NEHRU COLLEGE OF ENGINEERING AND RESEARCH CENTRE
(NAAC Accredited)
(Approved by AICTE, Affiliated to APJ Abdul Kalam Technological University, Kerala)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE MATERIALS

CS 364 MOBILE COMPUTING

VISION OF THE INSTITUTION

To mould true citizens who are millennium leaders and catalysts of change through excellence in education.

MISSION OF THE INSTITUTION

NCERC is committed to transform itself into a center of excellence in Learning and Research in Engineering and Frontier Technology and to impart quality education to mould technically competent citizens with moral integrity, social commitment and ethical values.

We intend to facilitate our students to assimilate the latest technological know-how and to imbibe discipline, culture and spiritually, and to mould them in to technological giants, dedicated research scientists and intellectual leaders of the country who can spread the beams of light and happiness among the poor and the underprivileged.

ABOUT DEPARTMENT

- Established in: 2002
- Course offered : B.Tech in Computer Science and Engineering

  M.Tech in Computer Science and Engineering

  M.Tech in Cyber Security
- Approved by AICTE New Delhi and Accredited by NAAC
- Affiliated to the University of A P J Abdul Kalam Technological University.

DEPARTMENT VISION

Producing Highly Competent, Innovative and Ethical Computer Science and Engineering Professionals to facilitate continuous technological advancement.

DEPARTMENT MISSION

1. To Impart Quality Education by creative Teaching Learning Process
2. To Promote cutting-edge Research and Development Process to solve real world problems with emerging technologies.
3. To Inculcate Entrepreneurship Skills among Students.
4. To cultivate Moral and Ethical Values in their Profession.

PROGRAMME EDUCATIONAL OBJECTIVES

PEO1: Graduates will be able to Work and Contribute in the domains of Computer Science and Engineering through lifelong learning.

PEO2: Graduates will be able to Analyse, design and development of novel Software Packages, Web Services, System Tools and Components as per needs and specifications.

PEO3: Graduates will be able to demonstrate their ability to adapt to a rapidly changing environment by learning and applying new technologies.

PEO4: Graduates will be able to adopt ethical attitudes, exhibit effective communication skills, Teamworkand leadership qualities.

PROGRAM OUTCOMES (POS)

Engineering Graduates will be able to:

1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOMES (PSO)

PSO1: Ability to Formulate and Simulate Innovative Ideas to provide software solutions for Real-time Problems and to investigate for its future scope.

PSO2: Ability to learn and apply various methodologies for facilitating development of high quality System Software Tools and Efficient Web Design Models with a focus on performance optimization.

PSO3: Ability to inculcate the Knowledge for developing Codes and integrating hardware/software products in the domains of Big Data Analytics, Web Applications and Mobile Apps to create innovative career path and for the socially relevant issues.

COURSE OUTCOMES

CO1	To acquire knowledge about various mobile computing applications, services and architecture
CO2	To Understand about various wireless communication systems and techniques
CO3	To use various routing protocols and acquire knowledge about protocol architectures used in Wireless LAN technology
CO4	To learn about concepts in mobile internet and IP.
CO5	To use key platforms and protocols for mobile application development
CO6	To understand various security issues in mobile computing and new technological trends for next generation cellular networks

MAPPING OF COURSE OUTCOMES WITH PROGRAM OUTCOMES

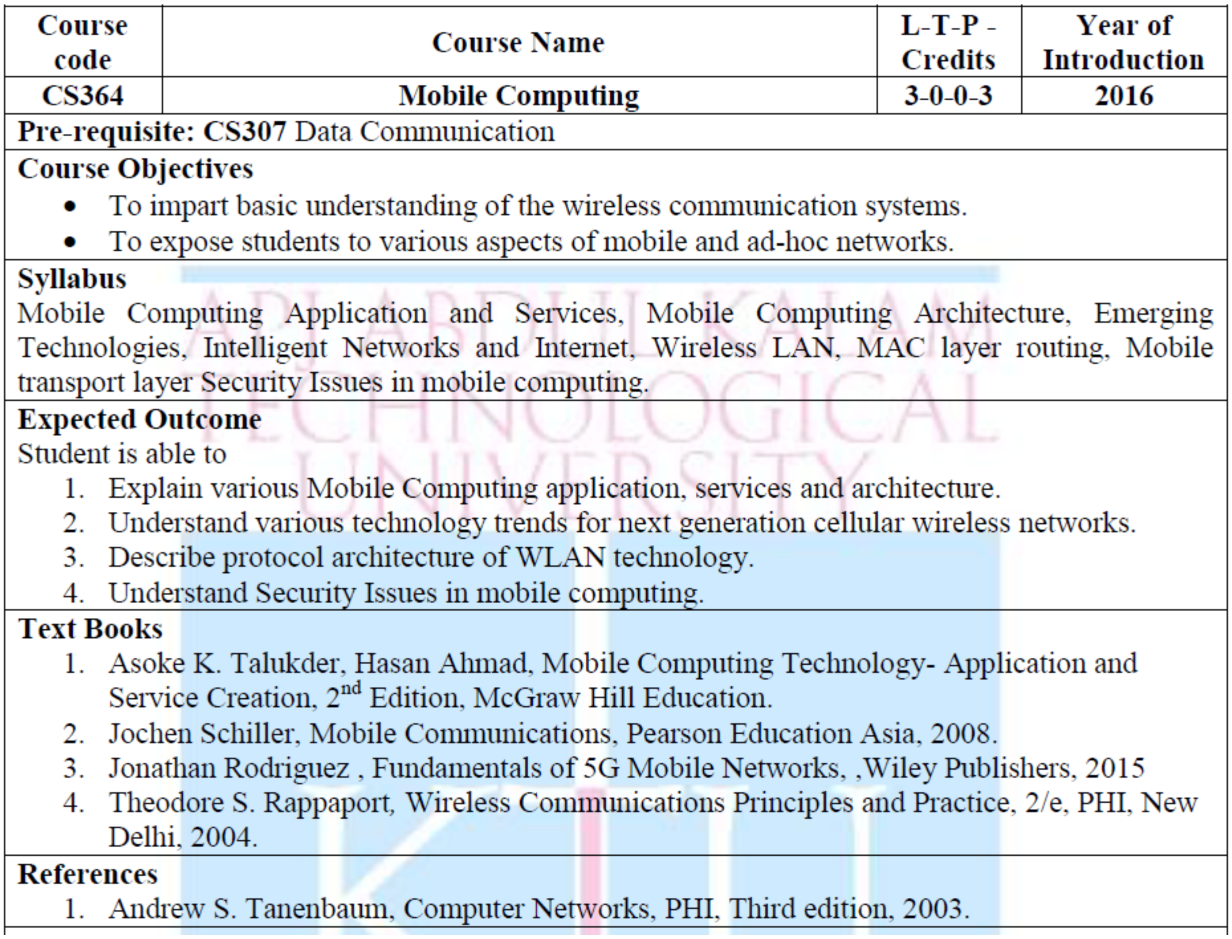
	PO 1	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PO 7**	**PO 8**	**PO 9**	**PO 10**	**PO 11**	**PO 12**
CO1	**3**	**-**	**-**	**-**								
CO2	**3**	**3**	**3**	**3**	**2**							
CO3	**3**	**2**	**-**	**2**	**2**							
CO4	**3**	**2**	**2**	**2**	**2**							
CO5	**3**	**3**	**3**	**3**	**3**							
CO6	**3**	**3**	**3**	**3**	**3**	**3**						

Note: H-Highly correlated=3, M-Medium correlated=2, L-Less correlated=1

MAPPING OF COURSE OUTCOMES WITH PROGRAM SPECIFIC OUTCOMES

	PSO1	PSO2	PSO3
CO1	-		-
CO2	2	-	-
CO3	-	-	-
CO4	-	-	2
CO5	2		3
CO6	2		-

SYLLABUS



Course code	Course Name	L-T-P - Credits	Year of Introduction
CS364	Mobile Computing	3-0-0-3	2016

Pre-requisite: CS307 Data Communication

Course Objectives

- To impart basic understanding of the wireless communication systems.
- To expose students to various aspects of mobile and ad-hoc networks.

Syllabus

Mobile Computing Application and Services, Mobile Computing Architecture, Emerging Technologies, Intelligent Networks and Internet, Wireless LAN, MAC layer routing, Mobile transport layer Security Issues in mobile computing.

Expected Outcome

Student is able to

1. Explain various Mobile Computing application, services and architecture.
2. Understand various technology trends for next generation cellular wireless networks.
3. Describe protocol architecture of WLAN technology.
4. Understand Security Issues in mobile computing.

Text Books

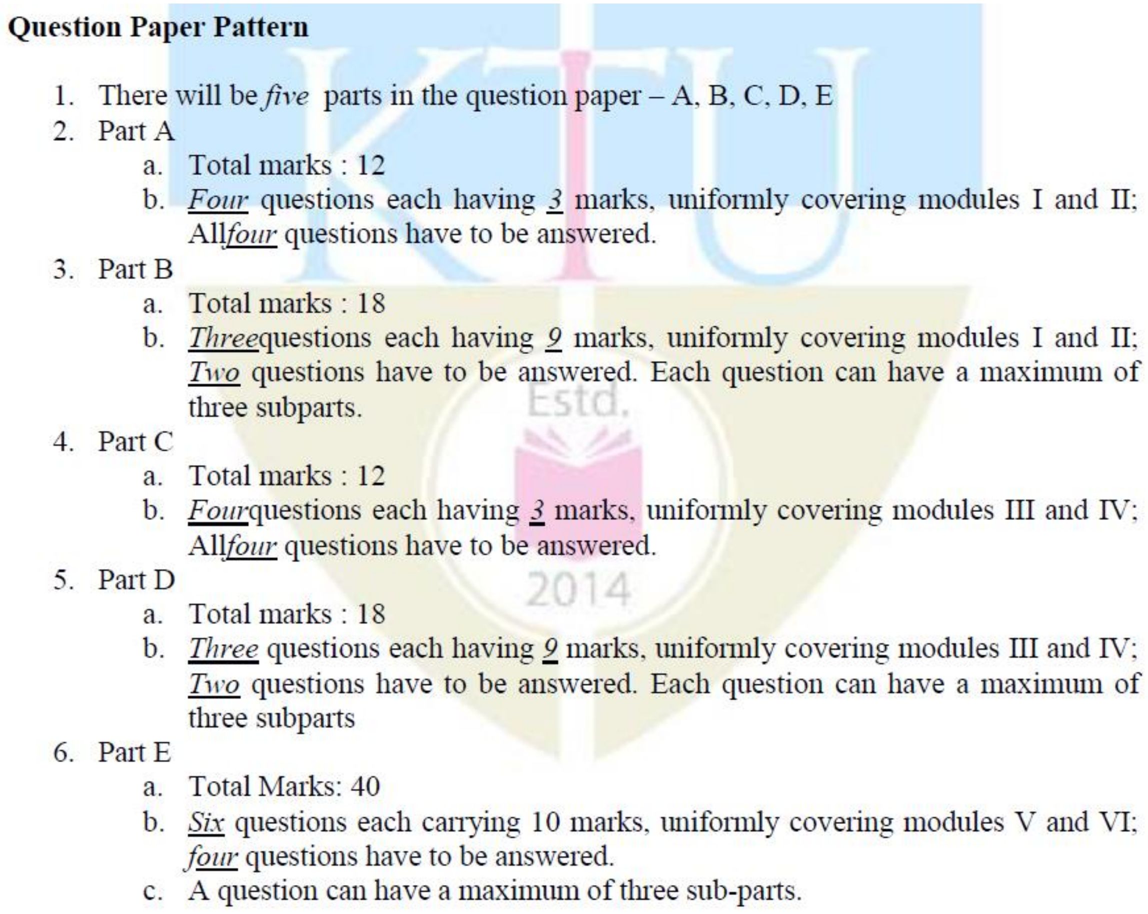
1. Asoke K. Talukder, Hasan Ahmad, Mobile Computing Technology- Application and Service Creation, 2nd Edition, McGraw Hill Education.
2. Jochen Schiller, Mobile Communications, Pearson Education Asia, 2008.
3. Jonathan Rodriguez , Fundamentals of 5G Mobile Networks, ,Wiley Publishers, 2015
4. Theodore S. Rappaport, Wireless Communications Principles and Practice, 2/e, PHI, New Delhi, 2004.

References

1. Andrew S. Tanenbaum, Computer Networks, PHI, Third edition, 2003.

Course Plan

Module	Contents	Hours	End Sem. Exam Marks
I	Introduction to mobile computing, Middleware and Gateways, Application and services, Internet-Ubiquitous networks, Architecture and three-tier architecture for Mobile Computing, Design consideration for Mobile Computing.	06	15%
II	Spread spectrum – Direct sequence, Frequency hoping. Medium Access Control - SDMA, FDMA, TDMA, CDMA, Cellular concepts- channel assignment strategy- hand off strategy interface and system capacity- improving coverage and capacity in cellular system, Satellite Systems-GEO, LEO, MEO. Wireless Communication Systems- Telecommunication Systems- GSM-GSM services & features, architecture -DECT features & characteristics, architecture.	06	15%
FIRST INTERNAL EXAM			
III	Wireless LANS: Wireless LAN Standards – IEEE 802 Protocol Architecture, IEEE 802.11 System Architecture, Protocol Architecture & Services, Cellular Networks: Channel allocation, multiple access, location management, Handoffs. MAC Layer & Management, Routing - Classification of Routing Algorithms, Algorithms such as DSR, AODV, DSDV, Mobile Agents, Service Discovery.	07	15%
IV	Mobile internet-mobile network layer-mobile IP-dynamic host configuration protocol-, mobile transport layer-implications of TCP on mobility-indirect TCP-snooping TCP- mobile TCP transmission-selective retransmission, Transaction oriented TCP- Support for mobility-file systems-WAP.	07	15%
SECOND INTERNAL EXAM			
V	Mobile Transport Layer - Conventional TCP/IP Protocols, Indirect TCP, Snooping TCP, Mobile TCP, Other Transport Layer Protocols for Mobile Networks. Protocols and Platforms for Mobile Computing - WAP, Bluetooth, XML, J2ME, JavaCard, PalmOS, Linux for Mobile Devices, Android.	08	20%
VI	Security issues in mobile computing, Information Security, Components of Information Security, Next Generation Networks-LTE – Architecture & Interface – LTE radio planning and tools, 5G architecture, MIMO, Super core concept, Features and Application Case Study – Setting up anadhoc network system, LiFi.	08	20%
END SEMESTER EXAM			



Question Paper Pattern

1. There will be *five* parts in the question paper – A, B, C, D, E
2. Part A
   a. Total marks : 12
   b. *Four* questions each having *3* marks, uniformly covering modules I and II; All*four* questions have to be answered.
3. Part B
   a. Total marks : 18
   b. *Three*questions each having *9* marks, uniformly covering modules I and II; *Two* questions have to be answered. Each question can have a maximum of three subparts.
4. Part C
   a. Total marks : 12
   b. *Four*questions each having *3* marks, uniformly covering modules III and IV; All*four* questions have to be answered.
5. Part D
   a. Total marks : 18
   b. *Three* questions each having *9* marks, uniformly covering modules III and IV; *Two* questions have to be answered. Each question can have a maximum of three subparts
6. Part E
   a. Total Marks: 40
   b. *Six* questions each carrying 10 marks, uniformly covering modules V and VI; *four* questions have to be answered.
   c. A question can have a maximum of three sub-parts.

QUESTION BANK

MODULE I			
Q:NO:	**QUESTIONS**	**CO**	**KL**
1	Define mobile computing.	CO1	K1
2	Discuss about functions of mobile compile computing	CO1	K2
3	Differentiate between middleware and gateways.	CO1	K4
4	Elaborate about various middleware's used.	CO1	K5
5	Define ubiquitous network with an example	CO1	K2

6	Elaborate about applications and services of mobile computing	CO1	K5
7	Discuss about three tier architecture used in mobile computing.	CO1	K2
8	Explain about application layer in three tier architecture	CO1	K2
9	Write a note on design consideration for mobile computing	CO1	K2
MODULE II			
1	Define spread spectrum.	CO2	K1
2	Discuss about DSSS technology	CO2	K2
3	Differentiate between DSSS and FHSS	CO2	K4
4	Elaborate about frequency hoping spread spectrum.	CO2	K5
5	Write a note on media access control.	CO2	K2
6	Illustrate with a diagram, the GSM Architecture	CO2	K4
7	Elaborate about SDMA & TDMA	CO2	K5
8	Discuss about FDMA.	CO2	K2
9	Describe about various channel assignment strategies.	CO2	K2
10	Elaborate about different handoff techniques	CO2	K5
11	Explain about GSM services & features	CO2	K2
12	Write a note on Satellite systems	CO2	K2
13	Compare GEO, LEO and MEO	CO2	K4
14	Write a note on DECT features & Architecture	CO2	K2
MODULE III			
1	Describe about wireless LAN & various Standards	CO3	K2
2	Write a note on IEEE 802 Architecture	CO3	K2
3	Discuss about IEEE 802.11 system architecture	CO3	K2
4	Differentiate between various wireless LAN standards	CO3	K4
5	Elaborate about IEEE 802.11 protocol architecture.	CO3	K5

6	Write a note on services provided by IEEE 802.11 architecture	CO3	K2
7	Write a note on AODV algorithm.	CO3	K2
8	Elaborate about Channel allocation strategies	CO3	K5
9	Discuss about multiple access & location management.	CO3	K2
10	Describe about various MAC layer &its management.	CO3	K2
11	Elaborate about various routing protocols.	CO3	K5
12	Explain about Mobile agents & Service discovery.	CO3	K2
13	Write a note on DSR algorithm.	CO3	K2
14	Compare AODV and DSDV.	CO3	K4
MODULE IV			
1	Describe about fast transmission and slow start techniques.	CO4	K2
2	Discuss about transaction-oriented TCP	CO4	K2
3	Differentiate between I-TCP & M-TCP	CO4	K4
4	Elaborate about indirect TCP	CO4	K5
5	Define mobile IP & Explain its Components	CO4	K1
6	Write a note on tunneling & Encapsulation	CO4	K2
7	Elaborate about TCP protocol with figures	CO4	K5
8	Discuss about IP packet delivery	CO4	K2
9	Explain about File system in mobile computing	CO4	K2
10	Write a note on implications on TCP mobility	CO4	K2
11	Describe about terms & Entities used in Mobile IP	CO4	K2
12	Explain the Wireless application protocol (WAP)	CO4	K2

MODULE V			
1	Describe about XML programming technique.	CO5	K2
2	Analyze the use of Indirect TCP in mobile devices	CO5	K4
3	Explain about Bluetooth technology used.	CO5	K2
4	Elaborate about J2ME programming technique.	CO5	K5
5	Define Mobile TCP & Explain about it.	CO5	K2
6	Elaborate about Java card with necessary figures.	CO5	K5
7	Discuss about usage of Linux in mobile devices	CO5	K2
8	Explain about Android in mobile devices	CO5	K2
9	Write a note on Palm OS.	CO5	K2
10	Describe about snooping TCP	CO5	K2
11	Elaborate about WAP protocol	CO5	K5
12	Write a note on conventional TCP protocol	CO5	K2
13	Describe about XML programming technique.	CO5	K2
1	Describe about fast transmission and slow start techniques.	CO4	K2
2	Discuss about transaction-oriented TCP	CO4	K2
3	Differentiate between I-TCP & M-TCP	CO4	K4
4	Elaborate about indirect TCP	CO4	K5
5	Define mobile IP & Explain its Components	CO4	K1
6	Write a note on tunneling & Encapsulation	CO4	K2
7	Elaborate about TCP protocol with figures	CO4	K5
8	Discuss about IP packet delivery	CO4	K2
9	Explain about File system in mobile computing	CO4	K2

10	Write a note on implications on TCP mobility	CO4	K2
11	Describe about terms & Entities used in Mobile IP	CO4	K2
12	Explain the Wireless application protocol (WAP)	CO4	K2
MODULE VI			
1	Point out the various security issues in mobile Computing	CO6	K2
2	Discuss about super core concept	CO6	K2
3	Elaborate about features & applications of 5G Technology	CO6	K5
4	Explain about LTE architecture with diagram.	CO6	K2
5	Elaborate about various components of information security.	CO6	K5
6	Elaborate out MIMO concept	CO6	K5
7	Elaborate about LiFi concept used in mobile computing.	CO6	K4
8	Illustrate with diagram, the 5G architecture in mobile computing	CO6	K4
9	Next Generation networks are future of mobile computing. Justify the statement	CO6	K3
10	Elaborate about LTE radio planning	CO6	K5
11	How to implement security measures in mobile computing systems	CO6	K3

APPENDIX 1	
CONTENT BEYOND THE SYLLABUS	
S:NO;	**TOPIC**
1	Current Mobile Technologies

MODULE NOTES

.

.

.

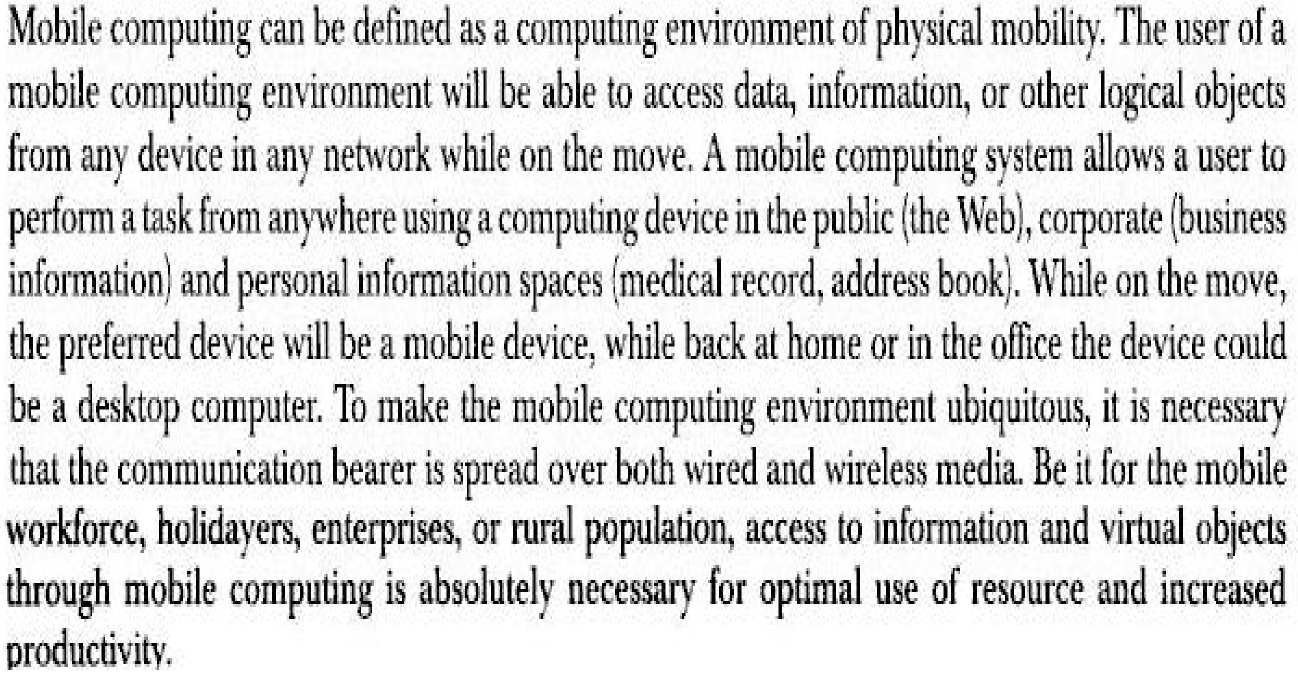
.

.

.

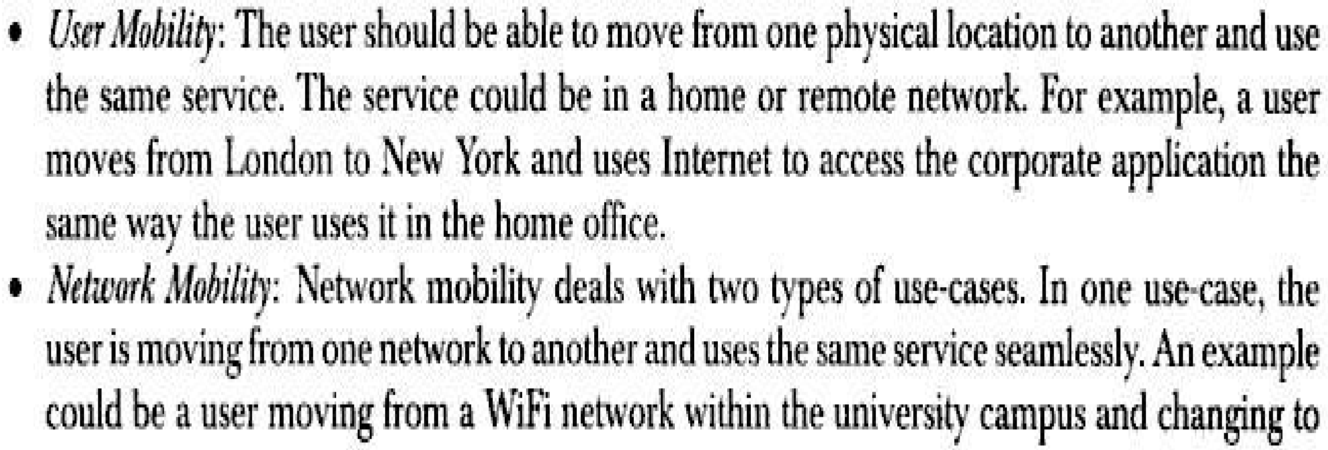
MODULE 1

INTRODUCTION TO MOBILE COMPUTING:

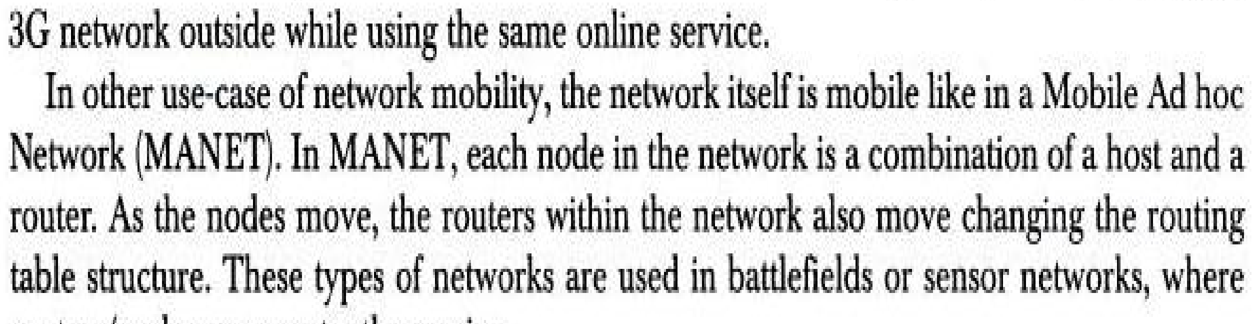


Mobile computing can be defined as a computing environment of physical mobility. The user of a mobile computing environment will be able to access data, information, or other logical objects from any device in any network while on the move. A mobile computing system allows a user to perform a task from anywhere using a computing device in the public (the Web), corporate (business information) and personal information spaces (medical record, address book). While on the move, the preferred device will be a mobile device, while back at home or in the office the device could be a desktop computer. To make the mobile computing environment ubiquitous, it is necessary that the communication bearer is spread over both wired and wireless media. Be it for the mobile workforce, holidayers, enterprises, or rural population, access to information and virtual objects through mobile computing is absolutely necessary for optimal use of resource and increased productivity.

MOBILE COMPUTING FUNCTIONS :



- *User Mobility*: The user should be able to move from one physical location to another and use the same service. The service could be in a home or remote network. For example, a user moves from London to New York and uses Internet to access the corporate application the same way the user uses it in the home office.
- *Network Mobility*: Network mobility deals with two types of use-cases. In one use-case, the user is moving from one network to another and uses the same service seamlessly. An example could be a user moving from a WiFi network within the university campus and changing to



3G network outside while using the same online service.

In other use-case of network mobility, the network itself is mobile like in a Mobile Ad hoc Network (MANET). In MANET, each node in the network is a combination of a host and a router. As the nodes move, the routers within the network also move changing the routing table structure. These types of networks are used in battlefields or sensor networks, where routers/nodes are constantly moving.

- *Bearer Mobility*: The user should be able to move from one bearer to another and use the same service. An example could be a user using a service through WAP bearer in his home network in Bangalore. He moves to Coimbatore where WAP is not supported and switches over to the voice or SMS (short message service) bearer to access the same application.
- *Device Mobility*: The user should be able to move from one device to another and use the same service. An example could be sales representatives using their desktop computer in their home office. During the day while they are on the street they would like to use their Palmtop to access the application.
- *Session Mobility*: A user session should be able to move from one user-agent environment to another. An example could be a user using his service through a CDMA (Code Division Multiple Access) 1X network. The user entered into the basement to park the car and got disconnected from his CDMA network. He goes to his home office and starts using the desktop. The unfinished session in the CDMA device moves from the mobile device to the desktop computer.
- *Agent Mobility*: The user-agent or the applications should be able to move from one node to another. Examples could be aglets, crawler software, or even a malicious worm or virus software that moves from one node to another. There is another use-case of mobile agent in the Cloud Computing paradigm, where applications will be moving from platform to platform and infrastructure to infrastructure depending on temporal and economic considerations. In Cloud Computing, there will not be any fixed association between the application and the host running it—software agents in the cloud will constantly be mobile.
- *Host Mobility*: The user device can be either a client or server. When it is a server or host, some of the complexities change. In case of host mobility, mobility of the IP needs to be taken care of.

The mobile computing functions can be logically divided into the following major segments (Fig. 1.1):

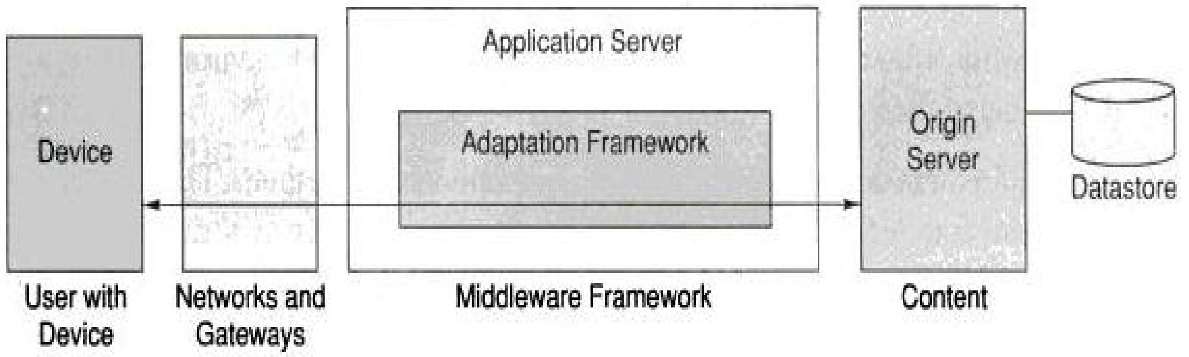
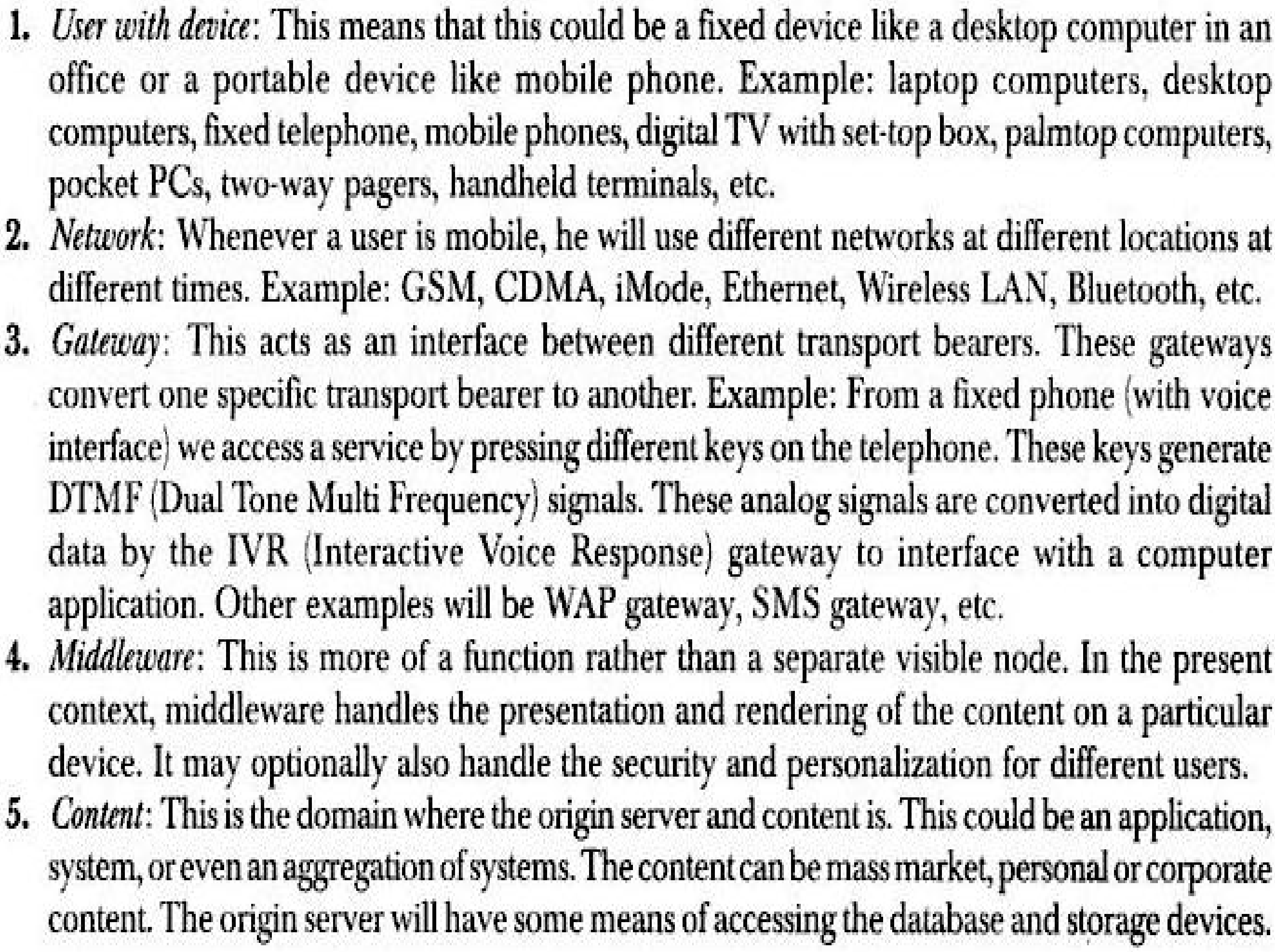
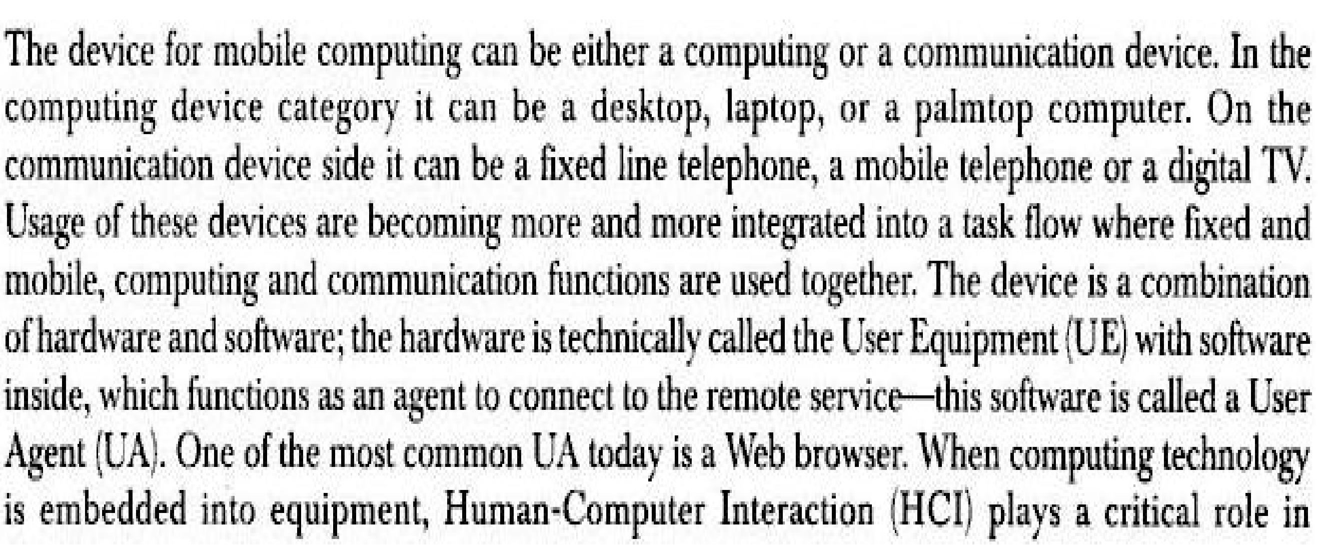


Figure 1.1 Mobile Computing Functions

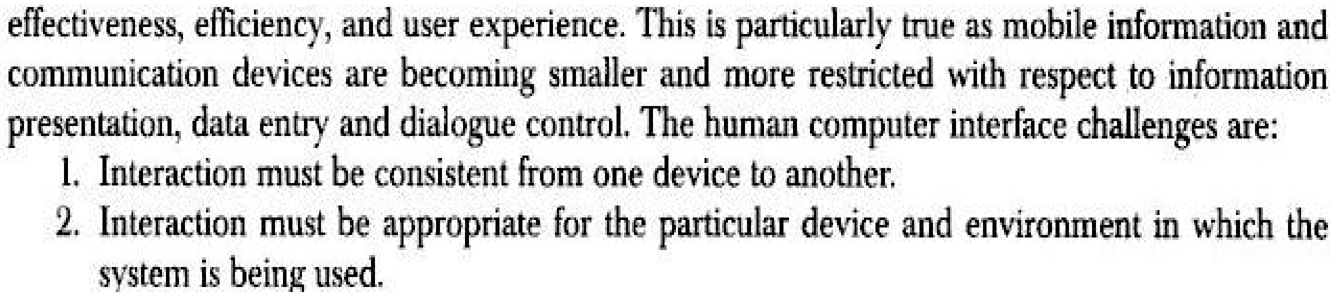


1. *User with device*: This means that this could be a fixed device like a desktop computer in an office or a portable device like mobile phone. Example: laptop computers, desktop computers, fixed telephone, mobile phones, digital TV with set-top box, palmtop computers, pocket PCs, two-way pagers, handheld terminals, etc.
2. *Network*: Whenever a user is mobile, he will use different networks at different locations at different times. Example: GSM, CDMA, iMode, Ethernet, Wireless LAN, Bluetooth, etc.
3. *Gateway*: This acts as an interface between different transport bearers. These gateways convert one specific transport bearer to another. Example: From a fixed phone (with voice interface) we access a service by pressing different keys on the telephone. These keys generate DTMF (Dual Tone Multi Frequency) signals. These analog signals are converted into digital data by the IVR (Interactive Voice Response) gateway to interface with a computer application. Other examples will be WAP gateway, SMS gateway, etc.
4. *Middleware*: This is more of a function rather than a separate visible node. In the present context, middleware handles the presentation and rendering of the content on a particular device. It may optionally also handle the security and personalization for different users.
5. *Content*: This is the domain where the origin server and content is. This could be an application, system, or even an aggregation of systems. The content can be mass market, personal or corporate content. The origin server will have some means of accessing the database and storage devices.

MOBILE COMPUTING DEVICES :



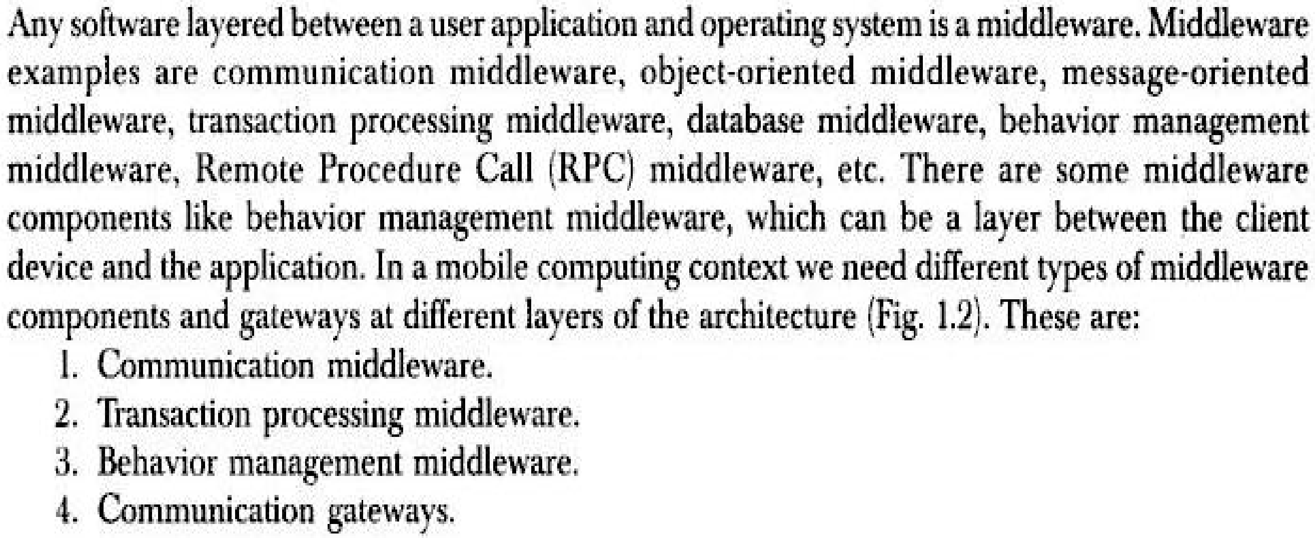
The device for mobile computing can be either a computing or a communication device. In the computing device category it can be a desktop, laptop, or a palmtop computer. On the communication device side it can be a fixed line telephone, a mobile telephone or a digital TV. Usage of these devices are becoming more and more integrated into a task flow where fixed and mobile, computing and communication functions are used together. The device is a combination of hardware and software; the hardware is technically called the User Equipment (UE) with software inside, which functions as an agent to connect to the remote service—this software is called a User Agent (UA). One of the most common UA today is a Web browser. When computing technology is embedded into equipment, Human-Computer Interaction (HCI) plays a critical role in



effectiveness, efficiency, and user experience. This is particularly true as mobile information and communication devices are becoming smaller and more restricted with respect to information presentation, data entry and dialogue control. The human computer interface challenges are:

1. Interaction must be consistent from one device to another.
2. Interaction must be appropriate for the particular device and environment in which the system is being used.

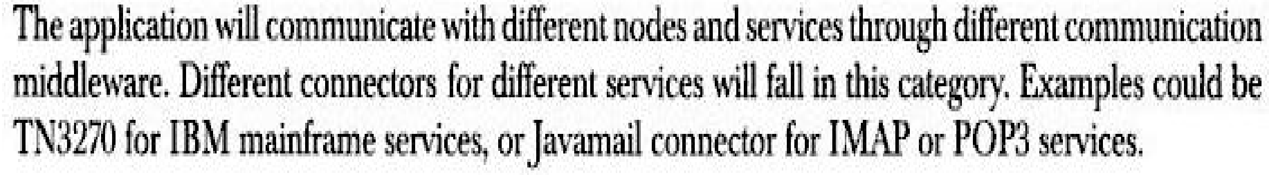
MIDDLEWARE AND GATEWAYS:



Any software layered between a user application and operating system is a middleware. Middleware examples are communication middleware, object-oriented middleware, message-oriented middleware, transaction processing middleware, database middleware, behavior management middleware, Remote Procedure Call (RPC) middleware, etc. There are some middleware components like behavior management middleware, which can be a layer between the client device and the application. In a mobile computing context we need different types of middleware components and gateways at different layers of the architecture (Fig. 1.2). These are:

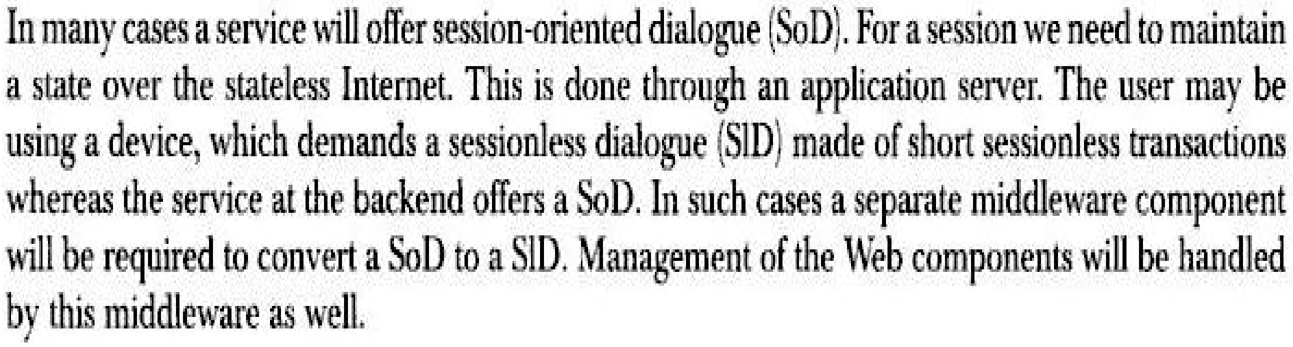
1. Communication middleware.
2. Transaction processing middleware.
3. Behavior management middleware.
4. Communication gateways.

