridge.

Ad hoc wireless networks also have the following drawbacks:

• Unlike devices in infrastructure mode, devices on an ad hoc network are unable to prevent SSID broadcasting. As a consequence, if an attacker is nearby and within signal range, they may discover and connect to an ad hoc device.

• Lack of network infrastructure services, such as access to a RADIUS (remote authentication dial-in user service) server for 802.1x authentication requirements, limits security alternatives.

• Without the installation of a dedicated network gateway, certain wireless ad hoc networks are unable to connect to the internet or bridge conventional LANs.

A cellular-connected smartphone using "hotspot" mode, which is a form of an ad hoc network, is one example of a device that can only access the internet if one of them is connected to it and sharing it with the others. The client using this function may have performance issues when internet sharing is enabled, particularly if there are numerous connected devices. Ad hoc mode calls for additional endpoint system resources because moving devices alters the actual network configuration, although an access point in infrastructure mode often seems fixed to endpoints.

Network security on-demand

Ad hoc networks sometimes suffer from the fact that they were designed to be ephemeral, as was previously said, and as a result lack many of the cutting-edge security measures frequently found in permanent, infrastructure WLANs. As a result, just the most basic security features may be enabled for many different kinds of ad hoc networks. Using a smartphone in ad hoc mode is a fantastic illustration of this. In this situation, a smartphone with ad hoc capability may be set up to broadcast a Wi-Fi SSID so that more users can connect. This SSID, however, cannot be kept secret from others. The smartphone device also is unable to function with more secure authentication protocols like WPA-Enterprise, which requires 802.1x authentication to a RADIUS server. The sole option is WPA-Personal, which needs a static private key to be used and exchanged in order to prevent unwanted access.

However, compared to a wireless infrastructure that is mobile and always operational, the risk of an attacker accessing this kind of ad hoc network is much lower since it is utilized momentarily, has a limited coverage area, and often moves.

Network Protocol

The Internet Mechanism (IP), often known as TCP/IP or the Internet Protocol Suite, is a protocol for exchanging data over packet-switched networks. IP, the main protocol in the Internet Layer of the Internet Protocol Suite, is responsible for sending distinct protocol datagrams (packets) based only on addresses from the source host to the destination host.

Data is exchanged from one computer to another over the Internet using the Internet Protocol (IP). A host, or computer, on the Internet is identified by at least one IP address that distinguishes it from every other host.

The message is broken up into smaller units called packets when you transmit or receive data (such as an email note or a Web page). These packets each include the Internet addresses of the sender and the recipient. Any packet is initially forwarded to a gateway computer, which only comprehends a portion of the Internet. Once the destination address has been read by the gateway computer, it transmits the packet to an adjacent gateway, which scans the destination address again, and so on throughout the Internet, until one gateway identifies the packet as coming from a computer in its immediate neighbourhood or domain. The packet is subsequently sent by that gateway straight to the designated machine.

A message is broken up into a number of packets, and each packet might travel the Internet through a separate path. The order in which packages arrive may vary from the order in which they were dispatched. They are simply sent using the Internet Protocol. The Transmission Control Protocol (TCP), a different protocol, is in charge of restoring the original order.

Since IP is a connectionless protocol, there is no ongoing connection established between the communicating end points. Every packet that moves via the Internet is viewed as a separate data unit with no connection to any other data unit. (TCP, the connection-oriented protocol that maintains track of the packet sequence in a message, is the reason the packets do get placed in the correct order.) The Network Layer is layer 3 in the Open Systems Interconnection (OSI) communication paradigm, where IP is located.

IPv4

The TCP/IP Protocols employ the Internet Protocol version 4 (IPv4) as its delivery method. A best-effort delivery service, IPv4 is an unstable and connectionless datagram protocol. Best-effort implies that IPv4 does not provide error correction or flow control (except for error detection on the header). IPv4 does its best to send a communication to its destination but makes no assurances due to the assumption that the underlying layers are unreliable.

IPv6

In 1995, the Internet Engineering Task Force (IETF) released a definition for IPng, the next generation of IP. The IPv6 standard was created in 1996 based on this specification. Compared to IPv4, IPv6 offers a number of functional improvements. It was built to enable high-speed networks, a variety of data streams, including graphics and video, among other things. Ipv6 specifies the source and destination using a 128-bit address.

Cellular IP

It is a standard communications protocol developed by the Internet Engineering Task Force (IETF) that enables users of mobile devices to switch between networks while keeping their permanent IP address. A protocol that enables users of mobile devices with IP addresses linked to one network to continue using their connections while switching to a network with a different IP address.

CHAPTER 9

WIRELESS APPLICATION PROTOCOL (WAP)

Sowmya M S, Assistant Professor
Department of Computer Science Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Karnataka – 562112
Email Id- ms.sowmya@jainuniversity.ac.in

In order to standardise how wireless devices, including cell phones and radio transceivers, may be used for internet access, including email, the web, newsgroups, and instant messaging, the Wireless Application Protocol (WAP) specification is a collection of communication protocols. At a gathering called the WAP Forum, Unwired Planet (now Enea Openwave Mobility), Motorola, Ericsson, and Nokia came up with the idea for WAP. Prior to the invention of WAP, it was possible to access the internet wirelessly, but different manufacturers used various methods, while WAP was designed to be an industry standard. WAP, however, is currently seen as being out of date since newer devices employ networks and browsers that are comparable to those on PCs.

Working of WAP

WAP refers to a set of protocols that are intended to facilitate communication between WAP-enabled web browsers and network technologies as well as WAP-compatible hardware, such as mobile phones. Before WAP, a user's device and mobile operator would determine how much access they may have to mobile data. The WAP protocol was developed to standardise mobile data access, but it also served as a method for getting over carrier and device restrictions that often gave mobile customers a terrible experience. It did it in a number of ways:

WAP was designed specifically for the high-latency, low-bandwidth mobile networks of the time, which were infamous for cutting the connection before a page could completely display. WAP supported established and widely used internet protocols including Internet Protocol, User Datagram Packets, and XML. Similar to how any browser can display HTML code regardless of the kind of hardware it is operating on, the Wireless Markup Language (WML) standard enabled websites to be produced without taking into account the user's mobile hardware.

Using WAP

The following advantages for wireless network operators, content producers, and end users were put out by WAP when it was first introduced in 1999:

Operators of wireless networks and mobile phones. WAP was created with the intention of enhancing already-existing wireless data services, such as voicemail, and facilitating the creation of new mobile apps. Without making any further infrastructure adjustments or phone modifications, these apps might be created.

Producers of content. For third-party application developers, WAP opened up a market for extra apps and mobile phone features. It was suggested that developers use the WML programming language to build apps for mobile devices.

Consumers. Access to internet services like banking, entertainment, messaging, and other information on mobile devices should be simple and safe for users of mobile phones. Access

to intranet data, including corporate databases and business applications, may also be possible through WAP.

Despite these alleged advantages, WAP was not widely adopted in many nations, and its usage sharply decreased about 2010 as a result of widespread HTML compatibility in mobile phones.

Mobile computing uses Wireless Application Protocol (WAP)

With a hierarchical structure highly reminiscent of the TCP/IP protocol stack, Wireless Application Protocol (WAP) is a programming paradigm, application environment, and collection of communication protocols based on the idea of the World Wide Web (WWW). See the Wireless Application Protocol's or WAP's most salient characteristics in mobile computing:

A De-Facto standard or protocol called WAP was created for micro-browsers, allowing mobile devices to communicate, share, and send data over the Internet. WAP is built on the idea of the World Wide Web (WWW), and its backend functionality is likewise comparable to that of the WWW. However, WAP accesses its services using the markup language Wireless Markup Language (WML), while the WWW uses HTML. As an XML 1.0 application, WML is referred to.

The WAP Forum was established in 1998 by many significant IT firms, including Ericson, Motorola, Nokia, and Unwired Planet, with the goal of standardising the diverse wireless technologies via protocols.

Following the creation of the WAP model, it became widely recognised as a wireless protocol capable of operating on a variety of wireless devices, including mobile, printers, pagers, etc. The WAP Forum joined with a number of other industry forums in 2002 to become the Open Mobile Alliance thanks to the work of the WAP Forum's different members (OMA).

WAP's capacity to enable the development of web applications for portable devices led to its selection as a De-Facto standard (Figure 9.1).

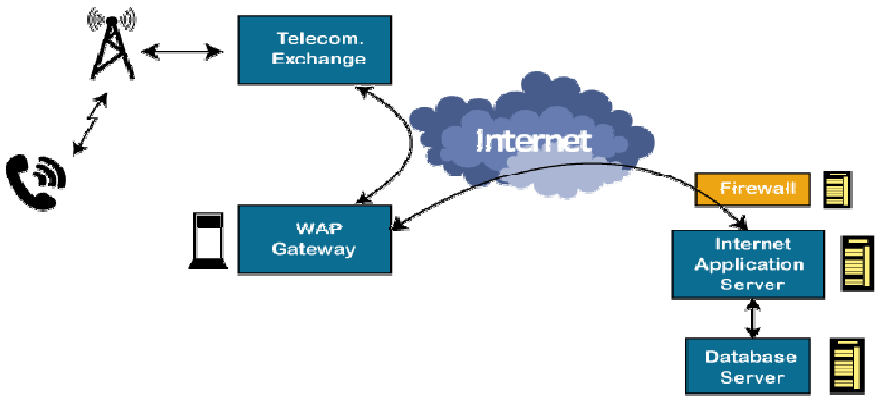


Figure 9.1: Mobile computing uses Wireless Application Protocol

Model WAP

The WAP paradigm functions similarly to the conventional client-server approach, but it also requires a separate element known as a WAP gateway. The function of this gateway is to serve as a bridge between mobile devices and the internet. The hardware capabilities present in modern handsets were absent in earlier mobile devices. As a result, smartphones often come

with tiny mobile browsers, commonly referred to as minibrowsers or microbrowsers. The request was made to a WAP gateway whenever a user typed a URL into the browser on their device. On behalf of the device, this gateway would access the website, obtain the required page, and then convert the page to WML format. The device that renders the website would then get the WML code.

Working of the WAP model, or wireless application protocol as shown in Figure 9.2.