Communication middleware:



The application will communicate with different nodes and services through different communication middleware. Different connectors for different services will fall in this category. Examples could be TN3270 for IBM mainframe services, or Javamail connector for IMAP or POP3 services.

Transaction processing middleware :



In many cases a service will offer session-oriented dialogue (SoD). For a session we need to maintain a state over the stateless Internet. This is done through an application server. The user may be using a device, which demands a sessionless dialogue (SID) made of short sessionless transactions whereas the service at the backend offers a SoD. In such cases a separate middleware component will be required to convert a SoD to a SID. Management of the Web components will be handled by this middleware as well.

Behavior management middleware

Different devices deliver differently. We can have applications which are developed specially to deliver in a certain manner. For example, we can have one application for the Web, another for WAP, and a different one for SMS. On the contrary, we may choose to have a middleware, which will manage device-specific rendering at run-time. This middleware will identify the device properly and handle all device-specific rendering independent of the application. The system may be required to have some context awareness, which will be handled by the behavior management middleware.

Communication gateways:

Between the device and the middleware there will be a system of networks. Gateways are deployed when there are different transport bearers or networks with dissimilar protocols. For example, we need an IVR gateway to interface Voice with a computer, or a WAP gateway to access Internet over a mobile phone.

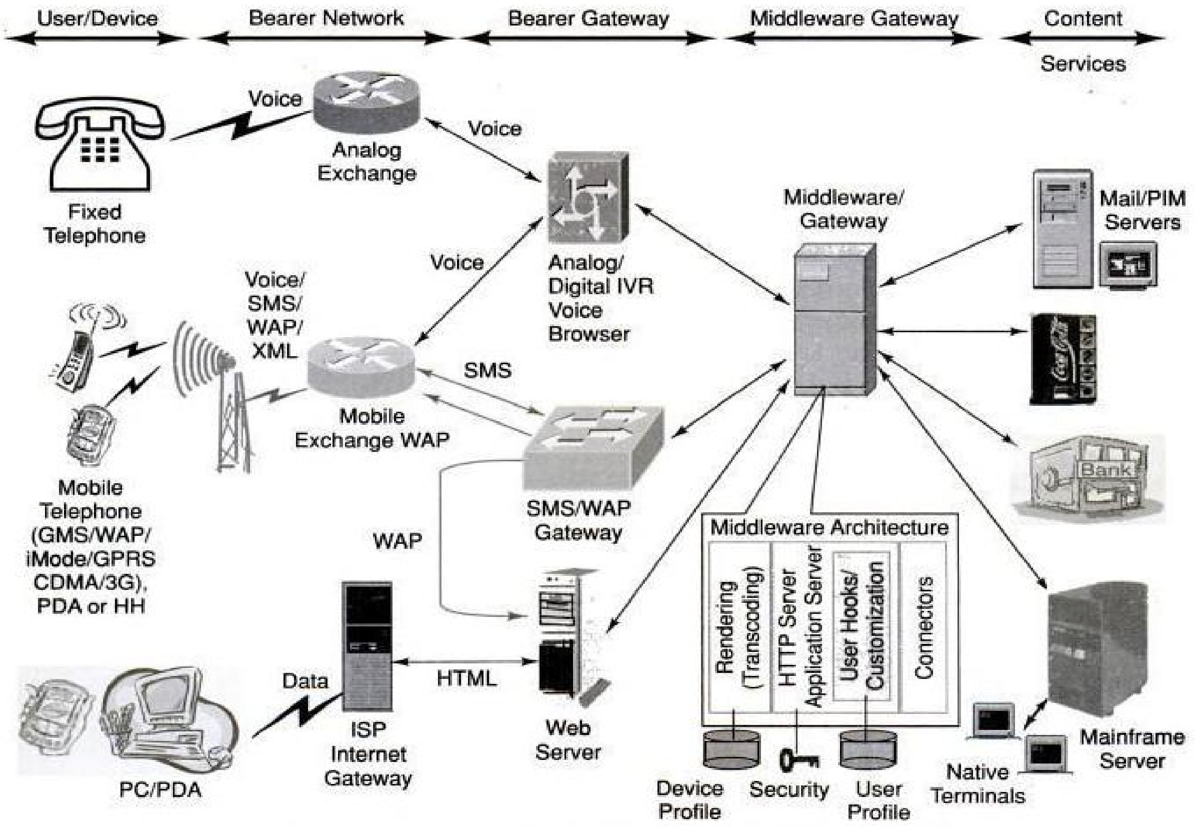
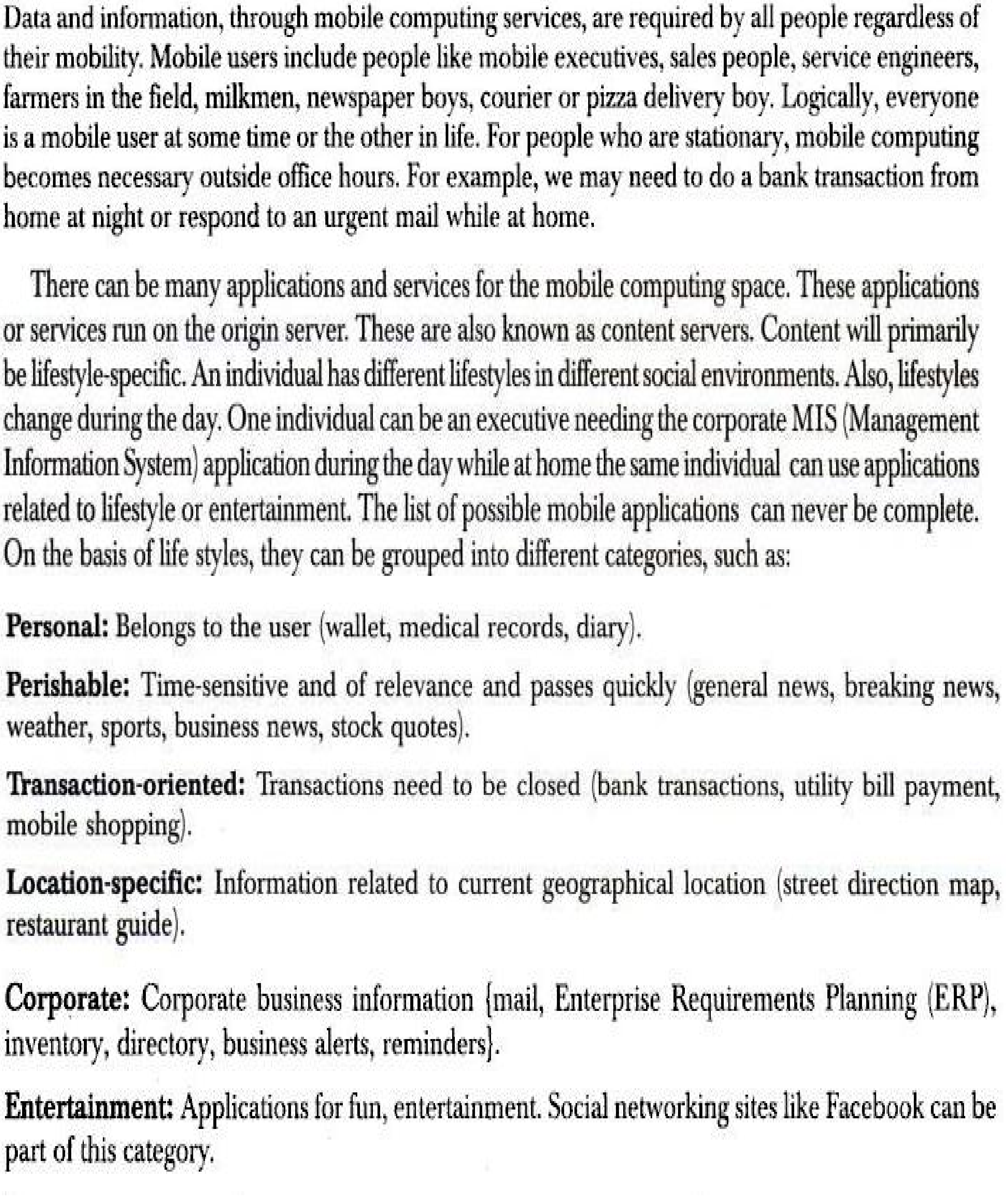


Figure 1.2 Schematic Representation of a Mobile Computing Environment

APPLICATION AND SERVICES :



Data and information, through mobile computing services, are required by all people regardless of their mobility. Mobile users include people like mobile executives, sales people, service engineers, farmers in the field, milkmen, newspaper boys, courier or pizza delivery boy. Logically, everyone is a mobile user at some time or the other in life. For people who are stationary, mobile computing becomes necessary outside office hours. For example, we may need to do a bank transaction from home at night or respond to an urgent mail while at home.

There can be many applications and services for the mobile computing space. These applications or services run on the origin server. These are also known as content servers. Content will primarily be lifestyle-specific. An individual has different lifestyles in different social environments. Also, lifestyles change during the day. One individual can be an executive needing the corporate MIS (Management Information System) application during the day while at home the same individual can use applications related to lifestyle or entertainment. The list of possible mobile applications can never be complete. On the basis of life styles, they can be grouped into different categories, such as:

Personal: Belongs to the user (wallet, medical records, diary).

Perishable: Time-sensitive and of relevance and passes quickly (general news, breaking news, weather, sports, business news, stock quotes).

Transaction-oriented: Transactions need to be closed (bank transactions, utility bill payment, mobile shopping).

Location-specific: Information related to current geographical location (street direction map, restaurant guide).

Corporate: Corporate business information {mail, Enterprise Requirements Planning (ERP), inventory, directory, business alerts, reminders}.

Entertainment: Applications for fun, entertainment. Social networking sites like Facebook can be part of this category.

Here are some examples:

News: This is a very big basket of applications having different types of news. News could be political, current affairs, breaking news, business news, sports news, community news, etc. While people are on the move, they can always be connected to their culture and community through news, using mobile computing.

Youth: This is a very high growth market with different applications to suit the lifestyles of the youth. These are primarily message-based applications like person-to-person messaging, chat, forums, dating, etc.

Weather: There are different types of applications and services where mobile computing can make a difference. Notification services on weather is a very sought after application. If we have access to information related to the weather while on vacation or while driving from one location to another then the global positioning system (GPS) can help locate a person or sometimes save lives in case of a natural calamity.

Corporate application: Standard corporate information is an important piece of information for mobile workers and includes corporate mail, address book, appointments, MIS applications, corporate Intranet, corporate ERP, etc.

Sales force automation: This group will offer many applications. It will cater to the large population of sales personnel. Applications will include sales order bookings, inventory enquiry, shipment tracking, logistics related applications, etc. These applications will be very effective over wireless devices.

m-broker: Getting correct and timely information related to different stocks are very important. Also, online trading of stocks while on the move is quite critical for certain lifestyles. Stock tickers, stock alerts, stock quotes, and stock trading can be made ubiquitous so that users can check their portfolio and play an active role in the market.

Telebanking: We need to access our bank information for different transactions. Earlier, people used to go to the bank, but things are changing. Banks are coming to customers through telebanking. If telebanking can be made ubiquitous it helps the customer as well as the bank. Many banks in India are today offering banking over Internet (web), voice and mobile phones through SMS.

m-shopping: This mobile application is used to shop with the help of mobile devices like Palm top, Pocket PC, mobile phones, etc. You can use this application to pay for a soft drink or soda from a vending machine in an airport or a movie theatre using a mobile phone, especially when you do not have sufficient cash.

Micropayment-based application: Micropayments involve transactions where the amount of money involved is not very high—it could be a maximum of Rs 1000 ($ 25) or so. Micropayment through mobile phones can help rural people to do business effectively.

Interactive games: Many mobile network operators have started offering different types of contests and interactive games that can be played through mobile phones. The applications could be similar to any quiz, housie, etc.

Interactive TV shows: Ma.iy TV companies around the world use email, SMS and Voice as a bearer for interactive TV or reality TV shows. In these shows viewers are encouraged to participate by asking questions, sharing opinions or even answering different quizzes. Nowadays viewers vote for their favorite TV stars using SMS.

Digital/Interactive TV: These are interactive TV programs through digital TV using set-top boxes and Internet. Video-on-demand, community programs, healthcare, and shopping applications are quite popular under this media category.

Experts on call: This is an application system for experts. Experts use these services to schedule their time and business with clients; clients use this to schedule business with the expert. A typical example could be to fix up an appointment with the tax consultant.

GPS-based systems: Applications related to location tracking come under this category. This could be a simple service like tracking a vehicle. Another example could be tracking an individual who got stuck due to bad weather while on a trekking trip. Fleet management companies and locations-aware software systems need GPS-based applications.

Remote monitoring: This is important for children at home where parents monitor where their children are or what are they doing. Also, monitoring and controlling of home appliances will be part of this application.

Entertainment: This contains a very large basket of applications starting from horoscope to jokes. Many people in some parts of Asia decide their day based on the planetary positions and horoscope information.

Directory services: This includes information related to movies, theatre, public telephones, restaurant guide, public information systems and Yellow pages.

Sports: This service offers online sports updates. In India live cricket score is the most popular mobile computing application. Getting scores of a live cricket match is the most popular mobile computer application. This service is available in India through Web, Voice, SMS, and WAP.

Maps/navigation guide: This is an application which has a lot of demand for traveling individuals. These services need to be location-aware to guide the user to use the most optimum path to reach a destination. The directions given by these applications also take traffic congestion, one way, etc., into consideration. GPS-based driving is becoming very popular in the US and advanced countries, where a user enters the postal address of the destination. The GPS-based system calculates the route, loads the right map and helps the driver navigate in real-time.

Virtual office: There are many people who are self-employed and do not have a physical office. Thus mobile and virtual office where they can check their mails, schedules, appointments, etc., while they are on the move are a must for them. Insurance agents and many other professions need these types of services.

m-exchange for industries: Manufacturing industry exchange from a mobile device can be a very cost effective solution for small/cottage industries. It may not be possible for a cottage industry to invest in a computer. However, accessing an exchange for a manufacturing company through a SMS may be affordable.

m-exchange for agricultural produce: Exchange for farmers on different type of agricultural products can be very useful for countries like India. If farmers can get information about where to get a good price for their product, it helps both farmers and consumers. There is a system *www.echoupal.com* to do exactly this. Think of this available over mobile phones.

Applications for speech/hearing challenged people: Telecommunication always meant communicating through Voice. There are people who cannot speak or hear. These include people with disabilities and senior citizens who lost their speech due to old age or after suffering a stroke. Text-based communication can help rehabilitate some of these disabled individuals.

Agricultural information: Think about a case where a farmer receives an alert in his local language through his mobile phone and immediately knows that the moisture content in air is 74%. He can then decide how much to water his harvest. This can save his money, the harvest (excess water is sometimes harmful), and the scarce water resource. Portable devices with voice interface can change the economics of rural India with this kind of application.

Corporate knowledge-based applications: Many corporates today have a knowledge base. Making this ubiquitous can reduce cost and increase productivity.

Community knowledge-based applications: Knowledge is equally important for a community. Making knowledge ubiquitous always help society.

Distance learning: Applications related to distance learning are a must for countries with limited or no access to digital and information technology. For virtual schools in Asia or Africa, it is possible to have access to good faculty through the distance learning mode.

Digital library: These are libraries which can be accessed from anywhere anytime because of the Internet. Digital libraries can go a long way in shortening the digital divide as they also have support of local language and are easy and cheaper to commission.

Telemedicine and healthcare: Making telemedicine and healthcare easily available can save many lives. For example, a person complains of chest pain while traveling and requires immediate medical attention. He has to be taken to a doctor in a remote town. In this case, access to the patient's record can help expedite diagnosis. Reminder services for medicines or checkups can be very useful. In rural India, virtual clinics can help those who otherwise do not have access to medical care.

Micro-credit schemes: Micro-credit has a distinct role to play for a country's microeconomy. Grameen Bank with all its applications in Bangladesh is the best example of micro-credit.

Environmental protection and management: Ubiquity is a must for applications on environmental protection and management. Applications related to industrial hygiene will be part of this category.

e-governance: These applications are very important to bridge the digital divide. The Bhoomi project of Karnataka government has computerized 20 million land records of 0.67 million farmers living in 30,000 villages in the state. Many such projects of the government can be made electronic, resulting in better and faster access to information managed by the government.

Virtual laboratories: There are many laboratories and knowledge repositories around the world which are made accessible to various cultures and countries through digital and information technology.

Community forums: There are different social and community meetings. In the case of India, panchayats can be made electronic. These may help increase the involvement of more people in community development work.

Law enforcements: Most of the time law enforcement staff are on the streets and need access to different types of services through wireless methods. These may be access to criminal records, information related to vehicles, or even a picture of the accident site taken through a MMS phone. This information can help insurance companies to resolve the claim faster.

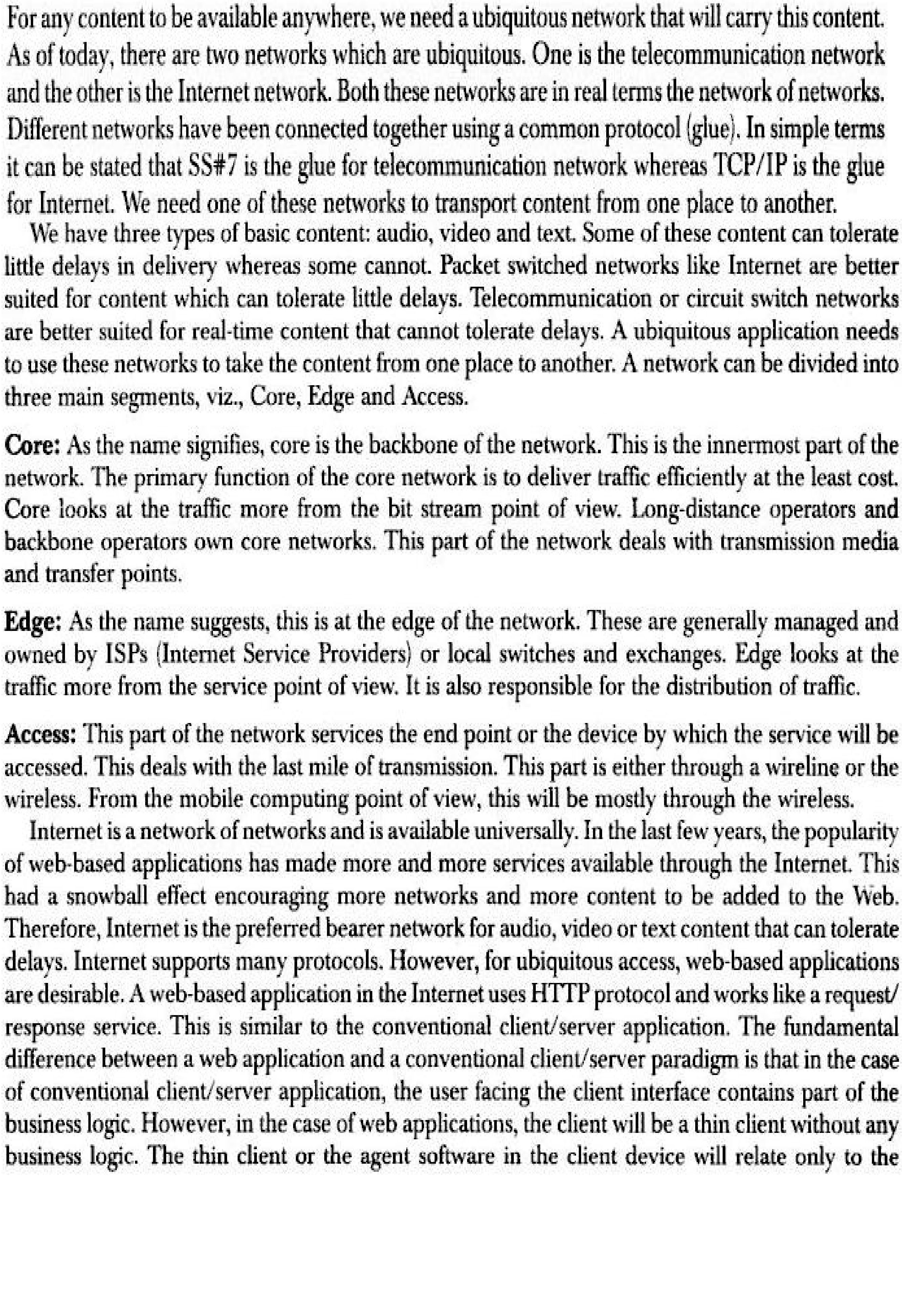
Job facilitator: These could be either proactive alerts or information related to jobs and employment opportunities.

Telemetric applications: Almost every industry and sphere of life has the need for telemetric applications. Examples could be monitoring and control in manufacturing industry; vehicle tracking; meter reading; health care and emergency services; vending machine monitoring; research (telemetric orthodontic); control and service request for different emergency services for utilities like power plants, etc.

Downloads: Different types of downloads starting from ring tones to pictures are part of this category. In many countries this type of application is very popular. It is estimated that the market for ring tone downloads is more than 1 billion dollars.

Alerts and notifications: This can be either business or personal alerts. Simple examples could be breaking news alerts from a newspaper. Complex examples of alert could be for a doctor when the patient is in critical condition. In India many mobile operators are offering cricket alerts. In this service, subscribers receive score information every 15 minutes, about every wicket fall!

INTERNET - UBIQUITOUS NETWORK



For any content to be available anywhere, we need a ubiquitous network that will carry this content. As of today, there are two networks which are ubiquitous. One is the telecommunication network and the other is the Internet network. Both these networks are in real terms the network of networks. Different networks have been connected together using a common protocol (glue). In simple terms it can be stated that SS#7 is the glue for telecommunication network whereas TCP/IP is the glue for Internet. We need one of these networks to transport content from one place to another.

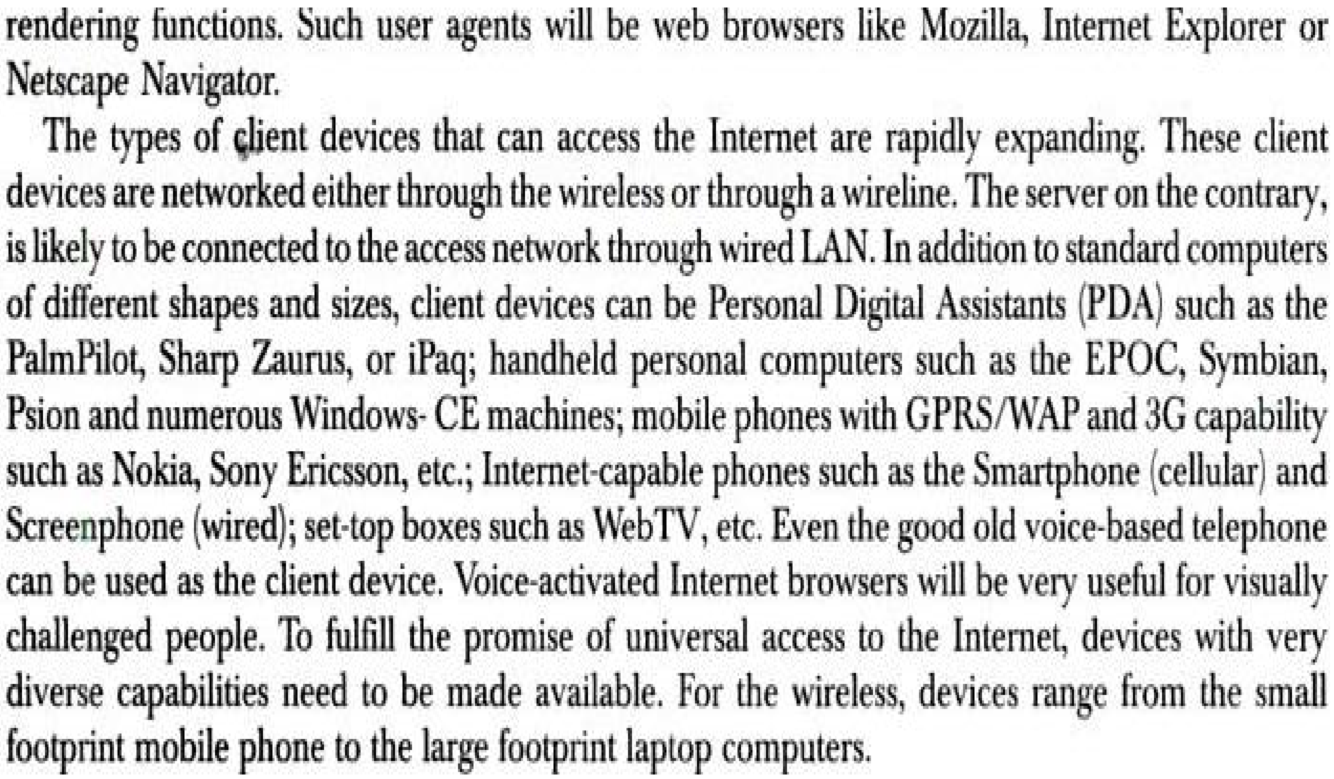
We have three types of basic content: audio, video and text. Some of these content can tolerate little delays in delivery whereas some cannot. Packet switched networks like Internet are better suited for content which can tolerate little delays. Telecommunication or circuit switch networks are better suited for real-time content that cannot tolerate delays. A ubiquitous application needs to use these networks to take the content from one place to another. A network can be divided into three main segments, viz., Core, Edge and Access.

Core: As the name signifies, core is the backbone of the network. This is the innermost part of the network. The primary function of the core network is to deliver traffic efficiently at the least cost. Core looks at the traffic more from the bit stream point of view. Long-distance operators and backbone operators own core networks. This part of the network deals with transmission media and transfer points.

Edge: As the name suggests, this is at the edge of the network. These are generally managed and owned by ISPs (Internet Service Providers) or local switches and exchanges. Edge looks at the traffic more from the service point of view. It is also responsible for the distribution of traffic.

Access: This part of the network services the end point or the device by which the service will be accessed. This deals with the last mile of transmission. This part is either through a wireline or the wireless. From the mobile computing point of view, this will be mostly through the wireless.

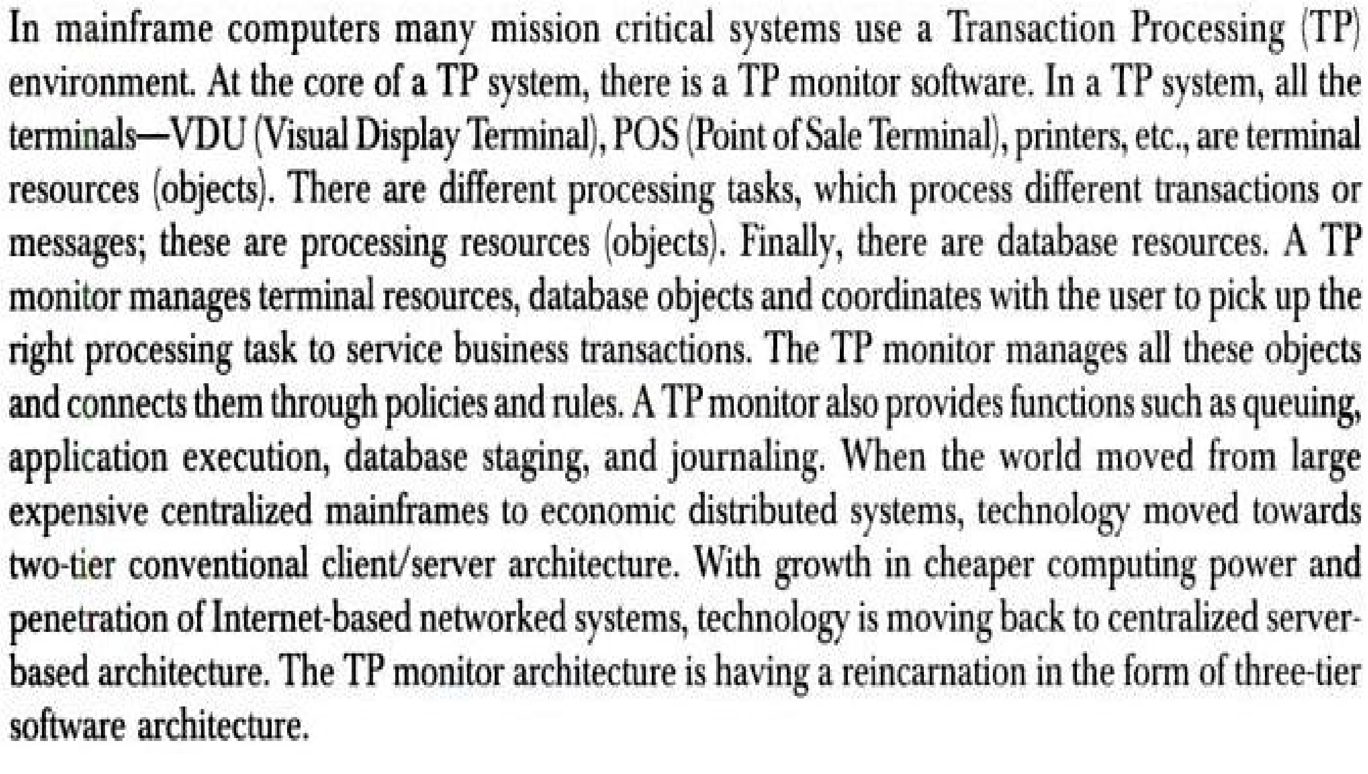
Internet is a network of networks and is available universally. In the last few years, the popularity of web-based applications has made more and more services available through the Internet. This had a snowball effect encouraging more networks and more content to be added to the Web. Therefore, Internet is the preferred bearer network for audio, video or text content that can tolerate delays. Internet supports many protocols. However, for ubiquitous access, web-based applications are desirable. A web-based application in the Internet uses HTTP protocol and works like a request/ response service. This is similar to the conventional client/server application. The fundamental difference between a web application and a conventional client/server paradigm is that in the case of conventional client/server application, the user facing the client interface contains part of the business logic. However, in the case of web applications, the client will be a thin client without any business logic. The thin client or the agent software in the client device will relate only to the



rendering functions. Such user agents will be web browsers like Mozilla, Internet Explorer or Netscape Navigator.

The types of client devices that can access the Internet are rapidly expanding. These client devices are networked either through the wireless or through a wireline. The server on the contrary, is likely to be connected to the access network through wired LAN. In addition to standard computers of different shapes and sizes, client devices can be Personal Digital Assistants (PDA) such as the PalmPilot, Sharp Zaurus, or iPaq; handheld personal computers such as the EPOC, Symbian, Psion and numerous Windows- CE machines; mobile phones with GPRS/WAP and 3G capability such as Nokia, Sony Ericsson, etc.; Internet-capable phones such as the Smartphone (cellular) and Screenphone (wired); set-top boxes such as WebTV, etc. Even the good old voice-based telephone can be used as the client device. Voice-activated Internet browsers will be very useful for visually challenged people. To fulfill the promise of universal access to the Internet, devices with very diverse capabilities need to be made available. For the wireless, devices range from the small footprint mobile phone to the large footprint laptop computers.

ARCHITECTURE AND THREE-TIER ARCHITECTURE FOR MOBILE COMPUTING :



In mainframe computers many mission critical systems use a Transaction Processing (TP) environment. At the core of a TP system, there is a TP monitor software. In a TP system, all the terminals—VDU (Visual Display Terminal), POS (Point of Sale Terminal), printers, etc., are terminal resources (objects). There are different processing tasks, which process different transactions or messages; these are processing resources (objects). Finally, there are database resources. A TP monitor manages terminal resources, database objects and coordinates with the user to pick up the right processing task to service business transactions. The TP monitor manages all these objects and connects them through policies and rules. A TP monitor also provides functions such as queuing, application execution, database staging, and journaling. When the world moved from large expensive centralized mainframes to economic distributed systems, technology moved towards two-tier conventional client/server architecture. With growth in cheaper computing power and penetration of Internet-based networked systems, technology is moving back to centralized server-based architecture. The TP monitor architecture is having a reincarnation in the form of three-tier software architecture.

The network-centric mobile computing architecture uses three-tier architecture as shown in Figure 2.1. In the three-tier architecture, the first layer is the User Interface or Presentation Tier. This layer deals with user facing device handling and rendering. This tier includes a user system interface where user services (such as session, text input, dialog and display management) reside. The second tier is the Process Management or Application Tier. This layer is for application programs or process management where business logic and rules are executed. This layer is capable of accommodating hundreds of users. In addition, the middle process management tier controls transactions and asynchronous queuing to ensure reliable completion of transactions. The third and final tier is the Database Management or Data Tier. This layer is for database access and management. The three-tier architecture is better suited for an effective networked client/server design. It provides increased *performance*, *flexibility*, *maintainability*, *reusability*, and *scalability*, while hiding the complexity of distributed processing from the user. All these characteristics have made three-tier architectures a popular choice for Internet applications and net-centric information systems. Centralized process logic makes administration and change management easier by localizing changes in a central place and using them throughout the system.

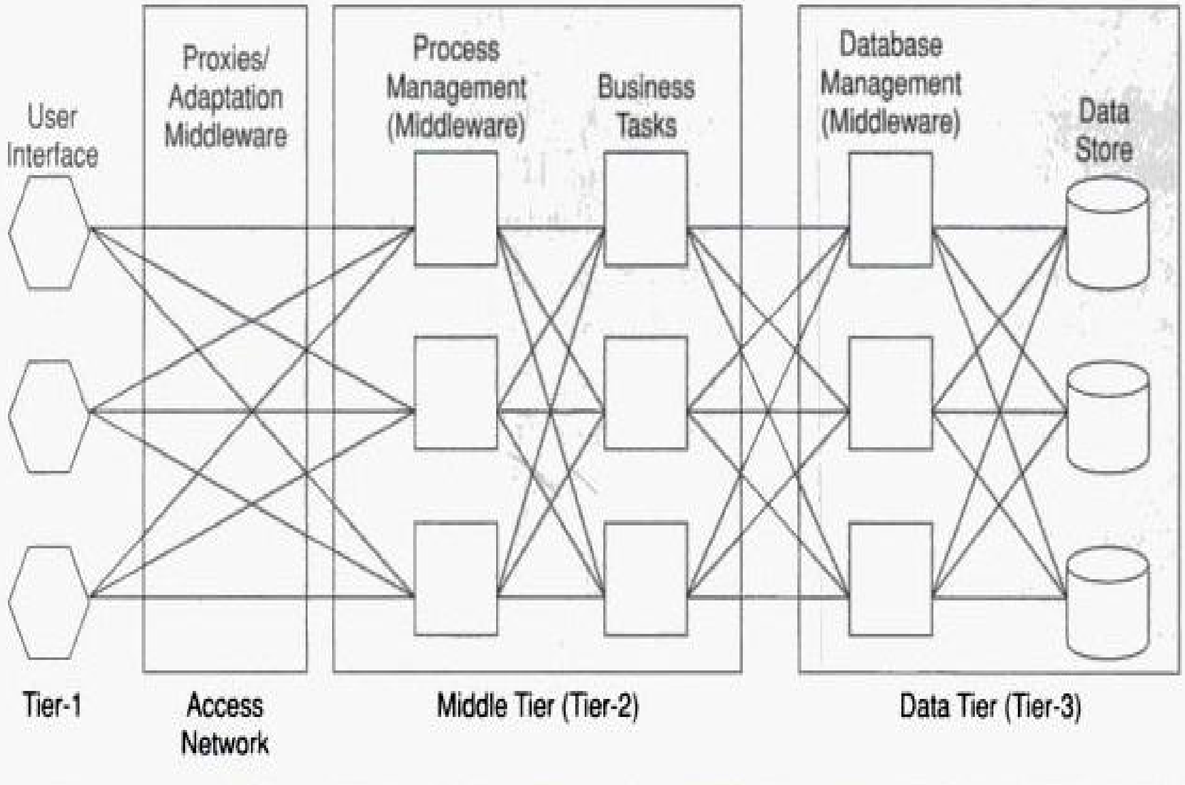


Figure 2.1 Three-tier Architecture for Mobile Computing

THREE-TIER ARCHITECTURE :

To design a system for mobile computing, we need to keep in mind that the system will be used through any network, bearer, agent and device. To have universal access, it is desirable that the server is connected to a ubiquitous network like the Internet. To have access from any device, a web browser is desirable. The reason is simple; web browsers are ubiquitous, they are present in any computer. The browser agent can be Internet Explorer or Netscape Navigator or Mozilla or any other standard agent. Also, the system should preferably be context aware. We will discuss context awareness later.

We have introduced the concept of three-tier architecture. We have also discussed why it is necessary to go for Internet and three-tier architecture for mobile computing. The important question is what a mobile three-tier application actually should consist of. Figure 2.2 depicts a three-tier architecture for a mobile computing environment. These tiers are presentation tier, application tier and data tier. Depending upon the situation, these layers can be further sublayered.

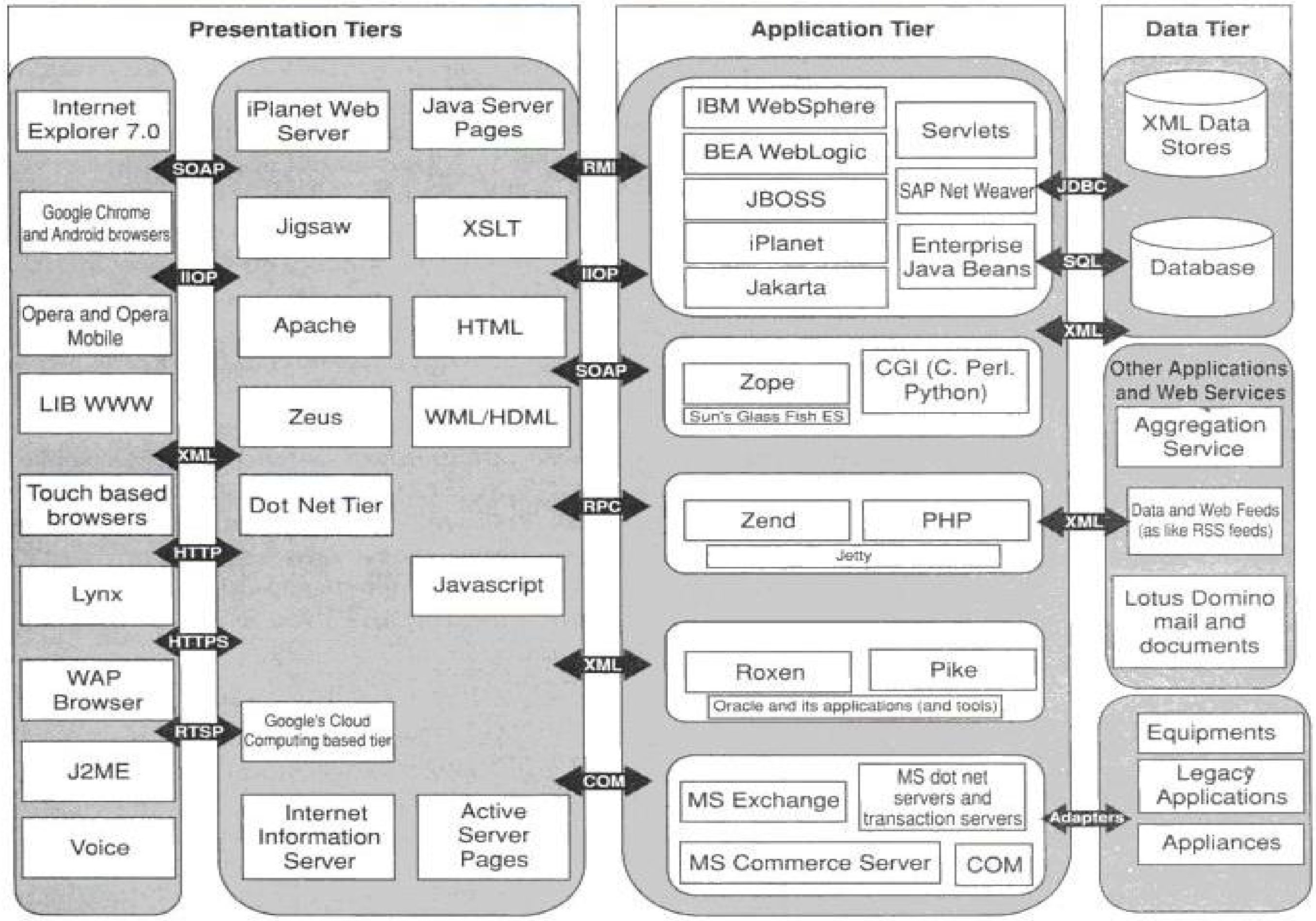
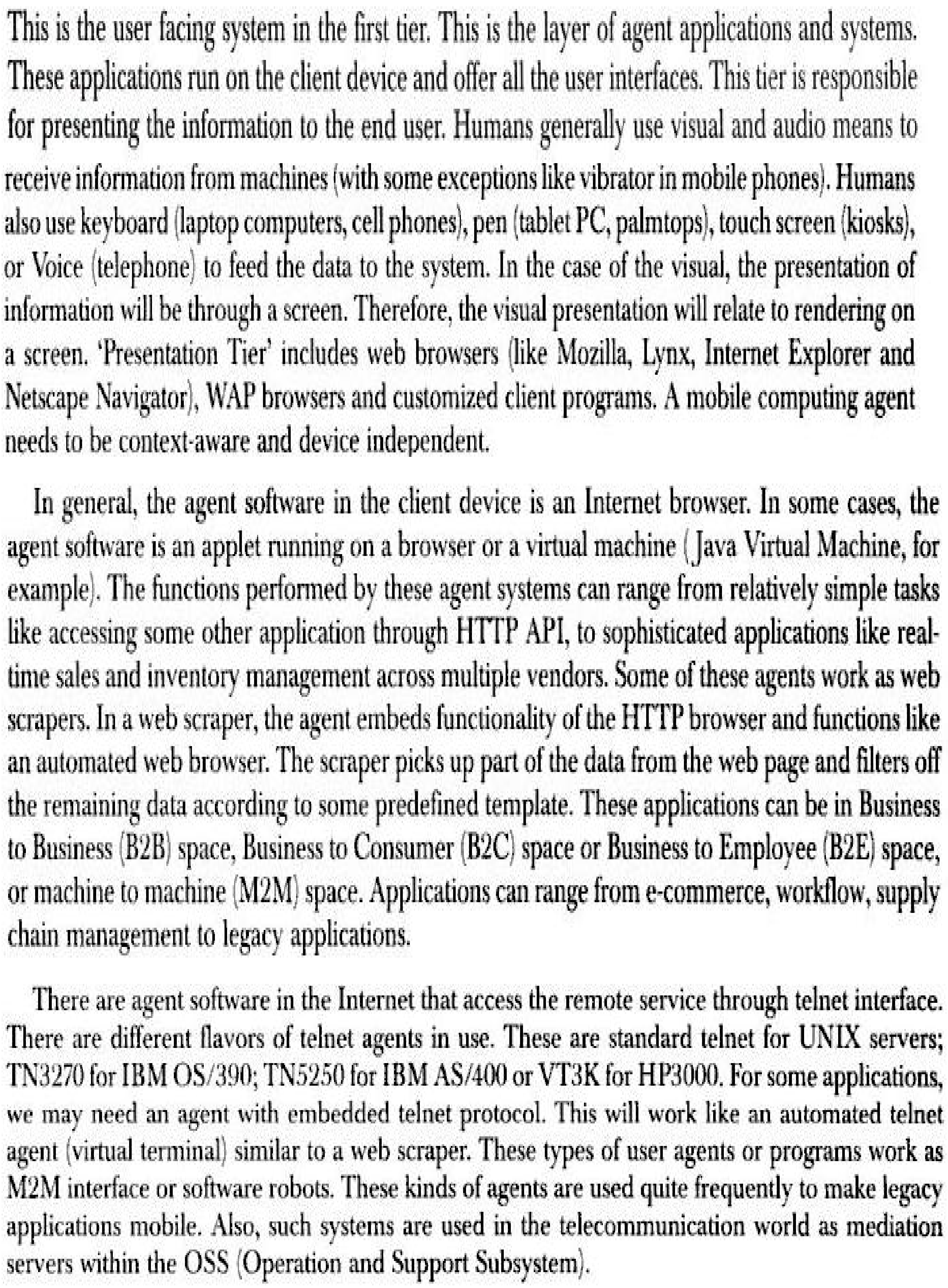


Figure 2.2 The Mobile Computing Architecture

PRESENTATION TIER :

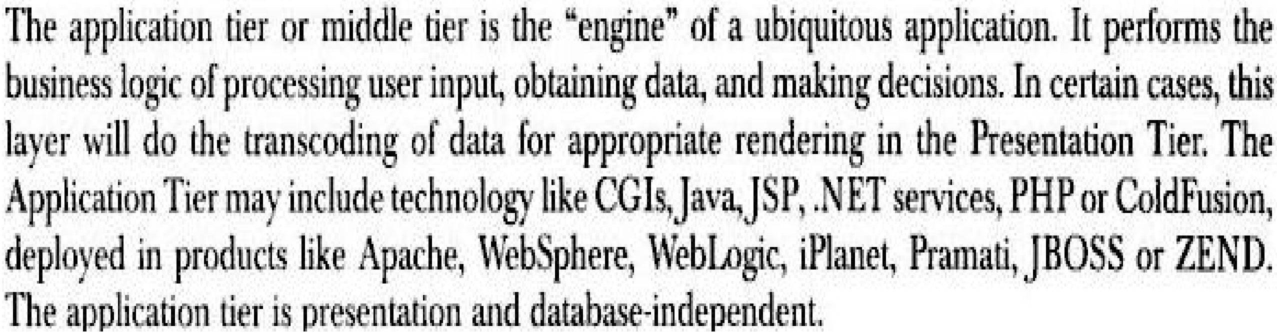


This is the user facing system in the first tier. This is the layer of agent applications and systems. These applications run on the client device and offer all the user interfaces. This tier is responsible for presenting the information to the end user. Humans generally use visual and audio means to receive information from machines (with some exceptions like vibrator in mobile phones). Humans also use keyboard (laptop computers, cell phones), pen (tablet PC, palmtops), touch screen (kiosks), or Voice (telephone) to feed the data to the system. In the case of the visual, the presentation of information will be through a screen. Therefore, the visual presentation will relate to rendering on a screen. 'Presentation Tier' includes web browsers (like Mozilla, Lynx, Internet Explorer and Netscape Navigator), WAP browsers and customized client programs. A mobile computing agent needs to be context-aware and device independent.

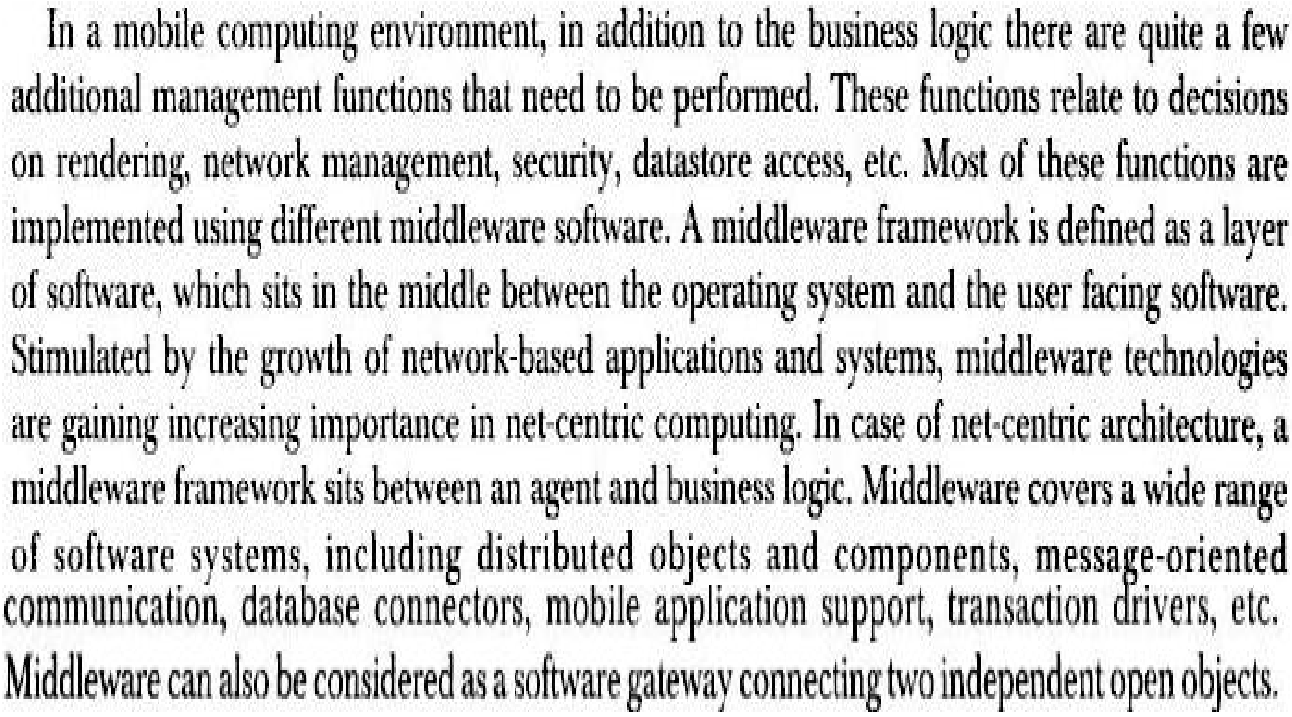
In general, the agent software in the client device is an Internet browser. In some cases, the agent software is an applet running on a browser or a virtual machine (Java Virtual Machine, for example). The functions performed by these agent systems can range from relatively simple tasks like accessing some other application through HTTP API, to sophisticated applications like real-time sales and inventory management across multiple vendors. Some of these agents work as web scrapers. In a web scraper, the agent embeds functionality of the HTTP browser and functions like an automated web browser. The scraper picks up part of the data from the web page and filters off the remaining data according to some predefined template. These applications can be in Business to Business (B2B) space, Business to Consumer (B2C) space or Business to Employee (B2E) space, or machine to machine (M2M) space. Applications can range from e-commerce, workflow, supply chain management to legacy applications.

There are agent software in the Internet that access the remote service through telnet interface. There are different flavors of telnet agents in use. These are standard telnet for UNIX servers; TN3270 for IBM OS/390; TN5250 for IBM AS/400 or VT3K for HP3000. For some applications, we may need an agent with embedded telnet protocol. This will work like an automated telnet agent (virtual terminal) similar to a web scraper. These types of user agents or programs work as M2M interface or software robots. These kinds of agents are used quite frequently to make legacy applications mobile. Also, such systems are used in the telecommunication world as mediation servers within the OSS (Operation and Support Subsystem).

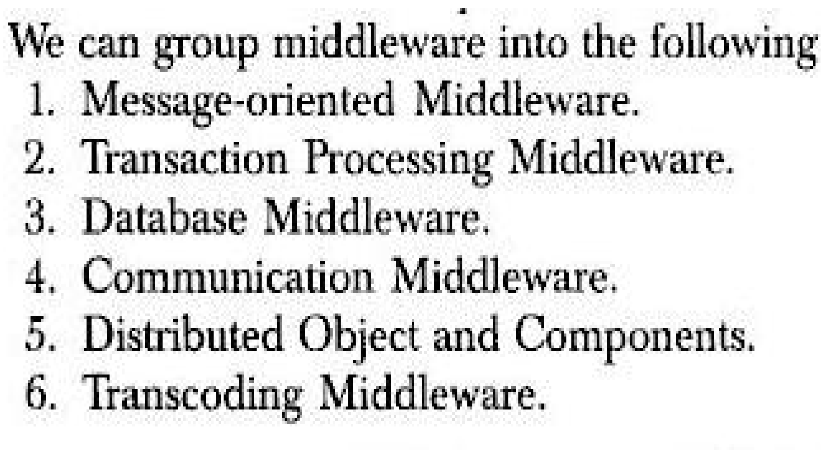
APPLICATION TIER :



The application tier or middle tier is the "engine" of a ubiquitous application. It performs the business logic of processing user input, obtaining data, and making decisions. In certain cases, this layer will do the transcoding of data for appropriate rendering in the Presentation Tier. The Application Tier may include technology like CGIs, Java, JSP, .NET services, PHP or ColdFusion, deployed in products like Apache, WebSphere, WebLogic, iPlanet, Pramati, JBOSS or ZEND. The application tier is presentation and database-independent.



In a mobile computing environment, in addition to the business logic there are quite a few additional management functions that need to be performed. These functions relate to decisions on rendering, network management, security, datastore access, etc. Most of these functions are implemented using different middleware software. A middleware framework is defined as a layer of software, which sits in the middle between the operating system and the user facing software. Stimulated by the growth of network-based applications and systems, middleware technologies are gaining increasing importance in net-centric computing. In case of net-centric architecture, a middleware framework sits between an agent and business logic. Middleware covers a wide range of software systems, including distributed objects and components, message-oriented communication, database connectors, mobile application support, transaction drivers, etc. Middleware can also be considered as a software gateway connecting two independent open objects.



We can group middleware into the following major categories:

1. Message-oriented Middleware.
2. Transaction Processing Middleware.
3. Database Middleware.
4. Communication Middleware.
5. Distributed Object and Components.
6. Transcoding Middleware.

Message-oriented Middleware (MOM)

Message-oriented Middleware is a middleware framework that loosely connects different applications through asynchronous exchange of messages. A MOM works over a networked environment without having to know what platform or processor the other application is resident on. The message can contain formatted data, requests for action, or unsolicited response. The MOM system provides a message queue between any two interoperating applications. If the destination process is out of service or busy, the message is held in a temporary storage location until it can be processed. MOM is generally asynchronous, peer-to-peer, and works in publish/subscribe fashion. In the publish/subscriber mode one or many objects subscribe to an event. As the event occurs, it will be published by the loosely coupled asynchronous object. The MOM will notify the subscribers about this event. However, most implementations of MOM support synchronous (request/response) message passing as well. MOM is most appropriate for event-driven applications. When an event occurs, the publisher application hands on to the messaging middleware application the responsibility of notifying subscribers that the event has happened. In a net-centric environment, MOM can work as the integration platform for different applications. An example of MOM is Message Queue from IBM known as MQ Series. The equivalent from Java is JMS (Java Message Service).

Transaction Processing (TP) Middleware

Transaction Processing Middleware provides tools and an environment for developing transaction-based distributed applications. An ideal TP system will be able to input data into the system at the point of information source and the output of the system is delivered at the point of information sink. In an ideal TP system, the device for input and output can potentially be different (Fig. 2.3). Also, the output can be an unsolicited message for a device. TP is used in data management, network access, security systems, delivery order processing, airline reservations, customer service, etc., to name a few. TP systems are generally capable of providing services to thousands of clients in a distributed client/server environment. CICS (Customer Information Control System) is one of the early TP application systems on IBM mainframe computers.

TP middleware maps numerous client requests through application-service routines to different application tasks. In addition to these processing tasks, TP middleware includes numerous management features, such as restarting failed processes, dynamic load balancing and ensuring consistency of distributed data. TP middleware is independent of the database architecture. TP middleware optimizes the use of resources by multiplexing many client functions on to a much smaller set of application-service routines. This also helps in reducing the response time. TP middleware provides a highly active system that includes services for delivery-order processing, terminal and forms management, data management, network access, authorization, and security. In the Java world and net-centric systems, transaction processing is done through the J2EE application server with the help of entity and session beans.

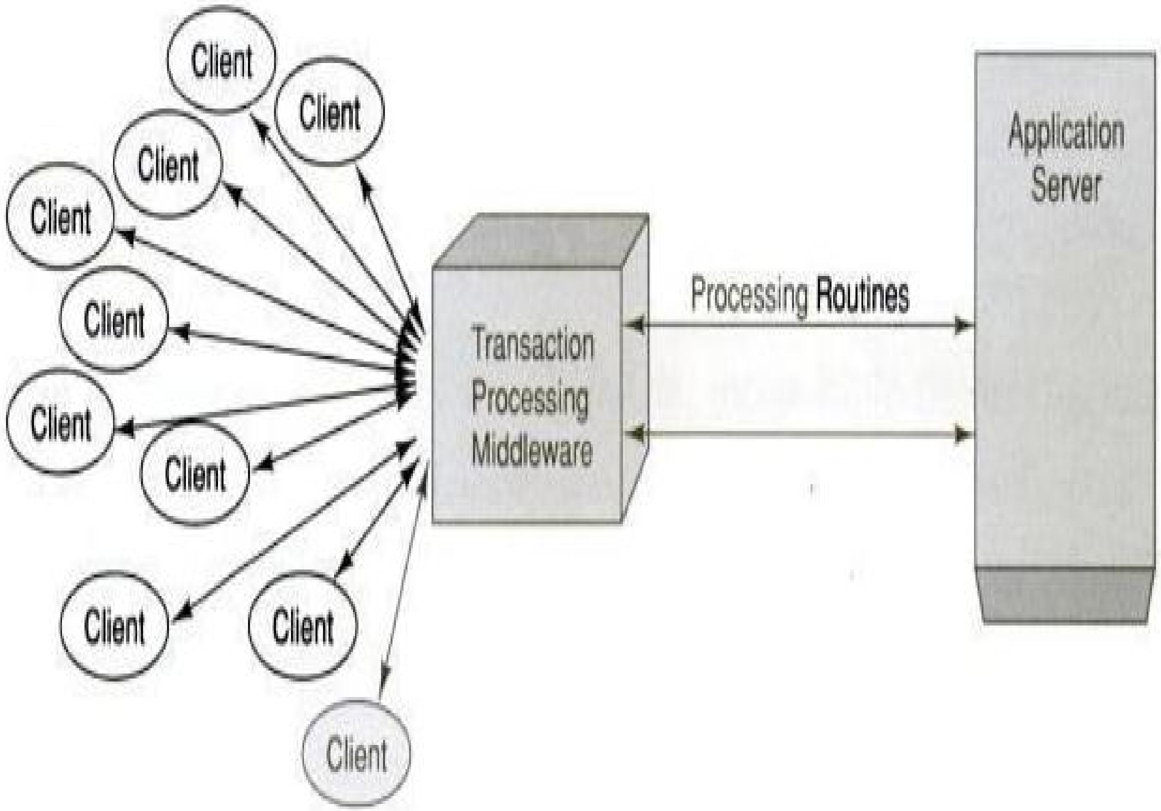


Figure 2.3 Transaction Processing Middleware

Model View Controller (MVC): Java uses the MVC architectural pattern which is an example of transaction processing system. It splits an application into separate layers, viz., presentation, domain logic, and data access. *Model* is the domain-specific representation of the information on which the application operates. Domain logic manipulates and adds meaning to the raw data. MVC does not specifically mention the data access layer because it is assumed to be encapsulated by the model. *View* is responsible for rendering the model into a form suitable for interaction and understood by the user, typically a user interface element. *Controller* manages processes and responds to events, typically user actions, and may invoke changes on the model. In the context of Web applications and J2EE, the MVC pattern is widely used. In Web applications, where the view is the actual HTML page, and the controller is the code which gathers dynamic data and generates the content within the HTML, the model is represented by the actual content, usually stored in a database.

Communication Middleware

Communication Middleware is used to connect one application to another through some communication middleware, like connecting one application to another through telnet. These types of middleware are quite useful in the telecommunication world. There are many elements in the core telecommunication network where the user interface is through telnet. A mediation server automates the telnet protocol to communicate with these nodes in the network. Another example could be to integrate legacy applications through proprietary communication protocols like TN5250 or TN3270.

Distributed Object and Components

An example of distributed objects and components is CORBA (Common Object Request Broker Architecture). CORBA is an open distributed object computing infrastructure being standardized by the Object Management Group (http://www.omg.org). CORBA simplifies many common network programming tasks used in a net-centric application environment. These are object registration, object location, and activation; request demultiplexing; framing and error-handling; parameter marshalling and demarshalling; and operation dispatching. CORBA is vendor-independent infrastructure. A CORBA-based program from any vendor on almost any computer, operating system, programming language and network, can interoperate with a CORBA-based program from the same or another vendor, on almost any other computer, operating system, programming language and network. CORBA is useful in many situations because of the easy way that CORBA integrates machines from so many vendors, with sizes ranging from mainframes through minis and desktops to hand-helds and embedded systems. One of its most important, as well as the most frequent uses is in servers that must handle a large number of clients, at high hit rates, with high reliability.

Transcoding Middleware

Transcoding Middleware is used to transcode one format of data to another to suit the need of the client. For example, if we want to access a web site through a mobile phone supporting WAP, we need to transcode the HTML page to WML page so that the mobile phone can access it. Another example could be accessing a map from a PDA. The same map, which can be shown in a computer, needs to be reduced in size to fit the PDA screen. Technically transcoding is used for content adaptation to fit the need of the device. Content adaptation is also required to meet the network bandwidth needs. For example, some frames in a video clip need to be dropped for a low bandwidth network. Content adaptation used to be done through proprietary protocols. To allow interoperability, IETF has accepted the Internet Content Adaptation Protocol (ICAP). ICAP is now standardized and described in RFC3507.

Internet Content Adaptation Protocol (ICAP)

Popular web servers are required to deliver content to millions of users connected at ever-increasing bandwidths. Progressively, content is being accessed through different devices and agents. A majority of these services have been designed keeping the desktop user in mind. Some of them are also available for other types of protocols. For example, there are a few sites that offer contents in HTML and WML to service desktop and WAP phones. However, the model of centralized services that are responsible for all aspects of every client's request seems to be reaching the end of its useful life. ICAP, the Internet Content Adaptation Protocol, is a protocol aimed at providing simple object-based content vectoring for HTTP services. ICAP is a lightweight protocol to do transcoding on HTTP messages. This is similar to executing a "remote procedure call" on a HTTP request. The protocol allows ICAP clients to pass HTTP messages to ICAP servers for some sort of transformation. The server executes its transformation service on messages and sends back responses to the client, usually with modified messages. The adapted messages may be either HTTP requests or HTTP responses. For example, before a document is displayed in the agent, it is checked for virus.

Web Services

As the need for peer-to-peer, application-to-application communication and interoperability grows, the use of Web services on the Internet will also grow. Web services provide a standard means of communication and information exchange among different software applications, running on a variety of platforms or frameworks. Web service is a software system identified by a URI, whose public interfaces and bindings are defined using XML (eXtensible Markup Language). Its definition can be discovered by other software systems connected to the network. Using XML-based messages these systems may then interact with the Web service in a manner prescribed by its definition.

DATA TIER

The Data Tier is used to store data needed by the application and acts as a repository for both temporary and permanent data. The data can be stored in any form of datastore or database. These can range from sophisticated relational database, legacy hierarchical database, to even simple text files. The data can also be stored in XML format for interoperability with other systems and datasources. A legacy application can also be considered as a data source or a document through a communication middleware.

Database Middleware

We have discussed that for a mobile computing environment, the business logic should be independent of the device capability. Likewise, though not essential, it is advised that business logic should be independent of the database. Database independence helps in maintenance of the system better. Database middleware allows the business logic to be independent and transparent of the database technology and the database vendor. Database middleware runs between the application program and the database. These are sometimes called database connectors as well. Examples of such middleware will be ODBC, JDBC, etc. Using these middleware, the application will be able to access data from any data source. Data sources can be text files, flat files, spreadsheets, or a network, relational, indexed, hierarchical, XML database, object database, etc., from vendors like Oracle, SQL, Sybase, etc.

SyncML

SyncML protocol is an emerging standard for synchronization of data access from different nodes. When we moved from the conventional client/server model of computing to the net-centric model of computing, we moved from distributed computing to centralized computing with networked access. The greatest benefit of this model is that resources are managed at a centralized level. All the popular mobile devices like handheld computers, mobile phones, pagers and laptops work in an occasionally connected computing mode and access these centralized resources from time to time. In an occasionally connected mode, some data are cached in the local device and accessed frequently. The ability to access and update information on the fly is key to the pervasive nature of mobile computing. Examples are emails and personal information like appointments, address book, calendar, diary, etc. Storing and accessing phone numbers of people from the phone address book is more user-friendly compared to accessing the same from a server. However, managing the appointments database is easier in a server, though caching the same on the mobile client is critical. Users will cache emails into the device for reference. We take notes or draft a mail in the mobile device. For workflow applications, data synchronization plays a significant role. The data in the mobile device and server need to be synchronized. Today vendors use proprietary technology for performing data synchronization. SyncML protocol is the emerging standard for synchronization of data across different nodes. SyncML is a new industry initiative to develop and promote a single, common data synchronization protocol that can be used industry-wide.

The ability to use applications and information on a mobile device, then to synchronize any updates with the applications and information back at the office or on the network, is key to the utility and popularity of mobile computing. The SyncML protocol supports naming and identification of records and common protocol commands to synchronize local and network data. It supports identification and resolution of synchronization conflicts. The protocol works over all networks used by mobile devices, both wireless and wireline. Since wireless networks employ different transport protocols and media, a SyncML will work smoothly and efficiently over:

- HTTP 1.1 (i.e., the Internet).
- WSP (the Wireless Session Protocol, part of the WAP protocol suite).
- OBEX (Object Exchange Protocol, i.e., Bluetooth, IrDA and other local connectivity).
- SMTP, POP3 and IMAP.
- Pure TCP/IP networks.
- Proprietary wireless communication protocols.

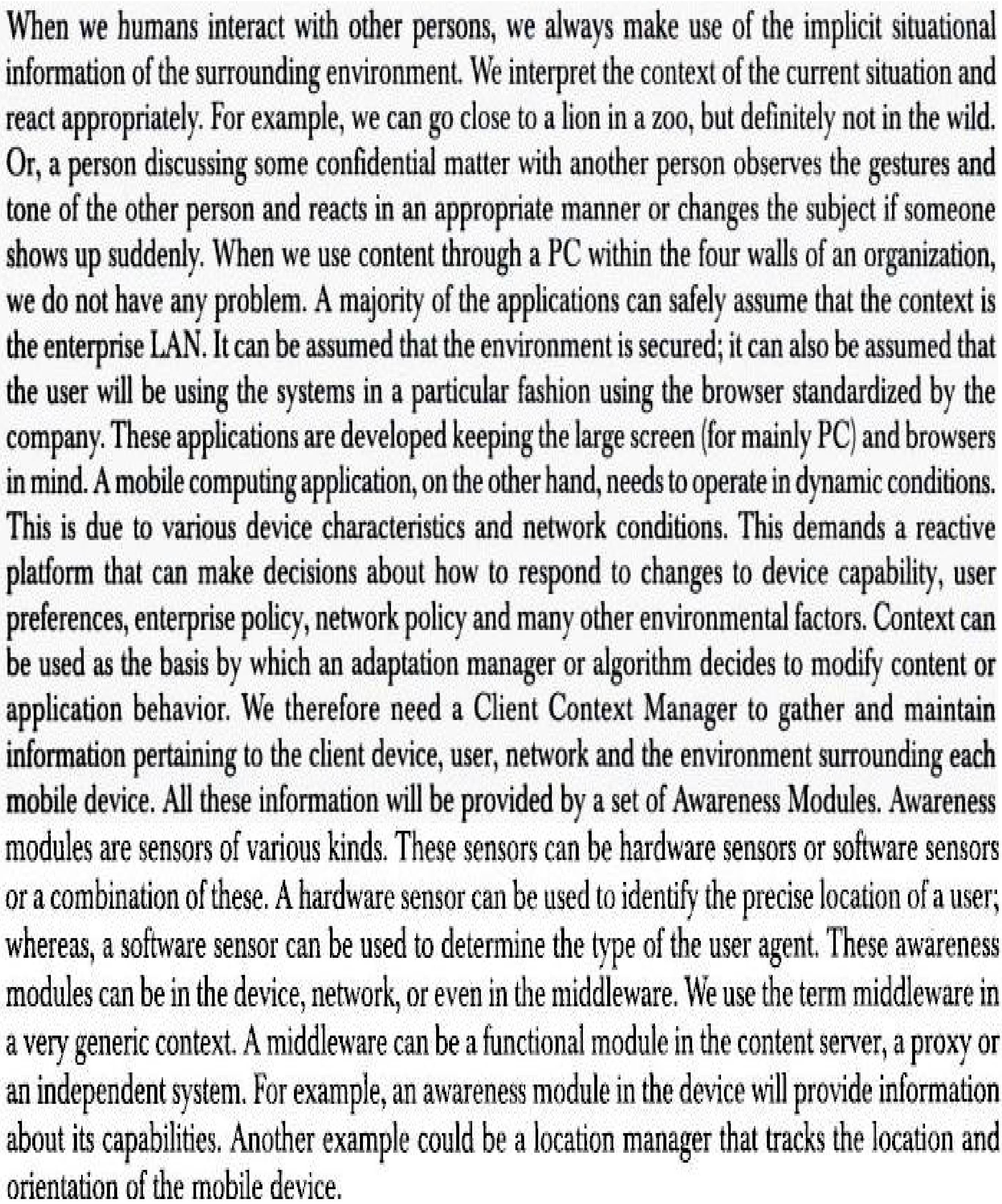
DESIGN CONSIDERATION FOR MOBILE COMPUTING:

The mobile computing environment needs to be context-independent as well as context-sensitive. Context information is the information related to the surrounding environment of an actor in that environment. The term "context" means, all the information that helps determine the state of an object (or actor). This object can be a person, a device, a place, a physical or computational object, the surrounding environment or any other entity being tracked by the system. In a mobile computing environment, context data is captured so that decisions can be made about how to adapt content or behavior to suit this context. Mobility implies that attributes associated with devices and users will change constantly. These changes mean that content and behavior of applications should be adapted to suit the current situation. There are many ways in which content and behavior can be adapted. Following are some examples:

1. **Content with context awareness:** Build each application with context awareness. There are different services for different client context (devices). For example, a bank decides to offer mobile banking application through Internet, PDA and mobile phone using WAP. These services are different and are http://www.mybank.com/inet.html, http://www.mybank.com/palm.html and http://www.mybank.com/wap.wml, respectively. The service http://www.mybank.com/inet.html assumes that the user will use computers to access this service. Therefore it is safe to offer big pages with text boxes and drop down menus. Also, it is fine to add a few animated pictures for the new product the bank is launching. We know that http://www.mybank.com/palm.html is a service for a PalmOS PDA. As the display size is small, we design the screen to be compact for the PDA and do not offer the same product animation. For the WAP service at http://www.mybank.com/wap.wml, we do a completely different user interface; we make all drop down options available through the option button in the mobile phone and remove all the graphics and animations.
2. **Content switch on context:** Another way is to provide intelligence for the adaptation of content within the service. This adaptation happens transparent to the client. In this case the service is the same for Internet, PDA and WAP. All access the bank's service through http://www.mybank.com/. An intelligent piece of code identifies the agent to decide what type of device or context it is. This intelligent code does the adaptation at runtime based upon the agent in hand. The simplest way to do this is to look at the user-agent value at the HTTP header and decide whether to route the request to http://mybank.com/inet.html or http://www.mybank.com/palm.html or http://www.mybank.com/wap.wml.
3. **Content transcoding on context:** Another way is to provide an underlying middleware platform that performs the adaptation of the content based on the context and behavior of the device. This adaptation happens transparent to the client and the application. The middleware platform is intelligent enough to identify the context either from the HTTP or additional customized parameters. In this case the service may be in html or XML, the middleware platform transcodes the code from html (or XML) to html, and wml on the fly. It can also do the transcoding based on policy so that the html generated for a computer is different from a PDA.

Following sections describe different types of context that can be enhance the usability ,reliability and security of the service :

CLIENT CONTEXT MANAGER



When we humans interact with other persons, we always make use of the implicit situational information of the surrounding environment. We interpret the context of the current situation and react appropriately. For example, we can go close to a lion in a zoo, but definitely not in the wild. Or, a person discussing some confidential matter with another person observes the gestures and tone of the other person and reacts in an appropriate manner or changes the subject if someone shows up suddenly. When we use content through a PC within the four walls of an organization, we do not have any problem. A majority of the applications can safely assume that the context is the enterprise LAN. It can be assumed that the environment is secured; it can also be assumed that the user will be using the systems in a particular fashion using the browser standardized by the company. These applications are developed keeping the large screen (for mainly PC) and browsers in mind. A mobile computing application, on the other hand, needs to operate in dynamic conditions. This is due to various device characteristics and network conditions. This demands a reactive platform that can make decisions about how to respond to changes to device capability, user preferences, enterprise policy, network policy and many other environmental factors. Context can be used as the basis by which an adaptation manager or algorithm decides to modify content or application behavior. We therefore need a Client Context Manager to gather and maintain information pertaining to the client device, user, network and the environment surrounding each mobile device. All these information will be provided by a set of Awareness Modules. Awareness modules are sensors of various kinds. These sensors can be hardware sensors or software sensors or a combination of these. A hardware sensor can be used to identify the precise location of a user; whereas, a software sensor can be used to determine the type of the user agent. These awareness modules can be in the device, network, or even in the middleware. We use the term middleware in a very generic context. A middleware can be a functional module in the content server, a proxy or an independent system. For example, an awareness module in the device will provide information about its capabilities. Another example could be a location manager that tracks the location and orientation of the mobile device.

Almost any information available at the time of an interaction can be seen as context information. Some examples are:

1. **Identity:** The device will be in a position to communicate its identity without any ambiguity.
2. **Spatial information:** Information related to the surrounding space. This relates to location, orientation, speed, elevation and acceleration.
3. **Temporal information:** Information related to time. This will be time of the day, date, time zone and season of the year.
4. **Environmental information:** This is related to the environmental surroundings. This will include temperature, air quality, moisture, wind speed, natural light or noise level. This also includes information related to the network and network capabilities.
5. **Social situation:** Information related to the social environment. This will include who you are with, and people that are nearby; whether the user is in a meeting or in a party.
6. **Resources that are nearby:** This will relate to the other accessible resources in the nearby surroundings like accessible devices, hosts or other information sinks.
7. **Availability of resources:** This will relate to information about the device in use. This will include battery power, processing power, persistence store, display, capabilities related to I/O (input/output) and bandwidth.
8. **Physiological measurements:** This relates to the physiological state of the user. This includes information like blood pressure, heart rate, respiration rate, muscle activity and tone of voice.
9. **Activity:** This relates to the activity state of the user. This includes information like talking, reading, walking and running.
10. **Schedules and agendas:** This relates to the schedules and agendas of the user.

A system is context-aware if it can extract, interpret and use context-related information to adapt its functionality to the current context. The challenge for such systems lies in the complexity of capturing, representing, filtering and interpreting contextual data. To capture context information generally some sensors are required. This context information needs to be represented in a machine-understandable format, so that applications can use this information. In addition to being able to obtain the context-information, applications must include some 'intelligence' to process the information and deduce the meaning. These requirements lead us to three aspects of context management:

1. **Context sensing:** The way in which context data is obtained.
2. **Context representation:** The way in which context information is stored and transported.
3. **Context interpretation:** The way in which meaning is obtained from the context representation.

W3C has proposed a standard for context information. This standard is called Composite Capabilities/Preference Profiles (CC/PP), for describing device capabilities and user preferences. All these context information are collated and made available to the management components.

Composite Capabilities/Preference Profiles (CC/PP)

Composite Capabilities/Preference Profiles (CC/PP) is a proposed W3C standard for describing device capabilities and user preferences. Special attention has been paid to wireless devices such as mobile phones and PDAs. In practice, the CC/PP model is based on RDF (resource description framework) and can be serialized using XML.

A CC/PP profile contains a number of attribute names and associated values that are used by an application to determine the appropriate form of a resource to deliver to a client. This is to help a client or proxy/middleware to describe their capabilities to an origin server or other sender of resource data. It is anticipated that different applications will use different vocabularies to specify application-specific properties within the scope of CC/PP. However, for different applications to interoperate, some common vocabulary is needed. The CC/PP standard defines all these.

CC/PP is designed in such a way that an origin server or proxy can perform some sort of content to device matching. CC/PP is designed to suit an adaptation algorithm. The sequence of steps in the general case would look something like the following (Fig. 2.6):

1. Device sends serialized profile model with request for content.
2. Origin server receives serialized RDF profile and converts it into an in-memory model.
3. The profile for the requested document is retrieved and an in-memory model is created.
4. The device profile model is matched against the document profile model.
5. A suitable representation of the document is chosen. At this stage the document to be returned can be chosen from a number of different versions of the same document (content switch on context) or it can be dynamically generated (content transcoding on context).
6. Document is returned to device and presented.

If a document or application is specific about how it should be displayed, or if there are several versions of the document or application for different devices, then the adaptation manager can ask the client context manager for detailed context information. The client context manager will enquire with the relevant awareness module and extract the necessary context information. This fine-grained approach allows a high level of adaptation to take place. In cases where the document does not provide profile information, or the profile is limited in description, the adaptation manager can obtain a general context class from the context manager and perform some limited adaptation. For example, some adaptation can still take place where the location of the user is important. The policy manager can specify some rules about how adaptation should take place when a user is at a certain location, regardless of the information provided in an application or document profile.

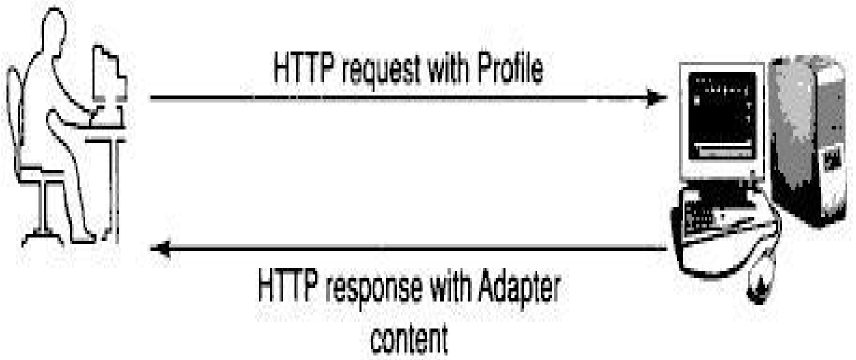


Figure 2.6 The Simplest Use of CC/PP

Policy Manager

The policy manager is responsible for controlling policies related to mobility. A policy is a set of rules; these rules need to be followed under different conditions. Introduction of mobility within an enterprise brings with it different types of challenges that are not normally seen in traditional computing environments. When we consider mobility, it is assumed that the data or information will be visible from outside the four walls of the enterprise. Organizations generally have policies regarding the disclosure of information. For example, documents from certain systems can be printed only on certain printers in the organization. Some hard copy documents may be viewed only at the office of the CEO. These kinds of policies must be transferable to a mobile computing environment. Mobile computing policy manager will be able to define policy for documents/services and assign roles to users. Each role will have permissions, prohibitions and obligations associated with it. Each policy will have access rights associated with respect to read, write, execute. A policy in combination with role and current context information will be able to determine what actions a user is allowed to perform, or what actions a user is obligated to perform.

Semantic Web

As mentioned earlier, policies are sets of rules. When we drive in the street we are expected to follow the right of way. In a party there are some etiquettes to be followed. We humans learn these rules, policies, laws, and etiquettes from documents or experienced people. This is to help us to behave correctly in the society. The question is how to make a machine understand policies and make them behave in the expected fashion? Data in the Web is generally hidden away in HTML files, how do we determine which content is useful in some contexts, but often not in others. Facilities to put machine understandable data on the Web are becoming a necessity. The Semantic Web is targeted to address this need. The idea is of having data on the Web defined and linked in a way that it can be used by machines not just for display, but for automation, security, filtering, integration and reuse of data across various applications.

Semantic Web technologies are still very much in their infancy. It is believed that a large number of Semantic Web applications can be used for a variety of different tasks, increasing the modularity of applications on the Web. The Semantic Web is generally built on syntaxes which use URIs to represent data, usually in tuple-based structures, i.e., many tuples of URI data that can be held in databases, or interchanged on the World Wide Web using a set of particular syntaxes developed especially for the task. These syntaxes are called RDF (Resource Description Framework) syntaxes.

Security Manager

The Security Manager provides a secure connection between the client device and the origin server. Depending on the security policies of an organization, if the security requirements are not met or some content is not be viewable the security manager will ensure security with respect to:

- *Confidentiality:* The message being transacted needs to be confidential. Nobody will be able to see it.
- *Integrity:* The message being transacted needs to be tamper-resistant. Nobody will be able to change any part of the message.
- *Availability:* The system will be available. Nobody will be able to stop the service.
- *Non-repudiation:* Users of the system can be identified. Nobody after using the system can claim otherwise.
- *Trust:* There are complex issues of knowing what resources, services or agents to trust. The system will be trusted.

Confidentiality is managed by encryption. Using encryption techniques we change the message to some other message so that it cannot be understood. There are different types of encryption algorithms and standards. In a defined environment like enterprise LAN or a VPN (Virtual Private Network), we can standardize some encryption algorithm like 128 bits AES to be used. However, in a ubiquitous environment, the environment is unpredictable with ad-hoc groups of devices. Also, the networks and their security level cannot be guaranteed all the time. Integrity can be managed using different hashing algorithms. Availability relates to peripheral security related to Web server, firewall, etc. Non-repudiation can be managed with digital signatures. For trust we may need to establish some sort of third-party recommendation system. Third-party rating system can also help establish trust. The security manager needs to manage all these aspects.

Platform for Privacy Preference Project (P3P)

The Platform for Privacy Preference Project (P3P) is an emerging standard defined by W3C. P3P enables web sites to express their privacy practices in a standardized format so that they can be retrieved and interpreted by user agents. With P3P, users need not read the privacy policies they visit; instead, key information about the content of the web site can be conveyed to the user. Any discrepancies between a site's practices and the user's preferences can be flagged as well. The goal of P3P is to increase user trust and confidence in the Web.

P3P provides a technical mechanism to inform users about privacy policies about the site. This will help users to decide whether to release personal information or not. However, P3P does not provide any mechanism for ensuring that sites act according to their policies. P3P is intended to be complementary to both legislative and self-regulatory programs that can help enforce web site policies.

Adaptability Manager

The Adaptability Manager is responsible for adapting content, behavior and other aspects according to context and policy. The adaptability manager may take any number of actions depending on the information passed to it by the context manager. This information may or may not be in the form of RDF. The most obvious action to perform is to transcode content so that it may be viewed on a particular device. Other actions might include appending location-specific information to documents.

Content Adaptation and Transcoding

In a ubiquitous situation, services are used from any device through any network. Therefore, the content should be able to adapt to these dynamic situations. The adaptation may be static or dynamic.

Content adaptation can be performed either at the content level at the server end or at the agent level in the client device. Content adaptation can be done at an intermediate level in a middleware framework as well. To do a good job of content adaptation, we need to go beyond the header. We need to consider the requirements of the entire Web page or relationships between its various components in different media. It also needs to look at adaptation within the scope of the same and a different modality. Modes can be audio, video, voice, image or text. We are differentiating between audio and voice by the characteristics that audio is a sound clip as an object like the audio part of a multimedia lecture, whereas voice is real-time and synthesized from some other form or representation. Content adaptation needs to consider the following attributes.

1. **Physical capabilities of the device:** Screen size, i.e., width and height in pixels, color and bits/pixel.
2. **Logical capabilities of the device:** Required for displaying video, image and playing audio.

3. **Effective network bandwidth**.
4. **Payload:** The total amounts of bits that can be delivered to the agent for the static parts. For streaming media this will be the initial buffer space required before the media starts playing. For storage constrained devices, the payload will be defined as the storage space.

Transcoding can be classified as the following:

- *Spatial transcoding* is transcoding in space or dimension. In this transcoding technique a standard frame is downscaled and reduced. The frame is changed from one size to a different size to suit the target device.
- *Temporal transcoding* copes with a reduction of number of frames in the time scale. This technique downscales the number of transferred frames to suit the target device and network bandwidth.
- *Color transcoding* is sometimes requested for monochrome clients. Using less bits for pixel can reduce bandwidth and sometimes modify the perception of images.
- *Code transcoding* is used to change coding from one standard to another. One such example could be compression of the data or transcode a BMP file to WBMP for wireless device.
- *Object or semantic transcoding* comprises some different techniques based on computer vision techniques. The goal is to extract semantically valuable objects from the scene and transfer them with the lower amount of compression in order to maintain both details and speed.

Server side content adaptation can be achieved through the concept of InfoPyramid. InfoPyramid creates context-aware content through static transcoding. The transcoding is done off-line at the content creation time. InfoPyramid is used to store multiple resolutions and modalities of the transcoded content, along with any associated meta-data. For server side adaptation, each atomic item of the document is analysed to determine its resource requirements. The types of resources considered are those that may differentiate different client devices. The resource requirement is determined by the following attributes.

1. Static content size in bits.
2. Display size such as height, width and area.
3. Streaming bit-rate.
4. Color requirements.
5. Compression formats.
6. Hardware requirements, such as display for images, support for audio and video.

This is very useful for enterprises whose users are likely to use the service from different networks and devices. For example, a bank or a courier company which has its customer base across the world and is likely to use the service from any device from any network. When the Web server receives a user request, it determines the capabilities of the requesting client device. A customization module (context-sensitive content switch) dynamically selects the page from the InfoPyramids. The selection is based on the resolutions or modalities that best meet the client capabilities. This selected content is then rendered in a suitable delivery format for delivery to the client. This type of transcoding is most suitable for enterprises where the content type is known.

In case of client-side adaptation, the adaptation is done by the agent application. The agent application does the adaptation based on its capabilities. For example, let us assume that the client device does not support color: therefore, a color image received by the agent will be displayed as a black and white image. Client-side adaptation can be quite effective for static images. However, it may not be very effective for streaming payload delivery.

The other technique of transcoding is through a middleware. One big benefit of the middleware approach is that it is totally transparent to the device and the content. Content providers do not have to change the way they author or serve content. However, there are a number of drawbacks to this approach:

1. Content providers have no control over how their content will appear to different clients.
2. There may be legal issues arising from copyright that may preclude or severely limit the transcoding by proxies.
3. HTML tags mainly provide formatting information rather than semantic information.
4. Transcoding sometimes could be difficult to apply to many media types such as video and audio.
5. Developing a general purpose transcoding engine is very difficult if not impossible.

Transcoding through middleware is transparent to both device and content. Therefore, this transcoding technique has to be very robust and universal. That is why this transcoding technique is the most difficult to engineer. It is most desirable for content aggregators and value-added service providers.

Content Rating and Filtering

Any city in the world has regions well marked like business district, residential area, shopping complex, so on and so forth. In Bangalore, for example, Commercial Street, Koramangala, and Shivaji Market signify commercial/shopping area, residential area and market place respectively. By looking at the name of a web site or the document header, can we make some judgement about the content? This is necessary for content filtering and personalization. If we want to make sure that children at home are not accessing some restricted material, how do we do this? In a bookstore, adult magazines are displayed on the topmost shelf so that children cannot reach them. Children below 18 are not allowed to buy cigarettes or alcohol from a shop. In Internet, everything is freely accessible. How do we enforce such social discipline in the electronic world?

W3C has proposed a standard called PICS (Platform for Internet Content Selection) for rating of web content. Filtering of the content can take place depending on this rating. PICS specification is a set of technical specifications for labels (meta-data) that help software and rating services to work together. Rating and labeling services choose their own criteria for proper identification and filtering of the content. Since rating will always involve some amount of subjective judgement, it is left to the service provider to define the ratings. Rating can be through self-labeling or third-party labeling of content. In third-party labeling some independent rating agency can be used. The rating of Internet sites was originally designed to help parents and teachers control what children access on the Internet, but it also facilitates other uses for labels, including code signing and privacy.

The RSACI (Recreational Software Advisory Council Internet) has a PICS-compliant rating system called Resaca. Web pages that have been rated with the Resaca system contain labels recognized by many popular browsers like Netscape and Internet Explorer. Resaca uses four categories—violence, nudity, sex, and language—and a number for each category indicating the degree or level of potentially offensive content. Each number can range from 0, meaning the page contains no potentially offensive content, to 4, meaning the page contains the highest levels of potentially offensive content. For example, a page with a Resaca language level of 0 contains no offensive language or slangs. A page with a language level of 4 contains crude, vulgar language or extreme hate speech. When an end-user asks to see a particular URL, the software filter fetches the document but also makes an inquiry to the label bureau to ask for labels that describe that URL. Depending on what the labels say, the filter may block access to that URL. PICS labels can describe anything that can be named with a URL. That includes FTP and Gopher. E-mail messages do not normally have URLs, but messages from discussion lists that are archived on the Web do have URLs and can thus be labeled. A label can include a cryptographic signature. This mechanism lets the user check that the label was authorized by the service provider.