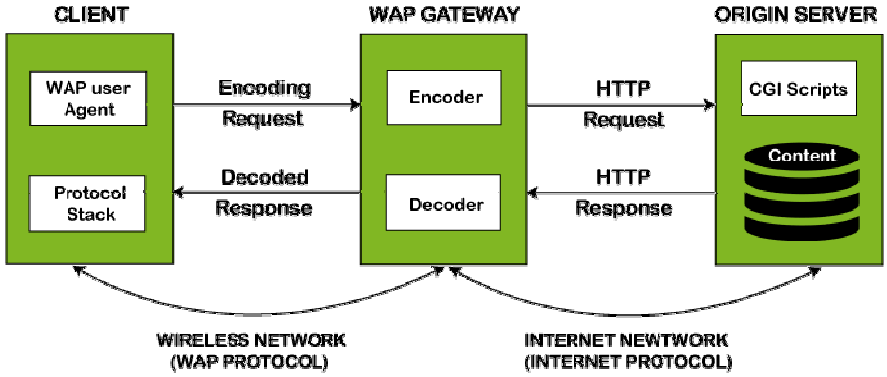


Figure 9.2: Working of the WAP model, or wireless application protocol

The Wireless Application Protocol, or WAP Model, functions as follows:

There are three tiers in the WAP model: Client, Gateway, and Origin Server. The mobile device transmits the URL-encoded request through a network to a WAP gateway using the WAP protocol when a user opens the browser on his or her mobile device and chooses a website that he or she wishes to see. He or she makes an encoding request to the WAP gateway using a mobile device. The submitted encoding request is translated by the WAP gateway before being delivered over the Internet as a regular HTTP URL request.

A specific Web server receives the request, processes it like it would any other request, and then uses a WAP gateway to transmit the result back to the mobile device. The mobile users' browsers may now display the final answer from the WML file. Mobile computing uses Wireless Application Protocol (WAP)

Stack WAP Protocol

It describes the many data transfer and communication layers utilised in the WAP model:

The Wireless Application Environment (WAE), mobile device standards, and content creation programming languages, such as WML, are all part of the application layer.

Session Layer: The Wireless Session Protocol is part of the session layer (WSP). It is in charge of quick disconnecting and reconnecting.

Transaction Layer: Running on top of UDP, the transaction layer consists of the Wireless Transaction Protocol (WTP) (User Datagram Protocol). This layer provides transaction functionality and is a component of TCP/IP.

Data integrity, privacy, and authentication are handled by the Wireless Transaction Layer Security (WTLS) component of the security layer during data transfer.

Transport Layer: Wireless Datagram Protocol makes up this layer (WDP). It gives upper tiers of the WAP protocol stack a uniform data format.

Protocol stack for WAP

For WAP devices, equipment, software, and other technologies to work together, the WAP standard specifies the following protocol stack, which comprises the following:

Wireless Transport Layer Security for managing privacy, authentication, and data integrity through public key cryptography; Wireless Application Environment for managing mobile device specifications and programming languages like WML; Wireless Session Protocol for managing connection suspensions and reconnections; Wireless Transaction Protocol for managing transaction support for requests and responses to servers; and Wireless Datagram Protocol, an adaptation of the Internet Datagram Protocol.

Benefits of WAP

The key advantage of WAP was that it enabled mobile devices to have extensive internet access. Prior to the adoption of WAP, mobile operators often provided proprietary and very limited mobile access. One or more specialised services from a carrier may include stock quotations, movie listings, weather updates, news headlines, and sports. However, there wasn't usually a way to access the whole web. Along with enabling widespread internet access, WAP increased access speeds via data compression and assisted in lowering the frequency of timeouts and connection errors that had previously hampered mobile access.

The drawbacks of WAP

WAP's main drawback was that it was never embraced by everyone. Mobile service providers in certain places prevented uptake by imposing substantial surcharges for data access. Another drawback of WAP was that early mobile browsers lacked features seen in more advanced browsers today. As a consequence, WAP sometimes had issues with how mobile devices displayed websites. More complicated and larger pages often couldn't be displayed at all.

Application protocol applications:

A. WAP makes it easier for you to use your mobile devices to access the Internet.
B. On wireless devices, you may play games on mobile devices.
C. It makes it easier for you to access emails on mobile networks.
D. Mobile devices may be used to fill out expense claims and access timesheets.
E. Nowadays, online mobile banking is increasingly common.
F. It may also be utilised in a variety of Internet-based services, including traffic updates, weather forecasts, flight information, movie and theatre information, and geographic location. All of them are made feasible by WAP technology.
G. WLL (Wireless in Local Loop)

In order to replace its wireline counterpart, wireless local loop (WLL) offers two-way calling services to the stationary or "fixed" customers. It is a system that uses wireless technology to link a customer to the PSTN and radio signals to provide regular phone service. It is a broadcast connection technology that delivers speech and data through high frequency radio channels instead of fibre optic cables.

The circuit linking a subscriber's station (such as a phone set) and the line termination equipment in a central office (a switch in the telephone network) is known as a loop in telephony. After a certain distance from the central office, the trunks in the loop are divided into multiple more compact bundles of circuits.

Eventually, these circuits are divided into distinct drops for the residential homes.

In the public switched telephone network, the central office switch is often the initial site of traffic concentration (PSTN). In more recent installations, traffic is concentrated using statistical multiplexers, and residential or commercial campuses are connected to the central office via fibre optics.

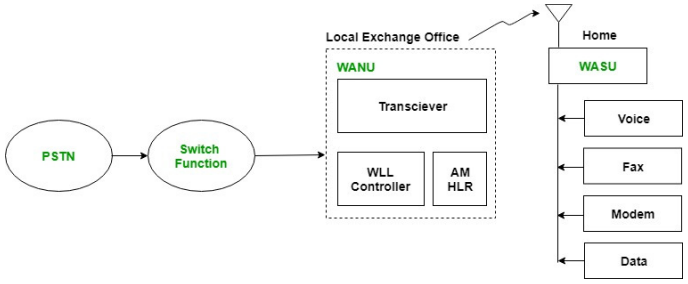


Figure 9.3: WLL **Architecture**

Figure 9.3 displays a condensed version of the WLL architectural reference model. The base station transceivers (BTS) or radio ports (RP), the radio controller (RPCU), an access manager (AM), and the home location registry (HLR), if necessary, make up the wireless access network unit (WANU) in this diagram. AWLL refers to the interface between the WANU and the switch. UWLL refers to the air interface between the WANU and the user side. The WANU should include features including radio resource management, restricted mobility management, air interface authentication and privacy, and over-the-air subscriber unit registration (SUs). Additionally, it can be needed to provide switching, routing, billing, and operation and maintenance (OAMP) services as appropriate or necessary. The WANU also offers protocol conversion and data and audio transcoding services. The wireless access subscriber unit (WASU) offers a TWLL interface to the subscriber and an air interface to the network. This interface has features for authentication, local power, OAMP, dual tone multi frequency (DTMF), and transcoding protocols. The switching fabric (SF) in this reference model may be an ISDN switch, a digital switch with or without Advanced Intelligent Network (AIN) functionality, or a mobile switching centre (MSC).

It is made up of three main parts.

WANU (Wireless Access Network Unit) (Wireless Access Network Unit)

Several BST or radio components, an RPCU (Radio port control unit), an AM (access Manager, responsible for RPCU operation), and an HLR make up this system. It is in charge of subscriber identification and air registration. OAM, routing, billing, switching, and protocol conversion may also be necessary. Data and voice transcoding.

WASU (Wireless access subscriber unit) (Wireless access subscriber unit)

It offers an "Uwll" air interface for n/w and a "PWLL" interface for subscribers. It is in charge of the signalling and voice trans-coding functions.

SF (Switching Fabric) (Switching Fabric)

It is connected to a switch, which might be a digital switch, an MSC switch, or an ISDN switch. Leased lines, microwaves, or optical Fibre may all be used for the transmission between WANU and SF.

WANU is linked to the switch through the "AWLL" interface.

Technologies for Wireless Local Loops

Usually, one of the four main technologies serves as the foundation for the WLL systems. Systems Based on Satellites: For remote places like islands and rural settlements, these systems provide telecommunication services. These systems come in two varieties:

Equipment created especially for WLL applications

Technology served as an add-on service for mobile satellite systems.

The former, albeit it could be pricey, delivers a quality and grade of service equivalent to wired access. The latter claims to be less expensive, but because of bandwidth limitations, it could not provide a level of service and quality that is similar to regular telephone service (POTS). The HNS telephony earth station (TES) technology is an example of a satellite-based technology created particularly for WLL. Almost any geostationary earth orbit (GEO) C-band or Ku-band satellite may be used with this technology. For many years, satellite technology has been utilised to provide telephone service to far-flung regions of the globe. When landlines are not economically viable or when an emergency backup is needed, these systems provide an alternative to terrestrial telephone infrastructure. The Inmarsat International Circular Orbit (ICO) system, Iridium, Globalstar, Odyssey, American Mobile Satellite Corporation (AMSC), Asia Cellular Satellite (ACeS), and Thuraya mobile satellite system are only a few of the numerous proposed mobile satellite service systems. These systems are designed specifically to accommodate low-cost mobile terminals, especially for voice and data applications that need low bit rates.

Cellular-based Systems: These systems provide WLL services with high power, high range, average subscriber density, and average circuit quality. The main purpose of cellular WLL technology is to increase the availability of basic phone services. They typically operate at 800-900 MHz, 1.8-1.9 GHz, and sometimes at 450 MHz or 1.5 GHz in the mobile frequency ranges.

This strategy provides fixed wireless connectivity from the same cellular platform together with mobility. High-tier coverage is ideal for cellular networks. It is necessary to support mobiles moving at speeds more than 100 mph and cells with a radius of up to 10 miles. Extensive signal processing is needed to accomplish the aforementioned objectives, which translates to significant latency, high overhead, and poor user bandwidth. The deployment of these devices inside and in picocells is not recommended. It is necessary to increase the air interface's complexity while maintaining the same low user bandwidth.

Low-Tier PCS or Microcellular-Based Systems: These systems provide WLL services with good circuit quality, low power, small range, and high subscriber density. These innovations are thought to speed up market entrance and increase the capacity of the current infrastructure. They are generally used in the frequency ranges of 800 MHz, 1.5 GHz, 1.8 GHz, and 1.9 GHz.

More base stations are needed to cover the same service area as with cellular-based WLL. Where backhaul from several base stations to the switch is supported by an existing infrastructure or when wireline-like services and quality are necessary, operators may take low-tier WLL technology into consideration. Systems for Fixed Wireless Access: These are exclusive radio systems intended for fixed wireless applications alone; they may or might not be expandable to PCS or cordless. The main drawbacks of the cellular method are signalling transparency and the limited availability of toll-quality speech (although new toll-quality vocoders created for cellular technology may solve this issue). Low-end PCS and microcellular strategies' range is their main drawback. Fixed wireless access (FWA) technology that isn't conventional can solve these problems and improve productivity. The local telephone area is

covered by FWA systems for zonal regions straight from the PSTN switches. The networks for rural regions connect end users at the furthest extremities of rural linkages.