While the motivation for PICS was concern over children accessing inappropriate materials, it is a general "meta-data" system, meaning that labels can provide any kind of descriptive information about Internet material. For example, a labeling vocabulary could indicate the literary quality of an item rather than its appropriateness for children. Most immediately, PICS labels could help in finding particularly desirable materials, and this is the main motivation for the ongoing work on a next generation label format that can include arbitrary text strings. More generally, the W3C is working to extend Web meta-data capabilities generally and is applying them specifically in the following areas:

1. Digital Signature: Coupling the ability to make assertions with a cryptographic signature block that ensures integrity and authenticity.
2. Intellectual Property Rights Management: Using a meta-data system to label Web resources with respect to their authors, owners and rights management information.
3. Privacy (P3): Using a meta-data system to allow sites to make assertions about their privacy practices and for users to express their preferences for the type of interaction they want to have with those sites.
4. Personalization: Based on some policy, the content can be personalized to suit the need of the user and the service.

Regardless of content control, meta-data systems such as PICS are going to be an important part of the Web, because they enable more sophisticated commerce (build and manage trust relationships), communication, indexing, and searching services. Content filtering can take place either at the client end or at the middleware proxy end.

Content Aggregation

Over a period, the dynamics associated with the content has changed considerably. Earlier, there was a requester requesting for content and a responder responding to the content requested. The game was simple with only two players, the requester and the responder. These contents were corporate content or content for the mass (primarily web sites). There was no concept of charging for the content. Today there is a concept of OEM (Original Equipment Manufacturer) in content. There are some organizations which create content like an OEM. There are other ASPs (Application Service Providers), MVNOs (Mobile Virtual Network Operators), and content aggregators who source content from these OEMs and provide the content as a value added service to different individuals, content providers, and network operators.

In the current scenario, there are primarily four parties involved; they are end user (EU), the content provider (CP), the content aggregator (CA), and the ISP (Internet Service Provider) or the wireless or wireline network operator (NO). The network operator will have routers, cache, gateways and other nodes to offer the service. In this scheme anybody can become a requester or a responder. There could be different parameters, which will determine the content. These parameters are of two types, static and dynamic. The static adaptation parameters are those which can be received before the service begins. The content is adapted, based on this parameter. The dynamic adaptation parameters are those which are required with every request. For example, a user may initiate a request for a MPEG stream. The NO will transcode the stream to suit the bandwidth of the end user and delivers the same to the user. However, through a dynamic parameter, the user can specify a different parameter for transcoding.

From the content aggregator's perspective we may classify the service into two categories:

1. Single service request: This works at the user level and works for only one user. For example, a user may request the proxy server at the NO to translate the page into Hindi and then deliver the same to the user. In this case, the end user buys the content and the translation service.
2. Group service request: This works for a group of users. This type of request is initiated either at the CA level or the NO level. For example, the content aggregator has some arrangement for advertisement. The content aggregator examines all the HTML pages and inserts an advertisement at an appropriate place.

Seamless Communication

The basic premise of a ubiquitous system is that the system will be available and accessible from anywhere, anytime and through any network or device. A user will be able to access the system after moving from one place to another place (foreign place). The user will also be able to access the system while on the move (traveling mode). Mobile healthcare professionals, for example, may need to seamlessly switch between different modes of communication when they move from indoors to outdoors. A corporate user requires a similar kind of facility as well. Also, what is necessary is, during the movement, the session needs to continue. If we take the example of healthcare sector, some data and information are exchanged between the patient and the hospital. While the patient is moved from home, to ambulance, to a helicopter, to the hospital, the information exchange has to continue without any interruption.

Seamless communication will combine seamless handoffs and seamless roaming. Handoff is the process by which the connection to the network (point of attachment) is moved from one base station (access point) to another base station within the same network. Whereas, roaming will involve the point of attachment moving from one base station of one network to another base station of another network. The basic challenge in handoff is that it has to work while a session is in progress. Cellular technology with respect to voice has reached a level of maturity where a seamless voice communication is possible through handoff and roaming. The data technology is yet to mature to provide a similar level of service. In some parts of the world, handoff is termed as handover.

Seamless communication offers users freedom to roam across different wireless networks. Roaming works within homogeneous networks, like GSM to GSM or CDMA2000 to CDMA2000.

Nowadays, roaming is also possible from GSM to CDMA2000 network and vice-versa provided the user device is dual band and can connect to both these networks. True seamless roaming will include handoff and roaming in a heterogeneous hybrid network. The user will move from a WiFi to 3G to wired LAN to GSM while the session is in progress. Users will be able to communicate using whatever wireless device is currently at hand. Thus, GPRS-enabled cell phones, PDAs and laptops will be able to roam and communicate freely and access the Internet across both WLANs and WWANs.

In seamless roaming, the following aspects need to be maintained and managed in a seamless fashion without any disruption of service:

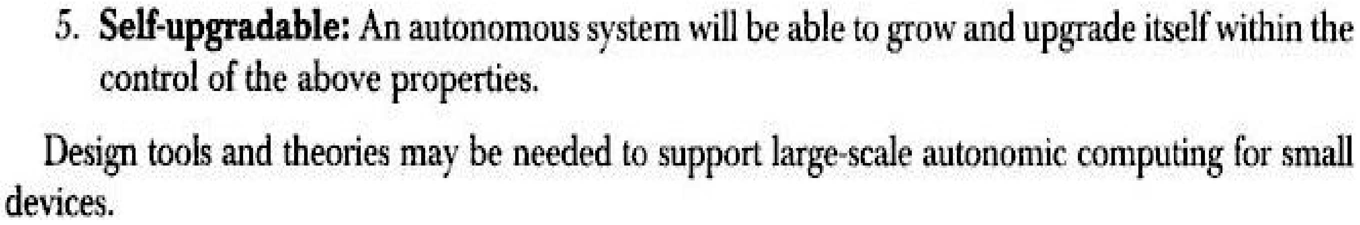
1. Authentication across network boundaries.
2. Authorization across network boundaries.
3. Billing and charging data collection.
4. End-to-end data security across roaming.
5. Handoff between wireless access points.
6. Roaming between networks.
7. Session migration.
8. IP mobility.

The task of managing authentication between client devices and networks, often involving multiple login names and passwords, will become automatic and invisible to the user, as will the configuration of various settings and preferences that accumulate with client devices.

Autonomous Computing

The world is heading for a software complexity crisis. Software systems are becoming bigger and more complex. Systems and applications cover millions of lines of code and require skilled IP professionals to install, configure, tune and maintain. New approaches are needed to provide flexible and adaptable software and hardware, both for mobile devices and the intelligent environment. Ease of use will have some effect on acceptance of a ubiquitous system. The scale of these ubiquitous systems necessitates "autonomic" systems. The purpose of autonomous system is to free users and system administrators from the details of system operation and maintenance complexity. Also, the system will run 24 × 7. The essence of autonomous system is self-management, which is a combination of the following functions:

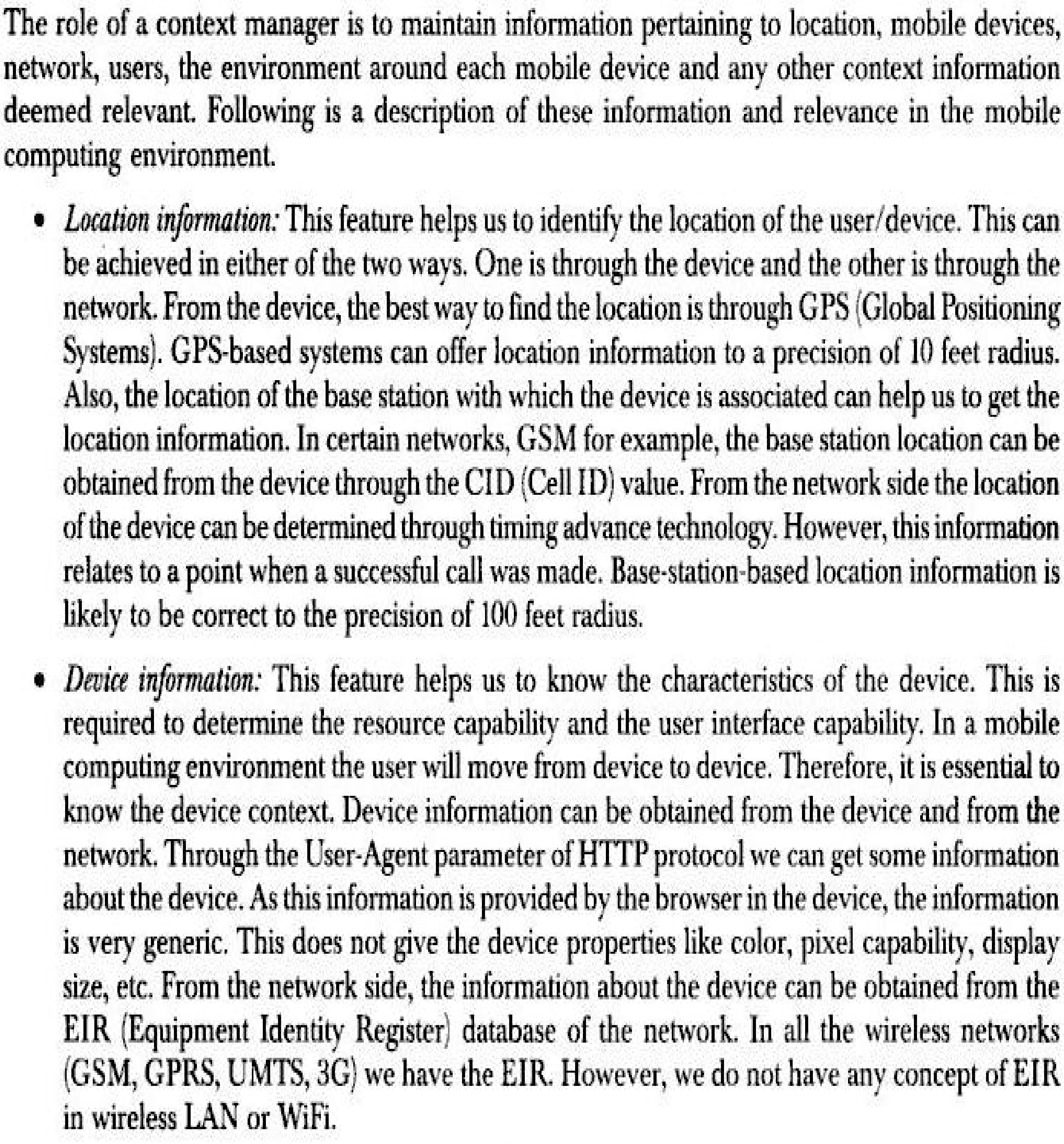
1. **Self-configurable:** An autonomous system will configure itself automatically in accordance with high-level policies. This will suit the functional requirement of the user.
2. **Self-optimizing:** An autonomous system will continuously look for ways to improve its operation with respect to resource, cost and performance. This will mean that an autonomous system will keep on tuning hundreds of tunable parameters to suit the user and the environment.
3. **Self-healing:** An autonomous system will detect, diagnose and repair localized problems resulting from bugs or failures. These failures could be the result of either software or hardware failure.
4. **Self-protecting:** An autonomous system will be self-protecting. This will be from two aspects. It will defend itself from external attacks; also, it will not propagate or cascade failure to other parts of the system.



5. **Self-upgradable:** An autonomous system will be able to grow and upgrade itself within the control of the above properties.

Design tools and theories may be needed to support large-scale autonomic computing for small devices.

CONTEXT AWARE SYSTEMS :



The role of a context manager is to maintain information pertaining to location, mobile devices, network, users, the environment around each mobile device and any other context information deemed relevant. Following is a description of these information and relevance in the mobile computing environment.

- *Location information:* This feature helps us to identify the location of the user/device. This can be achieved in either of the two ways. One is through the device and the other is through the network. From the device, the best way to find the location is through GPS (Global Positioning Systems). GPS-based systems can offer location information to a precision of 10 feet radius. Also, the location of the base station with which the device is associated can help us to get the location information. In certain networks, GSM for example, the base station location can be obtained from the device through the CID (Cell ID) value. From the network side the location of the device can be determined through timing advance technology. However, this information relates to a point when a successful call was made. Base-station-based location information is likely to be correct to the precision of 100 feet radius.
- *Device information:* This feature helps us to know the characteristics of the device. This is required to determine the resource capability and the user interface capability. In a mobile computing environment the user will move from device to device. Therefore, it is essential to know the device context. Device information can be obtained from the device and from the network. Through the User-Agent parameter of HTTP protocol we can get some information about the device. As this information is provided by the browser in the device, the information is very generic. This does not give the device properties like color, pixel capability, display size, etc. From the network side, the information about the device can be obtained from the EIR (Equipment Identity Register) database of the network. In all the wireless networks (GSM, GPRS, UMTS, 3G) we have the EIR. However, we do not have any concept of EIR in wireless LAN or WiFi.

- *Network information:* In a mobile computing environment, the user moves from network to network. Sometime they are even heterogeneous in nature. Network information is required to identify the capability of the network. Capability information will include security infrastructure, services offered by the networks, etc. For example, while roaming a user moves from a GPRS network to a GSM network. Therefore, the rendering may need an adaptation from WAP to SMS. In the future, some of these will be done through programmable networks.
- *User information:* This information is required to identify the user correctly. From the security point of view, the system needs to ensure that the user is genuine and is who he claims to be. We need to ensure that nobody else is impersonating. This information can be validated through authentication independent of device or network. However, user preferences' information need to be obtained from the network. For charging the user properly we need to refer to some subscriber information available in the network.
- *Environment information:* This includes ambient surrounding awareness. We need to know the temperature, elevation, moisture, and other ambient-related information which are necessary for sensor-based networks.

GPS

Global Positioning System (GPS) is a system that gives us the exact position on the Earth. GPS is funded by and controlled by the US Department of Defense. There are GPS satellites orbiting the Earth, which transmit signals that can be detected by anyone with a GPS receiver. Using the receiver, we can determine the location of the receiver. GPS has three parts: the space segment, the user segment, and the control segment.

The space segment consists of 24 satellites, each in its own orbit 11,000 nautical miles above the Earth. Each GPS setellite takes 12 hours to orbit the Earth. Each satellite is equipped with an accurate clock to let it broadcast signals coupled with a precise time message.

The user segment consists of receivers, which can be in the users' hand, embedded in a mobile device or mounted in a vehicle. The user segment receives the satellite signal which travels at the speed of light. Even at this speed, the signal takes a measurable amount of time to reach the receiver. The difference between the time the signal is sent and the time it is received, multiplied by the speed of light, enables the receiver to calculate the distance to the satellite. To measure precise latitude, longitude and altitude, the receiver measures the time it took for the signals from four separate satellites to get to the receiver. If we know our exact distance from a satellite in space, we know we are somewhere on the surface of an imaginary sphere with radius equal to the distance to the satellite radius. If we know our exact distance from four satellites, we know precisely where we are on the surface of the each.

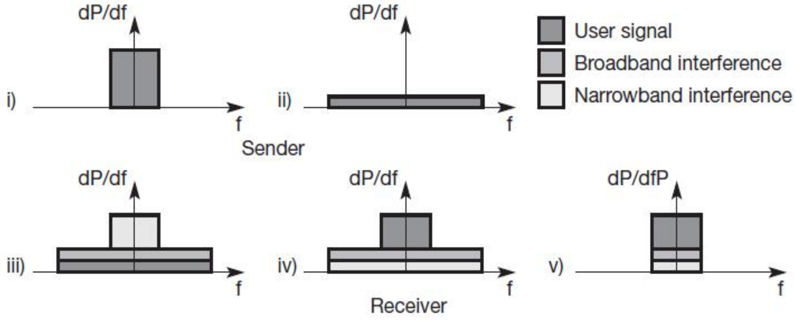
MODULE 2

Spread spectrum

Spread spectrum techniques involve spreading the bandwidth needed to transmit data .The main advantage is the resistance to narrowband interference.

diagram i) shows an idealized narrowband signal from a sender of user data (here power density dP/df versus frequency f).

ii), The sender now spreads the signal i.e., converts the narrowband signal into a broadband signal. The energy needed to transmit the signal (the area shown in the diagram) is the same, but it is now spread over a larger frequency range. The power level of the spread signal can be much lower than that of the original narrowband signal without losing data. Depending on the generation and reception of the spread signal, the power level of the user signal can even be as low as the background noise. This makes it difficult to distinguish the user signal from the background noise and thus hard to detect.



iii) During transmission, narrowband and broadband interference add to the Signal. The sum of interference and user signal is received.

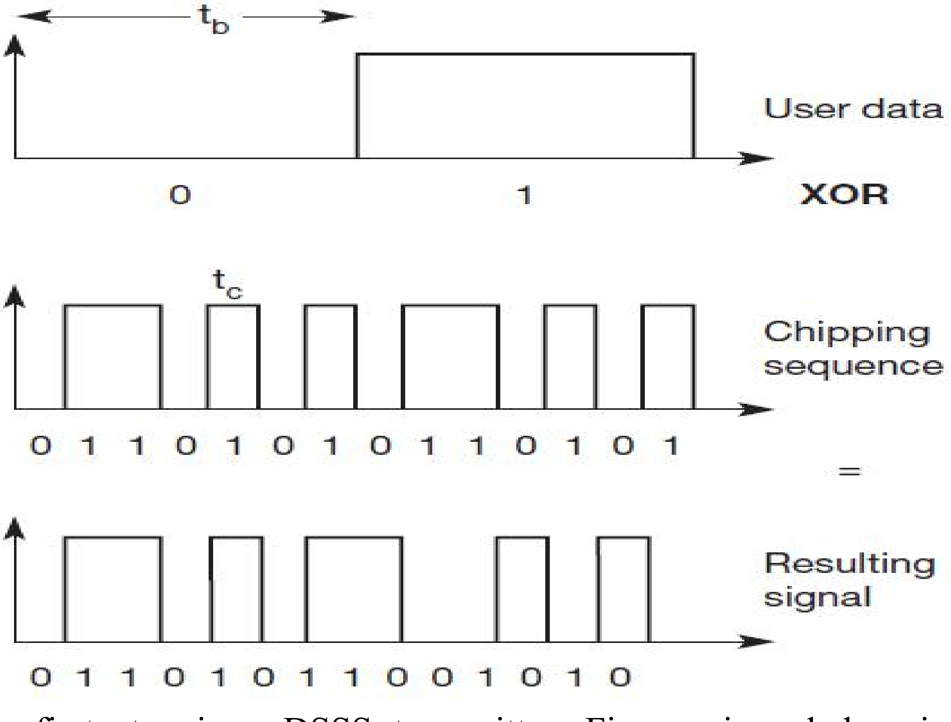
iv) The receiver now knows how to despread the signal, converting the spread user signal into a narrowband signal again, while spreading the narrowband interference and leaving the broadband interference.

v) The receiver applies a bandpass filter to cut off frequencies left and right of the narrowband signal.

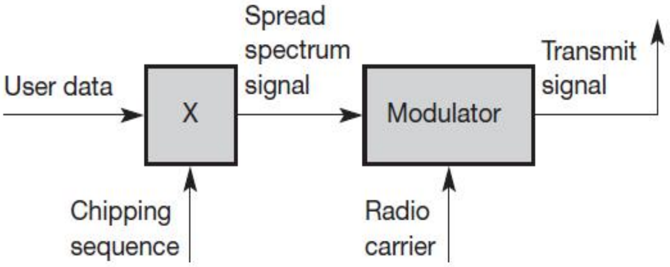
Finally, the receiver can reconstruct the original data because the power level of the user signal is high enough, i.e., the signal is much stronger than the remaining interference.

Direct sequence spread spectrum

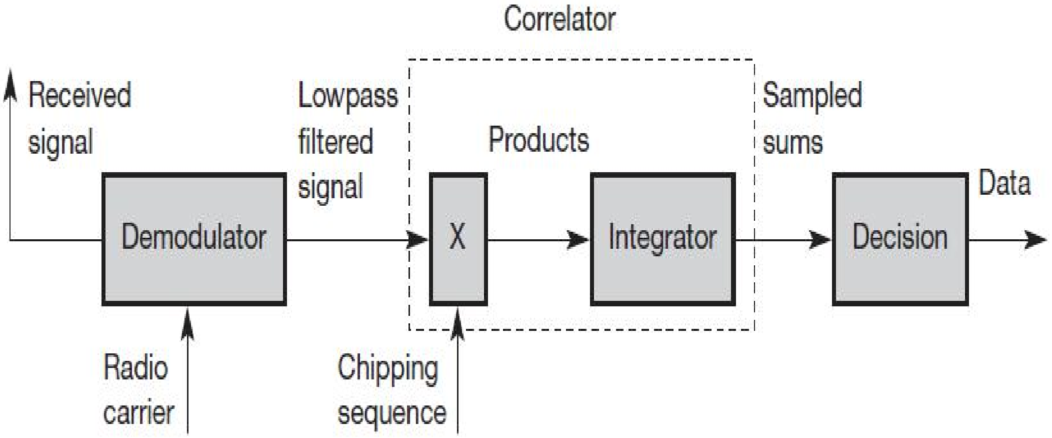
Direct sequence spread spectrum (DSSS) systems take a user bit stream and perform an (XOR) with a so-called chipping sequence as shown in Figure below. The example shows that the result is either the sequence 0110101 (if the user bit equals 0) or its complement 1001010 (if the user bit equals 1). While each user bit has a duration tb, the chipping sequence consists of smaller pulses, called chips, with a duration tc. If the chipping sequence is generated properly it appears as random noise: this sequence is also sometimes called pseudo-noise sequence. The spreading factor s = tb/tc determines the bandwidth of the resulting signal. If the original signal needs a bandwidth w, the resulting signal needs s·w after spreading. While the spreading factor of the very simple example is only 7 (and the chipping sequence 0110101 is not very random), civil applications use spreading factors between 10 and 100, military applications use factors of up to 10,000. Wireless LANs complying with the standard IEEE 802.11 use, for example, the sequence 10110111000, a so-called Barker code, if implemented using DSSS. Barker codes exhibit a good robustness against interference and insensitivity to multi-path propagation. Other known Barker codes are 11, 110, 1110, 11101, 1110010, and 1111100110101. Up to now only the spreading has been explained. However, transmitters and receivers using DSSS need additional components as shown in the simplified block diagrams in Figure given below.



The first step in a DSSS transmitter, Figure given below is the spreading of the user data with the chipping sequence (digital modulation). The spread signal is then modulated with a radio carrier (radio modulation). Assuming for example a user signal with a bandwidth of 1 MHz. Spreading with the above 11-chip Barker code would result in a signal with 11 MHz bandwidth. The radio carrier then shifts this signal to the carrier frequency (e.g., 2.4 GHz in the ISM band). This signal is then transmitted.



DSS Transmitter



DSS Receiver

The DSSS receiver is more complex than the transmitter. The receiver only has to perform the inverse functions of the two transmitter modulation steps. However, noise and multi-path propagation require additional mechanisms to reconstruct the original data. The first step in the receiver involves demodulating the received signal. This is achieved using the same carrier as the transmitter reversing the modulation and results in a signal with approximately the same bandwidth as the original spread spectrum signal. Additional filtering can be applied to generate this signal. While demodulation is well known from ordinary radio receivers, the next steps constitute a real challenge for DSSS receivers, contributing to the complexity of the system. The receiver has to know the original chipping sequence, i.e., the receiver basically generates the same pseudo random sequence as the transmitter. Sequences at the sender and receiver have to be precisely synchronized because the receiver calculates the product of a chip with the incoming signal. This comprises another XOR operation together with a medium access mechanism that relies on this scheme. During a bit period, which also has to be derived via synchronization, an integrator adds all these products. Calculating the products of chips and signal, and adding the products in an integrator is also called correlation, the device a correlator. Finally, in each bit period a decision unit samples the sums generated by the integrator and decides if this sum represents a binary 1 or a 0. If transmitter and receiver are perfectly synchronized and the signal is not too distorted by noise or multi-path propagation.

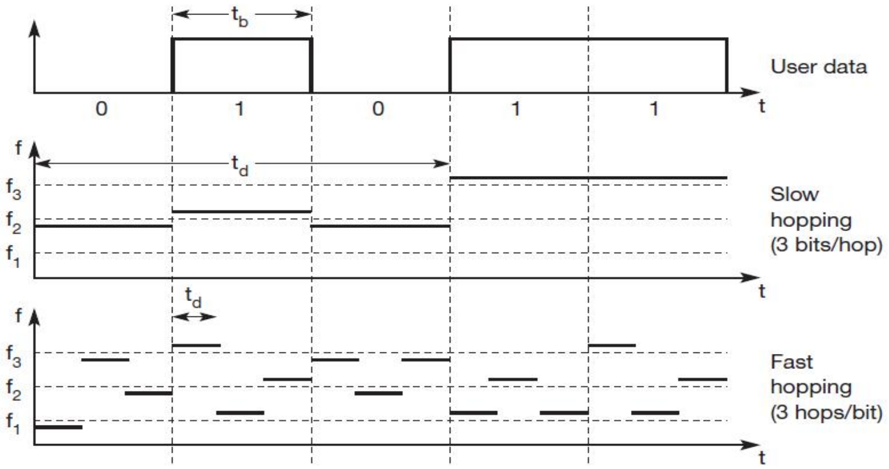
DSSS works perfectly well according to the simple scheme shown. Sending the user data 01 and applying the 11-chip Barker code 10110111000 results in the spread 'signal' 1011011100001001000111. On the receiver side, this 'signal' is XORed bit-wise after demodulation with the same Barker code as chipping sequence. This results in the sum of products equal to 0 for the first bit and to 11 for the second bit. The decision unit can now map the first sum (=0) to a binary 0, the second sum (=11) to a binary 1 – this constitutes the original user data.

In case of multi-path propagation several paths with different delays exist between a transmitter and a receiver. Additionally, the different paths may have different path losses. In this case, using so-called rake receivers provides a possible solution. A rake receiver uses n correlators for the n strongest paths. Each correlator is synchronized to the transmitter plus the delay on that specific path. As soon as the receiver detects a new path which is stronger than the currently weakest path, it assigns this new path to the correlator with the weakest path. The output of the correlators are then combined and fed into the decision unit. Rake receivers can even take advantage of the multi-path propagation by combining the different paths in a constructive way

Frequency hopping spread spectrum

For frequency hopping spread spectrum (FHSS) systems, the total available bandwidth is split into many channels of smaller bandwidth plus guard spaces between the channels. Transmitter and receiver stay on one of these channels for a certain time and then hop to another channel. This system implements FDM and TDM. The pattern of channel usage is called the hopping sequence, the time spend on a channel with a certain frequency is called the dwell time.

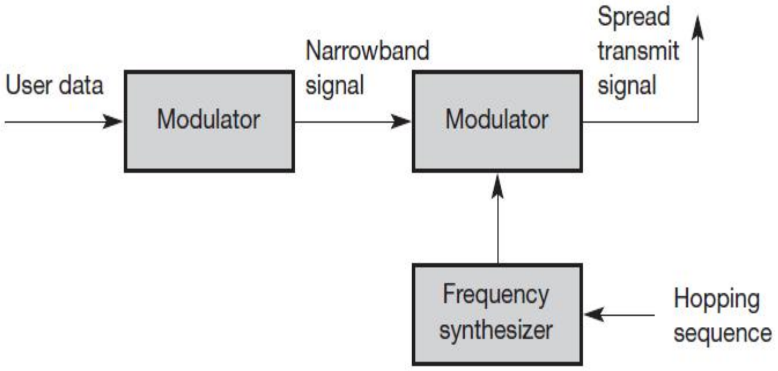
FHSS comes in two variants, slow and fast hopping In slow hopping, the transmitter uses one frequency for several bit periods Figure below shows five user bits with a bit period tb. Performing slow hopping, the transmitter uses the frequency f2 for transmitting the first three bits during the dwell time td. Then, the transmitter hops to the next frequency f3. Slow hopping systems are typically cheaper and have relaxed tolerances, but they are not as immune to narrowband interference as fast hopping systems. Slow frequency hopping is an option for GSM.



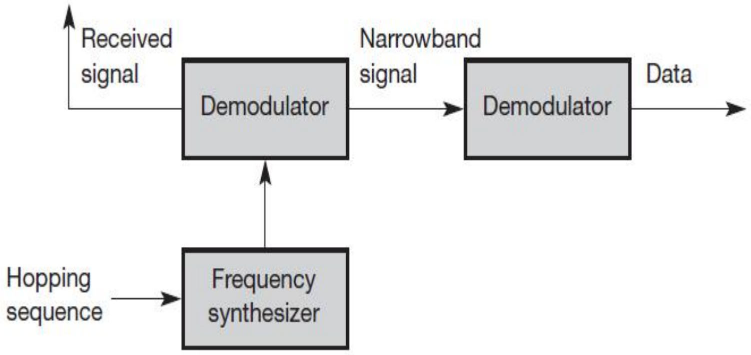
For fast hopping systems, the transmitter changes the frequency several times during the transmission of a single bit. In the example of Figure shown above the transmitter hops three times during a bit period. Fast hopping systems are more complex to implement because the transmitter and receiver have to stay synchronized within smaller tolerances to perform hopping at more or less the same points in time. However, these systems are much better at overcoming the effects of narrowband interference and frequency selective fading as they only stick to one frequency for a very short time. Another example of an FHSS system is Bluetooth

Figures shows below is a simplified block diagrams of FHSS transmitters and receivers respectively. The first step in an FHSS transmitter is the modulation of user data according to one of the digital-to-analog modulation schemes, e.g., FSK or BPSK, as discussed in section 2.6. This results in a narrowband signal, if FSK is used with a frequency f0 for a binary 0 and f1 for a binary 1. In the next step, frequency hopping is performed, based on a hopping sequence. The hopping sequence is fed into a frequency synthesizer generating the carrier frequencies fi. A second modulation uses the modulated narrowband signal and the carrier frequency to generate a new spread signal with frequency of fi+f0 for a 0 and fi+f1 for a 1 respectively. If different FHSS transmitters use hopping sequences that never overlap, i.e., if two transmitters never use the same frequency fi at the same time, then these two transmissions do not interfere. This requires the coordination of all transmitters and their hopping sequences.

As for DSSS systems, pseudo-random hopping sequences can also be used without coordination. These sequences only have to fulfill certain properties to keep interference minimal. Two or more transmitters may choose the same frequency for a hop, but dwell time is short for fast hopping systems, so interference is minimal. The receiver of an FHSS system has to know the hopping sequence and must stay synchronized. It then performs the inverse operations of the modulation to reconstruct user data. Several filters are also needed .



FHSS Transmitter



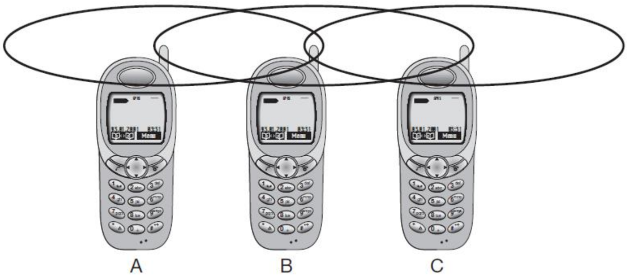
FHSS Receiver

Medium Access Control (MAC)

When sending data to another device on the network, the MAC block encapsulates higher-level frames into frames appropriate for the transmission medium (i.e. the MAC adds a syncword preamble and also padding if necessary), adds a frame check sequence to identify transmission errors, and then forwards the data to the physical layer as soon as the appropriate channel access method permits it. Controlling when data is sent and when to wait is necessary to avoid congestion and collisions, especially for topologies with a collision domain (bus, ring, mesh, point-to-multipoint topologies). Additionally, the MAC is also responsible for compensating for congestion and collisions by initiating retransmission if a jam signal is detected, and/or negotiating a slower transmission rate if necessary. When receiving data from the physical layer, the MAC block ensures data integrity by verifying the sender's frame check sequences, and strips off the sender's preamble and padding before passing the data up to the higher layers.

Hidden and exposed terminals

Consider the scenario with three mobile phones as shown in Figure. The transmission range of A reaches B, but not C (the detection range does not reach C either). The transmission range of C reaches B, but not A. Finally, the transmission range of B reaches A and C, i.e., A cannot detect C and vice versa. A starts sending to B, C does not receive this transmission. C also wants to send something to B and senses the medium. The medium appears to be free, the carrier sense fails. C also starts sending causing a collision at B. But A cannot detect this collision at B and continues with its transmission. A is hidden for C and vice versa.



While hidden terminals may cause collisions, the next effect only causes unnecessary delay. Now consider the situation that B sends something to A and C wants to transmit data to some other mobile phone outside the interference ranges of A and B. C senses the carrier and detects that the carrier is busy (B’s signal). C postpones its transmission until it detects the medium as being idle again. But as A is outside the interference range of C, waiting is not necessary. Causing a ‘collision’ at B does not matter because the collision is too weak to propagate to A. In this situation, C is exposed to B.

Near and far terminals

Consider the situation as shown in Figure . A and B are both sending with the same transmission power. As the signal strength decreases proportionally to the square of the distance, B’s signal drowns out A’s signal. As a result, C cannot receive A’s transmission. Now think of C as being an arbiter for sending rights (e.g., C acts as a base station coordinating media access). In this case, terminal B would already drown out terminal A on the physical layer. C in return would have no chance of applying a fair scheme as it would only hear B. The near/far effect is a severe problem of wireless networks using CDM. All signals should arrive at the receiver with more or less the same strength. Otherwise a person standing closer to somebody could always speak louder than a person further away.

Even if the senders were separated by code, the closest one would simply drown out the others. Precise power control is needed to receive all senders with the same strength at a receiver. For example, the UMTS system adapts power 1,500 times per second.

SDMA

Space Division Multiple Access (SDMA) is used for allocating a separated space to users in wireless networks. A typical application involves assigning an optimal base station to a mobile phone user. The mobile phone may receive several base stations with different quality. A MAC algorithm could now decide which base station is best, taking into account which frequencies (FDM), time slots (TDM) or code (CDM) are still available (depending on the technology). Typically, SDMA is never used in isolation but always in combination with one or more other schemes. The basis for the SDMA algorithm is formed by cells and sectorized antennas which constitute the infrastructure implementing space division multiplexing (SDM) .

FDMA

Frequency division multiple access (FDMA) comprises all algorithms allocating frequencies to transmission channels according to the frequency division multiplexing (FDM) scheme. Allocation can either be fixed (as for radio stations or the general planning and regulation of frequencies) or dynamic (i.e., demand driven). Allocation can either be fixed (as for radio stations or the general planning and regulation of frequencies) or dynamic (i.e., demand driven).

Furthermore, FDM is often used for simultaneous access to the medium by base station and mobile station in cellular networks. Here the two partners typically establish a duplex channel, i.e., a channel that allows for simultaneous transmission in both directions. The two directions, mobile station to base station and vice versa are now separated using different frequencies. This scheme is then called frequency division duplex (FDD). Again, both partners have to know the frequencies in advance; they cannot just listen into the medium. The two frequencies are also known as uplink, i.e., from mobile station to base station or from ground control to satellite, and as downlink, i.e., from base station to mobile station or from satellite to ground control.

TDMA

- Time division multiple access (TDMA) offers a much more flexible scheme, which comprises all technologies that allocate certain time slots for communication, i.e., controlling TDM.

- Tuning in to a certain frequency is not necessary, i.e., the receiver can stay at the same frequency the whole time and synchronization between sender and receiver has to be achieved in the time domain.

- Using only one frequency, and thus very simple receivers and transmitters, many different algorithms exist to control medium access.

Fixed TDM

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

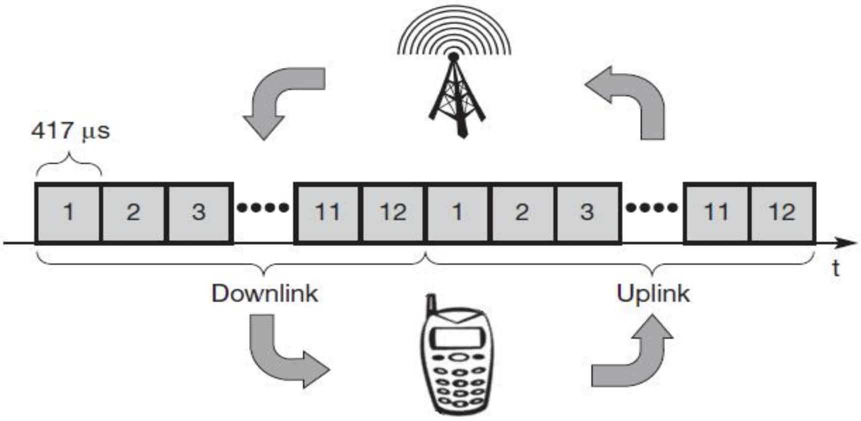
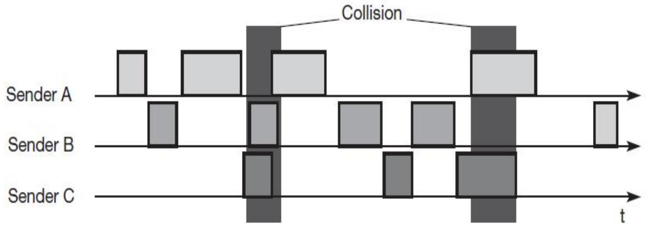


Figure shows how these fixed TDM patterns are used to implement multiple access and a duplex channel between a base station and mobile station. Assigning different slots for uplink and downlink using the same frequency is called time division duplex (TDD). As shown in the figure, the base station uses one out of 12 slots for the downlink, whereas the mobile station uses one out of 12 different slots for the uplink. Uplink and downlink are separated in time. Up to 12 different mobile stations can use the same frequency without interference using this scheme. Each connection is allotted its own up- and downlink pair. In the example below, which is the standard case for the DECT cordless phone system, the pattern is repeated every 10 ms, i.e., each slot has a duration of 417 μs.

This repetition guarantees access to the medium every 10 ms, independent of any other connections.

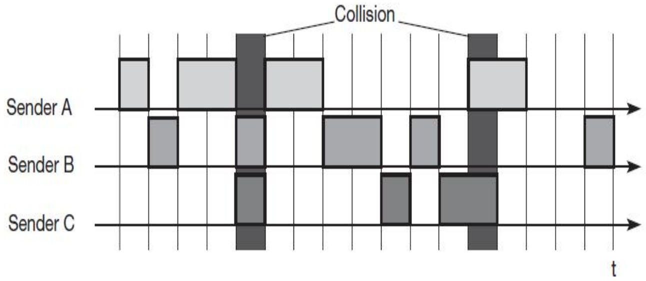
Classical Aloha

Aloha neither coordinates medium access nor does it resolve contention on the MAC layer. Instead, each station can access the medium at any time as shown in Figure given below. This is a random access scheme, without a central arbiter controlling access and without coordination among the stations. If two or more stations access the medium at the same time, a collision occurs and the transmitted data is destroyed. Resolving this problem is left to higher layers (e.g., retransmission of data).



Slotted Aloha

In this case, all senders have to be synchronized, transmission can only start at the beginning of a time slot as shown in Figure . Still, access is not coordinated. Under the assumption stated above, the introduction of slots raises the throughput from 18 per cent to 36 per cent, i.e., slotting doubles the throughput.



Carrier sense multiple access

- One improvement to the basic Aloha is sensing the carrier before accessing the medium. This is what carrier sense multiple access (CSMA) schemes generally do.

- Sensing the carrier and accessing the medium only if the carrier is idle decreases the probability of a collision. But, as already mentioned in the introduction, hidden terminals cannot be detected.

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

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

- In 1-persistent CSMA systems, all stations wishing to transmit access the medium at the same time, as soon as it becomes idle. This will cause many collisions if many stations wish to send and block each other. To create some fairness for stations waiting for a longer time, back-off algorithms can be introduced

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

Demand assigned multiple access

A general improvement of Aloha access systems can also be achieved by reservation mechanisms and combinations with some (fixed) TDM patterns. These schemes typically have a reservation period followed by a transmission period. During the reservation period, stations can reserve future slots in the transmission period. While, depending on the scheme, collisions may occur during the reservation period, the transmission period can then be accessed without collision

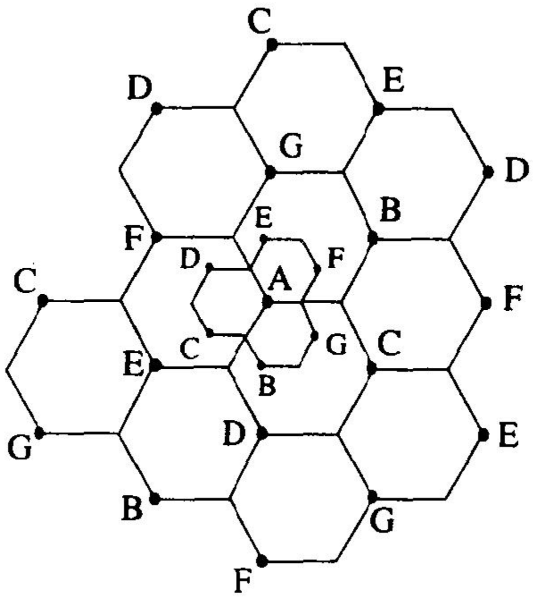
One basic scheme is demand assigned multiple access (DAMA) also called reservation Aloha, During a contention phase following the slotted Aloha scheme, all stations can try to reserve future slots. Collisions during the reservation phase do not destroy data transmission, but only the short requests for data transmission. If successful, a time slot in the future is reserved, and no other station is allowed to transmit during this slot.To maintain the fixed TDM pattern of reservation and transmission, the stations have to be synchronized from time to time. DAMA is an explicit reservation scheme. Each transmission slot has to be reserved explicitly.

improving coverage and capacity in cellular System

As the demand for wireless service increases, the number of channels assigned to a cell eventually becomes insufficient to support the required number of users. At this point, cellular design techniques are needed to provide more channels per unit coverage area. Techniques such as *cell splitting*, *sectoring*, and *coverage zone approaches* are used in practice to expand the capacity of cellular systems. Cell splitting allows an orderly growth of the cellular system. Sectoring uses directional antennas to further control the interference and frequency reuse of channels. The *zone microcell* concept distributes the coverage of a cell and extends the cell boundary to hard-to-reach places. While cell splitting increases the number of base stations in order to increase capacity, sectoring and zone microcells rely on base station antenna placements to improve capacity by reducing co-channel interference. Cell splitting and zone microcell techniques do not suffer the trunking inefficiencies experienced by sectored cells, and enable the base station to oversee all handoff chores related to the microcells, thus reducing the computational load at the MSC. These three popular capacity improvement techniques will be explained in detail.

Cell Splitting :

Cell splitting is the process of subdividing a congested cell into smaller cells, each with its own base station and a corresponding reduction in antenna height and transmitter power. Cell splitting increases the capacity of a cellular system since it increases the number of times that channels are reused. By defining new cells which have a smaller radius than the original cells and by installing these smaller cells (called *microcells*) between the existing cells, capacity increases due to the additional number of channels per unit area.

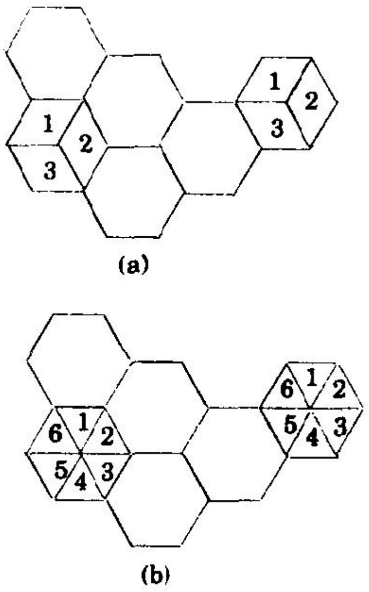


Imagine if everv cell were reduced in such a wav that the radius of every cell was cut in half. In order to cover the entire service area with smaller cells, approximately four times as many cells would be required. This can be easily shown by considering a circle with radius R. The area covered by such a circle is four times as large as the area covered by a circle with radius $R/2$. The increased number of cells would increase the number of clusters over the coverage region, which in turn would increase the number of channels, and thus capacity, in the coverage area. Cell splitting allows a system to grow by replacing large cells with smaller cells, while not upsetting the channel allocation scheme required to maintain the minimum co-channel reuse ratio Q between co-channel cells.

Sectoring:

The technique for decreasing co-channel interference and thus increasing system capacity by using directional antennas is called *sectoring*.

When sectoring is employed, the channels used in a particular cell are broken down *into sectored groups* and are used *only within* a particular sector, as illustrated in Figure 2.10(a) and (b). Assuming 7-cell reuse, for the case of 120° sectors, the number of interferers in the first tier is reduced from 6 to 2. This is because only 2 of the 6 co-channel cells receive interference with a particular sectored channel group.



(a)

(b)

Figure 2.10
(a) 120° sectoring.
(b) 60° sectoring.

A Novel Microcell Concept

The increased number of handoffs required when sectoring is employed results in an increased load on the switching and control link elements of the mobile system. A solution to this problem was presented by Lee [Lee91b]. This proposal is based on a microcell concept for 7 cell reuse, as illustrated in Figure 2.12. In this scheme, each of the three (or possibly more) zone sites (represented as Tx/Rx in Figure 2.12) are connected to a single base station and share the same radio equipment. The zones are connected by coaxial cable, fiberoptic cable, or microwave link to the base station. Multiple zones and a single base station make up a cell. As a mobile travels within the cell, it is served by the zone with the strongest signal. This approach is superior to sectoring since antennas are placed at the outer edges of the cell, and any base station channel may be assigned to any zone by the base station.

As a mobile travels from one zone to another within the cell, it retains the same channel. Thus, unlike in sectoring, a handoff is not required at the MSC when the mobile travels between zones within the cell. The base station simply switches the channel to a different zone site. In this way, a given channel is active only in the particular zone in which the mobile is traveling, and hence the base station radiation is localized and interference is reduced. The channels are distributed in time and space by all three zones and are also reused in co-channel cells in the normal fashion. This technique is particularly useful along highways or along urban traffic corridors.

The advantage of the zone cell technique is that while the cell maintains a particular coverage radius, the co-channel interference in the cellular system is reduced since a large central base station is replaced by several lower powered transmitters (zone transmitters) on the edges of the cell. Decreased co-channel interference improves the signal quality and also leads to an increase in capacity,

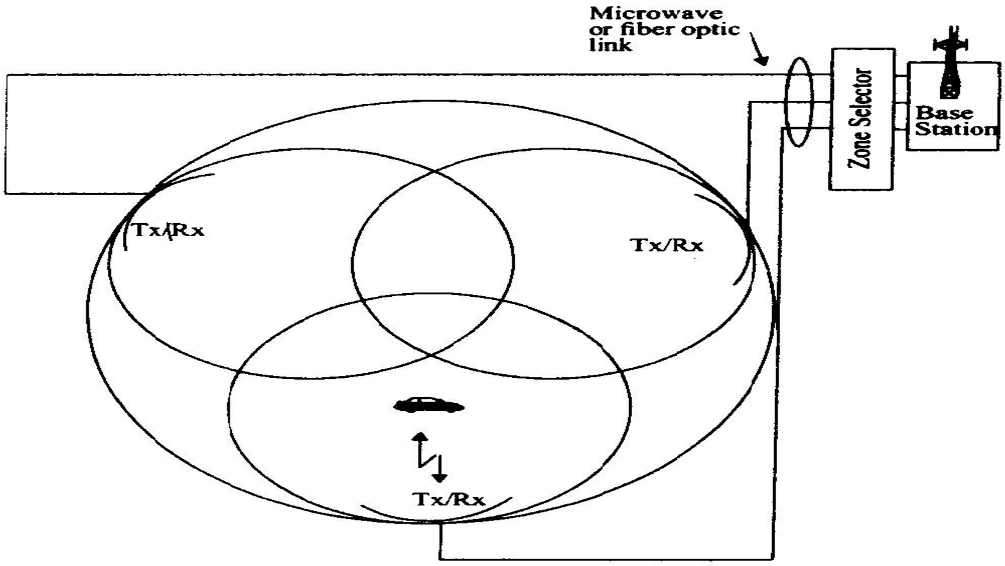


Figure 2.12
The microcell concept [adapted from [Lee91b] © IEEE].

GSM Architecture

The GSM architecture consists of three major interconnected subsystems that interact with themselves and with users through certain network interface. The subsystems are Base Station Subsystem (BSS), Network Switching Subsystem (NSS) and Operational Support Subsystem (OSS). Mobile Station (MS) is also a subsystem but it is considered as a part of BSS.

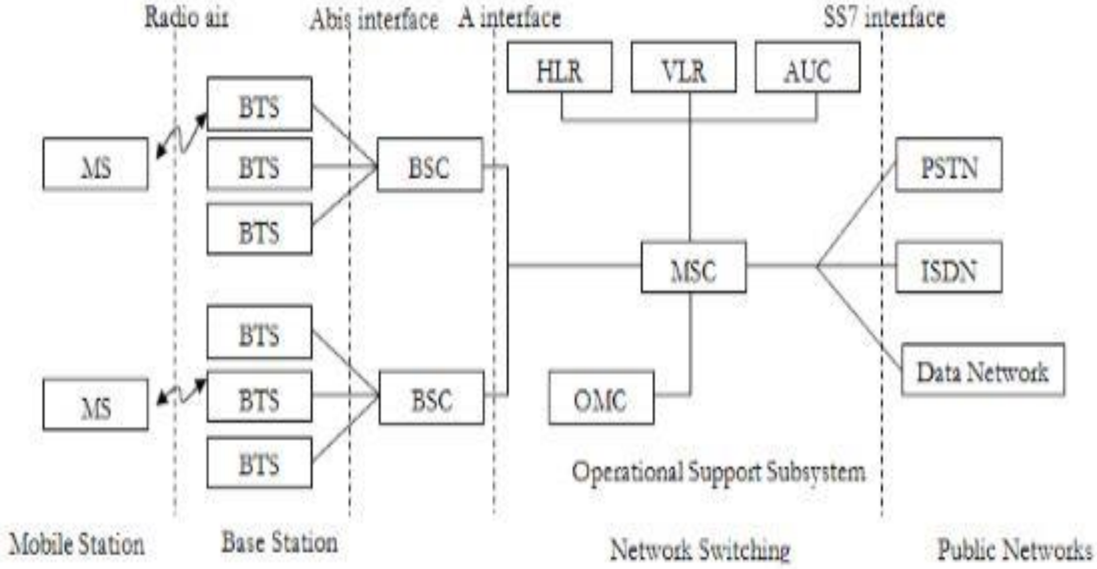


Fig: GSM Architecture

1. Mobile Station (MS): Mobile Station is made up of two entities.

A. Mobile equipment (ME):

It is a portable, vehicle mounted, hand held device. It is uniquely identified by an IMEI number. It is used for voice and data transmission. It also monitors power and signal quality of surrounding cells foe optimum handover. 160 characters long SMS can also be sent using Mobile Equipment.

B. Subscriber Identity module (SIM):

It is a smart card that contains the International Mobile Subscriber Identity (IMSI) number. Also allows users to send and receive calls and receive other subscriber services. - It is protected by password or PIN. It contains encoded network identification details. it has key information to activate the phone. It can be moved from one mobile to another.

2. Base Station Subsystem (BSS):

It is also known as radio subsystem, provides and manages radio transmission paths between the mobile station and the Mobile Switching Centre (MSC). BSS also manages interface between the mobile station and all other subsystems of GSM. It consists of two parts.

A. Base Transceiver Station (BTS):

It encodes, encrypts, multiplexes, modulates and feeds the RF signal to the antenna. BTS consists of transceiver units. It communicates with mobile stations via radio air interface and also communicates with BSC via Abis interface.

B. Base Station Controller (BSC):

It manages radio resources for BTS. It assigns frequency and time slots for all mobile stations in its area. Also handles call set up, transcoding and adaptation functionality handover for each MS radio power control. BSC communicates with MSC via A interface and also with BTS.

3. Network Switching Subsystem (NSS):

it manages the switching functions of the system and allows MSCs to communicate with other networks such as PSTN and ISDN. It consist of

A. Mobile switching Centre:

It is a heart of the network. It manages communication between GSM and other networks. It manages call set up function, routing and basic switching. Also performs mobility management including registration, location updating and inter BSS and inter MSC call handoff. It provides billing information. MSC does gateway function while its customers roam to other network by using HLR/VLR.

B. Home Location Registers (HLR): -

It is a permanent database about mobile subscriber in a large service area. - Its database contains IMSI, IMSISDN, prepaid/post-paid, roaming restrictions and supplementary services.

C. Visitor Location Registers (VLR): -

It is a temporary database which updates whenever new MS enters its area by HLR database. - It controls mobiles roaming in its area. It reduces number of queries to HLR. - Its database contains IMSI, TMSI, IMSISDN, MSRN, location, area authentication key.

D. Authentication Centre: -

It provides protection against intruders in air interface. - It maintains authentication keys and algorithms and provides security triplets (RAND, SRES, Ki).

E. Equipment Identity Registry (EIR):

It is a database that is used to track handset using the IMEI number. EIR is made up of three sub classes- the white list, the black list and the gray list.

4. Operational Support Subsystem (OSS):

It supports the operation and maintenance of GSM and allows system engineers to monitor, diagnose and troubleshoot all aspects of GSM system. It supports one or more Operation Maintenance Centres (OMC) which are used to monitor the performance of each MS, BS, BSC and MSC within a GSM system. It has three main functions:

- To maintain all telecommunication hardware and network operations with a particular market.
- To manage all charging and billing procedures
- To manage all mobile equipment in the system.

Interfaces used for GSM network :

1)UM Interface –Used to communicate between BTS with MS

2)Abis Interface— Used to communicate BSC TO BTS

3)A Interface-- Used to communicate BSC and MSC

4) Singling protocol (SS 7)- Used to communicate MSC with other network .

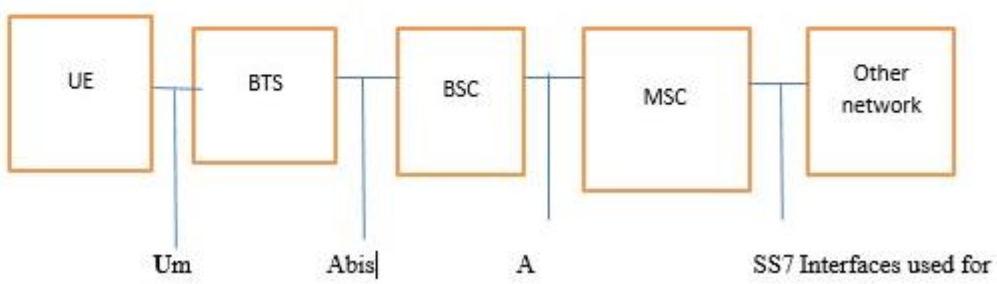


Fig: GSM Interfaces

DECT Characteristics & Architecture

DECT (Digital European Cordless Telephone) standardized by ETSI (ETS 300.175-x) for cordless telephones, DECT standard describes air interface between base-station and mobile phone. DECT has been renamed for international marketing reasons into Digital Enhanced Cordless Telecommunication

Characteristics

Frequency: 1880-1900 MHz

Channels: 120 full duplex

Duplex mechanism: TDD (Time Division Duplex) with 10 ms frame length

Multiplexing scheme: FDMA with 10 carrier frequencies and TDMA with 2x 12 slots

Modulation: digital, Gaussian Minimum Shift Key (GMSK)

Power: 10 mW average (max. 250 mW)

Range: Aprox. 50 m in buildings, 300 m open space

Architecture

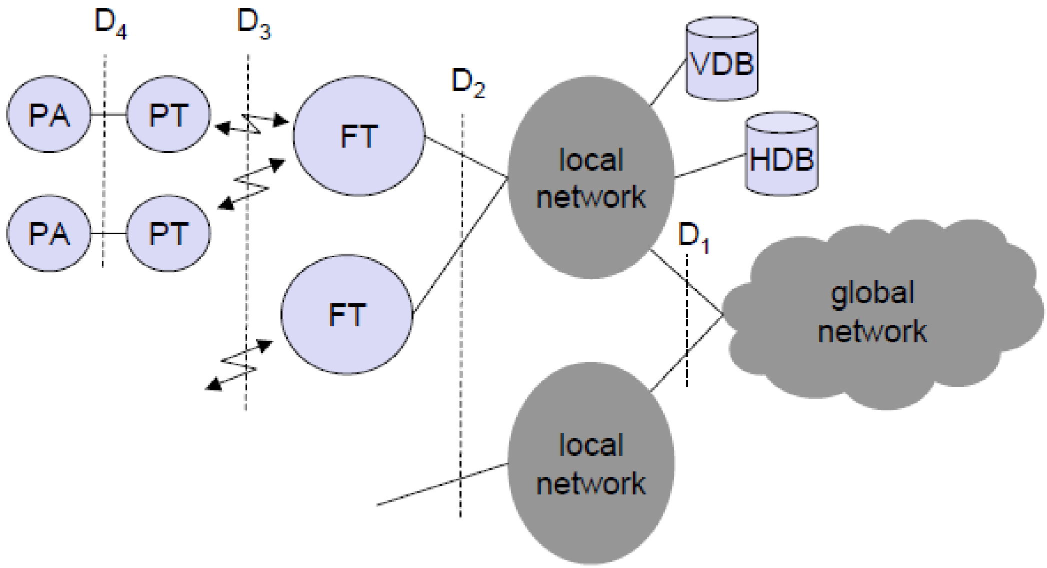


Fig: DECT Architecture

The main Components are

PA - Portable Application

PT - Portable radio Transmission

FT - Fixed radio Transmission

HDB - Home Data Base

VDB - Visitor Data Base (During Roaming)

MODULE 3

Wireless LAN Standards :

- **802.11** — applies to wireless LANs and provides 1 or 2 Mbps transmission in the 2.4 GHz band using either frequency hopping spread spectrum (FHSS) or direct sequence spread spectrum (DSSS).
- 802.11a — an extension to 802.11 that applies to wireless LANs and provides up to 54-Mbps in the 5GHz band. 802.11a uses an orthogonal frequency division multiplexing encoding scheme rather than FHSS or DSSS.
- 802.11b (also referred to as 802.11 High Rate or Wi-Fi) — an extension to 802.11 that applies to wireless LANS and provides 11 Mbps transmission (with a fallback to 5.5, 2 and 1-Mbps) in the 2.4 GHz band. 802.11b uses only DSSS. 802.11b was a 1999 ratification to the original 802.11 standard, allowing wireless functionality comparable to Ethernet.
- 802.11e — a wireless draft standard that defines the **Q**uality **o**f **S**ervice (QoS) support for LANs, and is an enhancement to the 802.11a and 802.11b wireless LAN (WLAN) specifications. 802.11e adds QoS features and multimedia support to the existing IEEE 802.11b and IEEE 802.11a wireless standards, while maintaining full backward compatibility with these standards.
- 802.11g — applies to wireless LANs and is used for transmission over short distances at up to 54-Mbps in the 2.4 GHz bands.
- 802.11n — 802.11n builds upon previous 802.11 standards by adding **m**ultiple-**i**nput **m**ultiple-**o**utput(MIMO). The additional transmitter and receiver antennas allow for increased data throughput through spatial multiplexing and increased range by exploiting the spatial diversity through coding schemes like Alamouti coding. The real speed would be 100 Mbit/s (even 250 Mbit/s in PHY level), and so up to 4-5 times faster than 802.11g.
- 802.11ac — 802.11ac builds upon previous 802.11 standards, particularly the 802.11n standard, to deliver data rates of 433Mbps per spatial stream, or 1.3Gbps in a three-antenna (three stream) design. The 802.11ac specification operates only in the 5 GHz frequency range and features support for wider channels (80MHz and 160MHz) and beamforming capabilities by default to help achieve its higher wireless speeds.
- 802.11ac Wave 2 — 802.11ac Wave 2 is an update for the original 802.11ac spec that uses MU-MIMOtechnology and other advancements to help increase theoretical maximum wireless speeds for the spec to 6.93 Gbps.
- 802.11ad — 802.11ad is a wireless specification under development that will operate in the 60GHz frequency band and offer

much higher transfer rates than previous 802.11 specs, with a theoretical maximum transfer rate of up to 7Gbps (Gigabits per second).

- 802.11ah— Also known as Wi-Fi HaLow, 802.11ah is the first Wi-Fi specification to operate in frequency bands below one gigahertz (900 MHz), and it has a range of nearly twice that of other Wi-Fi technologies. It's also able to penetrate walls and other barriers considerably better than previous Wi-Fi standards.
- 802.11r - 802.11r, also called Fast **B**asic **S**ervice **S**et (BSS) Transition, supports VoWi-Fi handoff between access points to enable VoIP roaming on a Wi-Fi network with 802.1X authentication.
- 802.1X — Not to be confused with 802.11x (which is the term used to describe the family of 802.11 standards) 802.1X is an IEEE standard for port-based Network Access Control that allows network administrators to restricted use of IEEE 802 LAN service access points to secure communication between authenticated and authorized devices.

IEEE 802 Protocol Architecture :

Protocol Architecture

Protocols defined specifically for LAN and MAN (metropolitan area network) transmission address issues relating to the transmission of blocks of data over the network. In OSI terms, higher-layer protocols (layer 3 or 4 and above) are independent of network architecture and are applicable to LANs, MANs, and WANs. Thus, a discussion of LAN protocols is concerned principally with lower layers of the OSI model.

Figure relates the LAN protocols to the OSI architecture.This architecture was developed by the IEEE 802 committee and has been adopted by all organizations working on the specification of LAN standards. It is generally referred to as the IEEE 802 reference model.1 Working from the bottom up, the lowest layer of the IEEE 802 reference model corresponds to the **physical layer** of the OSI model and includes such functions as

- Encoding/decoding of signals
- Preamble generation/removal (for synchronization)
- Bit transmission/reception

In addition, the physical layer of the 802 model includes a specification of the transmission medium and the topology. Generally, this is considered “below” the lowest layer of the OSI model. However, the choice of transmission medium and topology is critical in LAN design, and so a specification of the medium is included. Above the physical layer are the functions associated with providing service to LAN users. These include the following:

- On transmission, assemble data into a frame with address and error detection fields.
- On reception, disassemble frame, and perform address recognition and error detection.

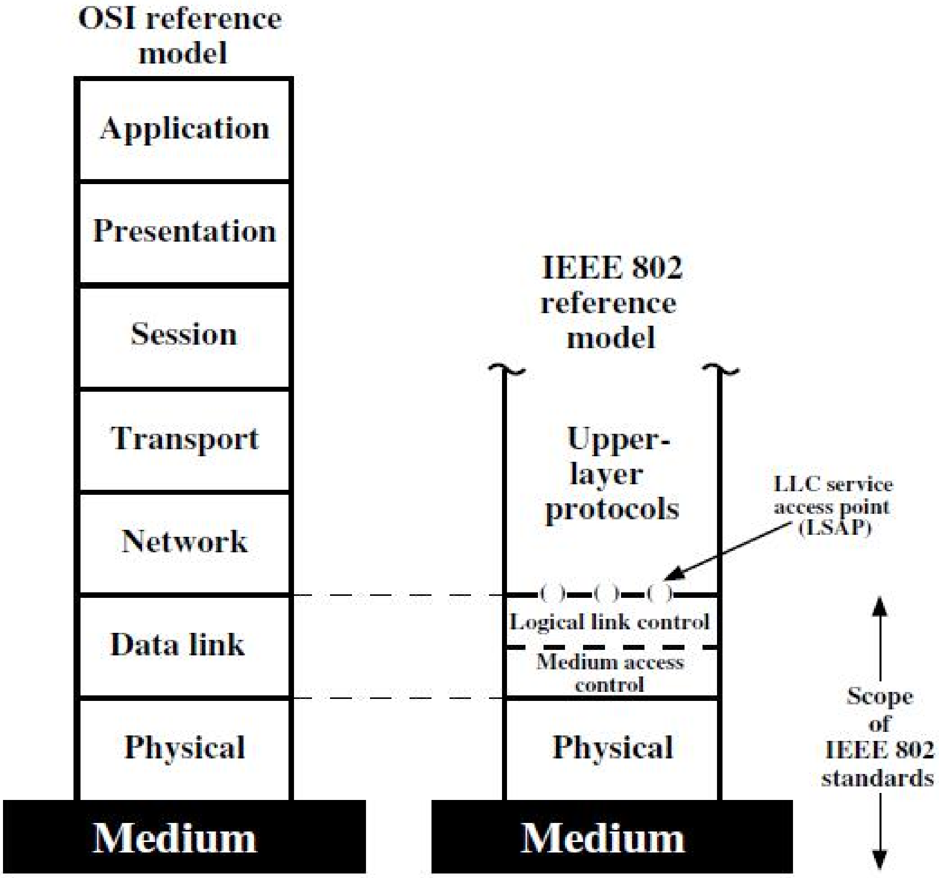


Figure 14.1 IEEE 802 Protocol Layers Compared to OSI Model

- Govern access to the LAN transmission medium.
- Provide an interface to higher layers and perform flow and error control.

These are functions typically associated with OSI layer 2. The set of functions in the last bullet item are grouped into a **logical link control (LLC)** layer. The functions in the first three bullet items are treated as a separate layer, called **medium access control (MAC)**. The separation is done for the following reasons:

- The logic required to manage access to a shared-access medium is not found in traditional layer 2 data link control.
- For the same LLC, several MAC options may be provided.

Higher-level data are passed down to LLC, which appends control information as a header, creating an *LLC protocol data unit (PDU)*. This control information is used in the operation of the LLC protocol. The entire LLC PDU is then passed down to the MAC layer, which appends control information at the front and back of the packet, forming a *MAC* *frame*. Again, the control information in the frame is needed for the operation of the MAC protocol.

For context, the figure also shows the use of TCP/IP and an application layer above the LAN protocols.

MAC Frame Format

The MAC layer receives a block of data from the LLC layer and is responsible for performing functions related to medium access and for transmitting the data. As with other protocol layers, MAC implements these functions making use of a protocol data unit at its layer. In this case, the PDU is referred to as a MAC frame. The exact format of the MAC frame differs somewhat for the various MAC protocols in use. The fields of this frame are as follows:

- **MAC control:** This field contains any protocol control information needed for the functioning of the MAC protocol. For example, a priority level could be indicated here.
- **Destination MAC address:** The destination physical attachment point on the LAN for this frame.
- **Source MAC address:** The source physical attachment point on the LAN for this frame.
- **Data:** The body of theMACframe. This may be LLC data from the next higher layer or control information relevant to the operatoin of the MAC protocol.
- **CRC:** The cyclic redundancy check field (also known as the frame check sequence, FCS, field). This is an error-detecting code, as we have seen in HDLC and other data link control protocols .In most data link control protocols, the data link protocol entity is responsible not only for detecting errors using the CRC but for recovering from those errors by retransmitting damaged frames.

In the LAN protocol architecture, these two functions are split between the MAC and LLC layers. The MAC layer is responsible for detecting errors and discarding any frames that are in error. The LLC layer optionally keeps track of which frames have been successfully received and retransmits unsuccessful frames.

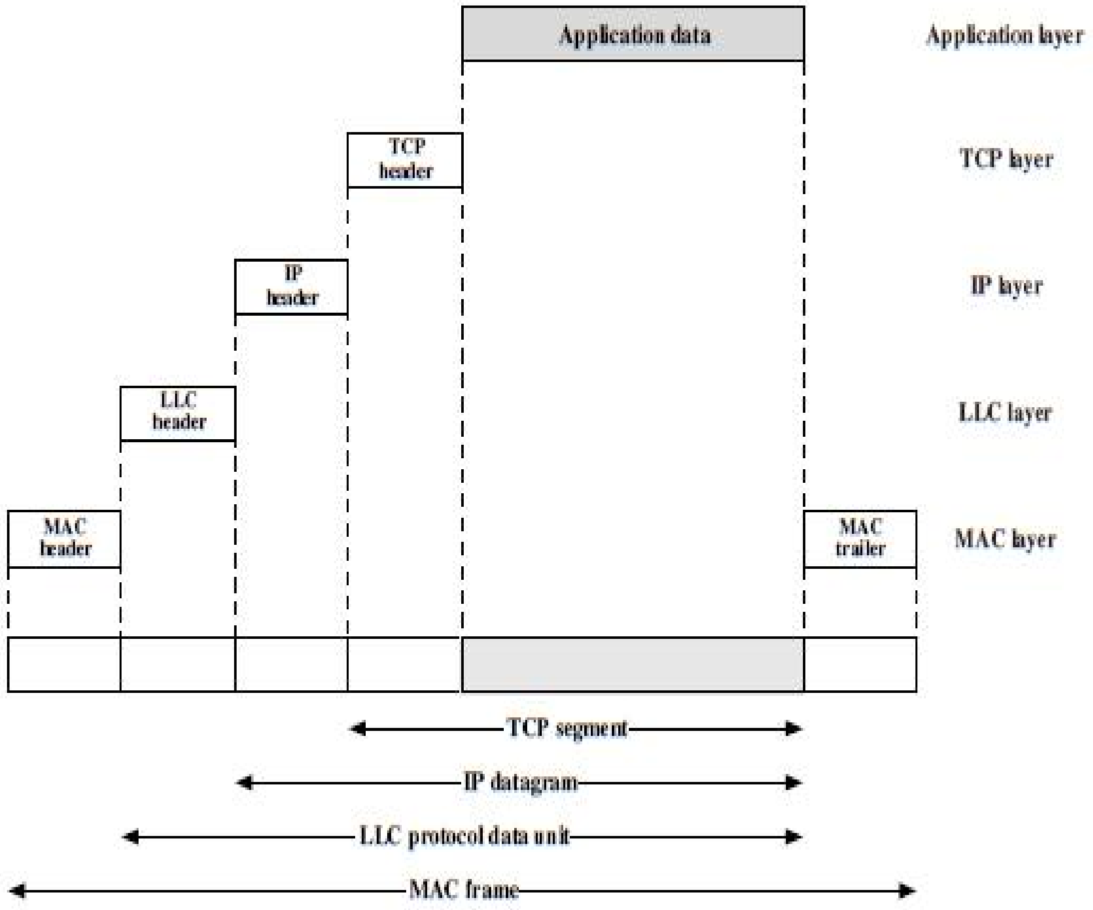


Figure 14.2 IEEE 802 Protocols in Context

Logical Link Control

The LLC layer for LANs is similar in many respects to other link layers in common use. Like all link layers, LLC is concerned with the transmission of a link-level PDU between two stations, without the necessity of an intermediate switching node. LLC has two characteristics not shared by most other link control protocols:

1. It must support the multiaccess, shared-medium nature of the link (this differs from a multidrop line in that there is no primary node).

2. It is relieved of some details of link access by the MAC layer. Addressing in LLC involves specifying the source and destination LLC users.

Typically, a user is a higher-layer protocol or a network management function in the station. These LLC user addresses are referred to as service access points (SAPs), in keeping with OSI terminology for the user of a protocol layer. We look first at the services that LLC provides to a higher-level user, and then at the LLC protocol.

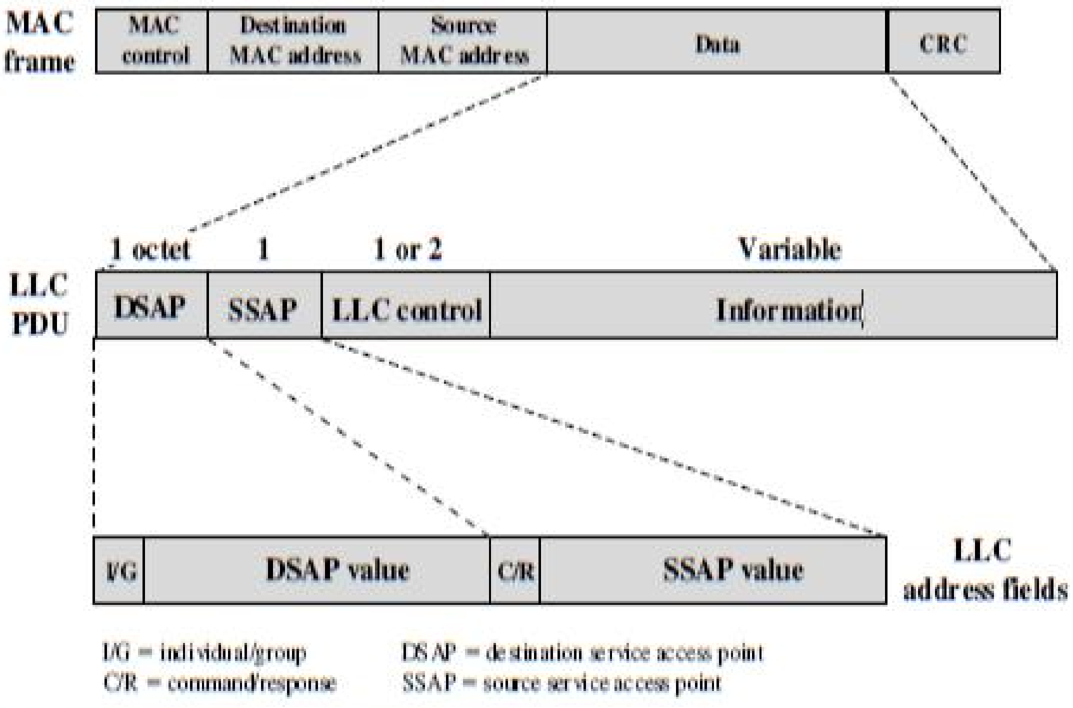


Figure 14.3 LLC PDU in a Generic MAC Frame Format

LLC Services

LLC specifies the mechanisms for addressing stations across the medium and for controlling the exchange of data between two users. The operation and format of this standard is based on HDLC.

Three services are provided as alternatives for attached devices using LLC:

- **Unacknowledged connectionless service:** This service is a datagram-style service. It is a very simple service that does not involve any of the flow- and errorcontrol mechanisms. Thus, the delivery of data is not guaranteed. However, in most devices, there will be some higher layer of software that deals with reliability issues.
- **Connection-mode service:** This service is similar to that offered by HDLC. A logical connection is set up between two users exchanging data, and flow control and error control are provided.
- **Acknowledged connectionless service:** This is a cross between the previous two services. It provides that datagrams are to be acknowledged, but no prior logical connection is set up.

Typically, a vendor will provide these services as options that the customer can select when purchasing the equipment. Alternatively, the customer can purchase equipment that provides two or all three services and select a specific service based on application.

The **unacknowledged connectionless service** requires minimum logic and is useful in two contexts. First, it will often be the case that higher layers of software will provide the necessary reliability and flow-control mechanism, and it is efficient to avoid duplicating them. For example, TCP could provide the mechanisms needed to ensure that data are delivered reliably. Second, there are instances in which the overhead of connection establishment and maintenance is unjustified or even counterproductive (for example, data collection activities that involve the periodic sampling of data sources, such as sensors and automatic self-test reports from security equipment or network components). In a monitoring application, the loss of an occasional data unit would not cause distress, as the next report should arrive shortly. Thus, in most cases, the unacknowledged connectionless service is the preferred option.

The **connection-mode service** could be used in very simple devices, such as terminal controllers, that have little software operating above this level. In these cases, it would provide the flow control and reliability mechanisms normally implemented at higher layers of the communications software.

The **acknowledged connectionless service** is useful in several contexts. With the connection-mode service, the logical link control software must maintain some sort of table for each active connection, to keep track of the status of that connection. If the user needs guaranteed delivery but there are a large number of destinations for data, then the connection-mode service may be impractical because of the large number of tables required. An example is a process control or automated factory environment where a central site may need to communicate with a large number of processors and programmable controllers. Another use of this is the handling of important and time-critical alarm or emergency control signals in a factory. Because of their importance, an acknowledgment is needed so that the sender can be assured that the signal got through. Because of the urgency of the signal, the user might not want to take the time first to establish a logical connection and then send the data.

LLC Protocol

The basic LLC protocol is modeled after HDLC and has similar functions and formats. The differences between the two protocols can be summarized as follows:

- LLC makes use of the asynchronous balanced mode of operation of HDLC, to support connection-mode LLC service; this is referred to as type 2 operation. The other HDLC modes are not employed.

- LLC supports an unacknowledged connectionless service using the unnumbered information PDU; this is known as type 1 operation.
- LLC supports an acknowledged connectionless service by using two new unnumbered PDUs; this is known as type 3 operation.
- LLC permits multiplexing by the use of LLC service access points (LSAPs). All three LLC protocols employ the same PDU format (Figure 14.3), which consists of four fields.

The DSAP and SSAP fields each contain a 7-bit address, which specify the destination and source users of LLC. One bit of the DSAP indicates whether the DSAP is an individual or group address. One bit of the SSAP indicates whether the PDU is a command or response PDU. The format of the LLC control field is identical to that of HDLC using extended (7-bit) sequence numbers.

For **type 1 operation**, which supports the unacknowledged connectionless service, the unnumbered information (UI) PDU is used to transfer user data. There is no acknowledgment, flow control, or error control. However, there is error detection and discard at the MAC level. Two other PDU types, XID and TEST, are used to support management functions associated with all three types of operation. Both PDU types are used in the following fashion.

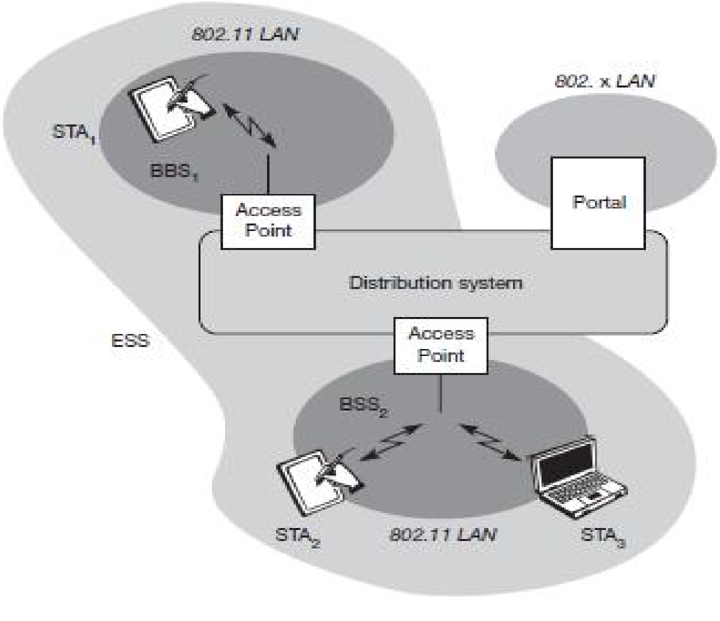
An LLC entity may issue a command (C/R bit 0) XID or TEST. The receiving LLC entity issues a corresponding XID or TEST in response. The XID PDU is used to exchange two types of information: types of operation supported and window size. The TEST PDU is used to conduct a loopback test of the transmission path between two LLC entities. Upon receipt of a TEST command PDU, the addressed LLC entity issues a TEST response PDU as soon as possible.

With **type 2 operation**, a data link connection is established between two LLC SAPs prior to data exchange. Connection establishment is attempted by the type 2 protocol in response to a request from a user. PDU2 to request a logical connection with the other LLC entity. If the connection is accepted by the LLC user designated by the DSAP, then the destination LLC entity returns an unnumbered acknowledgment (UA) PDU. The connection is henceforth uniquely identified by the pair of user SAPs. If the destination LLC user rejects the connection request, its LLC entity returns a disconnected mode (DM) PDU. Once the connection is established, data is exchanged using information PDUs, as in HDLC. The information PDUs include send and receive sequence numbers, for sequencing and flow control. The supervisory PDUs are used, as in HDLC, for flow control and error control. Either LLC entity can terminate a logical LLC connection by issuing a disconnect (DISC) PDU.

With **type 3 operation**, each transmitted PDU is acknowledged. A new (not found in HDLC) unnumbered PDU, the acknowledged connectionless (AC) information PDU, is defined. User data are sent in AC command PDUs and must be acknowledged using an AC response PDU. To guard against lost PDUs, a 1-bit sequence number is used. The sender alternates the use of 0 and 1 in its AC command PDU, and the receiver responds with an AC PDU with the opposite number of the corresponding command. Only one PDU in each direction may be outstanding at any time.

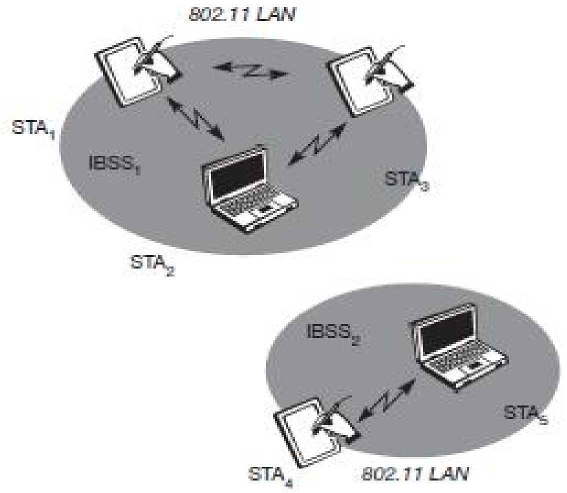
System architecture

Wireless networks can exhibit two different basic system architectures as infrastructure-based or ad-hoc.
Components of an infrastructure and a wireless part as specified for IEEE 802.11. Several nodes, called **stations (STAi)**, are connected to **access points (AP)**. Stations are terminals with access mechanisms to the wireless medium and radio contact to the AP. The stations and the AP which are within the same radio coverage form a **basic service set (BSSi)**. The example shows two BSSs – BSS1 and BSS2 – which are connected via a **distribution system**. A distribution system connects several BSSs via the AP to form a single network and thereby extends the wireless coverage area. This network is now called an **extended service set (ESS)** and has its own identifier, the ESSID. The ESSID is the 'name' of a network and is used to separate different networks. Without knowing the ESSID (and assuming no hacking) it should not be possible to participate in the WLAN. The distribution system connects the wireless networks via the APs with a **portal**, which forms the interworking unit to other LANs.



The architecture of the distribution system is not specified further in IEEE 802.11. It could consist of bridged IEEE LANs, wireless links, or any other networks.

However, **distribution system services** are defined in the standard Stations can select an AP and associate with it. The APs support roaming (i.e., changing access points), the distribution system handles data transfer between the different APs. APs provide synchronization within a BSS, support power management, and can control medium access to support time-bounded service. These and further functions are explained in the following sections.



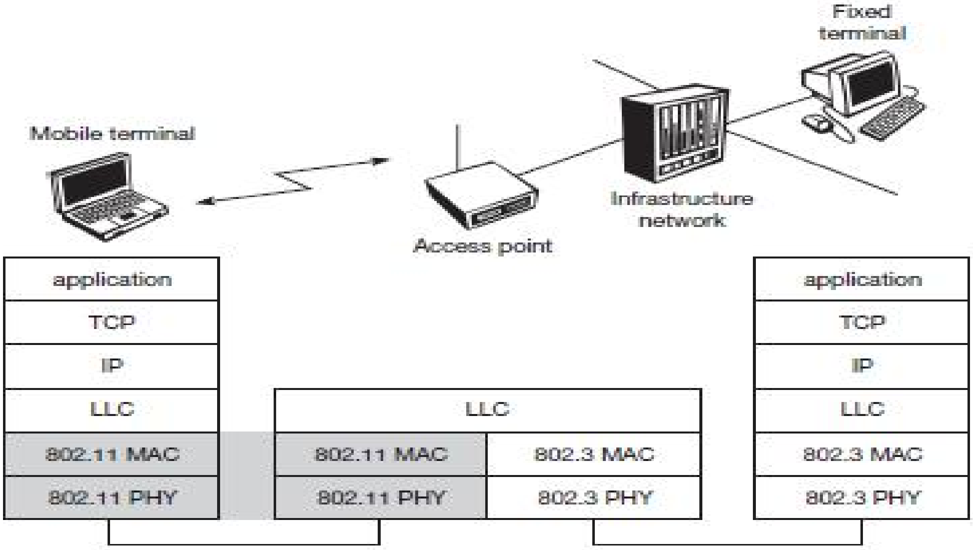
In addition to infrastructure-based networks, IEEE 802.11 allows the building of ad-hoc networks between stations, thus forming one or more independent BSSs (IBSS) as shown in Figure. In this case, an IBSS comprises a group of stations using the same radio frequency. Stations STA1, STA2, and STA3 are in IBSS1, STA4 and STA5 in IBSS2. This means for example that STA3 can communicate directly with STA2 but not with STA5. Several IBSSs can either be formed via the distance between the IBSSs or by using different carrier frequencies (then the IBSSs could overlap physically). IEEE 802.11 does not specify any special nodes that support routing, forwarding of data or exchange of topology information as, e.g., HIPERLAN 1 or Bluetooth

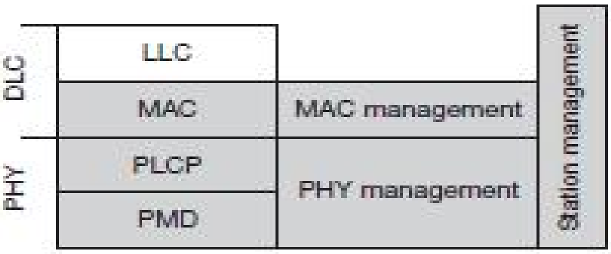
Protocol architecture

As indicated by the standard number, IEEE 802.11 fits seamlessly into the other 802.x standards for wired LANs (see Halsall, 1996; IEEE, 1990). Figure shows the most common scenario: an IEEE 802.11 wireless LAN connected to a switched IEEE 802.3 Ethernet via a bridge. Applications should not notice any difference apart from the lower bandwidth and perhaps higher access time from the wireless LAN. The WLAN behaves like a slow wired LAN. Consequently, the higher layers (application, TCP, IP) look the same for wireless nodes as for wired nodes.

The upper part of the data link control layer, the logical link control (LLC), covers the differences of the medium access control layers needed for the different media. In many of today's networks, no explicit LLC layer is visible.

The IEEE 802.11 standard only covers the physical layer **PHY** and medium access layer **MAC** like the other 802.x LANs do. The physical layer is subdivided into the **physical layer convergence protocol (PLCP)** and the **physical medium dependent** sublayer **PMD**. The basic tasks of the MAC layer comprise medium access, fragmentation of user data, and encryption. The





- PLCP sublayer provides a carrier sense signal, called clear channel assessment (CCA), and provides a common PHY service access point (SAP) independent of the transmission technology.
- PMD sublayer handles modulation and encoding/decoding of signals.

The PHY layer (comprising PMD and PLCP) and the MAC layer will be explained in more detail in the following sections. Apart from the protocol sublayers, the standard specifies management layers and the station management. The **MAC management** supports the association and re-association of a station to an access point and roaming between different access points. It also controls authentication mechanisms, encryption, synchronization of a station with regard to an access point, and power management to save battery power. MAC management also maintains the MAC management information base (MIB).

The main tasks of the **PHY management** include channel tuning and PHY MIB maintenance. Finally, **station management** interacts with both management layers and is responsible for additional higher layer functions (e.g., control of bridging and interaction with the distribution system in the case of an access point).

IEEE 802.11 Services

IEEE 802.11 defines nine services that need to be provided by the wireless LAN to provide functionality equivalent to that which is inherent to wired LANs. Table lists the services and indicates two ways of categorizing them.

1. The service provider can be either the station or the distribution system (DS). Station services are implemented in every 802.11 station, including access point (AP) stations. Distribition services are provided between basic service sets (BSSs); these services may be implemented in an AP or in another special- purpose device attaced to the distribution system.

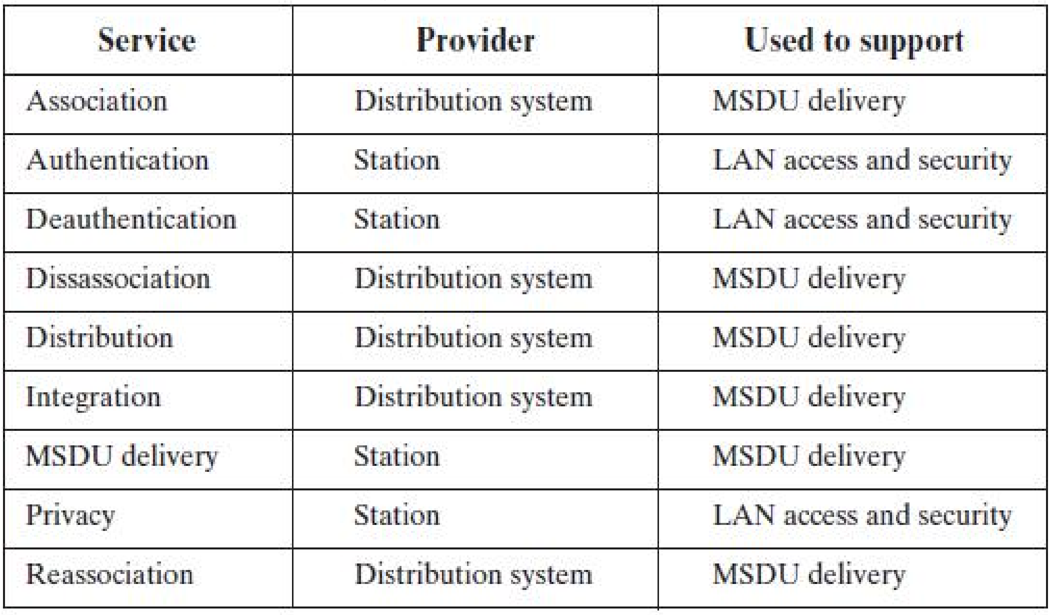
2. Three of the services are used to control IEEE 802.11 LAN access and confidentiality. Six of the services are used to support delivery of MAC service data units (MSDUs) between stations.