MOBILE AGENTS IN MOBILE COMPUTING

In mobile computing, mobile agents are sets of information and programmes that can move independently from one computer to another while continuing to function on the new gadget. In other words, a mobile agent is a piece of independent software that can roam across hosts in a network while interacting with resources and other agents. There is minimal chance of data loss during this process since the operating program's state is kept and then transferred to the new host. It makes it possible for the programme to resume where it left off prior to transfer and go on operating. The fundamental advantage of mobile agents is their capacity to shift complex processing processes to the location where enormous volumes of data need to be handled. Mobile agents are sometimes known as transportable agents [1]. They are divided into two categories:

Mobile Agents with a Static Migration Path: These agents have a pre-defined migration path.

Mobile Agents with an arbitrary migration path, such as Roamer: They have arbitrary migration paths. The mobile agents decide their course based on the state of the network.

Characteristics of Mobile Agents:

The mobility of the mobile agents is their most important trait. The mobile agents have intellect, social skills, and the capacity to learn on their own. They can communicate without a similar node since they are autonomous, independent, and stand alone. The user may continue to operate efficiently even after disconnecting from the network.

Intelligence: Mobile Agents may research topics related to their area and learn new things. They are referred to as intelligent agents since they have some domain knowledge. Additionally, they may move their state from one place to another without altering previously stored data and take the appropriate actions in the new context.

Autonomous: The mobile agents run their own business. This shows that the agents' performance and behaviour are influenced by both internal events and exterior actions done by users or the system. The mobile agents may make an independent decision when selecting a node.

Mobility: Mobile Agents are in some ways somewhat mobile. The agent has other nodes besides its home node. They may move between nodes while doing tasks in between. This function distributes the processing and balances the load. The fact that the agents will continue to operate even if the user logs out is another benefit of this capability.

Communicative: Mobile agents can interact with users, other agents, and systems successfully. For inter-agent communication, the mobile agents employ a communication language.

Life Cycle of Mobile Agents

The following conditions are guaranteed by the life cycle of mobile agents:

A. They can adapt to different environments.
B. They may switch from one node's location to another, for example, the environment at home or abroad.
C. They solely work for the objective and operate autonomously.

Mobile agents' Benefits:

Mobile agents have a number of benefits over traditional agents, including the following:

A. Mobile Agents have a self-sufficient and independent attitude.
B. They are simple to maintain or maintenance-friendly.
C. They can tolerate faults. It indicates that they can function without a live connection between the client and server.
D. They shorten the time for compilation.
E. They offer a network with minimal lag.
F. They put less strain on the network.
G. They make parallel processing possible. It indicates that they can run asynchronously on several heterogeneous network hosts.
H. They offer dynamic adaptability, in which the host environment's status affects how they behave.

Disadvantages of Mobile agents:

The biggest drawback of mobile agents is their security. They lack security.

Mobile Agents Applications:

Using mobile agents is employed in the subsequent applications:

Mobile Agents are used in a variety of fields, including robotics, data-intensive applications, traffic control, network management, and e-commerce. They are also utilized in computing processes such as grid computing, parallel computing, distributed computing, mobile computing, etc.

CHAPTER 10

MOBILE AD HOC NETWORKS (MANETS)

Rajapraveen.k.N, Assistant Professor
Department of Computer Science Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Karnataka – 562112
Email Id- p.raja@jainuniversity.ac.in

Wireless networks known as mobile ad hoc networks (MANETs) have dynamic topologies and no permanent infrastructure. Each computer that makes up a MANET is referred to as a node. Nodes in MANETs may be needed to perform dual roles as hosts and routers as well as to relay packets between nodes that are unable to connect with one another directly. The frequency spectrum requirements for each MANET node are much lower than those for a node in a fixed infrastructure network. A MANET is an independent group of mobile users who interact over wireless lines with limited capacity. The network architecture may change quickly and unexpectedly over time since the nodes are mobile. The network is decentralised, therefore all network activities, such as determining the topology and sending messages, must be carried out by the nodes themselves. Mobile nodes will therefore have routing capability (Figure 10.1).

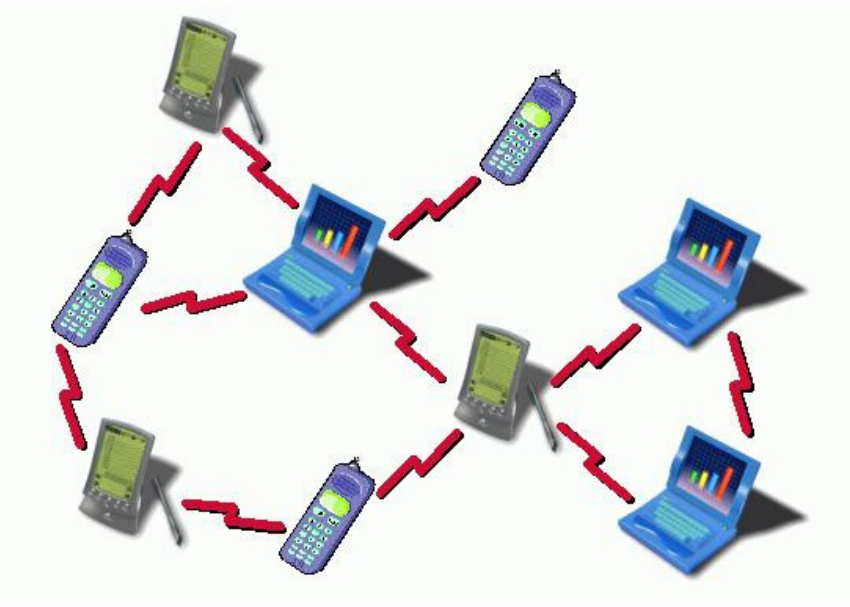


Figure 10.1: Mobile Ad hoc Networks (MANETs)

A mobile ad hoc network is made up of a number of wireless nodes that may be dynamically set up anywhere at any time without the aid of a fixed network infrastructure that has already been built. Bandwidth restrictions and changeable connection capacity, as well as energy-constrained nodes, are some characteristics of MANET. Limited security, distributed operation, autonomous terminal, lightweight terminal, multi-hop communications

Ad Hoc networks are required.

Fixed access points and backbone infrastructure cannot always be set up since they may not be available in disaster or conflict zones, and they are not always practicable for short-range radios like Bluetooth (10 m).

Ad hoc networks are advantageous when infrastructure is unavailable, damaged, or impracticable since They: - Do not need backbone infrastructure support - Are simple to instal

Feature of MANETs

MANET makes network construction quick. The sole prerequisite for setting up a new network is to provide a fresh set of nodes with restricted wireless communication range. A node has restricted capabilities since it can only link to other nodes in its immediate vicinity. As a result, it uses less electricity.

A MANET node is capable of finding a nearby node and service. A node connects with a distant node in the MANET by learning about the service of a local node using a service discovery protocol. MANET nodes are connected to one another peer-to-peer. MANET nodes possess autonomous switching, communication, and compute capabilities. Only closest node connection is included in the wireless connectivity range of MANETs. When an intermediate node fails, communication with the distant server is more delayed. The MANET is constrained by the little bandwidth available between two intermediary nodes. Because the node's power supply could be constrained, calculations must be energy-efficient.

In MANET, there is no demand for access points. There are just a few access points available for connecting to other networks or MANETs. IPods, Palm portable computers, Smartphones, PCs, smart labels, smart sensors, and car-embedded systems are just a few examples of MANET nodes. MANET nodes are capable of using a variety of protocols, including IrDA, Bluetooth, ZigBee, 802.11, GSM, and TCP/IP. Data caching, saving, and aggregation are performed by MANET nodes. In order to ensure smooth communication between the devices, MANET mobile device nodes move in unison with surrounding wireless nodes, sensor nodes, and embedded car components.

MANET difficulties

There are a number of difficulties that must be considered in the construction of a good wireless ad hoc network:

Dynamic Topology: Nodes are allowed to move about at will, which causes the topology to fluctuate at will. This trait necessitates network setup that is dynamic.

Limited security: Attacks on wireless networks are possible. Because any node should be able to join or exit the network at any moment, mobile ad hoc networks are more susceptible. Flexibility and more transparency are needed for this.

Wireless networks often have a limited amount of bandwidth. It is especially true in an ad hoc network since there isn't a backbone to manage or multiplex larger bandwidth.

Routing is difficult in a mobile ad hoc network. This is dependent on a number of variables, such as determining the routing path, choosing the right routers, topology, protocol, etc.

Use cases for MANETS

MANETs have a wide range of applications, from tiny, static networks limited by power sources to large-scale, mobile, highly dynamic networks. It is a difficult problem to build network protocols for these networks. Whatever the need, MANETs require effective distributed algorithms to decide network setup, link scheduling, and routing. The following are some of the key applications for MANETs:

Military battlefield– Soldiers, tanks, and aircraft in a military battlefield. Ad-hoc networking would enable the military to maintain an information network between the troops, vehicles, and military information headquarters while using standard network technology.

Sensor networks may be used to monitor the environment across a vast region.

Local level - Ad hoc networks may independently connect a momentary multimedia network utilising notebook or palmtop computers to disperse and distribute information among attendees at a conference or school, for example. Home networks, where gadgets may directly interact to share information, may be another suitable local level use.

Personal Area Network (PAN) - ubiquitous computing, which enables flexible communication between household equipment and personal electronic devices. Short-range MANET may make it easier for different mobile devices (such a PDA, laptop, and cell phone) to communicate with one another. Wireless communications have taken the role of cumbersome corded cords. With the use of technologies like Wireless LAN (WLAN), GPRS, and UMTS, such an ad hoc network may also increase access to the Internet or other networks.

Intelligent transportation using vehicular ad hoc networks, which enables real-time vehicle tracking and flexible traffic management. Civilian settings, including the taxicab network, conference rooms, sporting arenas, boats, and small aircraft. Search and rescue, law enforcement, and firefighting activities, as well as providing communication between far-off devices when the network infrastructure is not accessible.

Ad hoc may be utilised in rescue and emergency situations, such as a fire, flood, or earthquake. Where there is no or broken communication infrastructure, emergency rescue activities must be conducted and a communication network must be quickly deployed. Over a tiny hand carried device, information is sent from one member of the rescue team to another.

As mobile ad hoc networks lack a permanent infrastructure and routing calls for dispersed and coordinated activities from all nodes in the network, routing in these networks is a crucial challenge. Similar to Internet routing, MANETs provide point-to-point routing. The route discovery process is the primary distinction between routing in MANET and conventional internet.

The comparatively lengthy converge times of internet routing algorithms like RIP or OSPF are suitable for a wired network with rare topology changes. However, since node mobility causes a MANET's topology to alter quickly, conventional internet routing techniques are ineffective. Although MANET-specific routing protocols have been suggested, they have a high control cost and cannot scale to very large networks. The network address is another significant variation in the routing. The network address (IP address) used for internet routing is hierarchical and contains both a network ID and a machine ID from that network. In contrast, the network address for the majority of MANETs is only the node's ID and is not hierarchical. To determine the next hop, the routing protocol has to utilise the complete address.

Routing Protocols for MANET

Because the topology of an ad hoc network is dynamic, nodes in a mobile ad hoc network (MANET) are not aware of the topology of their network and must figure it out on their own. The fundamental guidelines state that anytime a new node joins an ad hoc network, it must make an announcement of its existence and must also pay attention to comparable announcement broadcasts from existing mobile nodes.

1. Table-driven routing protocols, commonly referred to as proactive routing methods. Every mobile node has a separate routing database that lists the paths to every potential destination mobile node.

These routing tables are updated frequently as and when the network topology changes since the mobile ad hoc network's topology is dynamic. Its weakness is that it struggles with huge networks since maintaining the route information to every potential node causes the routing table entries to become too large.

Destination Sequenced Distance Vector Routing Protocol (DSDV):

It is a proactive/table-driven routing protocol called the Destination Sequenced Distance Vector Routing Protocol (DSDV). As its name implies, it really expands the wired networks' distance vector routing technique. The Bellman-Ford routing method serves as its foundation. Due to the count-to-infinity issue, the distance vector routing protocol was not suitable for mobile ad hoc networks. Thus, the Destination Sequenced Distance Vector Routing Protocol (DSDV) was developed as a remedy.

Every routing item in the routing database kept by each node includes the addition of the destination sequence number. Only if the entry includes the newly updated route to the destination with a higher sequence number will a node include the new update in the database.

GSR: Global State Routing It is a table-driven, proactive routing mechanism. In actuality, it expands the wired networks' link state routing. The Dijkstra routing algorithm forms its foundation. Link state routing protocol was not designed for mobile ad hoc networks since each node directly floods the network with link state routing information, or global flooding, which may create control packet congestion.

Thus, Global State Routing Routing Protocol (GSR) was developed as a remedy. Link state routing packets are not universally flooded into the network via global state routing. Each mobile node in GSR keeps three tables, including an adjacency list, a topology table, a next hop table, and a distance table.

Protocols for reactive routing: Also called as on-demand routing protocol, they include. The path is only found in this sort of routing when it is necessary. Route request packets are sent around the mobile network to perform route discovery. Route discovery and route maintenance make up its two primary aspects.

Reactive/on-demand routing protocol called Dynamic Source Routing (DSR). The path is only found in this sort of routing when it is necessary. Route request packets are sent around the mobile network to perform route discovery. It is divided into two stages:

Route discovery: During this stage, the best route for transmitting data packets between the source and the destination mobile nodes is identified.

Route maintenance is carried out at this phase because mobile ad hoc networks have dynamic topologies and often have connection failures that cause the network between the mobile nodes to fail.

Protocol for Ad-Hoc On-Demand Vector Routing (AODV): It is an on-demand and reactive routing mechanism. It is an expansion of the dynamic source routing protocol (DSR) and aids in eradicating some of its drawbacks. After route discovery, the source mobile node in DSR includes the whole path in the header of the data packet it delivers to the destination mobile node. As a result, as the size of the network grows, so does the length of the total route and the size of the header in each data packet, which slows down the entire network.

Ad-Hoc on Demand Vector Routing protocol was developed as a result. The primary distinction is in how the route is stored; whereas DSR puts the path in the data packet's header, AODV stores the path in the routing table. Similar to how it works, it too has two phases: route discovery and route maintenance.

Hybrid Routing Protocol: This protocol essentially combines the benefits of reactive and proactive routing techniques. The source and destination mobile nodes' zones and positions are taken into account when these protocols adjust. Zone Routing Protocol is one of the most well-liked hybrid routing protocols (ZRP).

After segmenting the network into several zones, the locations of the source and destination mobile nodes are tracked. Proactive routing is used to transmit the data packets between the source and destination mobile nodes if they are both located in the same zone. Additionally, reactive routing is employed to transmit the data packets between the source and destination mobile nodes if they are situated in separate zones.

Introducing Dynamic Source Routing

Dynamic Source Routing (DSR), which can find the path from source to destination only when necessary and needed, falls under the reactive routing protocol category.

The "Route Discovery Mechanism" used by the Dynamic Source Routing protocol is able to determine the path taken by data packets from the source node to the destination nodes via intermediary nodes.

No distinct database is kept, similar to proactive routing methods like Global State Routing and Dynamic Sequence Distance Vector Routing. The main difference between DSR and GSR and DSDV is that in DSDV, after requesting a route from source to destination, the length of the path through intermediate nodes is examined. A "Re-Request" packet is then sent from source to destination using the network's shortest path. The "Re-Request" packet does provide its particular ID.

It is simpler for the sender to transmit the data packets on a set route rather than sending them on several pathways to calculate the total distance thanks to this procedure of sending a "Re-Request" packet independently from source to destination.

A straightforward and effective routing system created expressly for use in mobile node multi-hop wireless ad hoc networks is called the Dynamic Source Routing protocol (DSR). DSR eliminates the requirement for any pre-existing network management or infrastructure and enables the network to function entirely on its own. The protocol is made up of the "Route Discovery" and "Route Maintenance" mechanisms, which cooperate to let nodes find and keep track of routes to any destination in the ad hoc network. The protocol is totally demand-driven, enabling DSR's routing packet overhead to dynamically grow to just the amount required to respond to changes in the routes presently in use route finding. The source broadcasts a route request (RREQ) message indicating the destination node for which the route is requested if it does not already have a route to the destination in its route cache. A route record that details the order of nodes the RREQ message travelled is included in the message. An intermediate node checks to verify whether it is already in the route record when it gets an RREQ. If so, the message is lost. In order to avoid routing loops, this is done. The message is likewise dropped if the intermediary node has already received the RREQ. In accordance with the route indicated in the header, the intermediate node passes the RREQ to the next hop. The destination sends a route reply message in response to receiving the RREQ. A route response (RREP) message may be sent along a route if the destination has a route to the source in its route cache. If not,

the RREP message may be sent backwards to the source. To respond to RREQs, intermediate nodes may also utilise their route cache.

An intermediate node may add a route to the route record in the RREQ and send an RREP back to the source with the route if it already has one for the destination in its cache. The RREQ may not flood as much as a result. The source might, however, get stale routes if the cached route is outdated preserving the route. A route error (RERR) message is sent back to the source when a node discovers a broken connection while attempting to convey a packet to the next hop. All routes containing the link in error are removed at that node when an RERR message is received.

DSR has the following benefits: Routes are only maintained between nodes that need to communicate, which lowers the overhead of route maintenance Route caching can further lower the overhead of route discovery A single route discovery may result in multiple routes to the destination because intermediate nodes respond with information from local caches.

The following drawbacks of DSR must be taken into consideration: Packet header size increases with route length due to source routing Flood of route requests may potentially reach all nodes in the network Care must be taken to prevent 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 Response A storm issue. By forbidding a node from transmitting RREP if it hears another RREP with a shorter path, reply storm may be reduced.

An intermediary node might contaminate other caches by sending Route Reply using an outdated cached route. Route Caching is a technique for DSR improvement. Every node stores a new route it discovers in its cache. In the earlier instance, Node S learns route [S,E,F] to node F when it discovers route [S,E,F,J,D] to node D. Node K learns the route [K,G,C,S] to node S when node K gets a Route Request [S,C,G] directed for node. Node F learns the route [F,J,D] to node D when node S transmits Route Reply RREP [S,E,F,J,D]. Node E learns the path [E,F,J,D] to node D when it sends Data [S,E,F,J,D]. When a node overhears data packets, it may also pick up a route.

Utilizing a route cache may hasten route finding and slow down route request dissemination. Stale caches have the potential to negatively impact performance, which is a drawback. Cached routes may become invalid with the passage of time and host migration.

Working Dynamic Source Routing Protocol

While broadcasting the route to its neighbours, dynamic source routing avoids flooding the network with data. Only the total distance travelled or the number of nodes present between the source and destination nodes are used to determine the route.

Take into account a network of 10 nodes, with node N1 acting as the source and node N10 as the destination nodes. You may follow the steps below to learn how the DSR protocol works and how Re-Request packets are sent across the network.

Ad-hoc Distance Vector Routing on Demand (AODV)

Another responsive protocol is AODV, which only keeps active routes in caches or tables for a certain amount of time before expiring. A group of distant nodes that describe the route to a destination is referred to as a distance vector. The AODV algorithm may be seen as an ancestor of the DSR and DSDV algorithms. It makes use of the same route finding process as DSR. Performance may sometimes suffer as a consequence of DSR's inclusion of source routes in packet headers, especially when the data contents of a packet are minimal. By storing routing

tables at the nodes, AODV aims to outperform DSR by removing the need for routes to be included in data packets. The DSR characteristic that only maintains routes between nodes that need communication is retained by AODV. However, AODV employs hop-by-hop routing by keeping routing table entries at intermediary nodes as opposed to DSR, which uses source routing.

Route finding. When a source requires a route to a destination but does not have one in its routing database, the route discovery process is started. The source floods the network with RREQ packets that indicate the destination for which the route is requested to start route discovery. A node determines if it is the destination or has a route to the destination when it receives an RREQ packet. The node creates an RREP packet and sends it back to the source through the reverse way if any of the two conditions is true. A forward pointer to the node from whom it got the RREP is established by each node along the reverse route. As a result, a direct route from the source to the destination is established. The RREQ packet is rebroadcast if the node is not the destination and does not have a route there. Duplicate RREQ packets are dropped at intermediate nodes. The source node may start transmitting data to the destination as soon as it gets the first RREP. Each item in the node routing database and all RREQ and RREP packets are labelled with a destination sequence number to identify the relative degree of route out-of-dateness. A more recent (or current) route is indicated by a bigger destination sequence number. When a node receives an RREQ or RREP packet, it merely changes its routing information to establish the forward or reverse path, depending on whether the route in the RREQ or RREP packet is more recent than the node's own route.