The MSDU is a the block of data passed down from the MAC user to the MAC layer; typically this is a LLC PDU. If the MSDU is too large to be transmitted in a single MAC frame, it may be fragmented and transmitted in a series of MAC frames. **MSDU delivery**, which is the basic service, has already been mentioned.

Distribution of Messages Within a DS

The two services involved with the distribution of messages within a DS are distribution and integration. **Distribution** is the primary service used by stations to exchange MAC frames when the frame must traverse the DS to get from a station in one BSS to a station in another BSS. . If the two stations that are communicating are within the same BSS, then the distribution service logically goes through the single AP of that BSS.

IEEE 802.11 Services



Service	Provider	Used to support
Association	Distribution system	MSDU delivery
Authentication	Station	LAN access and security
Deauthentication	Station	LAN access and security
Dissassociation	Distribution system	MSDU delivery
Distribution	Distribution system	MSDU delivery
Integration	Distribution system	MSDU delivery
MSDU delivery	Station	MSDU delivery
Privacy	Station	LAN access and security
Reassociation	Distribution system	MSDU delivery

The **integration** service enables transfer of data between a station on an IEEE 802.11 LAN and a station on an integrated IEEE 802.x LAN. The term *integrated* refers to a wired LAN that is physically connected to the DS and whose stations may be logically connected to an IEEE 802.11 LAN via the integration service. The integration service takes care of any address translation and media conversion logic required for the exchange of data.

Association-Related Services :

The primary purpose of the MAC layer is to transfer MSDUs between MAC entities; this purpose is fulfilled by the distribution service. For that service to function, it requires information about stations within the ESS that is provided by the association-related services. Before the distribution service can deliver data to or accept data from a station, that station must be *associated*. Before looking at the concept of association, we need to describe the concept of mobility. The standard defines three transition types of based on mobility:

- **No transition:** A station of this type is either stationary or moves only within the direct communication range of the communicating stations of a single BSS.
- **BSS transition:** This is defined as a station movement from one BSS to another BSS within the same ESS. In this case, delivery of data to the station requires that the addressing capability be able to recognize the new location of the station.

- **ESS transition:** This is defined as a station movement from a BSS in one ESS to a BSS within another ESS. This case is supported only in the sense that the station can move.

To deliver a message within a DS, the distribution service needs to know where the destination station is located. Specifically, the DS needs to know the identity of the AP to which the message should be delivered in order for that message to reach the destination station. To meet this requirement, a station must maintain an association with the AP within its current BSS.

Three services relate to this requirement:

- **Association:** Establishes an initial association between a station and an AP. Before a station can transmit or receive frames on a wireless LAN, its identity and address must be known. For this purpose, a station must establish an association with an AP within a particular BSS. The AP can then communicate this information to other APs within the ESS to facilitate routing and delivery of addressed frames.
- **Reassociation:** Enables an established association to be transferred from one AP to another, allowing a mobile station to move from one BSS to another.
- **Disassociation:** A notification from either a station or an AP that an existing association is terminated. A station should give this notification before leaving an ESS or shutting down. However, the MAC management facility protects itself against stations that disappear without notification.

Access and Privacy Services

There are two characteristics of a wired LAN that are not inherent in a wireless LAN.

1. In order to transmit over a wired LAN, a station must be physically connected to the LAN. On the other hand, with a wireless LAN, any station within radio range of the other devices on the LAN can transmit. In a sense, there is a form of authentication with a wired LAN, in that it requires some positive and presumably observable action to connect a station to a wired LAN.

2. Similarly, in order to receive a transmission from a station that is part of a wired LAN, the receiving station must also be attached to the wired LAN. On the other hand, with a wireless LAN, any station within radio range can receive. Thus, a wired LAN provides a degree of privacy, limiting reception of data to stations connected to the LAN.

IEEE 802.11 defines three services that provide a wireless LAN with these two features:

• **Authentication:** Used to establish the identity of stations to each other. In a wired LAN, it is generally assumed that access to a physical connection conveys authority to connect to the LAN. This is not a valid assumption for a wireless LAN, in which connectivity is achieved simply by having an attached antenna that is properly tuned.

The authentication service is used by stations to establish their identity with stations they wish to communicate with. IEEE 802.11 supports several authentication schemes and allows for expansion of the functionality of these schemes. However, IEEE 802.11 requires mutually acceptable, successful authentication before a station can establish an association with an AP.

• **Deathentication:** This service is invoked whenever an existing authentication is to be terminated.

• **Privacy:** Used to prevent the contents of messages from being read by other than the intended recipient. The standard provides for the optional use of encryption to assure privacy. The algorithm specified in the standard is WEP,

Medium access control layer

- it has to control medium access, but it can also offer support for roaming, authentication, and power conservation.

- The basic services provided by the MAC layer are the mandatory **asynchronous data service** and an optional **time-bounded service**.

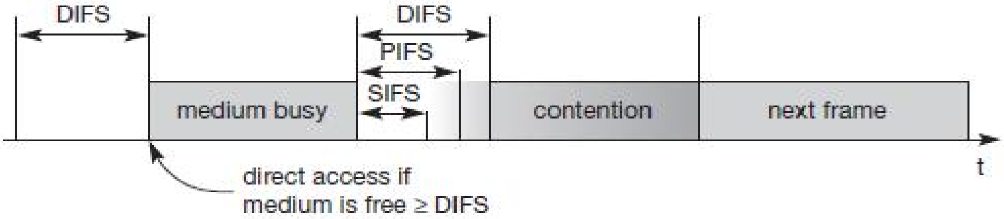
- While 802.11 only offers the asynchronous service in ad-hoc network mode, both service types can be offered using an infrastructure-based network together with the access point coordinating medium access.

- The asynchronous service supportsbroadcast and multi-cast packets, and packet exchange is based on a ‘best effort’ model, i.e., no delay bounds can be given for transmission.

The following three basic access mechanisms have been defined for IEEE 802.11: the mandatory basic method based on a version of CSMA/CA, an optional method avoiding the hidden terminal problem, and finally a contention- free polling method for time-bounded service. The first two methods are also summarized as **distributed coordination function (DCF)**, the third method is called **point coordination function (PCF**). DCF only offers asynchronous service, while PCF offers both asynchronous and time-bounded service but needs an access point to control medium access and to avoid contention. The MAC mechanisms are also called **distributed foundation wireless medium access control (DFWMAC)**.

For all access methods, several parameters for controlling the waiting time before medium access are important. Figure shows the three different parameters that define the priorities of medium access. The values of the parameters depend on the PHY and are defined in relation to a **slot** time. Slot time is derived from the medium propagation delay, transmitter delay, and other PHY dependent parameters. Slot time is 50 μs for FHSS and 20 μs for DSSS.

The medium, as shown, can be busy or idle (which is detected by the CCA). If the medium is busy this can be due to data frames or other control frames. During a contention phase several nodes try to access the medium.



● **Short inter-frame spacing (SIFS):** The shortest waiting time for medium access (so the highest priority) is defined for short control messages, such as acknowledgements of data packets or polling responses. For DSSS SIFS is 10 μs and for FHSS it is 28 μs.

● **PCF inter-frame spacing (PIFS):** A waiting time between DIFS and SIFS (and thus a medium priority) is used for a time-bounded service. An access point polling other nodes only has to wait PIFS for medium access PIFS is defined as SIFS plus one slot time.

● **DCF inter-frame spacing (DIFS):** This parameter denotes the longest waiting time and has the lowest priority for medium access. This waiting time is used for asynchronous data service within a contention period DIFS is defined as SIFS plus two slot times.

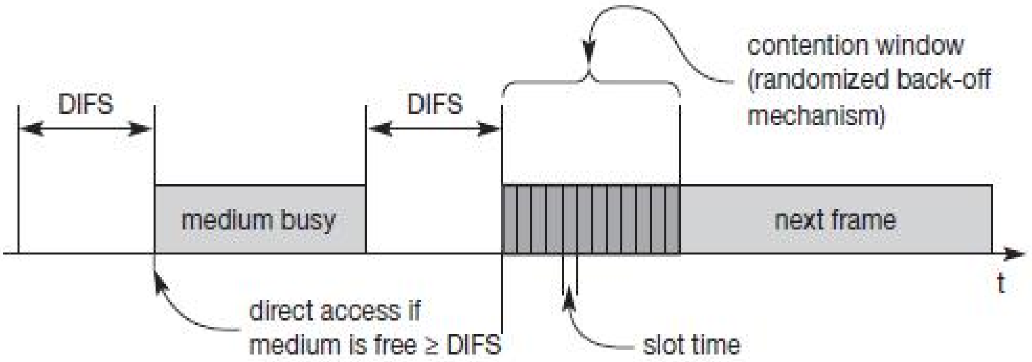
Basic DFWMAC-DCF using CSMA/CA :

- The mandatory access mechanism of IEEE 802.11 is based on **carrier sense multiple access with collision avoidance** (CSMA/CA)

- which is a random access scheme with carrier sense and collision avoidance through random backoff.

- If the medium is idle for at least the duration of DIFS (with the help of the CCA signal of the physical layer), a node can access the medium at once.

- This allows for short access delay under light load. But as more and needed.



- If the medium is busy, nodes have to wait for the duration of DIFS, entering a contention phase afterwards.

- Each node now chooses a **random backoff time** within a **contention window** and delays medium access for this random amount of time.

- The node continues to sense the medium. As soon as a node senses the channel is busy, it has lost this cycle and has to wait for the next chance, i.e., until the medium is idle again for at least DIFS.

- But if the randomized additional waiting time for a node is over and the medium is still idle, the node can access the medium immediately (i.e., no other node has a shorter waiting time).

- The additional waiting time is measured in multiples of the above-mentioned slots.

- This additional randomly distributed delay helps to avoid collisions – otherwise all stations would try to transmit data after waiting for the medium becoming idle again plus DIFS.

- While this process describes the complete access mechanism for broadcast frames, an additional feature is provided by the standard for unicast data transfer.

- But now, the receiver answers directly with an **acknowledgement (ACK)**.

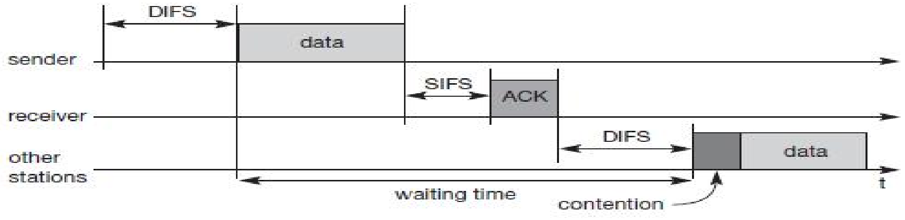
- The receiver accesses the medium after waiting for a duration of SIFS so no other station can access the medium in the meantime and cause a collision.

- The other stations have to wait for DIFS plus their backoff time.

- This acknowledgement ensures the correct reception (correct checksum CRC at the receiver) of a frame on the MAC layer, which is especially important in error-prone environments such as wireless connections.

- If no ACK is returned, the sender automatically retransmits the frame.

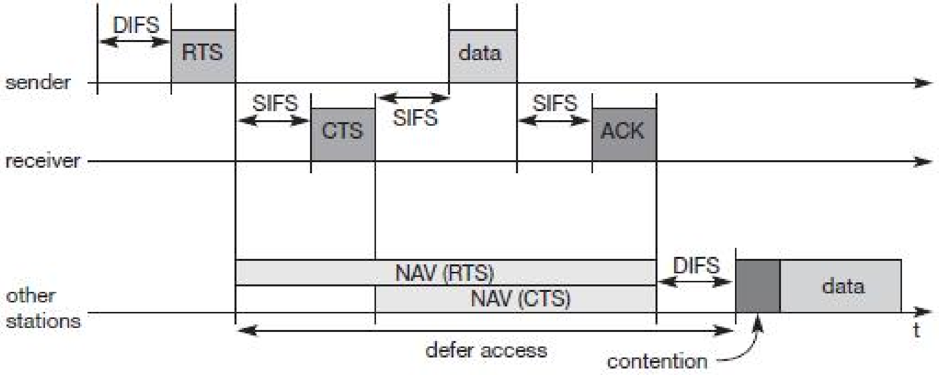
- The number of retransmissions is limited, and final failure is reported to the higher layer.



DFWMAC-DCF with RTS/CTS extension :

The problem of hidden terminals, a situation that can also occur in IEEE 802.11 networks. This problem occurs if one station can receive two others, but those stations cannot receive each other. The two stations may sense the channel is idle, send a frame, and cause a collision at the receiver in the middle. To deal with this problem, the standard defines an additional mechanism using two control packets, RTS and CTS. The use of the mechanism is optional; however, every 802.11 node has to implement the functions to react properly upon reception of RTS/CTS control packets. After waiting for DIFS (plus a random backoff time if the medium was busy), the sender can issue a **request to send (RTS)** control packet. The RTS packet thus is not given any higher priority compared to other data packets. The RTS packet includes the receiver of the data transmission to come and the duration of the whole data transmission. This duration specifies the time interval necessary to transmit the whole data frame and the acknowledgement related to it.

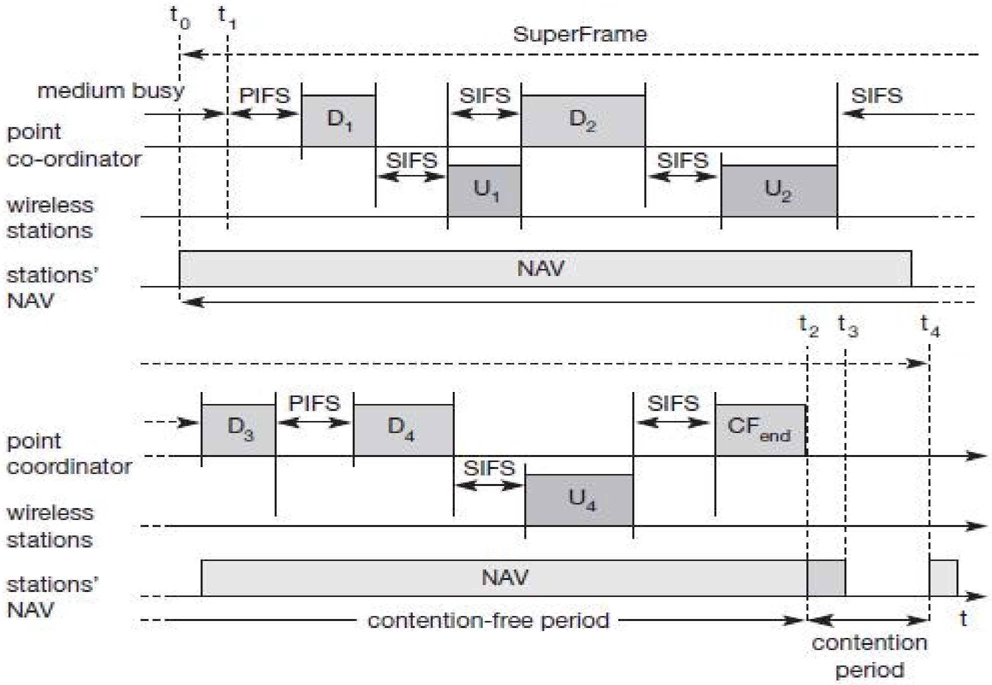
Every node receiving this RTS now has to set its **net allocation vector (NAV)** in accordance with the duration field. The NAV then specifies the earliest point at which the station can try to access the medium again. If the receiver of the data transmission receives the RTS, it answers with a **clear to send (CTS)** message after waiting for SIFS. This CTS packet contains the duration field again and all stations receiving this packet from the receiver of the intended data transmission have to adjust their NAV. The latter set of receivers need not be the same as the first set receiving the RTS packet. Now all nodes within receiving distance around sender and receiver are informed that they have to wait more time before accessing the medium. Basically, this mechanism reserves the medium for one sender exclusively (this is why it is sometimes called a virtual reservation scheme).



Finally, the sender can send the data after SIFS. The receiver waits for SIFS after receiving the data packet and then acknowledges whether the transfer was correct. The transmission has now been completed, the NAV in each node marks the medium as free and the standard cycle can start again.

DFWMAC-PCF with polling :

The two access mechanisms presented so far cannot guarantee a maximum access delay or minimum transmission bandwidth. To provide a time-bounded service, the standard specifies a **point coordination function (PCF)** on top of the standard DCF mechanisms. Using PCF requires an access point that controls medium access and polls the single nodes. Ad-hoc networks cannot use this function so, provide no QoS but ‘best effort’ in IEEE 802.11 WLANs. The **point co-ordinator** in the access point splits the access time into super frame periods as shown in Figure. A **super frame** comprises a **contention free period** and a **contention period**. The contention period can be used for the two access mechanisms presented above. The figure also shows several wireless stations (all on the same line) and the stations’ NAV (again on one line).



At time t0 the contention-free period of the super frame should theoretically start, but another station is still transmitting data (i.e., the medium is busy). This means that PCF also defers to DCF, and the start of the super frame may be postponed. The only possibility of avoiding variations is not to have any contention period at all. After the medium has been idle until t1, the point coordinator has to wait for PIFS before accessing the medium. As PIFS is smaller than DIFS, no other station can start sending earlier. The point coordinator now sends data D1 downstream to the first wireless station. This station can answer at once after SIFS. After waiting for SIFS again, the point coordinator can poll the second station by sending D2. This station may answer upstream to the coordinator with data U2. Polling continues with the third node. This time the node has nothing to answer and the point coordinator will not receive a packet after SIFS. After waiting for PIFS, the coordinator can resume polling the stations. Finally, the point coordinator can issue an end marker (CFend), indicating that the contention period may start again. Using PCF automatically sets the NAV, preventing other stations from sending. In the example, the contention-free period planned initially would have been from t0 to t3. However, the point coordinator finished polling earlier, shifting the end of the contention-free period to t2. At t4, the cycle starts again with the next super frame.

The transmission properties of the whole wireless network are now determined by the polling behavior of the access point. If only PCF is used and polling is distributed evenly, the bandwidth is also distributed evenly among all polled nodes. This would resemble a static, centrally controlled time division multiple access (TDMA) system with time division duplex (TDD) transmission. This method comes with an overhead if nodes have nothing to send, but the access point polls them permanently.

MAC frames

Figure shows the basic structure of an IEEE 802.11 MAC data frame together with the content of the frame control field. The fields in the figure refer to the following:

● **Frame control:** The first 2 bytes serve several purposes. They contain several sub-fields as explained after the MAC frame.

● **Duration/ID:** If the field value is less than 32,768, the duration field contains the value indicating the period of time in which the medium is occupied (in μs). This field is used for setting the NAV for the virtual reservation mechanism using RTS/CTS and during fragmentation. Certain values above 32,768 are reserved for identifiers.

● **Address 1 to 4:** The four address fields contain standard IEEE 802 MAC addresses (48 bit each), as they are known from other 802.x LANs. The meaning of each address depends on the DS bits in the frame control field and is explained in more detail in a separate paragraph.

● **Sequence control:** Due to the acknowledgement mechanism frames may be duplicated. Therefore a sequence number is used to filter duplicates.

● **Data:** The MAC frame may contain arbitrary data (max. 2,312 byte), which is transferred transparently from a sender to the receiver(s).

● **Checksum (CRC):** Finally, a 32 bit checksum is used to protect the frame as it is common practice in all 802.x networks. The frame control field shown in Figure contains the following fields:

● **Protocol version:** This 2 bit field indicates the current protocol version and is fixed to 0 by now. If major revisions to the standard make it incompatible with the current version, this value will be increased.

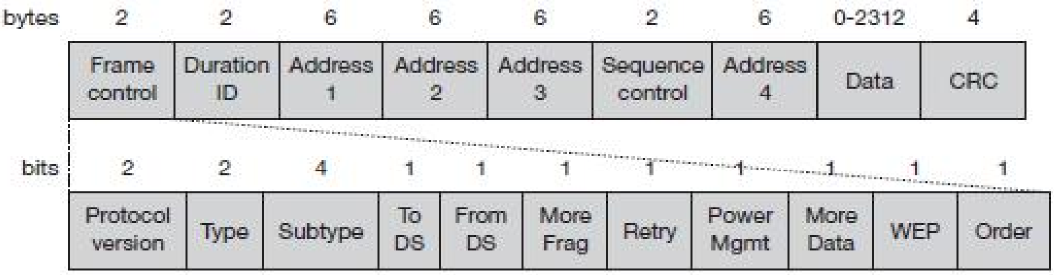
● **Type:** The type field determines the function of a frame: management (=00), control (=01), or data (=10). The value 11 is reserved. Each type has several subtypes as indicated in the following field.

● **Subtype:** Example subtypes for management frames are: 0000 for association request, 1000 for beacon. RTS is a control frame with subtype 1011, CTS is coded as 1100. User data is transmitted as data frame with subtype 0000. All details can be found in IEEE, 1999.

● **To DS/From DS** .

● **More fragments:** This field is set to 1 in all data or management frames that have another fragment of the current MSDU to follow.

- **Retry:** If the current frame is a retransmission of an earlier frame, this bit is set to 1. With the help of this bit it may be simpler for receivers to eliminate duplicate frames.
- **Power management:** This field indicates the mode of a station after successful transmission of a frame. Set to 1 the field indicates that the station goes into power-save mode. If the field is set to 0, the station stays active.
- **More data:** In general, this field is used to indicate a receiver that a sender has more data to send than the current frame. This can be used by an access point to indicate to a station in power-save mode that more packets are buffered. Or it can be used by a station to indicate to an access point after being polled that more polling is necessary as the station has more data ready to transmit.
- **Wired equivalent privacy (WEP):** This field indicates that the standard security mechanism of 802.11 is applied. However, due to many weaknesses found in the WEP algorithm higher layer security should be used to securean 802.11 network
- **Order:** If this bit is set to 1 the received frames must be processed in strict order.



bytes	2	2	6	6	6	2	6	0-2312	4
	Frame control	Duration ID	Address 1	Address 2	Address 3	Sequence control	Address 4	Data	CRC

bits	2	2	4	1	1	1	1	1	1	1	1
	Protocol version	Type	Subtype	To DS	From DS	More Frag	Retry	Power Mgmt	More Data	WEP	Order

MAC frames can be transmitted between mobile stations; between mobile stations and an access point and between access points over a DS .Two bits within the Frame Control field, ‘**to DS**’ and ‘**from DS**’, differentiate these cases and control the meaning of the four addresses used.

MAC management :

MAC management plays a central role in an IEEE 802.11 station as it more or less controls all functions related to system integration, i.e., integration of a wireless station into a BSS, formation of an ESS, synchronization of stations etc.

- **Synchronization:** Functions to support finding a wireless LAN, synchronization of internal clocks, generation of beacon signals.

● **Power management:** Functions to control transmitter activity for power conservation, e.g., periodic sleep, buffering, without missing a frame.
● **Roaming:** Functions for joining a network (association), changing access points, scanning for access points.
● **Management information base (MIB):** All parameters representing the current state of a wireless station and an access point are stored within a MIB for internal and external access. A MIB can be accessed via standardized protocols such as the simple network management protocol (SNMP).

Synchronization:

Each node of an 802.11 network maintains an internal clock. To synchronize the clocks of all nodes, IEEE 802.11 specifies a **timing synchronization function (TSF)**. synchronized clocks are needed for power management, but also for coordination of the PCF and for synchronization of the hopping sequence in an FHSS system. Using PCF, the local timer of a node can predict the start of a super frame, i.e., the contention free and contention period. FHSS physical layers need the same hopping sequences so that all nodes can communicate within a BSS. Within a BSS, timing is conveyed by the (quasi)periodic transmissions of a beacon frame. A **beacon** contains a timestamp and other management information used for power management and roaming (e.g., identification of the BSS). The timestamp is used by a node to adjust its local clock. The node is not required to hear every beacon to stay synchronized; however, from time to time internal clocks should be adjusted. The transmission of a beacon frame is not always periodic because the beacon frame is also deferred if the medium is busy. Within **infrastructure-based** networks, the access point performs synchronization by transmitting the (quasi)periodic beacon signal, whereas all other wireless nodes adjust their local timer to the time stamp. The access point is not always able to send its beacon periodically if the medium is busy. However, the access point always tries to schedule transmissions according to the expected beacon interval (**target beacon transmission time**), i.e., beacon intervals are not shifted if one beacon is delayed. The timestamp of a beacon always reflects the real transmit time, not the scheduled time.

For ad-hoc networks, the situation is slightly more complicated as they do not have an access point for beacon transmission. In this case, each node maintains its own synchronization timer and starts the transmission of a beacon frame after the beacon interval. where multiple stations try to send their beacon. However, the standard random backoff algorithm is also applied to the beacon frames so only one beacon wins.

All other stations now adjust their internal clocks according to the received beacon and suppress their beacons for this cycle. If collision occurs, the beacon is lost. In this scenario, the beacon intervals can be shifted slightly because all clocks may vary as may the start of a beacon interval from a node's point of view. However, after successful synchronization all nodes again have the same consistent view.

Power management

Wireless devices are battery powered (unless a solar panel is used). Therefore, power-saving mechanisms are crucial for the commercial success of such devices. Standard LAN protocols assume that stations are always ready to receive data, although receivers are idle most of the time in lightly loaded networks. However, this permanent readiness of the receiving module is critical for battery life as the receiver current may be up to 100 mA (Woesner, 1998). The basic idea of IEEE 802.11 power management is to switch off the transceiver whenever it is not needed. For the sending device this is simple to achieve as the transfer is triggered by the device itself. However, since the power management of a receiver cannot know in advance when the transceiver has to be active for a specific packet, it has to 'wake up' the transceiver periodically. Switching off the transceiver should be transparent to existing protocols and should be flexible enough to support different applications. However, throughput can be traded-off for battery life. Longer off-periods save battery life but reduce average throughput and vice versa.

The basic idea of power saving includes two states for a station: **sleep** and **awake**, and buffering of data in senders. If a sender intends to communicate with a power-saving station it has to buffer data if the station is asleep. The sleeping station on the other hand has to wake up periodically and stay awake for a certain time. During this time, all senders can announce the destinations of their buffered data frames. If a station detects that it is a destination of a buffered packet it has to stay awake until the transmission takes place. Waking up at the right moment requires the **timing synchronization function (TSF).** All stations have to wake up or be awake at the same time.

Roaming

Typically, wireless networks within buildings require more than just one access point to cover all rooms. Depending on the solidity and material of the walls, one access point has a transmission range of 10–20 m if transmission is to be of decent quality. Each storey of a building needs its own access point(s) as quite often walls are thinner than floors. If a user walks around with a wireless station, the station has to move from one access point to another to provide uninterrupted service. Moving between access points is called **roaming**. The term “handover” or “handoff” as used in the context of mobile or cellular phone systems would be more appropriate as it is simply a change of the active cell. However, for WLANs roaming is more common.

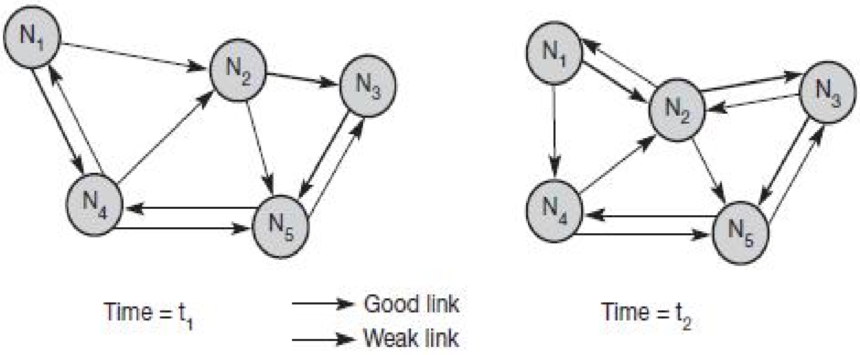
The steps for roaming between access points are:
● A station decides that the current link quality to its access point AP1 is too poor. The station then starts **scanning** for another access point.
● Scanning involves the active search for another BSS and can also be used for setting up a new BSS in case of ad-hoc networks.

IEEE 802.11 specifies scanning on single or multiple channels (if available at the physical layer) and differentiates between passive scanning and active scanning. **Passive scanning** simply means listening into the medium to find other networks, i.e., receiving the beacon of another network issued by the synchronization function within an access point. **Active scanning** comprises sending a **probe** on each channel and waiting for a response. Beacon and probe responses contain the information necessary to join the new BSS.

● The station then selects the best access point for roaming based on, e.g., signal strength, and sends an **association request** to the selected access point AP2.
● The new access point AP2 answers with an **association response**. If the response is successful, the station has roamed to the new access point AP2. Otherwise, the station has to continue scanning for new access points.
● The access point accepting an association request indicates the new station in its BSS to the distribution system (DS). The DS then updates its database, which contains the current location of the wireless stations. This database is needed for forwarding frames between different BSSs, i.e. between the different access points controlling the BSSs, which combine to form an ESS. Additionally, the DS can inform the old access point AP1 that the station is no longer within its BSS.

Routing

While in wireless networks with infrastructure support a base station always reaches all mobile nodes, this is not always the case in an ad-hoc network. A destination node might be out of range of a source node transmitting packets. Routing is needed to find a path between source and destination and to forward the packets appropriately. In wireless networks using an infrastructure, cells have been defined. Within a cell, the base station can reach all mobile nodes without routing via a broadcast. In the case of ad-hoc networks, each node must be able to forward data for other nodes. This creates many additional problems that are discussed in the following paragraphs. Figure 8.20 gives a simple example of an ad-hoc network. At a certain time t1 the network topology might look as illustrated on the left side of the figure. Five nodes, N1 to N5, are connected depending on the current transmission characteristics between them. In this snapshot of the network, N4 can receive N1 over a good link, but N1 receives N4 only via a weak link. Links do not necessarily have the same characteristics in both directions. The reasons for this are, e.g., different antenna characteristics or transmit power. N1 cannot receive N2 at all, N2 receives a signal from N1.



This situation can change quite fast as the snapshot at t2 shows. N1 cannot receive N4 any longer, N4 receives N1 only via a weak link. But now N1 has an asymmetric but bi-directional link to N2 that did not exist before.

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

inefficient. A hop counter is needed in each packet to avoid looping, and the diameter of the ad-hoc network, i.e., the maximum number of hops, should be known.

The following sections give three examples for routing algorithms that

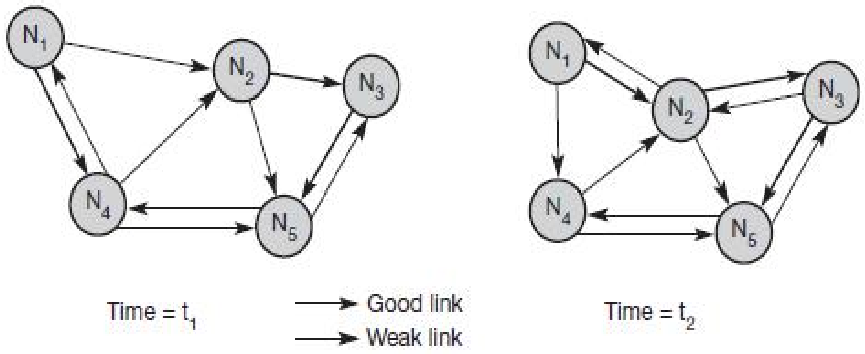
Destination sequence distance vector (DSDV):

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

DSDV now adds two things to the distance vector algorithm:

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

● **Damping:** Transient changes in topology that are of short duration should not destabilize the routing mechanisms. Advertisements containing changes in the topology currently stored are therefore not disseminated further. A node waits with dissemination if these changes are probably unstable. Waiting time depends on the time between the first and the best announcement of a path to a certain destination.



For each node N1 stores the next hop toward this node, the metric (here number of hops), the sequence number of the last advertisement for this node, and the time at which the path has been installed first. The table contains flags and a settling time helping to decide when the path can be assumed stable. Router advertisements from N1 now contain destination address, metric, and sequence number. Besides being loop-free at all times, DSDV has low memory requirements and a quick convergence via triggered updates.

Dynamic source routing (DSR) :

Imagine what happens in an ad-hoc network where nodes exchange packets from time to time, i.e., the network is only lightly loaded, and DSDV or one of the traditional distance vector or link state algorithms is used for updating routing tables. Although only some user data has to be transmitted, the nodes exchange routing information to keep track of the topology. These algorithms maintain routes between all nodes, although there may currently be no data exchange at all. This causes unnecessary traffic and prevents nodes from saving battery power.

Dynamic source routing (DSR), therefore, divides the task of routing into two separate problems :

● **Route discovery:** A node only tries to discover a route to a destination if it has to send something to this destination and there is currently no known route.

● **Route maintenance:** If a node is continuously sending packets via a route, it has to make sure that the route is held upright. As soon as a node detects problems with the current route, it has to find an alternative. The basic principle of source routing is also used in fixed networks, e.g. token rings.

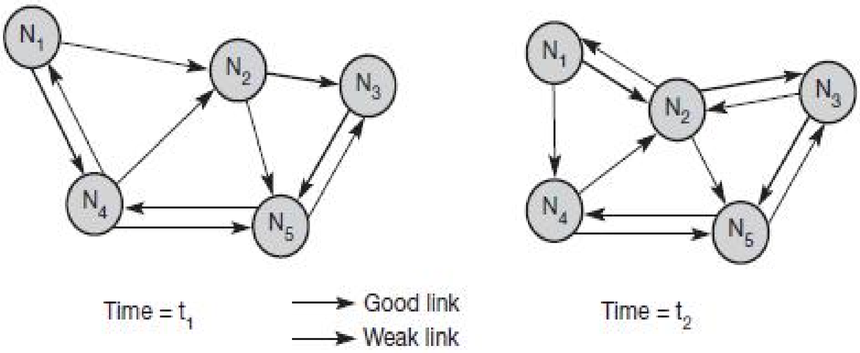
Dynamic source routing eliminates all periodic routing updates and works as follows. If a node needs to discover a route, it broadcasts a route request with a unique identifier and the destination address as parameters. Any node that receives a route request does the following.

● If the node has already received the request (which is identified using the unique identifier), it drops the request packet.

● If the node recognizes its own address as the destination, the request has reached its target.

● Otherwise, the node appends its own address to a list of traversed hops in the packet and broadcasts this updated route request.

Using this approach, the route request collects a list of addresses representing a possible path on its way towards the destination. As soon as the request reaches the destination, it can return the request packet containing the list to the receiver using this list in reverse order. One condition for this is that the links work bi-directionally. If this is not the case, and the destination node does not currently maintain a route back to the initiator of the request, it has to start a route discovery by itself. The destination may receive several lists containing different paths from the initiator. It could return the best path, the first path, or several paths to offer the initiator a choice. Applying route discovery to the example in Figure for a route from N1 to N3 at time t1 results in the following.



● N1 broadcasts the request ((N1), id = 42, target = N3), N2 and N4 receive this request.
● N2 then broadcasts ((N1, N2), id = 42, target = N3), N4 broadcasts ((N1, N4), id = 42, target = N3). N3 and N5 receive N2's broadcast, N1, N2, and N5 receive N4's broadcast.
● N3 recognizes itself as target, N5 broadcasts ((N1, N2, N5), id = 42, target = N3). N3 and N4 receive N5's broadcast. N1, N2, and N5 drop N4's broadcast packet, because they all recognize an already received route request (and N2's broadcast reached N5 before N4's did).
● N4 drops N5's broadcast, N3 recognizes (N1, N2, N5) as an alternate, but longer route.
● N3 now has to return the path (N1, N2, N3) to N1. This is simple assuming symmetric links working in both directions. N3 can forward the information using the list in reverse order.

The basic algorithm for route discovery can be optimized in many ways.

● To avoid too many broadcasts, each route request could contain a counter. Every node rebroadcasting the request increments the counter by one. Knowing the maximum network diameter (take the number of nodes if nothing else is known), nodes can drop a request if the counter reaches this number.
● A node can cache path fragments from recent requests. These fragments can now be used to answer other route requests much faster (if they still reflect the topology!).
● A node can also update this cache from packet headers while forwarding other packets.
● If a node overhears transmissions from other nodes, it can also use this information for shortening routes

Ad hoc On-Demand Distance Vector (AODV) Routing :

•AODV enables “dynamic, self-starting, multi-hop routing between mobile nodes wishing to establish and maintain an ad hoc network”
•AODV allows for the construction of routes to specific destinations and does not require that nodes keep these routes when they are not in active communication.
•AODV avoids the “counting to infinity” problem by using destination sequence numbers. This makes AODV loop-free.

•AODV defines 3 message types:
–Route Requests (RREQs)

–Route Replies (RREPs)

–Route Errors (RERRs)

•RREQ messages are used to initiate the route finding process.
•RREP messages are used to finalize the routes.
•RERR messages are used to notify the network of a link breakage in an active route.

•The AODV protocol is only used when two endpoints do not have a valid active route to each other.
•Nodes keep a “precursor list” that contains the IP address for each of its neighbors that are likely to use it for a next hop in their routing table.
•Route table information must be kept for all routes even short-lived routes.

•The routing table fields used by AODV are:
–Destination IP Address

–Destination Sequence Number

–Valid Destination Sequence number flag

–Other state and routing flags

–Network Interface

–Hop Count

–Next Hop

–List of Precursors

–Lifetime

• AODV protocol is designed for mobile ad hoc networks of tens to thousands of nodes.

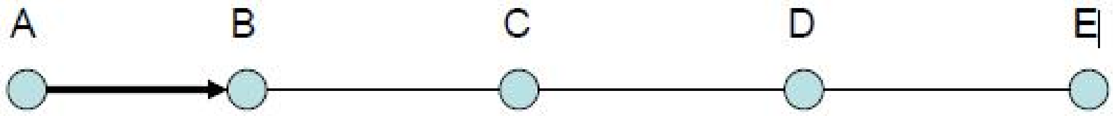
•The protocol was also designed to work in a network where all the nodes trust each other.

•Node A wants to send a message to node E.

•A valid route must be created between A and E.

•Node A generates a RREQ message with initial TTL of 1 and broadcast it to its neighbors. (In this case node B)

•The Message contains among other items node A's IP address and the IP address of node E.



•IF node B has an active route to node E then B will send a RREP message back to node A.

•If A sets a special flag in the RREQ message, node B will also send a "gratuitous" RREP message to node E.

•This will be necessary if node B will need to send packets back to A, i.e. TCP connection.

•RREP messages are unicast to the next hop toward the originator or destination if it is a gratuitous RREP.

•If A does not receive a RREP message within a certain time, it will re-broadcast the RREQ message with an incremented TTL value.

•Default increment is 2

•"Reverse" routes to the originator, in this case node A, are created as RREQ messages are forwarded.

•Active route is established when A receives a RREP message.

•This behavior (Incrementing TTL) keeps network utilization down.

•The proper maintenance of sequence numbers is crucial to keeping AODV loop-free and thereby avoiding the "counting to infinity" problem.

MOBILE AGENT :

An agent is a program that is autonomous enough to act independently even when the user or application that launched it is not available to provide guidance and handle errors.A mobile agent is an agent that can move through a heterogeneous network under its own control migrating from host to host and interacting with other agents and resources on each typically returning to its home site when its task is done

Mobile agents are programs that can move through a network under their own control mi grating from host to host and interacting with other agents and resources on each. It performs a distinct operation on user's behalf, the user who assigned that task to the MA, attempting in the completion of that task with no additional involvement of that user

MA independently transmits itself from one AP to other AP on which MA will communicate like a local agent with other object, data.

For many applications, MA becomes an efficient solution, for several points

Intelligence: Agents has an ability to learn, search with a domain knowledge that is the reason they are called as intelligent agent to possess a degree of domain knowledge. The ability of understanding determines the learning and adapting to the requirement behavior of agent in the logic so as to handle new situation very effectively . Thus in this research multi database in catastrophe healthcare system, agents plays an important role in making critical decision on behalf of the patient, paramedic in order to select the hospitals

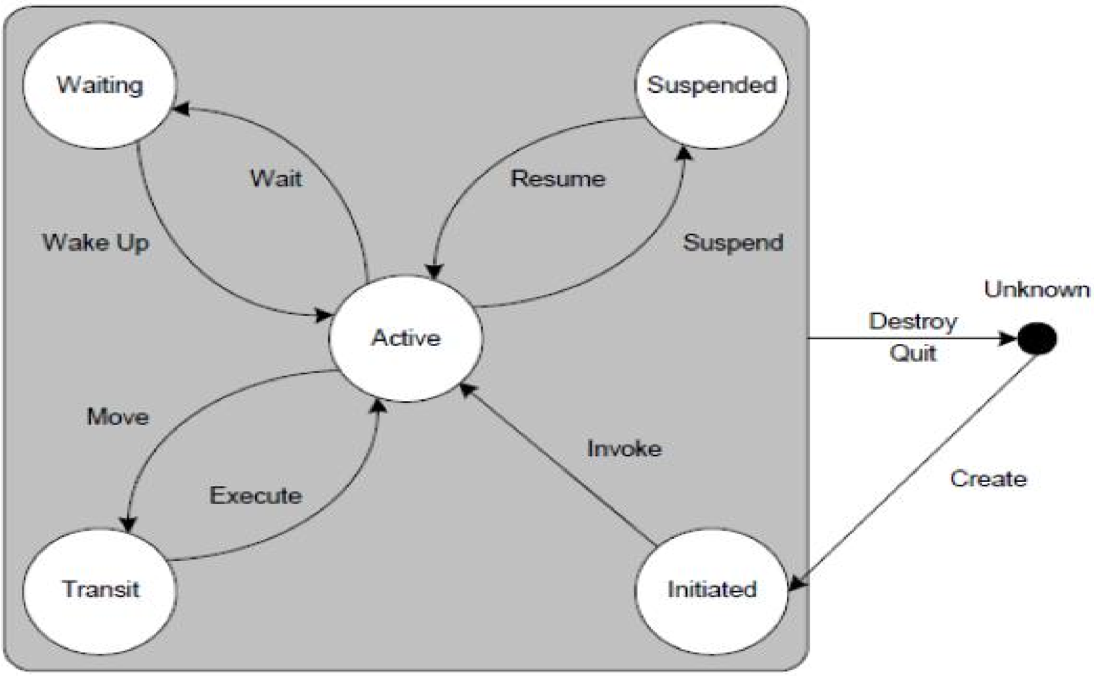
Autonomy: Autonomous means, agents are not only unreceptively motivated by outside actions initiated by the user or systems, but also an agent has internal events witch decided the agent performance and behavior. The mobile agents take an autonomous decision while selecting a perfect hospital in an emergency situation, is proven by research.

Mobility: Intelligent agents possess some degree of mobility. The agent is not limited to its home node. Migrating to host platform where it can carry out tasks locally is its capability, thus reducing processing load on its home platform, and reducing communication overhead. This is an attractive capability, in the context of distributed processing and balancing a load. Adding to this, there is another benefit that is the opportunity that even after the location it had originated from having gone offline; the agents will still keep functioning. Movable agents are called mobile agents.

Communicative: Communicating effectively with other agents, users and systems is a capability of Intelligent Agent. An Agent Communication Language (ACL) (for example KQML) helps carry out inter-agent communication. The KQML language assumes the specific anthologies for specific agent systems.

Mobile agent is autonomous in nature and should be having the capability to operate with no straight external input. In other words, some degree of control over their data and states should be held by them. While communicating with the environment and other agents, MAs should be interactive and adaptive. In other words, they should have the ability to respond to other agents or their environment. Mobility is the center possessions in mobile agent theory, is where the agent having talent to transfer it to various host within the same platform or on different environment autonomously.

A life-cycle Model



INITIATED- Even if the object agent is created, doesn't get a name and address unless it is registered with the AMS, and not able to talk other agents or systems.

ACTIVE - After registering with the Amsted agent object get a regular name and address and now able to use all JADE features which are FIPA compliant.

SUSPENDED - this is state which indicates that Agent object is stopped at present. So agent behavior is not executed.

DELETED - the Agent is definitely dead. The MA will not remain registered with Agent Managing Service and the internal thread has terminated its execution

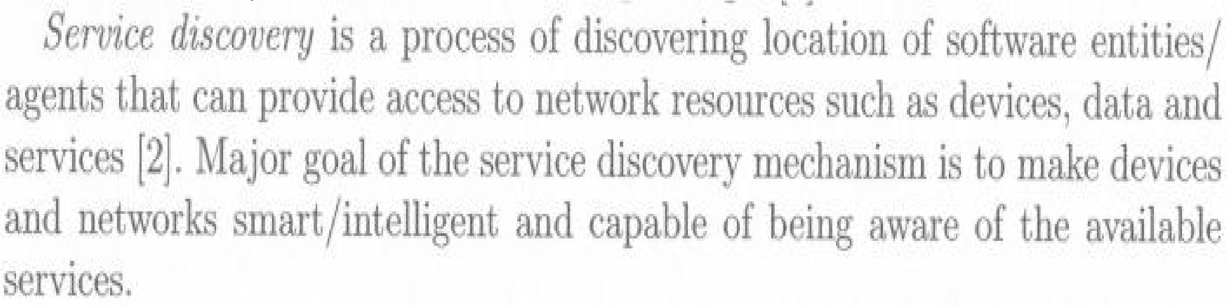
WAITING - when the object agent wants some resources, it is blocked. This is called as a waiting stage.