Maintaining the route. A node sends an RERR message to all sources utilising the broken link when it discovers a broken connection while trying to forward a packet to the next hop. All routes that use the connection along the way are deleted by the RERR packet. A source starts a new route discovery process if it gets an RERR packet and a route to the destination is still needed. If a route is not utilized for a certain period of time, it is also removed from the routing database

Routing types in AODV:

It comprises of the following three different kinds of routing messages.Route Request: RREQ: In order to start the route discovery process, a node that wants to send or transmit a packet but doesn't know how to get there sends an RREQ multicast message. Prior to the message reaching the destination node, neighboring nodes forward it to their neighbors while keeping track of where the message originated.

RREP: Route Reply: The destination node answers with an RREP, which follows the path taken by the RREQ to return to the source. Forward routes are created in the intermediary nodes as the RREP returns to the source. Nodes can enter an established route by sending an RREP in response to a received RREQ if an intermediary node, who understands the route to the destination, does so. Once the RREP reaches the source and the route is established, communication between the source and the destination will start.

Route Error: REER: As a reactive protocol, AODV often has less overhead than a proactive protocol (fewer route maintenance messages). A RERR message is sent through a node that detects the link interruption in the event of a connection disruption that causes the path to cease functioning, i.e., messages cannot be sent. Other nodes recast the message. The unreachable destination is indicated by the RERR message. The route becomes inactive at message receiving nodes.

A hop-by-hop vector routing technique called Destination Sequenced Distance Vector (DSDV) necessitates that each node regularly broadcast routing changes. The Bellman-Ford routing technique has been modified to create this table-driven algorithm. Every node in the network has a routing table with entries for all of the network's destinations and the amount of hops needed to get there. Each item is assigned a sequence number that aids in spotting outdated entries. By using this approach, the protocol is able to prevent routing loops from developing. To advertise its position, each node regularly sends updates that are tagged with an even sequence number that increases monotonically throughout the network. The address of the target, the quantity of hops needed to get there, the sequence number of the information received about the destination, as well as a new unique identifier specific to the broadcast, are all included in new route broadcasts. It is always taken the path designated by the most recent sequence number. The neighbours of the transmitting node learn that they are one hop distant from the source node when they get this update, and they factor this knowledge into their distance vectors. Every node's routing database contains the "next routing hop" for each destination that is accessible. The most current route, or one with the greatest sequence number, is the one that is taken.

CHAPTER 11

GLOBAL MOBILE SATELLITE SYSTEMS

Gowrishankar J, Assistant Professor
Department of Computer Science Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Karnataka – 562112

Mobile communications are supported by satellite communication. Satellites provide worldwide coverage without the expense of installing wire for base stations and are mostly unaffected by changes in population concentrations. Space Segments: mostly made up of communication satellites in different orbits, including GEO, LEO, MEO, and HEO, depending on the demands of the customers. Ground Segments: mostly made up of user terminals for Hub or Gateway Earth Stations. A Channel of Communication Depending on the user application, ITU approval, band width, geographic location, data rate, data volume, and traffic density requirements, communication is established by uplink and downlink frequencies in UHF, L band, S band, normal C-band, extended C-band, Ku band, X band, Ka band, Q-band, V band, W band, etc. The fast speed of low-altitude satellites creates additional issues with routing, locating mobile users, and switching across communication lines.

History using satellite technology

After World War II, satellite communications were first developed. Arthur C. Clarke's article on "Extra Terrestrial Relays" was published in 1945. However, it wasn't until 1957, in the midst of the Cold War, when the Soviet Union abruptly launched the first satellite SPUTNIK, shocking the Western world. SPUTNIK was essentially a tiny transmitter that sent a sporadic "beep," which is in no way analogous to a satellite today. Geostationary satellites are the mainstay of modern satellite news transmission. Their permanent location in the sky is a huge benefit. They seem to be anchored to a certain position because their spin is synchronised with the earth's rotation. In 1965, INTELSAT 1 (commonly referred to as "Early Bird"), the first commercial geostationary communication satellite, began operations. It weighed 68 kg, had 240 duplex telephone channels, or a single TV channel, and was in use for one and a half years. Following in 1967, INTELSAT 2 supplied 1,200 telephone channels, and INTELSAT 3 followed in 1969. While utilising cables is always an option for communication on land, this is not the case for ships at sea. In 1976, MARISAT launched three satellites that provided global marine connectivity. However, the ships with big antennas still needed to have sender and receiver fitted (1.2 m antenna, 40 W transmit power). In 1982, INMARSAT-A, the first mobile satellite phone system, was launched. Inmarsat-C was the first satellite system to provide mobile phone and data services six years later. (Interfaces to the X.25 packet data network exist, with data speeds of roughly 600 bit/s.) With the introduction of INMARSAT-M in 1993, satellite telephone networks become entirely digital.

With the development of international satellite systems for tiny mobile phones, such as Iridium and Globalstar, in the year 1989, a new era of satellite data transmission began. The fact that there are already more than 2,000 geostationary satellites in operation demonstrates the satellite communication industry's phenomenal expansion over the last 50 years. The Great Visionary Dr. Vikram Sarabhai launched the Indian Space Program in 1969. In 1981, INSAT-1A, the first Indian satellite, was sent into orbit. Satellites from the INSAT and GSAT series 42 are currently in use in orbit.

Traditionally used satellite systems

Modern satellites more closely resemble flying routers than the basic transponders of the past. In essence, transponders take in a signal on one frequency, magnify it, and then send it on a different frequency. Signal regeneration is now possible because to the usage of digital signals, when once only analogue amplification was feasible. The satellite converts the signal from a bit stream into a signal after decoding it. The greater quality of the signal received on the earth is a benefit of digital regeneration over pure analogue amplification. Intersatellite routing, error correction, and other tasks of higher communication layers are all provided by today's communication satellites.

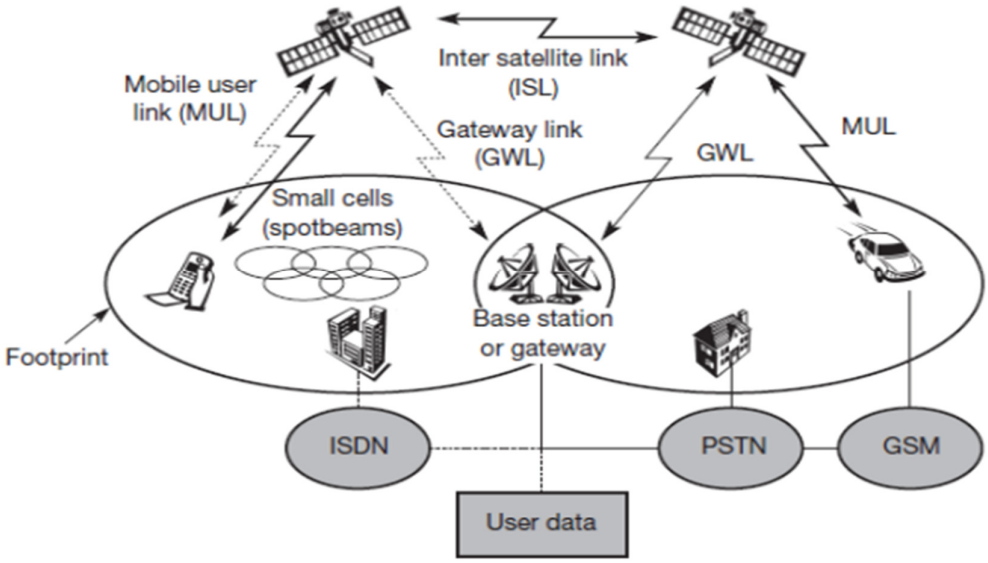


Figure 11.1: Traditionally used satellite systems

A typical scenario for satellite systems allowing worldwide mobile communication is shown in Figure 11.1. Each satellite may cover a certain region of the world with its beam, or "footprint," depending on its kind. Mobile users may communicate with the satellite inside the footprint through a mobile user link (MUL), and the base station that controls the satellite and serves as a gateway to other networks can do so via a gateway connection (GWL). Through intersatellite connectivity, satellites could be able to interact directly with one another (ISL).

Direct communication between users within various footprints is made possible thanks to this without the need for base stations or other earthly networks. Reducing the number of satellite-to-earth connections helps lower voice and data latency. Spot beams, such as the 163 spot beams per satellite in the ICO system (ICO, 2002), are used by certain satellites' unique antennas to divide larger cells into smaller ones. The necessary terrestrial service infrastructure, including the connections for control, between satellites.

The many networks that currently exist on earth will continue to benefit greatly from the inclusion of satellite systems. Users may communicate over cellular networks like GSM and UMTS as well as ISDN and other PSTN options. There are several gateways that enable smooth connection between these various networks. The flawless handover between a cellular network and a satellite system (known as a vertical handover), for instance, poses a significant problem since it is already well-known from inside cellular networks (horizontal handover). Users shouldn't realise when a satellite network replaces, say, a GSM network during a discussion.

Applications Satellites have historically been applied in the following fields:

Weather forecasting: To forecast the weather, satellites take pictures of the world and transmit them to a base station. Predicting abnormalities like hurricanes earlier allows for the taking of preventative actions, which is incredibly important.

Satellites for radio and TV transmission are a quick, low-cost alternative to cable networks for radio and television broadcasting.

Military satellites: Satellite-based communication networks are used for military purposes since they are more secure against enemy assault. The majority of military satellites use the X-band frequency.

Satellites for navigation: The Global Positioning System (GPS) offers very accurate localisation data with a few-meter accuracy. GPS is the standard navigation system for all ships and aeroplanes. For fleet management, vehicle localization, and other purposes, information gathered from GPS receivers placed in cars is employed.

With seven GEO/GSO satellites, the Indian Space Research Organization (ISRO) has created the "NavIC" deshi GPS, an indigenous Indian regional navigation satellite system. It has started to work. It may take the place of the most widely used GPS systems in the US, Russia, China, and Europe, including Glonass and Beidou. The NavIC System can create potential applications for road navigation, rail navigation, and safety, generate warnings for unmanned level crossings, track vessels for costal surveillance, conduct land and marine surveys, and correct differentials using geodetic receivers. It can also provide emergency calling, disaster management and warning, and time and frequency synchronisation for Internet and intranet applications.

Global telephone backbones: Satellite communications are becoming a viable alternative to cables for this purpose.

The fibre optical cables that bridge the seas are rapidly replacing the satellites. The major cause of this is fibre optical networks' enormous capacity (commercially, about 10 Gbit/s via wavelength division multiplexing; several Tbit/s in laboratories), especially considering how much faster they are than satellites. While the distance for a signal to travel from a sender to a geostationary satellite and back is around 72,000 km, if a fiber-optic connection across the Pacific or Atlantic Ocean is employed, the distance is often less than 10,000 km. Geostationary satellites experience a one-way, single-hop time delay of 0.25 s due to the unfortunate limitation of the speed of light. It may sometimes be irritating to use satellites for telephone conversations.

Satellites provide an easy and speedy link to global networks to locations that are unreachable owing to their geographic location. This is especially useful for developing or distant regions.

Global mobile communication: The most recent benefit of satellites is the transmission of mobile data. Because of their high latency rate, geostationary satellites are not the best choice for this, necessitating the employment of satellites in lower orbits. The use of satellites for mobile communication is an addition to current systems rather than a replacement. They are superior to current cellular systems like AMPS and GSM since they offer global coverage. For the satellite part of the UMTS system, frequency ranges that are immediately next to the terrestrial bands have been designated (S-Band: 1980–2010 MHz uplink, 2170–2200 MHz downlink).

Indigenous communication satellites provide INDIAN Critical Services.

The INSAT and GSAT series of communication satellites are operated by India and provide UHF, S-band, Normal C band, Cext C band, Ku-Band and Ka band multi band multipurpose Transponders for Telecommunication, Broadcasting, DTH, VSAT/Business Communication Societal Development, Dedicated satellite for Tele-education, e-learning, Telemedicine, METSAT and Disaster services, mobile services, advance communication services for internet, high throughput satellite for more channels and higher speeds.

Ford Aerospace US produced the INSAT-1 series of multifunctional satellites based on Indian design and service requirements. These satellites were sent into orbit by the foreign launchers Delta and Ariane. All satellites in the INSAT-2 series and later were produced domestically by ISRO. The GSAT series began after the INSAT series. Several. Satellites from the GSAT and INSAT-2,-4,-3 series that weigh less than 3000 kg have been launched by PSLV or GSLV launch vehicles from the Indian Launch Pad at SHAR. The older satellites were intended to operate for ten years. Today's geostationary satellites have planned lives of over 15 years.

Indian spacecraft in orbit that are presently operating include: 18 Earth observation satellites (including meteorological ones), 15 communication satellites, 7 navigation satellites, 2 space science satellites, and 26 satellites that are currently in different phases of development.

Principal Applications

Monitoring of resources, planning of infrastructure, assistance for disaster management, enabling weather forecasting, location-based services, and a variety of other social applications, such as those required for satellite communication. Agriculture, forestry, and the environment; water resources; urban and rural planning; asset mapping; prospecting for minerals; ocean resources; meteorology; location-based services; tele-education; tele-medicine; and assistance with disaster management.

Wireless Enterprise Networks

In order to link all users and systems on a local area network (LAN) to applications in the data centre and cloud and to make network data and analytics accessible, an enterprise network is made up of physical and virtual networks and protocols. In a LAN, many local computer devices are linked together to exchange data and applications via switches, routers, and ethernet or WiFi connections. For safe access, users normally need to create accounts. In order to protect user information while connecting to websites or servers outside of a LAN, businesses often utilise VPN software. Additionally, firewall software is used to monitor and manage network traffic, both inside the network itself and between the LAN and the outside world (north-south) (east-west).

Enterprise networking

Enterprise networking offers programmes and end users quick and dependable connection. In the modern network, applications are being spread more widely, making it necessary for businesses to simplify networking and security across wired and wireless infrastructure. Network administrators want network automation frameworks that streamline day 1 and day 2 network operations as well as corporate networking solutions that provide a single pane of glass across data centres and clouds. Enterprise network managers also have a significant duty in the area of security. Firewall setup is crucial to business networking since internal and perimeter firewalls are intended to protect applications and data from external assaults. Security administrators look for cutting-edge ways to scan data packets for viruses and malware in order

to enhance business network security and stop infections that may spread as a result of phishing scams and ransomware.

Advantages of corporate networking

Every business need a special networking solution that supports its workflow, manufacturing procedures, customer demand, logistics, etc. Organizations may do the following with the correct network: Collaboration may increase productivity since workers can collaborate on shared resources from a distance as well as in an office, industrial, or college setting. Access to business resources might be restricted. Organizations can provide connection to programmes and data that are monitored and protected by internal and external firewalls increased productivity Modern networking may significantly increase staff productivity, from faster test/dev with collaboration tools and version control to private cloud orchestration with cloud-based apps and an agile internal firewall. Lower prices Businesses may optimize the effective distribution of resources across on-premises and cloud infrastructure by combining server and network virtualization. To further enhance current corporate operations, enterprise networking comprises solutions for analytics, monitoring, and security that may be added.

Network Virtualization

Network virtualization is the technique of conceptually classifying physical networks to run as a single or a number of separate networks referred to as Virtual Networks.

General Network Virtualization Architecture

Network virtualization tools:

A. OS for a physical switch
B. OS for a physical switch

The network virtualization features of a hypervisor are used in conjunction with either built-in networking or third-party applications. The OS's fundamental job is to provide a straightforward set of instructions to the programme or process that is currently running. Similar to the service primitives provided at the application and network interface via the SAP, system calls created by the OS and performed through the libc library (Service Access Point). To build a virtual switch and set up virtual networks on it, utilise the hypervisor. The hypervisor's inherent networking capabilities are replaced by the third-party software, which is installed on the hypervisor. We may have a number of virtual machines (VMs) running efficiently on a single piece of computer hardware thanks to a hypervisor.

Network virtualization's purposes:

A. It makes it possible for nodes in a virtual network to be functionally grouped.
B. The virtual network may now share available network resources.
C. It enables frameless communication between nodes in a virtual network.
D. Traffic for management is constrained.
E. For communication across virtual networks, it enforces routing.
F. Virtual Data Center Network Virtualization

Physical Network 1.

Network adapters, switches, bridges, repeaters, routers, and hubs are examples of physical components.

Provides communication between physical servers running a hypervisor, as well as between physical servers and storage systems and clients.

VM Network 2.

A. Virtual switches make up.
B. Connects you to the hypervisor kernel.
C. Ties with the actual network.
D. Is located within the actual server.
E. Virtualization of the Network In VDC

Network virtualization benefits:

A. Enhanced manageability
B. Node grouping and regrouping are made easier.
C. Using management software, VM configuration is possible from a centralised management workstation.
D. Decreases CAPEX
E. There is less need to configure distinct physical networks for each node groupings.
F. Enhances Utilisation
G. The ability for several VMs to share a single physical network improves network resource consumption.
H. Improves Performance
I. Network broadcast is limited, and virtual machine performance is enhanced.
J. Increases security

In a hybrid environment with integration of the cloud, it must cohabit with physical devices.

A. A rise in complexity.
B. Upfront price.
C. Maybe a learning curve
D. Network virtualization examples include:
E. VLAN (Virtual LAN)
F. VLAN may enhance the performance and speed of congested networks.
G. Any network modifications or additions may be made easier using VLAN.

Overlays on the network

VXLAN, an encapsulation technology, offers a foundation for layer 3 networks to be overlaid with virtualized layer 2 networks. A novel method of encapsulation that is intended to ensure control-plane independence between the tunnel's ends is offered by the Generic Network Virtualization Encapsulation protocol (GENEVE).

Platform for Network Virtualization: VMware NSX

Switching, firewalling, and routing are examples of networking and security components that are transported by VMware NSX Data Center and specified and used in software. It transfers a virtual machine's (VM) operating model for the network.

Network virtualization applications

The development of application testing may make use of network virtualization to simulate real-world hardware and system software. We may combine numerous physical networks into a single network or divide a single physical network into many analytical networks with its assistance. Network virtualization enables the modelling of connections between applications, services, dependencies, and end-users for software testing in the area of application performance engineering. We can launch apps more quickly thanks to it, facilitating a speedier

go-to-market. Network virtualization enables software testing teams to provide accurate findings in a networked environment with anticipated instances and congestion problems.

Virtual Network Operates

With the use of contemporary technology, a virtual network may expand an existing wireless network. This comprises:

Virtualization software on host servers that enables the setup and configuration of a virtual network is known as "vSwitch software."

Virtual network adapter: Constructs a network gateway.

As a host for the virtual network architecture, the physical network is necessary.

Devices and virtual machines: Tools that connect to the network and provide a range of functions.

The network host infrastructure includes servers.

Security and firewalls: Designed to track and thwart security threats.

Virtual networks fall into one of three categories: VPN, VLAN, or VXLAN.

VPN Virtual private network is referred to as VPN. In essence, a VPN connects two or more active networks over the internet. Users may access the linked physical networks by logging in to this internet-based virtual network from anywhere. VPNs are frequently used to hide internet activity on public WiFi networks and to promote secure surfing. Data attached to packets that specifies routing information that directs users to the appropriate address creates a VPN. This results in the creation of a tunnel of addresses that encrypts browsing history and enables distant information access. VPNs provide a limited-scope, entirely virtual network that connects users through the internet.

VLAN

A virtual LAN network, or VLAN, divides the devices on a LAN network into domains with their own resources and settings. Better security, monitoring, and control of the devices and servers within a given domain are made possible by using a VLAN. This is particularly true for big networks, which may be more open to attack if domains are not utilised and individually kept track of.

VXLAN

Virtual extensible local area network is referred to as VXLAN. Your level 3 network architecture in this network offers a tunnel into level 2. Each tunnel's endpoints are created by virtual switches, and data may be sent between endpoints using a different piece of hardware called a physical or virtual base case.

Virtual networking advantages

The advantages of virtual networking are many and include:

Working remotely is possible because to virtual networking, which enables access to networks from anywhere in the globe.

Digital security: By using technologies like tunnelling encryption and domain segmentation, you may use virtual networking to make your networks more secure. Simplifies hardware Enterprise firms may decrease the amount of hardware they need to access, maintain, and

monitor by employing vSwitches to route operations from one location to another. Flexibility and scalability: Because it's virtual and just a small amount of hardware is needed to build a virtual network, scaling is simpler and more affordable. A small amount of software and configuration changes are needed for scaling, although more hardware is not always necessary.

Cost savings: Companies gain from lower hardware expenses by spending less on repairs and maintenance.