TRANSIT – while migrating, it within different node of same AP, the MA is enters in this state. All the buffered messages will be sent to the new place of agent by the system. To perform transitions between the various states, Agent class comes with public methods. For example, the method named do Wait() is used to put an agent into a WAITING state when it is in a state ACTIVE.

Some major advantages of using mobile agent technologies are :

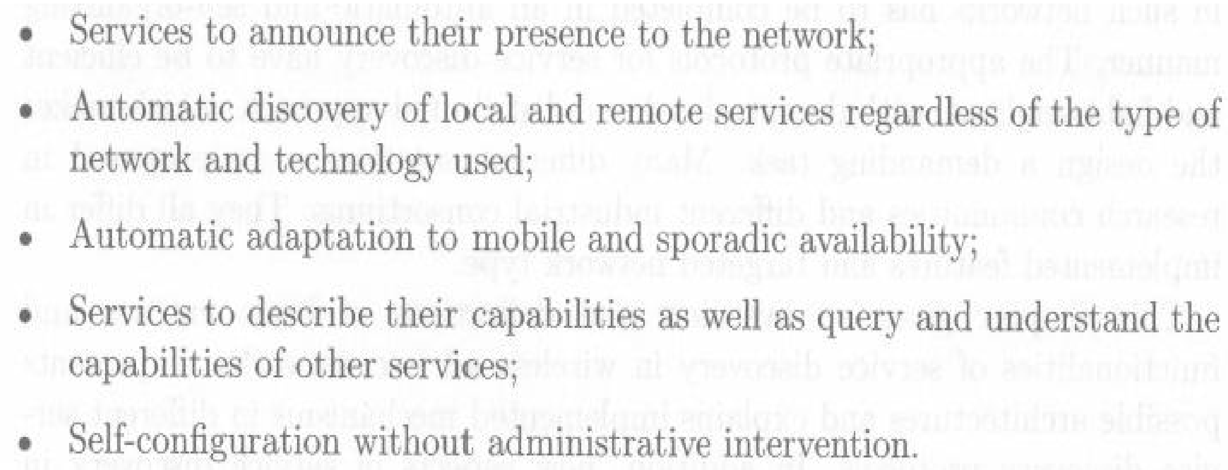
- With the use of mobile agent, it becomes a reduction in network traffic, because a mobile agent transfer itself with a state information, which is often very small than a data so reduces the network traffic.
- Even in case of disconnected operation mobile agent can perform in an asynchronous autonomous, as an agent can act on behalf of the user when the user is not present.
- It is possible for a mobile agent to interaction in real-time systems, which may prevent delays caused by network failure.
- MA can execute on single node at a time, CPU consumption is limited thus it saves an Efficiency.
- Mobile agent always Support for heterogeneous environments:
- A mobile agent can be exchanged virtually, so it is very easy software upgrades**.**

Service discovery :

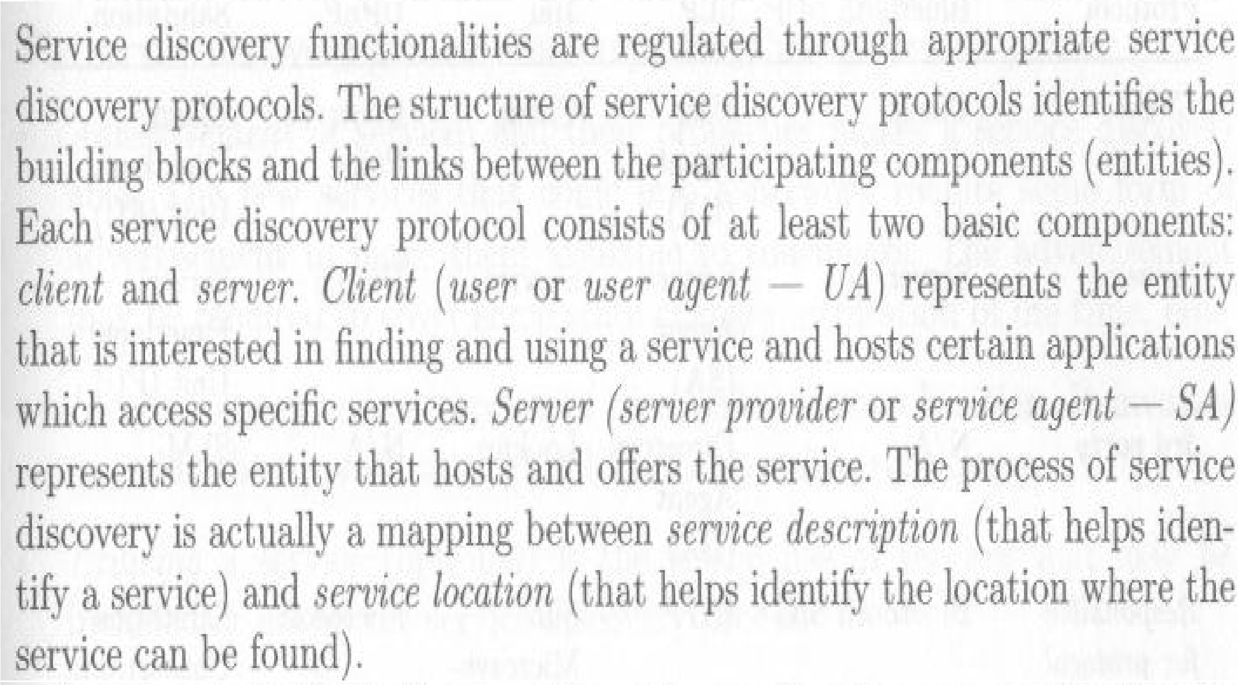


Service discovery is a process of discovering location of software entities/ agents that can provide access to network resources such as devices, data and services [2]. Major goal of the service discovery mechanism is to make devices and networks smart/intelligent and capable of being aware of the available services.

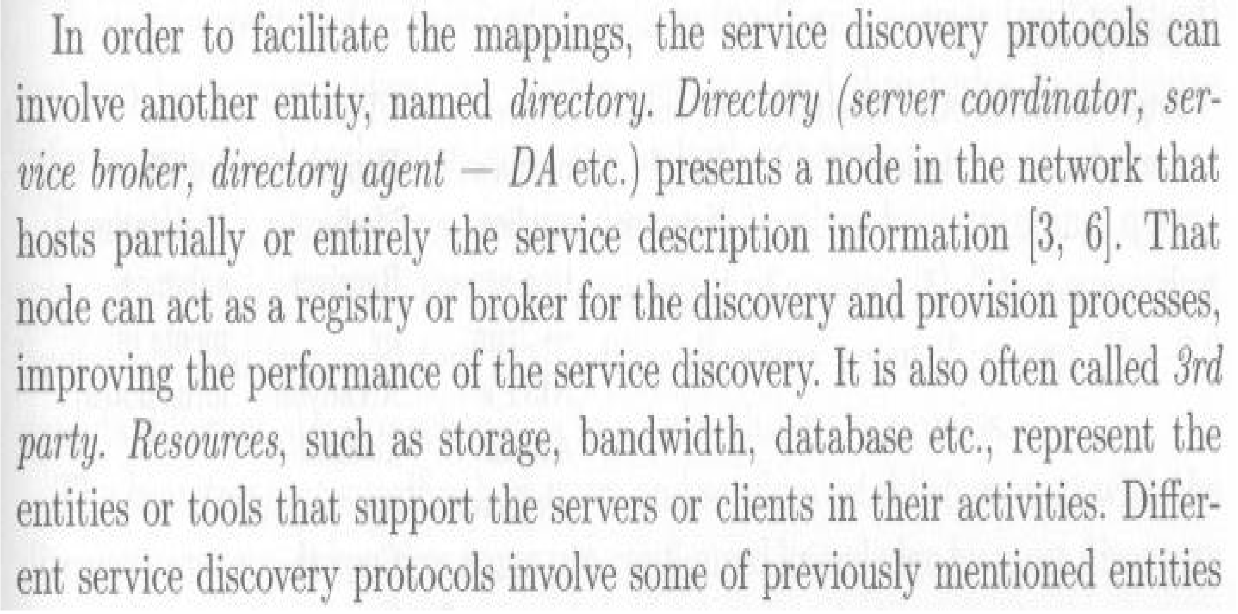
The Service discovery in ad-hoc network should facilitate:



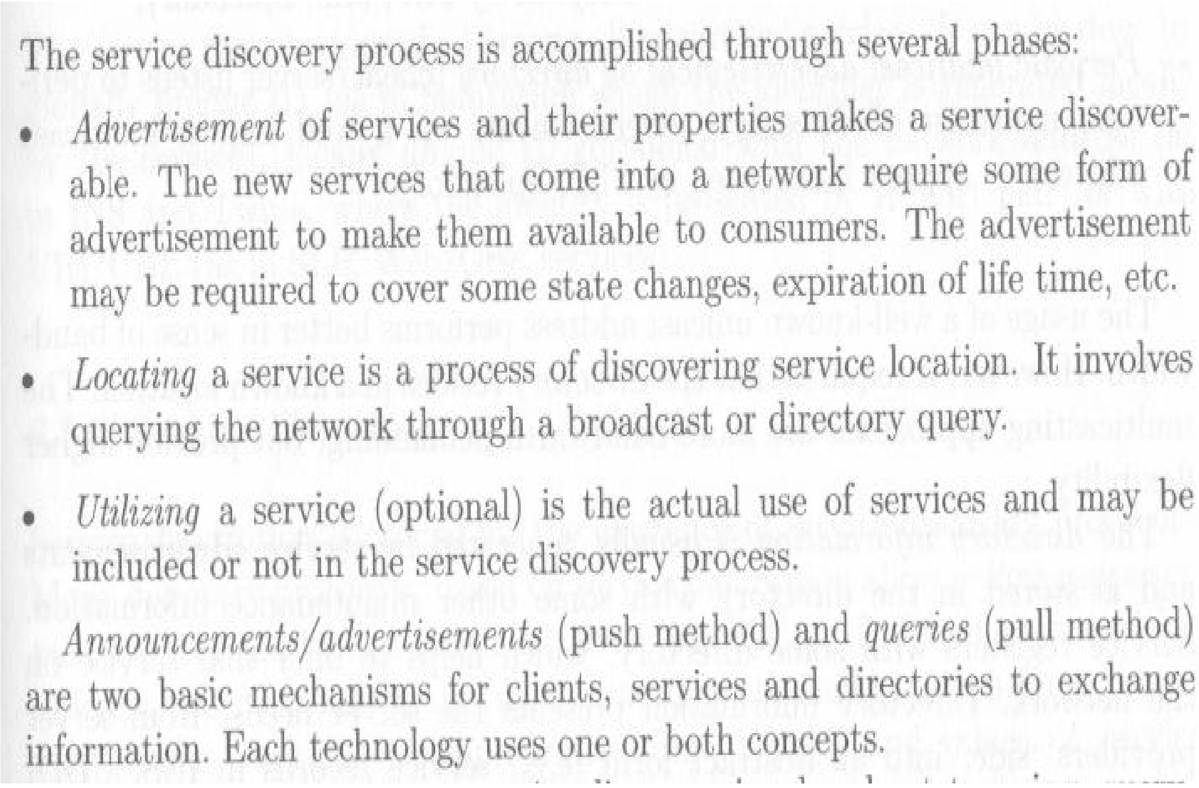
- Services to announce their presence to the network;
- Automatic discovery of local and remote services regardless of the type of network and technology used;
- Automatic adaptation to mobile and sporadic availability;
- Services to describe their capabilities as well as query and understand the capabilities of other services;
- Self-configuration without administrative intervention.



Service discovery functionalities are regulated through appropriate service discovery protocols. The structure of service discovery protocols identifies the building blocks and the links between the participating components (entities). Each service discovery protocol consists of at least two basic components: *client* and *server*. *Client* (*user* or *user agent – UA*) represents the entity that is interested in finding and using a service and hosts certain applications which access specific services. *Server (server provider* or *service agent – SA)* represents the entity that hosts and offers the service. The process of service discovery is actually a mapping between *service description* (that helps identify a service) and *service location* (that helps identify the location where the service can be found).



In order to facilitate the mappings, the service discovery protocols can involve another entity, named *directory*. *Directory (server coordinator, service broker, directory agent – DA* etc.) presents a node in the network that hosts partially or entirely the service description information [3, 6]. That node can act as a registry or broker for the discovery and provision processes, improving the performance of the service discovery. It is also often called *3rd party*. *Resources*, such as storage, bandwidth, database etc., represent the entities or tools that support the servers or clients in their activities. Different service discovery protocols involve some of previously mentioned entities

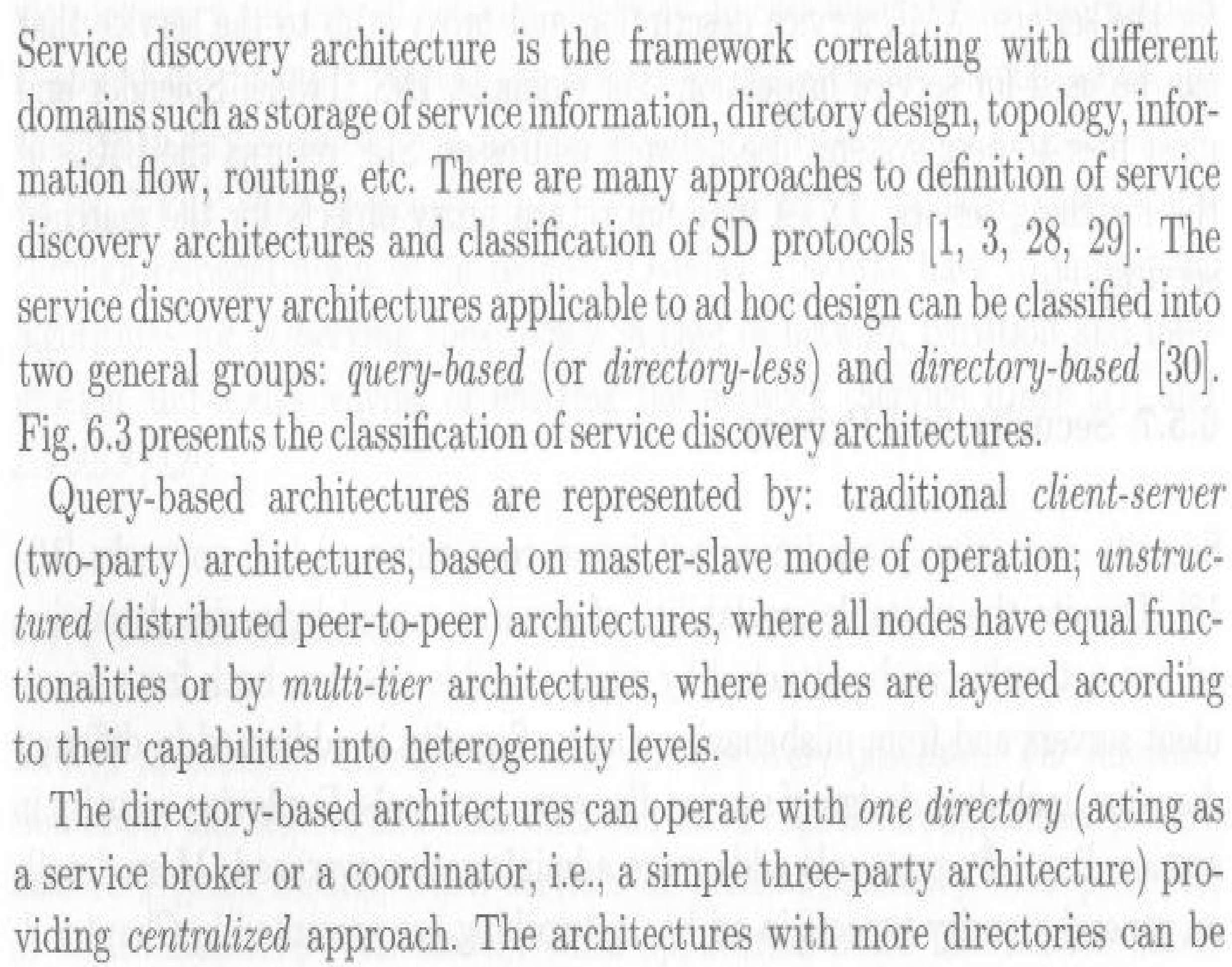


The service discovery process is accomplished through several phases:

- *Advertisement* of services and their properties makes a service discoverable. The new services that come into a network require some form of advertisement to make them available to consumers. The advertisement may be required to cover some state changes, expiration of life time, etc.
- *Locating* a service is a process of discovering service location. It involves querying the network through a broadcast or directory query.
- *Utilizing* a service (optional) is the actual use of services and may be included or not in the service discovery process.

Announcements/advertisements (push method) and *queries* (pull method) are two basic mechanisms for clients, services and directories to exchange information. Each technology uses one or both concepts.

Service Discovery Architecture



Service discovery architecture is the framework correlating with different domains such as storage of service information, directory design, topology, information flow, routing, etc. There are many approaches to definition of service discovery architectures and classification of SD protocols [1, 3, 28, 29]. The service discovery architectures applicable to ad hoc design can be classified into two general groups: *query-based* (or *directory-less*) and *directory-based* [30]. Fig. 6.3 presents the classification of service discovery architectures.

Query-based architectures are represented by: traditional *client-server* (two-party) architectures, based on master-slave mode of operation; *unstructured* (distributed peer-to-peer) architectures, where all nodes have equal functionalities or by *multi-tier* architectures, where nodes are layered according to their capabilities into heterogeneity levels.

The directory-based architectures can operate with *one directory* (acting as a service broker or a coordinator, i.e., a simple three-party architecture) providing *centralized* approach. The architectures with more directories can be

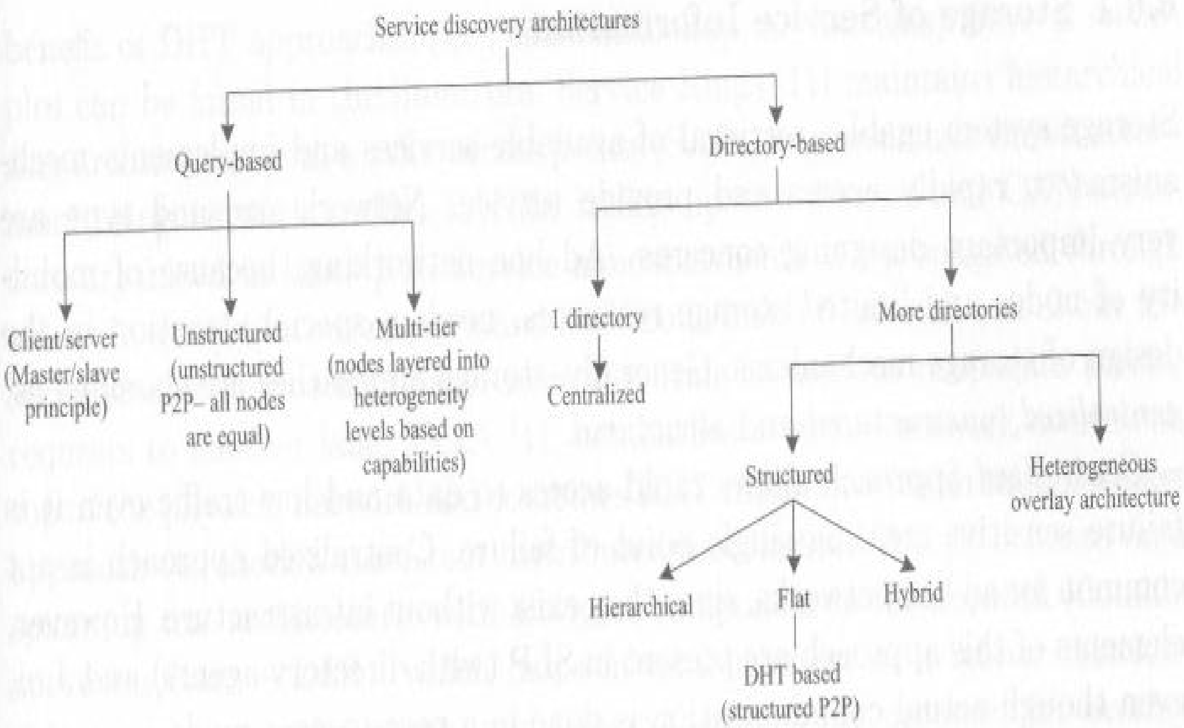

Service discovery architectures
Query-based
Directory-based
Client/server (Master/slave principle)
Unstructured (unstructured P2P– all nodes are equal)
Multi-tier (nodes layered into heterogeneity levels based on capabilities)
1 directory
Centralized
More directories
Structured
Heterogeneous overlay architecture
Hierarchical
Flat
Hybrid
DHT based (structured P2P)

Fig. 6.3. Classification of service discovery architectures

organized as *structured (hierarchical, flat or hybrid)* or *heterogeneous overlay* architectures, designed for heterogeneous environment with more discovery domains. Hierarchical architectures adopt parent-child relations between the nodes, often leading to tree-like structures or implementing some node clustering. Clustering architectures differentiate nodes functionality within the nodes in a cluster and between the nodes in different clusters. They can be combined with the multi-tier approach resulting in hierarchical multi-layer architectures, such as in [31]. Flat architectures are usually distributed DHT based structures (structured peer-to-peer), while hybrid architectures combine the elements of flat and hierarchical approaches as well as of query-based and directory-based elements.

Peer-to-peer networks recently gain on popularity. Especially applicable to networks with a large number of nodes, they seem to offer appropriate mechanisms to ad hoc mobile network design, capable of following the nodes dynamism and mobility. They will be addressed in one of the following subsections.

Cellular Networks:

Handoffs

• When a user/call moves to a new cell, then a new base station and new channel should be assigned (handoff)

• Handoffs should be transparent to users, while their number should be kept to minimum

• A threshold in the received power (Pr, handoff) should be determined to trigger the handoff process. This threshold value should be larger than the minimum acceptable received power (Pr, acceptable)

• Define: Δ=Pr,handoff - Pr,acceptable

– If Δ is large then too many handoffs

– If Δ is small then insufficient time to complete a handoff

• In order to correctly determine the beginning of handoff, we need to determine that a drop in the signal strength is not due to the momentary (temporary) bad channel condition, but it is due to the fact that the mobile is moving away from BS.

• Thus the BS needs to monitor the signal level for a certain period of time before initiating a handoff. The length of the time (running average measurements of signal) and handoff process depends on speed and moving pattern.

• First generation systems typical time interval to make a handoff was 10 seconds (large Δ). Second generations and after typical time interval to make a handoff is 1-2 seconds (small Δ).

• First generation systems: handoff decision was made by BS bymeasuring the signal strength in reverse channels.

• Second generation and after: Mobile Assisted Hand-Off (MAHO). Mobiles measure the signal strength from different neighboring BSs. Handoff is initiated if the signal strength from a neighboring BS is higher than the current BS's signal strength.

Cell Dwell Time

• It is the time over which a call maybe maintained within a cell (without handoff).

• It depends on: propagation, interference, distance between BS and MS, speed and moving pattern (direction), etc.

• Highway moving pattern: the cell dwell time is a r.v. with distribution highly concentrated around the mean.

• Other micro-cell moving patterns mix of different user types with large variations of dwell time (around the mean).

Prioritizing Handoffs:

• Guard Channels:

Fraction of total bandwidth in a cell is reserved for exclusive use of handoff calls. Therefore, total carried traffic is reduced if fixed channel assignment is used. However, if dynamic channel assignment is used the guard channel mechanisms may offer efficient spectrum utilization.

– Number of channels to be reserved: If it is low (under-reservation) the QoS on handoff call blocking probability can not be met. If reservation is high (over-reservation) may result in waste of resources and rejection of large number of new calls.

– Static and Dynamic schemes: Advantage of static scheme is its simplicity since no communication and computation overheads are involved. However problems of under-reservation and over reservations may occur if traffic does not conform to prior knowledge. Dynamic schemes may adjust better to changing traffic conditions.

• Queuing Handoffs:

The objective is to decrease the probability of forced determination of a call due to lack of available channels. When a handoff call (and in some schemes a new call) can not be granted the required resources at the time of its arrival, the request is put in a queue waiting for its admitting conditions to be met.

– This is achieved because there is a finite time interval between the time that the signal of a call drops below the handoff threshold, and the time that the call is terminated due to low (unacceptable) signal level. Queuing and size of buffer depends on traffic and QoS. Queueing in wireless systems is possible because signaling is done on separate control channels (without affecting the data transmission channels).

• According to the types of calls that are queued, queuing priority schemes are classified as: handoff call queuing, new call queuing and handoff/new call queuing (handoff calls are given non-preemptive priority over new calls).

Channel Allocation or Channel Assignment Strategies :

For efficient utilization of the radio spectrum, a frequency reuse scheme that is consistent with the objectives of increasing capacity and minimizing interference is required. A variety of channel assignment strategies have been developed to achieve these objectives. Channel assignment strategies can be classified as either *fixed* or *dynamic*. The choice of channel assignment strategy impacts the performance of the system, particularly as to how calls are managed when a mobile user is handed off from one cell to another [Tek91], [LiC93], [Sun94], [Rap93b].

In a fixed channel assignment strategy, each cell is allocated a predetermined set of voice channels. Any call attempt within the cell can only be served by the unused channels in that particular cell. If all the channels in that cell are occupied, the call is *blocked* and the subscriber does not receive service. Several variations of the fixed assignment strategy exist. In one approach, called the *borrowing strategy*, a cell is allowed to borrow channels from a neighboring cell if all of its own channels are already occupied. The mobile switching center (MSC) supervises such borrowing procedures and ensures that the borrowing of a channel does not disrupt or interfere with any of the calls in progress in the donor cell.

In a dynamic channel assignment strategy, voice channels are not allocated to different cells permanently. Instead, each time a call request is made, the serving base station requests a channel from the MSC. The switch then allocates a channel to the requested cell following an algorithm that takes into account the likelihood of future blocking within the cell, the frequency of use of the candidate channel, the reuse distance of the channel, and other cost functions.

Accordingly, the MSC only allocates a given frequency if that frequency is not presently in use in the cell or any other cell which falls within the minimum restricted distance of frequency reuse to avoid co-channel interference. Dynamic channel assignment reduce the likelihood of blocking, which increases the trunking capacity of the system, since all the available channels in a market are accessible to all of the cells. Dynamic channel assignment strategies require the MSC to collect real-time data on channel occupancy, traffic distribution, and *radio signal strength indications* (RSSI) of all channels on a continuous basis. This increases the storage and computational load on the system but provides the advantage of increased channel utilization and decreased probability of a blocked call.

Location Management :

- Location management deals with how to keep track of an active mobile station within the cellular network.
- A mobile station is active if it is powered on. Since the exact location of a mobile station must be known to the network during a call.
- location management usually means how to track an active mobile station between two consecutive phone calls.
- There are two basic operations involved with location management: location update and paging.
- The paging operation is performed by the cellular network. When an incoming call arrives for a mobile station, the cellular network will page the mobile station in all possible cells to find out the cell in which the mobile station is located so the incoming call can be routed to the corresponding base station. This process is called paging.
- The number of all possible cells to be paged is dependent on how the location update operation is performed.
- The location update operation is performed by an active mobile station. A location update scheme can be classified as either global or local.
- A location update scheme is global if all subscribers update their locations at the same set of cells, and a scheme is local if an individual subscriber is allowed to decide when and where to perform location update. A local scheme is also called individualized or per-user based.
- From another point of view, a location update scheme can be classified as either static or dynamic. A location update scheme is static if there is a predetermined set of cells at which location updates must be generated by a mobile station regardless of it mobility. A scheme is dynamic if a location update can be generated by a mobile station in any cell depending on its mobility.

Pagging :

In the attempt to locate recipients of calls as quickly as possible, multiple methods of paging have been created. The most basic method used is Simultaneous Paging, where every cell in the user's Location Areas (LAs) is paged at the same time in order to find the user. Unless there are a relatively low number of cells within the LA, this will cause excessive amounts of paging. Although this method will find the user quicker than the following scheme of Sequential Paging, the costs make Simultaneous Paging rather inefficient. An alternative scheme is Sequential Paging, where each cell within an LA is paged in succession, with one common theory suggesting the polling of small cell areas in order of decreasing user dwelling possibility. Unfortunately, this was found to have poor performance in some situations, as if the user was in an infrequently occupied location, not only might every cell be paged, but a large delay could occur in call establishment. Additionally, this method requires accurate data gathering concerning common user locations, which necessitates more frequent Location Updates (LUs) and thereby increased costs. Consequently, most real-world Sequential Paging methods simply poll the cells nearest to the cell of the most recent LU, and then continue outward if the user is not immediately found. However, such a method will still be inefficient if the user's velocity is high or an Location Management (LM) scheme is used which specifies infrequent LUs.

Multiple Access :

A cellular system divides any given area into cells where a mobile unit in each cell communicates with a base station. The main aim in the cellular system design is to be able to increase the capacity of the channel i.e. to handle as many calls as possible in a given bandwidth with a sufficient level of quality of service. There are several different ways to allow access to the channel. These includes mainly the following:

1) Frequency division multiple-access (FDMA)
2) Time division multiple-access (TDMA)
3) Code division multiple-access (CDMA)
4) Space Division Multiple access (SDMA)

Frequency Division Multiple Access : This was the initial multiple-access technique for cellular systems in which each individual user is assigned a pair of frequencies while making or receiving a call. One frequency is used for downlink and one pair for uplink. This is called frequency division duplexing (FDD). That allocated frequency pair is not used in the same cell or adjacent cells during the call so as to reduce the co channel interference. Even though the user may not be talking, the spectrum cannot be reassigned as long as a call is in place. Different users can use the same frequency in the same cell except that they must transmit at different times.

The features of FDMA are as follows:

- The FDMA channel carries only one phone circuit at a time.
- If an FDMA channel is not in use, then it sits idle and it cannot be used by other users to increase share capacity.
- After the assignment of the voice channel the BS and the MS transmit simultaneously and continuously.
- The bandwidths of FDMA systems are generally narrow i.e. FDMA is usually implemented in a narrow band system .

- The symbol time is large compared to the average delay spread.
- The complexity of the FDMA mobile systems is lower than that of TDMA mobile systems.
- FDMA requires tight filtering to minimize the adjacent channel interference.

Time Division Multiple Access: In digital systems, continuous transmission is not required because users do not use the allotted bandwidth all the time. In such cases, TDMA is a complimentary access technique to FDMA. Global Systems for Mobile communications (GSM) uses the TDMA technique. In TDMA, the entire bandwidth is available to the user but only for a finite period of time. In most cases the available bandwidth is divided into fewer channels compared to FDMA and the users are allotted time slots during which they have the entire channel bandwidth at their disposal TDMA requires careful time synchronization since users share the bandwidth in the frequency domain. The number of channels are less, inter channel interference is almost negligible. TDMA uses different time slots for transmission and reception. This type of duplexing is referred to as Time division duplexing(TDD).

The features of TDMA includes the following:

- TDMA shares a single carrier frequency with several users where each users makes use of non overlapping time slots.
- The number of time slots per frame depends on several factors such as modulation technique, available bandwidth etc.
- Data transmission in TDMA is not continuous but occurs in bursts. This results in low battery consumption since the subscriber transmitter can be turned OFF when not in use.
- Because of a discontinuous transmission in TDMA the handoff process is much simpler for a subscriber unit, since it is able to listen to other base stations during idle time slots.
- TDMA uses different time slots for transmission and reception thus duplexers are not required.
- TDMA has an advantage that is possible to allocate different numbers of time slots per frame to different users. Thus bandwidth can be supplied on demand to different users by concatenating or reassigning time slot based on priority

Space Division Multiple Access SDMA: utilizes the spatial separation of the users in order to optimize the use of the frequency spectrum. A primitive form of SDMA is when the same frequency is reused in different cells in a cellular wireless network. The radiated power of each user is controlled by Space division multiple access. SDMA serves different users by using spot beam antenna. These areas may be served by the same frequency or different frequencies. However for limited co-channel interference it is required that the cells be sufficiently separated. This limits the number of cells a region can be divided into and hence limits the frequency re-use factor. A more advanced approach can further increase the capacity of the network. This technique would enable frequency re-use within the cell. In a practical cellular environment it is improbable to have just one transmitter fall within the receiver beam width. Therefore it becomes imperative to use other multiple access techniques in conjunction with SDMA. When different areas are covered by the antenna beam, frequency can be re-used, in which case TDMA or CDMA is employed, for different frequencies FDMA can be used

Code Division Multiple Access: In CDMA, the same bandwidth is occupied by all the users, however they are all assigned separate codes, which differentiates them from each other. CDMA utilize a spread spectrum technique in which a spreading signal (which is uncorrelated to the signal and has a large bandwidth) is used to spread the narrow band message signal.

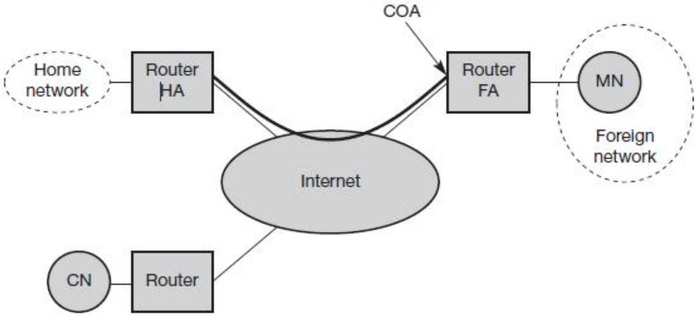
MODULE 4

MOBILE IP :

The goal of a mobile IP can be summarized as: 'supporting end-system mobility while maintaining scalability, efficiency, and compatibility in all respects with existing applications and Internet protocols.

Entities and terminology

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



● **Mobile node (MN):** A mobile node is an end-system or router that can change its point of attachment to the internet using mobile IP. The MN keeps its IP address and can continuously communicate with any other system in the internet as long as link-layer connectivity is given. Mobile nodes are not necessarily small devices such as laptops with antennas or mobile phones; a router onboard an aircraft can be a powerful mobile node.

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

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

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

● **Foreign agent (FA):** The FA can provide several services to the MN during its visit to the foreign network. The FA can have the COA, acting as tunnel endpoint and forwarding packets to the MN. The FA can be the default router for the MN. FAs can also provide security services because they belong to the foreign network as opposed to the MN which is only visiting. For mobile IP functioning, FAs are not necessarily needed. Typically, an FA is implemented on a router for the subnet the MN attaches to.

● **Care-of address (COA):** The COA defines the current location of the MN from an IP point of view. All IP packets sent to the MN are delivered to the COA, not directly to the IP address of the MN. Packet delivery toward the MN is done using a tunnel.

To be more precise, the COA marks the tunnel endpoint, i.e., the address where packets exit the tunnel. There are two different possibilities for the location of the COA:

● **Foreign agent COA:** The COA could be located at the FA, i.e., the COA is an IP address of the FA. The FA is the tunnel end-point and forwards packets to the MN. Many MN using the FA can share this COA as common COA.

● **Co-located COA:** The COA is co-located if the MN temporarily acquired an additional IP address which acts as COA. This address is now topologically correct, and the tunnel endpoint is at the MN. Co-located addresses can be acquired using services such as DHCP . One problem associated with this approach is the need for additional addresses if MNs request a COA. This is not always a good idea considering the scarcity of IPv4 addresses.

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

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

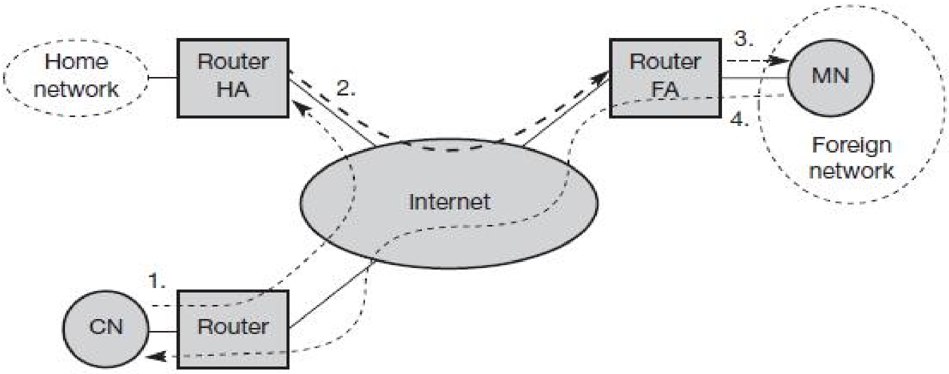
● If changing the router's software is not possible, the HA could also be implemented on an arbitrary node in the subnet. One disadvantage of this solution is the double crossing of the router by the packet if the MN is in a foreign network. A packet for the MN comes in via the router; the HA sends it through the tunnel which again crosses the router.

● Finally, a home network is not necessary at all. The HA could be again on the 'router' but this time only acting as a manager for MNs belonging to a virtual home network. All MNs are always in a foreign network with this solution.

IP packet delivery

Figure illustrates packet delivery to and from the MN using the example network. A correspondent node CN wants to send an IP packet to the MN. One of the requirements of mobile IP was to support hiding the mobility of the MN. CN does not need to know anything about the MN's current location and sends the packet as usual to the IP address of MN (step 1). This means that CN sends an IP packet with MN as a destination address and CN as a source address. The internet, not having information on the current location of MN, routes the packet to the router responsible for the home network of MN. This is done using the standard routing mechanisms of the internet.

The HA now intercepts the packet, knowing that MN is currently not in its home network. The packet is not forwarded into the subnet as usual, but encapsulated and tunnelled to the COA. A new header is put in front of the old IP header showing the COA as new destination and HA as source of the encapsulated packet (step 2). The foreign agent now decapsulates the packet, i.e., removes the additional header, and forwards the original packet with CN as source and MN as destination to the MN (step 3). Again, for the MN mobility is not visible. It receives the packet with the same sender and receiver address as it would have done in the home network

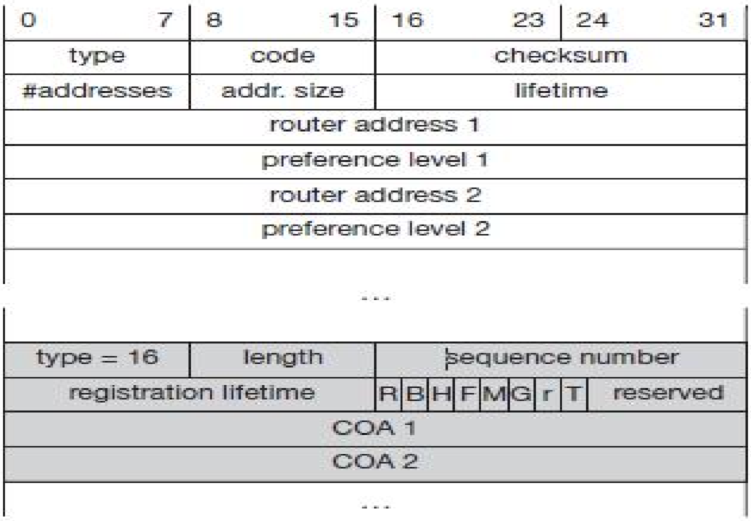


<u>Agent Discovery</u> :

One initial problem of an MN after moving is how to find a foreign agent. How does the MN discover that it has moved? For this purpose mobile IP describes two methods: agent advertisement and agent solicitation,

<u>Agent advertisement</u>

For the first method, foreign agents and home agents advertise their presence periodically using special **agent advertisement** messages. These advertisement messages can be seen as a beacon broadcast into the subnet. For these advertisements Internet control message protocol (ICMP) messages according to RFC 1256 are used with some mobility extensions. Routers in the fixed network implementing this standard also advertise their routing service periodically to the attached links. The agent advertisement packet according to RFC 1256 with the extension for mobility is shown in Figure . The upper part represents the ICMP packet while the lower part is the extension needed for mobility. The fields necessary on lower layers for the agent advertisement are not shown in this figure. Clearly, mobile nodes must be reached with the appropriate link layer address. The TTL field of the IP packet is set to 1 for all advertisements to avoid forwarding them. The IP destination address according to standard router advertisements can be either set to 224.0.0.1, which is the multicast address for all systems on a link , or to the broadcast address 255.255.255.255. The fields in the ICMP part are defined as follows. The **type** is set to 9, the **code** can be 0, if the agent also routes traffic from non-mobile nodes, or 16, if it does not route anything other than mobile traffic. Foreign agents are at least required to forward packets from the mobile node. The number of addresses advertised with this packet is in **#addresses** while the **addresses** themselves follow as shown. **Lifetime** denotes the length of time this advertisement is valid. **Preference** levels for each address help a node to choose the router that is the most eager one to get a new node.The difference compared with standard ICMP advertisements is what happens after the router addresses. This extension for mobility has the following fields defined: **type** is set to 16, **length** depends on the number of COAs provided with the message and equals 6 + 4*(number of addresses). An agent shows the total number of advertisements sent since initialization in the **sequence number**. By the **registration lifetime** the agent can specify the maximum lifetime in seconds a node can request during registration.



0 7	8 15	16 23	24 31
type	code	checksum	
#addresses	addr. size	lifetime	
router address 1			
preference level 1			
router address 2			
preference level 2			
...			
type = 16	length	sequence number	
registration lifetime		R B H F M G r T	reserved
COA 1			
COA 2			
...			

The following bits specify the characteristics of an agent in detail. The **R** bit (registration) shows, if a registration with this agent is required even when using a colocated COA at the MN. If the agent is currently too busy to accept new registrations it can set the **B** bit. The following two bits denote if the agent offers services as a home agent (**H**) or foreign agent (**F**) on the link where the advertisement has been sent. Bits M and G specify the method of encapsulation used for the tunnel. While IP-in-IP encapsulation is the mandatory standard, **M** can specify minimal encapsulation and **G** generic routing encapsulation. Now the field **r** at the same bit position is set to zero and must be ignored. The new field **T** indicates that reverse tunneling is supported by the FA. The following fields contain the **COAs** advertised. A foreign agent setting the F bit must advertise at least one COA. A mobile node in a subnet can now receive agent advertisements from either its home agent or a foreign agent. This is one way for the MN to discover its location.

Agent solicitation

If no agent advertisements are present or the inter-arrival time is too high, and an MN has not received a COA by other means, e.g., DHCP, the mobile node must send **agent solicitations**. These solicitations are again based on RFC 1256 for router solicitations. Care must be taken to ensure that these solicitation messages do not flood the network, but basically an MN can search for an FA endlessly sending out solicitation messages. Typically, a mobile node can send out three solicitations, one per second, as soon as it enters a new network. It should be noted that in highly dynamic wireless networks with moving MNs and probably with applications requiring continuous packet streams even one second intervals between solicitation messages might be too long. Before an MN even gets a new address many packets will be lost without additional mechanisms. If a node does not receive an answer to its solicitations it must decrease the rate of solicitations exponentially to avoid flooding the network until it reaches a maximum interval between solicitations (typically one minute). Discovering a new agent can be done anytime, not just if the MN is not connected to one. Consider the case that an MN is looking for a better connection while still sending via the old path. This is the case while moving through several cells of different wireless networks. After these steps of advertisements or solicitations the MN can now receive a COA, either one for an FA or a co-located COA. The MN knows its location (home network or foreign network) and the capabilities of the agent (if needed). .

Tunneling and encapsulation :

A **tunnel** establishes a virtual pipe for data packets between a tunnel entry and a tunnel endpoint. Packets entering a tunnel are forwarded inside the tunnel and leave the tunnel unchanged. Tunneling, i.e., sending a packet through a tunnel, is achieved by using encapsulation.

Encapsulation is the mechanism of taking a packet consisting of packet header and data and putting it into the data part of a new packet. The reverse operation, taking a packet out of the data part of another packet, is called **decapsulation**. Encapsulation and decapsulation are the operations typically performed when a packet is transferred from a higher protocol layer to a lower layer or from a lower to a higher layer respectively. Here these functions are used within the same layer.

TCP (transmission control protocol) :

TCP offers connections between two applications. Within a connection TCP can give certain guarantees, such as in-order delivery or reliable data transmission using retransmission techniques. TCP has built-in mechanisms to behave in a 'network friendly' manner. If, for example, TCP encounters packet loss, it assumes network internal congestion and slows down the transmission rate. This is one of the main reasons to stay with protocols like TCP.