Networking virtually and modern business

Virtual networking is crucial to every digital company strategy in the modern era. It is a technological advancement that takes into account the demands for cost effectiveness, flexibility, scalability, and remote accessibility. Like many other services that large firms may outsource, doing so provides advantages in terms of time, money, and precious resources that can be better used to make sure your technology is completely up to date and fulfilling your demands.

Virtual networking and NaaS services will continue to be more and more crucial to all enterprises as social demands force more employees to work remotely. For firms that have already undertaken the process of becoming a digital company, expanding virtual networking capabilities may be the next stage in digital transformation. One method for organisations to continue to develop in the digital age is to extend their virtual networks to include more than just a basic VPN for the extra productivity benefit.

Bluetooth

Bluetooth is an open standard for a radio system that offers the necessary network infrastructure to support speech and data transmission over short distances. It is made up of both hardware and software components. The standard also includes user profiles and use model descriptions. Harald Blatand, a Danish king who united Norway and Denmark in the tenth century, was known as Bluetooth. The idea of Bluetooth wireless technology was to combine the computer and telecom sectors. Without the need of a cable, Bluetooth technology enables ad hoc wireless connections between gadgets like cell phones, desktop, and laptop computers. Within a range of 50 metres (150 feet) or more, devices with Bluetooth-enabled processors may readily send data at a rate of roughly 720 Kbps through walls, clothes, and even baggage bags.

Protocol for Bluetooth

Switching techniques are combined in the Bluetooth protocol. The channel is slotted, and synchronous packets may reserve spaces. Asynchronous connection-less (ACL) connections for data and up to three simultaneous synchronous connection-oriented (SCO) links for voice are supported by the Bluetooth protocol stack, as well as a mix of asynchronous data and synchronous voice (DV packet type). A 64 Kb/s synchronous channel in each direction is supported for each voice channel. Maximum data rates for the asynchronous channel are 723.2 Kbps uplink and 57.6 Kbps downlink (or vice versa), or 433.9 Kbps for symmetric lines. The baseband serves as the physical layer of the stack, while the link manager and controller serve as the link layer. The way these two levels are built and utilised with apps affects the top layer interface. Here is a picture of the stack.

Scatternet and Piconet

Point-to-point (unicast) and point-to-multipoint (multicast) communications are both supported by Bluetooth. The master and slave model is used by the Bluetooth protocol. In a masterslave protocol, a device cannot communicate whenever and whenever it pleases. They must wait

until the master gives them permission to speak. Together, the master and slaves make up a piconet. The piconet's hopping pattern is set by the master, and the slaves must synchronise to it. There are also two more device types: Parked (P) and Standby (SB). Although parked devices are identified and may be awakened in a matter of milliseconds, they cannot actively participate in the piconet. Piconet is not used by devices that are in standby. There is precisely one master and a maximum of seven simultaneous slaves in each piconet. There is space for more than 200 devices. Ad hoc connections between a number of these piconets may create a bigger network. One way to conceptualise the topology is as a flexible, many piconet structure. Scatternet is the name of this network of piconets. When a device from one piconet participates in another piconet, a scatternet is created. According to this plan, a device that is a master in one piconet may also be a slave in the other.

The Bluetooth protocol is a synthesis of many protocols. The majority of Bluetooth devices need the Bluetooth Core protocols and Bluetooth radio protocols, although other protocols are utilised as needed by various applications. Bluetooth employs spread spectrum methods at the physical layer. It makes use of frequency hopping spread spectrum as well as direct sequence spread spectrum. Both connection-oriented (SCO-Synchronous Connection-oriented Link) and connectionless (ACLAsynchronous Connectionless Link) connections are used by Bluetooth. Application-oriented protocols are created by the combination of the Cable Replacement layer, Telephony Control layer, and Adopted protocol layer. These layers allow programmes to run over the Bluetooth Core protocols.

The Use of Bluetooth in Mobile Computing

High speed and low power wireless technology called Bluetooth is used to link phones and other portable devices for file transfers or conversation. This is based on technology for mobile computing. The list of some key characteristics of Bluetooth technology is as follows:

In order to connect phones, laptops, and other network devices across a short distance without the usage of any form of connecting cables, Bluetooth is also known as the IEEE 802.15 standard or protocol.

The 2.4 to 2.485 GHz frequency range is used by Bluetooth, an open wireless technology standard, to transmit or receive data to linked devices that are spread out across a certain distance.

Typically up to 30 feet or 10 metres, Bluetooth technology uses wireless signals to transfer data and files across short distances.

A collection of five firms called as Special Interest Group, founded in 1998, created the Bluetooth technology. The organisations include Toshiba, IBM, Ericsson, Intel, Nokia, and IBM.

In earlier device versions, the Bluetooth technology's data exchange range was up to 10 metres, while Bluetooth 5.0, the most recent version, can transmit data over a range of around 40–400 metres.

In the very first iteration of Bluetooth technology, the typical data transfer speed was roughly 1 Mbps. The second version offered a 3Mbps data rate speed and was called 2.0+ EDR. The third option offered a speed of 24 Mbps and was called 3.0+HS. This technology is currently at version 5.0.

Background of Bluetooth

The development of Bluetooth technology has a fascinating history. The wireless Bluetooth technology has the name of Harald Blatand, a Danish monarch. His last name in English is "Bluetooth." The term "Bluetooth" was given to this technology since the Danish King Harald Blatand unified Norway and Denmark, much as Bluetooth wireless technology connects two dissimilar devices for data transfer or communication.

In 1994, Ericsson Mobile Communications began to develop Bluetooth technology. The primary goal of this incredible technological advancement was to eliminate the need for wires for communication between mobile phones and other devices. The Bluetooth Special Interest Group (SIG), which was founded in 1998 by four major corporations at the time Ericsson, IBM, Nokia, and Toshiba released the first Bluetooth technology in 1999. Four more versions have now been made available. Bluetooth 5.0 is the most recent iteration of this technology.

The Bluetooth Technology Architecture

Because it is made up of many networks, the Bluetooth architecture is sometimes known as a "Piconet," and the Bluetooth network is made up of Personal Area Networks. A minimum of 2 and a maximum of 8 Bluetooth peer devices are present. It typically has one master and up to seven slaves. The technology that Piconet offers, based on its Master and Slave Nodes, allows for data transfer. The slave nodes are utilised to receive the data, while the master node is in charge of delivering the data. Short-wavelength radio waves at ultra-high frequencies are used in Bluetooth technology to transmit data. Spread spectrum and multiplexing are used by the Piconet. It combines the frequency hopping spread spectrum (FHSS) and code division multiple access (CDMA) technologies.

As previously mentioned, a Bluetooth connection may have one master and up to seven slaves. The device that starts communication with other devices is known as the master. The communications connection and traffic between the master device and the slave devices connected to it are managed by the master device. The slave devices must react to the master device and match the master device's set time with their broadcast and receive timing. Conditions for Successful Data Transmission using Bluetooth Technology in Mobile Computing The list of prerequisites for a successful data transfer using Bluetooth technology is as follows:

A. Maximum Master Node Number: 1
B. Maximum Slave Node Number: 7
C. Piconets may have a maximum of 8 nodes.
D. 28 - 1 = 255 is the maximum number of devices that may be associated.
E. Infinite (infinite) number of devices that can be parked

Explanation of Bluetooth Technology in Mobile Computing

The standby node is a sort of node that may either become a slave or parked node, or it can stay idle or unconnected. The parked node is a type of node that is prepared to be attached. Data transmission in Bluetooth technology is limited to interactions between master and slave nodes. Between slave and slave nodes, it cannot be done. But there is a way to join two master nodes.

The whole Piconet is unplugged if the link to the master node is severed. A network is referred to be a scatter-net if there is a link between two master nodes. In other words, scatter-nets are formed when a device joins many Piconets and actively participates in each one while sharing its time slots with the next device. The data transmission speed in a Piconet will drop if the

number of slaves or devices is raised, and it will rise if the number of slaves or devices is lowered.

Bluetooth Technology Specifications

Two kinds of Bluetooth technology are available:

A. Specifying the Core
B. Profession Specification
C. The Core Requirements

The Bluetooth protocol stack and the specifications for the evaluation and certification of Bluetooth-based devices are defined in the core standard.

There are 5 levels in the fundamental Bluetooth Technology specification:

Radio: It is used to describe the specifications for a Bluetooth transceiver's frequency, modulation, and power characteristics.

In the baseband layer, physical and logical channels, voice or data connection types, different packet formats, transmit and receive time, channel control, frequency hopping, and device addressing are all defined. Additionally, point-to-point or point-to-multipoint linkages are specified. A packet's length may range from 68 bits to a maximum of 3071 bits.

The methods for connection setup and ongoing link management are defined by the Link Manager Protocol (LMP).

L2CAP, or the Logical Link Control and Adaptation Protocol, is used to translate baseband-layer protocols from upper-layer protocols.

The Service Discovery Protocol (SDP) enables a Bluetooth device to ask other Bluetooth devices for information about their equipment, the services they provide, and the details of those services.

Here, the host is represented by the last two levels, while the first three layers represent the Bluetooth module. The Host Controller Interface connects these two logical groupings together.

Profession Specification

It offers use models that give specific information on how the Bluetooth protocol may be used for different kinds of applications.

Bluetooth Technology Benefits

The benefits of Bluetooth technology are listed below:

A. Wireless technology is the foundation of Bluetooth technology. It is affordable since it doesn't need any transmission line, which lowers the price.
B. In Bluetooth technology, creating a piconet is fairly straightforward.
C. The Speed Frequency Hopping method is used to solve the radio interference issue.
D. The usage of energy or power is incredibly minimal, at about 0.3mW. It enables the lowest possible battery life consumption.
E. It is strong because it ensures bit-level security. A 128-bit key is used to regulate the authentication process.
F. You may use it to transmit data and communicate verbally since Bluetooth supports data channels with up to three identical speech channels.

G. Unlike other wireless communication methods like infrared, it doesn't need line of sight or one-to-one communication.

Bluetooth Technology Drawbacks

The bandwidth of Bluetooth technology is limited. Because it is also shorter, the data transmission range might be a problem.

Uses for Bluetooth technology

Numerous communication and entertainment gadgets employ Bluetooth technology. The following are some of the most popular uses for Bluetooth technology:

The cordless desktop makes use of Bluetooth technology. It denotes the absence of a wired connection between the desktop and the peripheral devices, such as a mouse, keyboard, printer, speakers, etc. The Use of Bluetooth in Mobile Computing. The exchange of multimedia files including music, films, photographs, and other types of files that may be exchanged between devices via Bluetooth uses it. Bluetooth Speakers are another item that uses this technology.

A. Headphones with Bluetooth.
B. Bluetooth headsets used for making calls.
C. Gaming systems with Bluetooth, etc.

Bluetooth network technology creates a personal area network by wirelessly coupling mobile devices over a short distance (PAN). Instead of adopting the common OSI model or TCP/IP model, the Bluetooth design has its own separate model with a stack of protocols. Another distinctive feature of the Bluetooth system is that not every device has to utilise every protocol in the stack. This is thus because several programmes may utilise Bluetooth, and each application specifies which layer of the protocol stack to employ.

Bluetooth radio, Baseband, Link Manager Protocol (LMP), Logical Link Control and Adaptation Protocol (L2CAP), and Service Discovery Protocol are among the protocols that make up the Bluetooth Protocol Stack (SDP).

Radio Frequency Communications (RFComm) protocol is a part of the cable replacement protocol. It stands for Radio Frontend Component in abbreviation. With WAP, it offers a serial interface.

Protocols that have been adopted from standard models fall under this category. Point-to-Point Protocol (PPP), Internet Protocol (IP), User Datagram Protocol (UDP), Transmission Control Protocol (TCP), and Wireless Application Protocol are the most often used protocols in Bluetooth (WAP).

Attention command set; AT commands.

The Bluetooth protocol stack may be seen in the following diagram

Core Protocols' Functions

Radio: This physical layer-equivalent protocol provides the physical framework and technical requirements for radio wave transmission. It outlines the air interface, frequency ranges, requirements for frequency hopping, and modulation methods.

Baseband: This protocol utilises radio protocol services. It specifies the timing, power control, and addressing methods as well as the packet frame format.

The Link Manager Protocol (LMP) creates and maintains logical linkages between Bluetooth devices so that communications may take place. Device authentication, message encryption, and packet size negotiation are some of LMP's other key features.

L2CAP, or the Logical Link Control and Adaptation Protocol, allows for baseband layer frame format adaptation between upper layer frame formats. Both connection-oriented and connectionless services are supported by L2CAP.

SDP handles service-related inquiries, such as those pertaining to device information, in order to create a connection between competing Bluetooth devices.

Exchange of Information and Request

A Bluetooth link manager can ask another link manager for information such as the clock offset (master asking slave to tell it current clock offset stored by it which slave itself got from master during some packet exchange), slot offset (slot offset is the time in microseconds between the start of the master's transmission slot in the piconet where the PDU is transmitted and the start of the master's transmission slot where the BD ADDR device in the PDU is master), and other link manager-specific information. Timing accuracy (clock drift and jitter), link management version, and details on supporting capabilities like support for authentication, SCO packets, etc. are all helpful in master-slave switch and interpiconet connections.

Logical Link Control and Adaptation Protocol (L2CAP) L2CAP has the capacity to multiplex protocols, perform segmentation and reassembly operations, and use group abstractions to deliver connection-oriented and connectionless data services to higher layer protocols. Higher level protocols and applications are able to send and receive L2CAP packets up to 64 kilobits in length thanks to L2CAP. L2CAP only accepts links with ACLs. To create diverse connections between various Bluetooth apps, L2CAP leverages the idea of channels. Channel IDentifiers (CIDs), which indicate a logical end point of a connection for each application on a device, are used to identify these channels.

Reassembling and Segmenting

To increase efficiency, segmentation and reassembly (SAR) processes enable a maximum transmission unit (MTU) size bigger than the biggest Baseband packet. By distributing the network and transport packets utilised by upper layer protocols across a number of baseband packets, this lowers the overhead. L2CAP breaks up upper layer packets into "chunks" that can be sent by the Link Manager, then reassembles those chunks into L2CAP packets using data from the packet header and HCI.

Protocol for Service Discovery (SDP)

To connect to a piconet, a Bluetooth device has to use the Service Discovery Protocol (SDP).

A device uses SDP to find out what services are offered in a piconet and how to access them. SDP employs a client-server architecture in which the server maintains a list of services listed in service records. The attributes of one service are described in one service record on a server. There can only be one SDP server inside of a Bluetooth device. One SDP server represents all of the services that a device offers if it does so. Similar to this, several apps on a device may ask servers for service records using a single SDP client.

Protocol for Cable Replacement

The single component of this protocol stack is Radio Frequency Communication (RFCOMM).

RFCOMM: The ETSI 07.10 standard serves as the foundation for this serial line communication protocol. Over Bluetooth baseband protocol, the "cable replacement" protocol simulates RS-232 control and data streams. It offers numerous concurrent connections, flow control, serial cable line settings, and a dependable data stream.

Telephony Control Protocol This protocol stack consists of the AT-Commands and the Telephony Control Specification Binary (TCS BIN).

TCS Binary, often known as TCS BIN, is a bit-oriented telephony control protocol.

The call control signalling protocol for setting up voice and data conversations between Bluetooth devices is defined by TCS BIN. Additionally, it specifies the methods for managing groups of Bluetooth TCS devices.

The AT-commands defined by this protocol allow a mobile phone to be used and managed as a modem for fax and data transfers. A computer or DTE (Data Terminal Equipment) is used to provide AT (short for attention) instructions to "manage a modem or DCE" (Data Circuit terminating Equipment). The ITU-T Recommendation is the foundation for Bluetooth AT-commands.

PPP and TCP/IP are two adopted Internet standards that were created by the IETF. The WAP stack uses these as its bottom layer protocols.

OBEX: IrDA has outlined this session protocol. This protocol is used by Bluetooth as well, giving applications the option of using either the Bluetooth radio or the IrDA technology.

WAP/WAE: Bluetooth may be utilised as a carrier technology to transmit data between a nearby WAP server and a WAP client. PPP and the TCP/IP protocol suite are used by WAP to run on top of the Bluetooth stack.

DATA REPLICATION IN MOBILE COMPUTING

Data Replication:

In mobile computing, data replication refers to the sharing of data to assure data consistency between software and hardware resources connected over the internet, improving data reliability, availability, fault-tolerance, and accessibility. Data replication, to put it simply, is the practice of keeping several copies of the database at two or more locations in order to increase data availability in less time and at a lower cost. A common fault tolerance method for distributed databases is data replication in mobile computing [1].

Advantages from Data Replication:

Scenario data replication has become widely used in contemporary mobile computing as a reliable method of ensuring data availability, integrity, and a successful means of achieving fault tolerance. Data replication increases data sharing, lowers transmission costs, and improves the security of sensitive data in addition to ensuring the availability of the data. In mobile computing, data replication also chooses where and when to store the replica of the data, managing many data replicas across a network for effective use of the network resources [2].

Benefits from Data Replication:

A. Data replication ensures the accuracy of the information. The database system continues to function even in the event that one site fails since a backup copy is kept at a different location (s).
B. Since local copies of the data are accessible through data replication, the network load is decreased. As a result, query processing can be accomplished with less network utilization, especially during peak times.
C. Data updating is possible even outside of peak times thanks to data replication.
D. Having access to local copies of the data guarantees speedy query processing and, as a result, short response time.
E. Transactions require fewer joins of tables that are situated at various places and little network cooperation. Consequently, their nature becomes simpler.

Replication of Data Has Drawbacks:

Increased Storage Needs Keeping numerous copies of data results in higher storage expenses. The amount of storage needed is multiplied by the amount of storage needed for a centralized system. Increased Cost and Complexity of Data Updating. Every time a data item is updated, all copies of the data at the various sites must also reflect the modification. Complex synchronization methods and protocols are needed for this. Unwanted Application-Database Coupling, Removing data inconsistency necessitates intricate coordination at the application level if complicated updating mechanisms are not implemented.

Replication of Data for Mobile Computing:

Making duplicate copies of data saved across several locations helps to increase data reliability, efficiency, resilience, ease of transaction, fault tolerance, and reduces network burden [2].

Replication of Data Objectives:

Replication of data is done for two reasons: to increase data availability and to expedite query evaluation.

Data Replication Types:

Data replication comes in two different types.

Synchronous replication: Synchronous replication modifies the database replica as soon as changes are made to the relation table. Therefore, the duplicated data table and the original data table are identical.

Asynchronous Replication: In asynchronous replication, the replica will be changed following the firing of a commit action on the database.

Replication Protocols

These are the three replication plans:

Full Replication scheme: The database is accessible at all locations in a comprehensive replication strategy to facilitate user interaction with the communication network.

Benefits of complete replication

A. It offers great data availability.
B. The database is accessible at each site under this plan.
C. It enables quicker query execution.

Cons of complete replication

A. Concurrency control is challenging to implement in a comprehensive replication strategy.
B. Since every side needs to be updated during updating, the process is slower.

No Replication: Each component is stored exclusively at one location if there is no duplication.

Benefits of No Replication

A. Concurrency may be reduced with ease.
B. Data recovery becomes simple

Drawbacks of No Replication

- Data availability issues.

Because numerous clients are attempting to access the same data on the same server, this slows down the query execution process

---

Questions for Practice & Revision

1. What do you understand by Mobile Computing?
2. What are the three main parts of GSM architecture?
3. What Does Spatial Division Multiple Access (SDMA) Mean?
4. How does Mobile IP work in mobile computing?
5. What are the different data management issues in mobile computing?
6. When is it appropriate to utilize a wireless ad hoc network?
7. What are main features of WAP?
8. How does dynamic source routing handle routing?
9. Which communication is used in MANET?
10. How does Bluetooth function?

Reference Books for further Reading

1. Introduction to Mobile Computing; T. Imielinski, H.F. Korth.

2. The PARCTAB Ubiquitous Computing Experiment; R. Want, et al.

3. Scalable Support for Transparent Mobile Internetworking; D.B. Johnson.

4. Location Management for Networks with Mobile Users; B.R. Badrinath, T. Imielinski.

5. Dynamic Source Routing in Ad Hoc Wireless Networks; D.B. Johnson, D.A. Maltz.

6. Routing over Multi-Hop Wireless Network of Mobile Computers; C.E. Perkins, P. Bhagwat.

7. Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments; R. Caceres, L. Iftode.

8. Indirect Transport Layer Protocols for Mobile Wireless Environment; A.V. Bakre, B.R. Badrinath.

9. Connecting Mobile Workstations to the Internet over a Digital Cellular Telephone Network; M. Kojo, et al.

10. Asynchronous Video: Coordinated Video Coding and transport for Heterogeneous Networks with Wireless Access; J.M. Reason, et al.

11. Wireless Publishing: Issues and Solutions; T. Imielinski, S. Viswanathan

---