Congestion control

A transport layer protocol such as TCP has been designed for fixed networks with fixed end-systems. Data transmission takes place using network adapters, fiber optics, copper wires, special hardware for routers etc. This hardware typically works without introducing transmission errors. If the software is mature enough, it will not drop packets or flip bits, so if a packet on its way from a sender to a receiver is lost in a fixed network, it is not because of hardware or software errors. The probable reason for a packet loss in a fixed network is a temporary overload some point in the transmission path, i.e., a state of congestion at a node. Congestion may appear from time to time even in carefully designed networks. The packet buffers of a router are filled and the router cannot forward the packets fast enough because the sum of the input rates of packets destined for one output link is higher than the capacity of the output link. The only thing a router can do in this situation is to drop packets. A dropped packet is lost for the transmission, and the receiver notices a gap in the packet stream. Now the receiver does not directly tell the sender which packet is missing, but continues to acknowledge all in-sequence packets up to the missing one. The sender notices the missing acknowledgement for the lost packet and assumes a packet loss due to congestion. Retransmitting the missing packet and continuing at full sending rate would now be unwise, as this might only increase the congestion. Although it is not guaranteed that all packets of the TCP connection take the same way through the network, this assumption holds for most of the packets. To mitigate congestion, TCP slows down the transmission rate dramatically. All other TCP connections experiencing the same congestion do exactly the same so the congestion is soon resolved. This cooperation of TCP connections in the internet is one of the main reasons for its survival as it is today. Even under heavy load, TCP guarantees at least sharing of the bandwidth

Slow start

TCP's reaction to a missing acknowledgement is quite drastic, but it is necessary to get rid of congestion quickly. The behavior TCP shows after the detection of congestion is called **slow start**

The sender always calculates a **congestion window** for a receiver. The start size of the congestion window is one segment (TCP packet). The sender sends one packet and waits for acknowledgement. If this acknowledgement arrives, the sender increases the congestion window by one, now sending two packets (congestion window = 2). After arrival of the two corresponding acknowledgements, the sender again adds 2 to the congestion window, one for each of the acknowledgements. Now the congestion window equals 4. This scheme doubles the congestion window every time the acknowledgements come back, which takes one round trip time (RTT). This is called the exponential growth of the congestion window in the slow start mechanism. It is too dangerous to double the congestion window each time because the steps might become too large. The exponential growth stops at the **congestion threshold**. As soon as the congestion window reaches the congestion threshold, further increase of the transmission rate is only linear by adding 1 to the congestion window each time the acknowledgements come back. Linear increase continues until a time-out at the sender occurs due to a missing acknowledgement, or until the sender detects a gap in transmitted data because of continuous acknowledgements for the same packet. In either case the sender sets the congestion threshold to half of the current congestion window. The congestion window itself is set to one segment and the sender starts sending a single segment. The exponential growth starts once more up to the new congestion threshold, then the window grows in linear fashion.

Fast retransmit/fast recovery

Two things lead to a reduction of the congestion threshold. One is a sender receiving continuous acknowledgements for the same packet. This informs the sender of two things. One is that the receiver got all packets up to the acknowledged packet in sequence. In TCP, a receiver sends acknowledgements only if it receives any packets from the sender. Receiving acknowledgements from a receiver also shows that the receiver continuously receives something from the sender. The gap in the packet stream is not due to severe congestion, but a simple packet loss due to a transmission error. The sender can now retransmit the missing packet(s) before the timer expires. This behavior is called **fast Retransmit**

The receipt of acknowledgements shows that there is no congestion to justify a slow start. The sender can continue with the current congestion window. The sender performs a **fast recovery** from the packet loss. This mechanism can improve the efficiency of TCP dramatically.

The other reason for activating slow start is a time-out due to a missing acknowledgement. TCP using fast retransmit/fast recovery interprets this congestion in the network and activates the slow start mechanism

Implications on mobility

While slow start is one of the most useful mechanisms in fixed networks, it drastically decreases the efficiency of TCP if used together with mobile receivers or senders. The reason for this is the use of slow start under the wrong assumptions. From a missing acknowledgement, TCP concludes a congestion situation. While this may also happen in networks with mobile and wireless end-systems, it is not the main reason for packet loss.

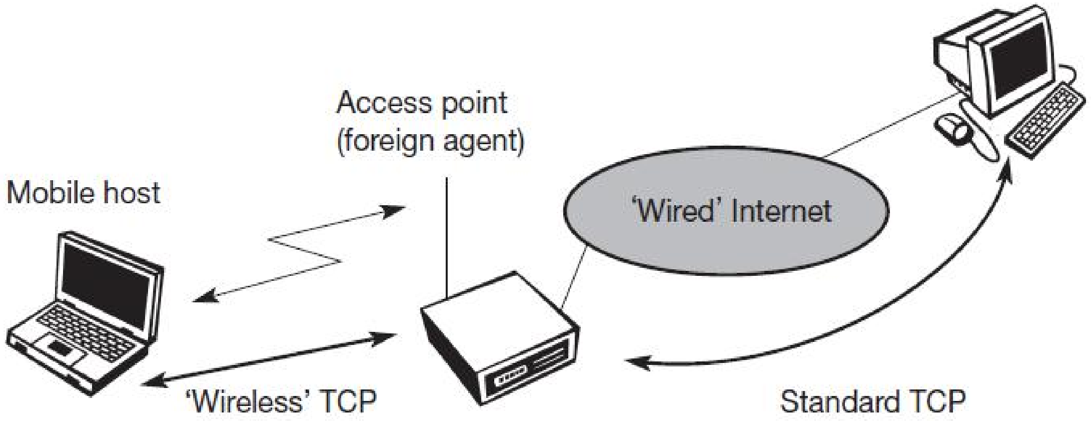
Error rates on wireless links are orders of magnitude higher compared to fixed fiber or copper links. Packet loss is much more common and cannot always be compensated for by layer 2 retransmissions (ARQ) or error correction (FEC). Trying to retransmit on layer 2 could, for example, trigger TCP retransmission if it takes too long. Layer 2 now faces the problem of transmitting the same packet twice over a bad link. Detecting these duplicates on layer 2 is not an option, because more and more connections use end-to-end encryption, making it impossible to look at the packet.

Mobility itself can cause packet loss. There are many situations where a soft handover from one access point to another is not possible for a mobile endsystem. For example, when using mobile IP, there could still be some packets in transit to the old foreign agent while the mobile node moves to the new foreign agent. The old foreign agent may not be able to forward those packets to the new foreign agent or even buffer the packets if disconnection of the mobile node takes too long. This packet loss has nothing to do with wireless access but is caused by the problems of rerouting traffic.

The TCP mechanism detecting missing acknowledgements via time-outs and concluding packet loss due to congestion cannot distinguish between the different causes. This is a fundamental design problem in TCP: An error control mechanism (missing acknowledgement due to a transmission error) is misused for congestion control (missing acknowledgement due to network overload). In both cases packets are lost (either due to invalid checksums or to dropping in routers). However, the reasons are completely different. TCP cannot distinguish between these two different reasons. Standard TCP reacts with slow start if acknowledgements are missing, which does not help in the case of transmission errors over wireless links and which does not really help during handover. This behavior results in a severe performance degradation of an unchanged TCP if used together with wireless links or mobile nodes.

Indirect TCP

Two competing insights led to the development of indirect TCP (I-TCP). One is that TCP performs poorly together with wireless links; the other is that TCP within the fixed network cannot be changed. I-TCP segments a TCP connection into a fixed part and a wireless part. Figure shows an example with a mobile host connected via a wireless link and an access point to the 'wired' internet where the correspondent host resides. The correspondent node could also use wireless access. The following would then also be applied to the access link of the correspondent host. Standard TCP is used between the fixed computer and the access point. No computer in the internet recognizes any changes to TCP. Instead of the mobile host, the access point now terminates the standard TCP connection, acting as a proxy. This means that the access point is now seen as the mobile host for the fixed host and as the fixed host for the mobile host. Between the access point and the mobile host, a special TCP, adapted to wireless links, is used. However, changing TCP for the wireless link is not a requirement. Even an unchanged TCP can benefit from the much shorter round trip time, starting retransmission much faster. A good place for segmenting the connection between mobile host and correspondent host is at the foreign agent of mobile IP. The foreign agent controls the mobility of the mobile host anyway and can also hand over the connection to the next foreign agent when the mobile host moves on. However, one can also imagine separating the TCP connections at a special server, e.g., at the entry point to a mobile phone network (e.g., IWF in GSM, GGSN in GPRS). The correspondent host in the fixed network does not notice the wireless link or the segmentation of the connection. The foreign agent acts as a proxy and relays all data in both directions. If the correspondent host sends a packet, the foreign agent acknowledges this packet and tries to forward the packet to the mobile host. If the mobile host receives the packet, it acknowledges the packet. However, this acknowledgement is only used by the foreign agent. If a packet is lost on the wireless link due to a transmission error, the correspondent host would not notice this. In this case, the foreign agent tries to retransmit this packet locally to maintain reliable data transport. Similarly, if the mobile host sends a packet, the foreign agent acknowledges this packet and tries to forward it to the correspondent host. If the packet is lost on the wireless link, the mobile hosts notice this much faster due to the lower round trip time and can directly retransmit the packet. Packet loss in the wired network is now handled by the foreign agent. I-TCP requires several actions as soon as a handover takes place. The access point acts as a proxy buffering packets for retransmission. After the handover, the old proxy must forward buffered data to the new proxy because it has already acknowledged the data. After registration with the new foreign agent, this new foreign agent can inform the old one about its location to enable packet forwarding. Besides buffer content, the sockets of the proxy, too, must migrate to the new foreign agent located in the access point. The socket reflects the current state of the TCP connection, i.e., sequence number, addresses, ports etc. No new connection may be established for the mobile host, and the correspondent host must not see any changes in connection state.



There are several advantages with I-TCP:

● I-TCP does not require any changes in the TCP protocol as used by the hosts in the fixed network or other hosts in a wireless network that do not use this optimization. All current optimizations for TCP still work between the foreign agent and the correspondent host.

● Due to the strict partitioning into two connections, transmission errors on the wireless link, i.e., lost packets, cannot propagate into the fixed network. Without partitioning, retransmission of lost packets would take place between mobile host and correspondent host across the whole network. Now only packets in sequence, without gaps leave the foreign agent.

● It is always dangerous to introduce new mechanisms into a huge network such as the internet without knowing exactly how they will behave. However, new mechanisms are needed to improve TCP performance (e.g., disabling slow start under certain circumstances), but with I-TCP only between the mobile host and the foreign agent. Different solutions can be tested or used at the same time without jeopardizing the stability of the internet. Furthermore, optimizing of these new mechanisms is quite simple because they only cover one single hop.

● The authors assume that the short delay between the mobile host and foreign agent could be determined and was independent of other traffic streams. An optimized TCP could use precise time-outs to guarantee retransmission as fast as possible. Even standard TCP could benefit from the short round trip time, so recovering faster from packet loss. Delay is much higher in a typical wide area wireless network than in wired networks due to FEC and MAC. GSM has a delay of up to 100 ms circuit switched, 200 ms and more packet switched (depending on packet size and current traffic). This is even higher than the delay on transatlantic links.

● Partitioning into two connections also allows the use of a different transport layer protocol between the foreign agent and the mobile host or the use of compressed headers etc. The foreign agent can now act as a gateway to translate between the different protocols.

But the idea of segmentation in I-TCP also comes with some **disadvantages**:

● The loss of the end-to-end semantics of TCP might cause problems if the foreign agent partitioning the TCP connection crashes. If a sender receives an acknowledgement, it assumes that the receiver got the packet. Receiving an acknowledgement now only means (for the mobile host and a correspondent host) that the foreign agent received the packet. The correspondent node does not know anything about the partitioning, so a crashing access node may also crash applications running on the correspondent node assuming reliable end-to-end delivery.

● In practical use, increased handover latency may be much more problematic. All packets sent by the correspondent host are buffered by the foreign agent besides forwarding them to the mobile host (if the TCP connection is split at the foreign agent). The foreign agent removes a packet from the buffer as soon as the appropriate acknowledgement arrives. If the mobile host now performs a handover to another foreign agent, it takes a while before the old foreign agent can forward the buffered data to the new foreign agent. During this time more packets may arrive. All these packets have to be forwarded to the new foreign agent first, before it can start forwarding the new packets redirected to it.

● The foreign agent must be a trusted entity because the TCP connections end at this point. If users apply end-to-end encryption, e.g., according to RFC 2401 the foreign agent has to be integrated into all security mechanisms.

Mobile TCP

The **M-TCP (mobile TCP)**1 approach has the same goals as I-TCP and snooping TCP: to prevent the sender window from shrinking if bit errors or disconnection but not congestion cause current problems. M-TCP wants to improve overall throughput, to lower the delay, to maintain end-to-end semantics of TCP, and to provide a more efficient handover.

- Additionally, M-TCP is especially adapted to the problems arising from lengthy or frequent disconnections.

- M-TCP splits the TCP connection into two parts as I-TCP does. An unmodified TCP is used on the standard host-**supervisory host (SH)** connection, while an optimized TCP is used on the SH-MH connection.

- The supervisory host is responsible for exchanging data between both parts similar to the proxy in ITCP

- The M-TCP approach assumes a relatively low bit error rate on the wireless link. Therefore, it does not perform caching/retransmission of data via the SH.

- If a packet is lost on the wireless link, it has to be retransmitted by the original sender. This maintains the TCP end-to-end semantics.

- The SH monitors all packets sent to the MH and ACKs returned from the MH.
- If the SH does not receive an ACK for some time, it assumes that the MH is disconnected. It then chokes the sender by setting the sender's window size to 0.

- Setting the window size to 0 forces the sender to go into **persistent mode**, i.e., the state of the sender will not change no matter how long the receiver is disconnected.

- This means that the sender will not try to retransmit data. As soon as the SH (either the old SH or a new SH) detects connectivity again, it reopens the window of the sender to the old value. The sender can continue sending at full speed. This mechanism does not require changes to the sender's TCP.

- The wireless side uses an adapted TCP that can recover from packet loss much faster. This modified TCP does not use slow start, thus, M-TCP needs a **bandwidth manager** to implement fair sharing over the wireless link.

The **advantages** of M-TCP are the following:

- It maintains the TCP end-to-end semantics. The SH does not send any ACK itself but forwards the ACKs from the MH.
- If the MH is disconnected, it avoids useless retransmissions, slow starts or breaking connections by simply shrinking the sender's window to 0.
- Since it does not buffer data in the SH as I-TCP does, it is not necessary to forward buffers to a new SH. Lost packets will be automatically retransmitted to the new SH.

The lack of buffers and changing TCP on the wireless part also has some **disadvantages:**

● As the SH does not act as proxy as in I-TCP, packet loss on the wireless link due to bit errors is propagated to the sender. M-TCP assumes low bit error rates, which is not always a valid assumption.

● A modified TCP on the wireless link not only requires modifications to the MH protocol software but also new network elements like the bandwidth manager.

Selective retransmission

A very useful extension of TCP is the use of selective retransmission. TCP acknowledgements are cumulative, i.e., they acknowledge in-order receipt of packets up to a certain packet. If a single packet is lost, the sender has to retransmit everything starting from the lost packet (go-back-n retransmission). This obviously wastes bandwidth, not just in the case of a mobile network, but for any network (particularly those with a high path capacity, i.e., bandwidthdelay- product).

Using RFC 2018 (Mathis, 1996), TCP can indirectly request a selective retransmission of packets. The receiver can acknowledge single packets, not only trains of in-sequence packets. The sender can now determine precisely which packet is needed and can retransmit it.

The **advantage** of this approach is obvious: a sender retransmits only the lost packets. This lowers bandwidth requirements and is extremely helpful in slow wireless links. The gain in efficiency is not restricted to wireless links and mobile environments. Using selective retransmission is also beneficial in all other networks. However, there might be the minor **disadvantage** of more complex software on the receiver side, because now more buffer is necessary to resequence data and to wait for gaps to be filled. But while memory sizes and CPU performance permanently increase, the bandwidth of the air interface remains almost the same. Therefore, the higher complexity is no real disadvantage any longer as it was in the early days of TCP.

Transaction-oriented TCP

Assume an application running on the mobile host that sends a short request to a server from time to time, which responds with a short message. If the application requires reliable transport of the packets, it may use TCP (many applications of this kind use UDP and solve reliability on a higher, application-oriented layer).

Using TCP now requires several packets over the wireless link. First, TCP uses a three-way handshake to establish the connection. At least one additional packet is usually needed for transmission of the request, and requires three more packets to close the connection via a three-way handshake. Assuming connections with a lot of traffic or with a long duration, this overhead is minimal. But in an example of only one data packet, TCP may need seven packets altogether.

In the internet, TCP is used for this purpose. Before a HTTP request can be transmitted the TCP connection has to be established. This already requires three messages. If GPRS is used as wide area transport system, one-way delays of 500 ms and more are quite common. The setup of a TCP connection already takes far more than a second.

This led to the development of a transaction-oriented TCP (T/TCP). T/TCP can combine packets for connection establishment and connection release with user data packets. This can reduce the number of packets down to two instead of seven. Similar considerations led to the development of a transaction service in WAP

The obvious **advantage** for certain applications is the reduction in the overhead which standard TCP has for connection setup and connection release. However, T/TCP is not the original TCP anymore, so it requires changes in the mobile host and all correspondent hosts, which is a major **disadvantage**. This solution no longer hides mobility. Furthermore, T/TCP exhibits several security problems.

File System :

The general goal of a file system is to support efficient, transparent, and consistent access to files, no matter where the client requesting files or the server(s) offering files are located. **Efficiency** is of special importance for wireless systems as the bandwidth is low so the protocol overhead and updating operations etc. should be kept at a minimum. **Transparency** addresses the problems of location- dependent views on a file system. To support mobility, the file system should provide identical views on directories, file names, access rights etc., independent of the current location. The main problem is **consistency**

Consistency :

The basic problem for distributed file systems that allow replication of data for performance reasons is the consistency of replicated objects (files, parts of files, parts of a data structure etc.). What happens, for example, if two portable devices hold copies of the same object, then one device changes the value of the object and after that, both devices read the value? Without further mechanisms, one portable device reads an old value.

To avoid inconsistencies many traditional systems apply mechanisms to maintain a permanent consistent view for all users of a file system. This **strong consistency** is achieved by atomic updates similar to database systems. A writer of an object locks the object, changes the object, and unlocks the object after the change. If an object is locked, no other device can write the object. Cached objects are invalidated after a change. Maintaining strong consistency is not only very expensive in terms of exchanging updates via the wireless link, but is also sometimes impossible. Assume a temporarily disconnected device with several objects in its cache. It is impossible to update the objects or invalidate them. Locking the cached objects may not be visible to other users. One solution is to forbid access to disconnected objects. This would prohibit any real application based on the file system. Mobile systems have to use a **weak consistency** model for file systems. Weak consistency implies certain periods of inconsistency that have to be tolerated for performance reasons.However, the overall file system should remain consistent so conflict resolution strategies are needed for reintegration. **Reintegration** is the process of merging objects from different users resulting in one consistent file system. A user could hold a copy of an object, disconnect from the network, change the object, and reconnect again. The changed object must then be reintegrated. A **conflict** may occur, e.g., if an object has been changed by two users working with two copies. During reintegration the file system may notice that both copies differ, the conflict resolution strategy has to decide which copy to use or how to proceed. The system may detect conflicts based on time stamps, version numbering, hash values, content comparison etc.

Coda:

The predecessor of many distributed file systems that can be used for mobile operation is the Andrew file system (AFS). Coda is the successor of AFS and offers two different types of replication: server replication and caching on clients.

Disconnected clients work only on the cache, i.e., applications use only cached replicated files. Figure shows the cache between an application and the server. Coda is a transparent extension of the client's cache manager. This very general architecture is valid for most of today's mobile systems that utilise a cache.

To provide all the necessary files for disconnected work, Coda offers extensive mechanisms for pre-fetching of files while still connected, called **hoarding.** If the client is connected to the server with a strong connection, hoarding transparently pre-fetches files currently used. This automatic data collection is necessary for it is impossible for a standard user to know all the files currently used. While standard programs and application data may be familiar to a user, he or she typically does not know anything about the numerous small system files needed in addition (e.g., profiles, shared libraries, drivers, fonts).

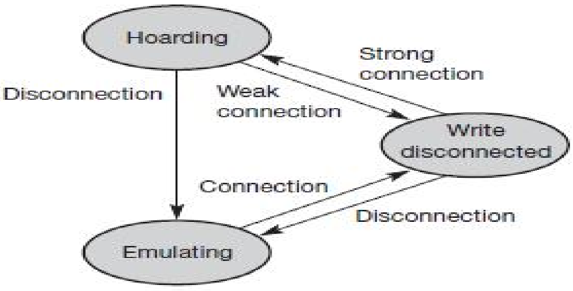
A user can pre-determine a list of files, which Coda should explicitly pre-fetch. Additionally, a user can assign priorities to certain programs. Coda now decides on the current cache content using the list and a least-recently-used (LRU) strategy. As soon as the client is disconnected, applications work on the replicates (see Figure , **emulating**).

Coda follows an optimistic approach and allows read and write access to all files. The system keeps a record of changed files, but does not maintain a history of changes for each file. The cache always has only onereplicate (possibly changed). After reconnection, Coda compares the replicates with the files on the server. If Coda notices that two different users have changed a file, reintegration of this file fails and Coda saves the changed file as a copy on the server to allow for manual reintegration.

Another problem of Coda is the definition of a conflict. Coda detects only write conflicts, i.e., if two or more users change a file. Now consider two files f1 and f2. One client uses values from files f1 and f2 to calculate something and stores the result in file f1. The other client uses values from files f1 and f2 to calculate something else and stores the result in file f2. Coda would not detect any problem during reintegration of the files. However, the results may not reflect the correct values based on the files. The order of execution plays an important role.

To solve this problem, a simple transaction mechanism was introduced into Coda as an option, the so-called isolation-only transactions. IOT allows grouping certain operations and checks them for serial execution. While in the beginning Coda simply distinguished the two states "hoarding" while connected and "emulating" while disconnected, the loosely connected state **write disconnected** was later integrated, . If a client is only weakly connected, Coda decides if it is worthwhile to fetch a file via this connection or to let the user wait until a better connection is available. In other words, Coda models the patience of a user and weighs it against the cost of fetching the file required by the user.

Figure illustrates the three states of a client in Coda. The client only performs hoarding while a strong connection to the server exists. If the connection breaks completely, the client goes into emulating and uses only the cached replicates. If the client loses the strong connection and only a weak connection remains, it does not perform hoarding, but decides if it should fetch the file in case of a cache miss considering user patience and file type. The weak connection, however, is not used for reintegration of files.



Little Work :

The distributed file system Little Work is, like Coda, an extension of AFS. Little Work only requires changes to the cache manager of the client and detects write conflicts during reintegration. Little Work has no specific tools for reintegration and offers no transaction service. However, Little Work uses more client states to maintain consistency.

● **Connected:** The operation of the client is normal, i.e., no special mechanisms from Little Work are required. This mode needs a continuous high bandwidth as available in typical office environments using, e.g., a WLAN.

● **Partially connected:** If a client has only a lower bandwidth connection, but still has the possibility to communicate continuously, it is referred to as partially connected. Examples for this type of network are packet radio networks. These networks typically charge based on the amount of traffic and not based on the duration of a connection. This client state allows to use cache consistency protocols similar to the normal state, but with a delayed write to the server to lower communication cost if the client changes the file again. This helps to avoid consistency problems, although no highbandwidth connection is available.

● **Fetch only:** If the only network available offers connections on demand, the client goes into the fetch only state. Networks of this type are cellular networks such as GSM with costs per call. The client uses the replicates in the cache in an optimistic way, but fetches files via the communication link if they are not available in the cache. This enables a user to access all files of the server, but this also tries to minimize communication by working on replicates and reintegrate after reconnection using a continuously high bandwidth link.

● **Disconnected:** Without any network, the client is disconnected. Little Work now aborts if a cache mis-occurs, otherwise replicates are used.

Ficus

Ficus is a distributed file system, which is not based on a client/server approach. Ficus allows the optimistic use of replicates, detects write conflicts, and solves conflicts on directories. Ficus uses so-called **gossip protocol**s, an idea many other systems took over later. A mobile computer does not necessarily need to have a direct connection to a server. With the help of other mobile computers, it can propagate updates through the network until it reaches a fixed network and the server. Thus, changes on files propagate through the network step-by-step. Ficus tries to minimize the exchange of files that are valid only for a short time, e.g. temporary files. A critical issue for gossip protocols is how fast they propagate to the client that needs this information and how much unnecessary traffic it causes to propagate information to clients that are not interested.

MIo-NFS

The system mobile integration of NFS (MIo–NFS) is an extension of the Network File System (NFS). In contrast to many other systems, MIo-NFS uses a pessimistic approach with tokens controling access to files. Only the token-holder for a specific file may change this file, so MIo-NFS avoids write conflicts. Read/write conflicts cannot be avoided.

MIo-NFS supports three different modes:

- **Connected:** The server handles all access to files as usual.
- **Loosely connected:** Clients use local replicates, exchange tokens over the network, and update files via the network.
- **Disconnected:** The client uses only local replicates. Writing is only allowed if the client is token-holder.

Rover

Compared to Coda, the Rover platform uses another approach to support mobility Instead of adapting existing applications for mobile devices, Rover provides a platform for developing new, mobility aware applications. Two new components have been introduced in Rover. **Relocatable dynamic objects** are objects that can be dynamically loaded into a client computer from a server (or vice-versa) to reduce client-server communication. A trade-off between transferring objects and transferring only data for objects has to be found. If a client needs an object quite often, it makes sense to migrate the object. Object migration for a single access, on the other hand, creates too much overhead. **Queued remote procedure calls** allow for non-blocking RPCs even when a host is disconnected. Requests and responses are exchanged as soon as a connection is available again. Conflict resolution is done in the server and is application specific.

Wireless application protocol :

The basic objectives of the WAP Forum and now of the OMA (**open mobile alliance**)are to bring diverse internet content (e.g., web pages, push services) and other data services (e.g., stock quotes) to digital cellular phones and other wireless, mobile terminals (e.g., PDAs, laptops). Moreover, a protocol suite should enable global wireless communication across different wireless network technologies, e.g., GSM, CDPD, UMTS etc. The forum is embracing and extending existing standards and technologies of the internet wherever possible and is creating a framework for the development of contents and applications that scale across a very wide range of wireless bearer networks and wireless device types. All solutions must be:

- **interoperable,** i.e., allowing terminals and software from different vendors to communicate with networks from different providers;
- **scaleable,** i.e., protocols and services should scale with customer needs and number of customers;
- **efficient,** i.e., provision of QoS suited to the characteristics of the wireless and mobile networks;
- **reliable,** i.e., provision of a consistent and predictable platform for deploying services; and
- **secure,** i.e., preservation of the integrity of user data, protection of devices and services from security problems.

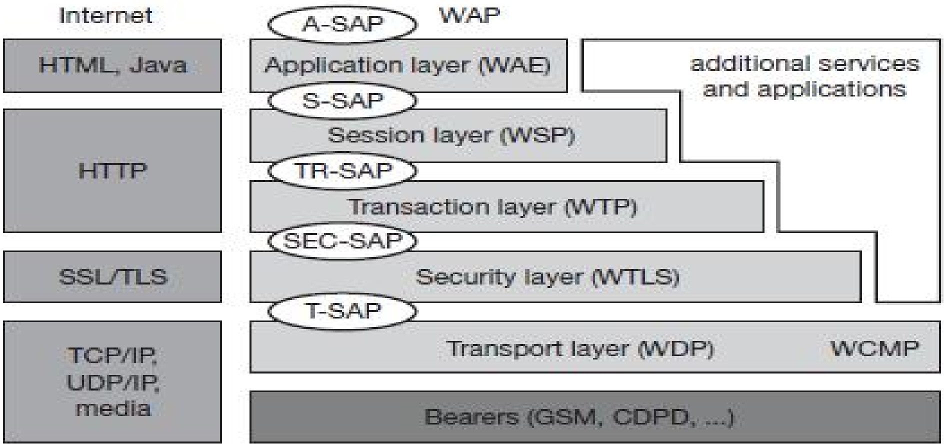
Architecture

Figure gives an overview of the WAP architecture, its protocols and components, and compares this architecture with the typical internet architecture when using the world wide web.

The basis for transmission of data is formed by different **bearer services**. WAP does not specify bearer services, but uses existing data services and will integrate further services. Examples are message services, such as short message service (SMS) of GSM, circuit-switched data, such as high-speed circuit switched data (HSCSD) in GSM, or packet switched data, such as general packet radio service (GPRS) in GSM. Many other bearers are supported, such as CDPD, IS-136, PHS. No special interface has been specified between the bearer service and the next higher layer, the **transport layer** with its **wireless datagram protocol (WDP)** and the additional **wireless control message protocol (WCMP),** because the adaptation of scripting languages, special markup languages, interfaces to telephony applications, and many content formats adapted to the special requirements of small, handheld, wireless devices.

Figure not only shows the overall WAP architecture, but also its relation to the traditional internet architecture for www applications. The WAP transport layer together with the bearers can be (roughly) compared to the services offered by TCP or UDP over IP and different media in the internet. If a bearer in the WAP architecture already offers IP services (e.g., GPRS, CDPD) then UDP is used as WDP. The TLS/SSL layer of the internet has also been adopted for the WAP architecture with some changes required for optimization. The functionality of the session and transaction layer can roughly be compared with the role of HTTP in the web architecture. However, HTTP does not offer all the additional mechanisms needed for efficient wireless, mobile access (e.g., session migration, suspend/resume). Finally, the application layer offers similar features as HTML and Java. Again, special formats and features optimized for the wireless scenario have been defined and telephony access has been added. WAP does not always force all applications to use the whole protocol architecture.

Applications can use only a part of the architecture as shown in Figure For example, this means that, if an application does not require security but needs the reliable transport of data, it can **directly** use a service of the transaction layer. Simple applications can directly use WDP.



- The **wireless datagram protocol (WDP)** operates on top of many different bearer services capable of carrying data. At the T-SAP WDP offers a consistent datagram transport service independent of the underlying bearer
- WDP offers **source** and **destination port numbers** used for multiplexing and demultiplexing of data respectively. The service primitive to send a datagram is **TDUnitdata. req** with the **destination address (DA), destination port (DP), Source address (SA), source port (SP)**, and **user data (UD)** as mandatory parameters

- **WDP management entity** supports WDP and provides information about changes in the environment, which may influence the correct operation of WDP.

- **wireless control message protocol (WCMP)** provides error handling mechanisms for WDP

- WCMP contains control messages that resemble the internet control message protocol (ICMP (Postel, 1981b) for IPv4, (Conta, 1998) for IPv6) messages and can also be used for diagnostic and informational purposes.

- WCMP can be used by WDP nodes and gateways to report errors.

- WTLS can provide different levels of security (for privacy, data integrity, and authentication) and has been optimized for low bandwidth, high-delay bearer networks. WTLS takes into account the low processing power and very limited memory capacity of the mobile devices for cryptographic algorithms. WTLS supports datagram and connection-oriented transport layer protocols.

- WTP offers several advantages to higher layers, including an improved reliability over datagram services, improved efficiency over connection-oriented services, and support for transaction-oriented services such as web browsing. In this context, a transaction is defined as a request with its response, e.g. for a web page.

- WTP offers many features to the higher layers. The basis is formed from three **classes of transaction service** as explained in the following paragraphs. Class 0 provides unreliable message transfer without any result message. Classes 1 and 2 provide reliable message transfer, class 1 without, class 2 with, exactly one reliable result message (the typical request/response case).

- WTP achieves reliability using **duplicate removal, retransmission, acknowledgement**s and unique **transaction identifiers**. No WTP-class requires any connection set-up or tear-down phase. This avoids unnecessary overhead on the communication link. WTP allows for **asynchronous transactions, abort of transactions, concatenation of messages**, and can **report success or failure** of reliable messages

- WSP offers : **Session management:** WSP introduces sessions that can be **established** from a client to a server and may be long lived. Sessions can also be **released** in an orderly manner. The capabilities of **suspending** and **resuming** a session are important to mobile applications.

- **Capability negotiation:** Clients and servers can agree upon a common level of protocol functionality during session establishment. Example parameters to negotiate are maximum client SDU size, maximum outstanding requests, protocol options, and server SDU size.
- **Content encoding:** WSP also defines the efficient binary encoding for the content it transfers. WSP offers content typing and composite objects, as explained for web browsing.

- The main idea behind the **wireless application environment (WAE)** is to create a general-purpose application environment based mainly on existing technologies and philosophies of the world wide web

- This environment should allow service providers, software manufacturers, or hardware vendors to integrate their applications so they can reach a wide variety of different wireless platforms in an efficient way.

MODULE 5

Conventional TCP /IP Protocol:

• Application data is first encoded using the application layer protocol header words by prefixing them over the data • Then the encoded data from the application layer is encoded again using the transport layer protocol header words by prefixing them over the previously encoded data

• At the receiver end the reverse process of decoding at each layer to retrieve back the application data takes place

• The data transported from the transport layer to next layer (L3) using TCP (or UDP in case of datagram)

• TCP— a connection-oriented protocol

• TCP— a transport layer protocol for the Internet

TCP

• A connection oriented less protocol • Session for establishment, data flow and congestion control, and session termination in TCP

Function of the transport layer

• To transport the port data

• UDP header specifies the ports • Used at the subsequent layers (from the transmitter transport layer up to the receiver transport layer) during transmission of port data (application layer data) to the receiver

User Datagram Protocol

• A connectionless protocol— there is no session for establishment, data flow and congestion control, and session termination in UDP

• Transmits like a person using a phone who just speaks without waiting, irrespective of whether the receiver at the other end is listening or not, replying or not

• Useful in transmitting datagrams

• One datagram length ≤ 216 words

• Usage Examples— as those for multicasting, registration request and registration reply

Other Transport Layer Protocols :

DCCP(Datagram congestion control protocol)

DCCP is a transport layer protocol which is message oriented. It offers certain features like congestion control mechanism, reliable connection setup, feature negotiation, and Explicit Congestion Notification.

DCCP handles setup and teardown of reliable connections while generating ECP messages when congestion occurs. Since DCCP is at the Transport Layer, applications do not need to be programmed to use DCCP. DCCP is usually implemented for multicast operations such as telephony, streaming media and online gaming. DCCP uses UDP for a faster transport mechanism since lost packets and resending those packets are not necessary.

SCTP(Stream control transmission protocol)

SCTP is also a standard protocol of the transport layer it is also same function like TCP and UDP.SCTP both have the features of TCP which is to provide ordered delivery of message and UDP message oriented.It sends multiple streams though one stream e.g mostly browsers treat each image on a web page as a connection and then one connection for the text with SCTP these can be sent as
one connection.

SCTP is also support connection between two hosts which is two connection In a network local network and in it if one connection goes down then the other one is used as a redundant connection to resume the transmission.

RSVP(Resource reservation protocol)

RSVP is an transport layer protocol and has a function of resource reservation across the network for an integrated services internet.it also provide the setup of resources reservation which is initiated at a receivers side for a multi-casting(one-host-to-many-host) or uni-costing(one host-to-one-host).It can be utilized on routers to provide the QOS to the hosts

.

RIP(Routing information protocol)

RIP as a name mention routing protocol which is use for the hop count as a metric for the routing, it implements a limit constrain on the number of hops allow in apath from the source to a destination to prevent routing loops.

Extensible Markup Language (XML)

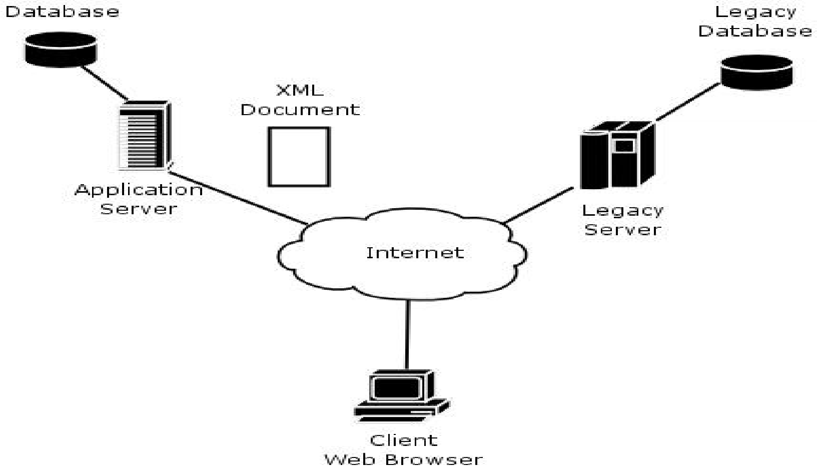
- Data and documents in a Web services environment are expressed and described by a family of technologies known as XML.
- XML was developed by the XML Working Group formed under the auspices of the W3C in 1996 and provides the foundation for many of the open standards of today.
- XML was created to form a standardized format for representing data and documents, concentrating on structure and thereby making it independent of the delivery medium; thus, XML is the perfect standard for today's utility computing needs.
- This is particularly true of needs that relate to interoperability between applications and components.

Following is an example of an XML document:

```
<?xml version=”1.0”>
<PhoneDirList>
<Person>
<Name>Jane Doe</Name>
<Phone>123-123-1234</Phone>
<Address1>1234 Green St</Address1>
<City>Austin</City>
<State>TX</State>
<Zip>78758</Zip>
</Person>
</PhoneList>
```

- Below Figure shows how XML can be used for transferring data between Web browsers and software applications.

- A Web client using browser technology requests data from an application server. With the appropriate software application, data from the database is queried and represented as an XML document and transferred through the Internet to the client browser.
- In the same fashion, the legacy server can request data from the application server as would the client Web browser. Only this time, instead of the XML data being displayed in a Web browser, the data is used by the application running at the legacy server as input that is processed and later merged into the legacy database.



J2ME

J2ME is a family of specifications that defines various downsized versions of the standard Java 2 platform; these downsized versions can be used to program consumer electronic devices ranging from cell phones to highly capable Personal Data Assistants (PDAs), smart phones, and set-top boxes.

J2ME uses configurations and profiles to customize the Java Runtime Environment (JRE). As a complete JRE, J2ME is comprised of a configuration, which determines the JVM used, and a profile, which defines the application by adding domain-specific classes.

Configuration :

Connected Limited Device Configuration (CLDC) is used specifically with the KVM for 16-bit or 32-bit devices with limited amounts of memory. This is the configuration (and the virtual machine) used for developing small J2ME applications. Its size limitations make CLDC more interesting and challenging (from a development point of view) than CDC. CLDC is also the configuration that we will use for developing our drawing tool application. An example of a small wireless device running small applications is a Palm hand-held computer.

CLDC defines the following requirements:
* Full Java language support (except for floating pointer support, finalization, and error handling)
* Full JVM support
* Security for CLDC
* Limited internationalization support
* Inherited classes -- all classes not specific to CLDC must be subsets of J2SE 1.3 classes

* Classes specific to CLDC are in javax.microedition package and subpackages
In addition to the javax.microedition package, the CLDC API consists of subsets of the J2SE java.io, java.lang, and java.util packages.

Connected Device Configuration (CDC) is used with the C virtual machine (CVM) and is used for 32-bit architectures requiring more than 2 MB of memory. An example of such a device is a Net TV box.CDC, also developed by the Java Community Process, provides a standardized, portable, full-featured Java 2 virtual machine building block for consumer electronic and embedded devices, such as smartphones, two-way pagers, PDAs, home appliances, point-of-sale terminals, and car navigation systems. These devices run a 32-bit microprocessor and have more than 2 MB of memory, which is needed to store the C virtual machine and libraries. While the K virtual machine supports CLDC, the C virtual machine (CVM) supports CDC. CDC is associated with the Foundation Profile, which is beyond the scope of this tutorial.

Types of Profiles in J2ME

A profile complements a configuration by adding additional classes that provide features appropriate to a particular type of device or to a specific vertical market segment. Both J2ME configurations have one or more associated profiles, some of which may themselves rely on other profiles. These processes are described in the following list:

Mobile Information Device Profile (MIDP)

This profile adds networking, user interface components, and local storage to CLDC. This profile is primarily aimed at the limited display and storage facilities of mobile phones, and it therefore provides a relatively simple user interface and basic networking based on HTTP 1.1. MIDP is the best known of the J2ME profiles because it is the basis for Wireless Java and is currently the only profile available for PalmOSbased handhelds.

The Mobile Information Device Profile, or MIDP for short, is one such profile, intended for use on small footprint devices with a limited user interface in the form of a small screen with some kind of input capability.

Java applications that run on MIDP devices are known as MIDlets. A MIDlet consists of atleast one Java class

PDA Profile (PDAP)

The PDA Profile is similar to MIDP, but it is aimed at PDAs that have better screens and more memory than cell phones. The PDA profile, which is not complete at the time of writing, will offer a more sophisticated user interface library and a Java-based
1 It could be argued that CLDC breaks this rule with its networking classes, because

there is no usable subset of the java.net package that would fit into the restricted memory available to a CLDC-based device. This problem is solved by creating a new package that contains a more lightweight set of networking classes. API for accessing useful features of the host operating system. When this profile becomes available, it is likely to take over from MIDP as the J2ME platform for small handheld computers such as those from Palm and Handspring.

Foundation Profile

The Foundation Profile extends the CDC to include almost all of the core Java 2 Version 1.3 core libraries. As its name suggests, it is intended to be used as the basis for most of the other CDC profiles.

Personal Basis and Personal Profiles

The Personal Basis Profile adds basic user interface functionality to the Foundation Profile. It is intended to be used on devices that have an unsophisticated user interface capability, and it therefore does not allow more than one window to be active at any time. Platforms that can support a more complex user interface will use the Personal Profile instead. At the time of writing, both these profiles are in the process of being specified.

RMI Profile

The RMI Profile adds the J2SE Remote Method Invocation libraries to the Foundation Profile. Only the client side of this API is supported.

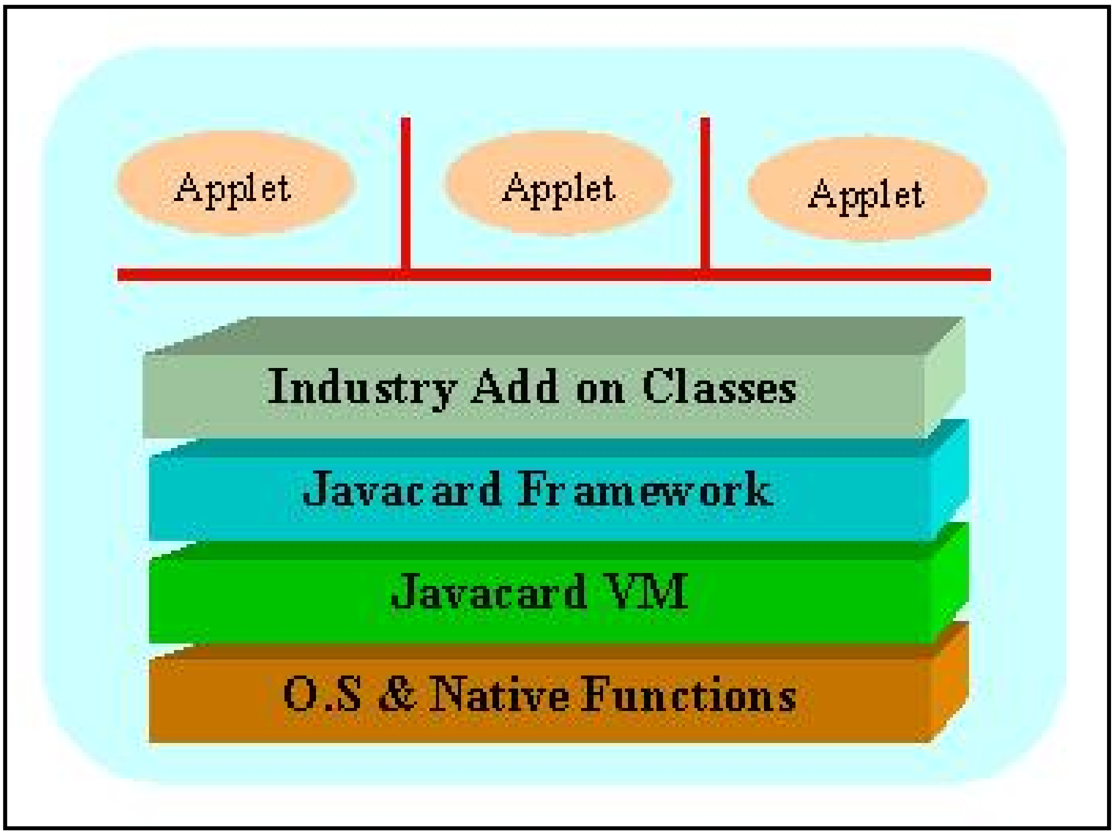
Game Profile

The Game Profile, which is still in the process of being defined, will provide a platform for writing games software on CDC devices. At the time of writing, it is not certain whether this profile will be derived from the Foundation Profile or based directly on CDC.

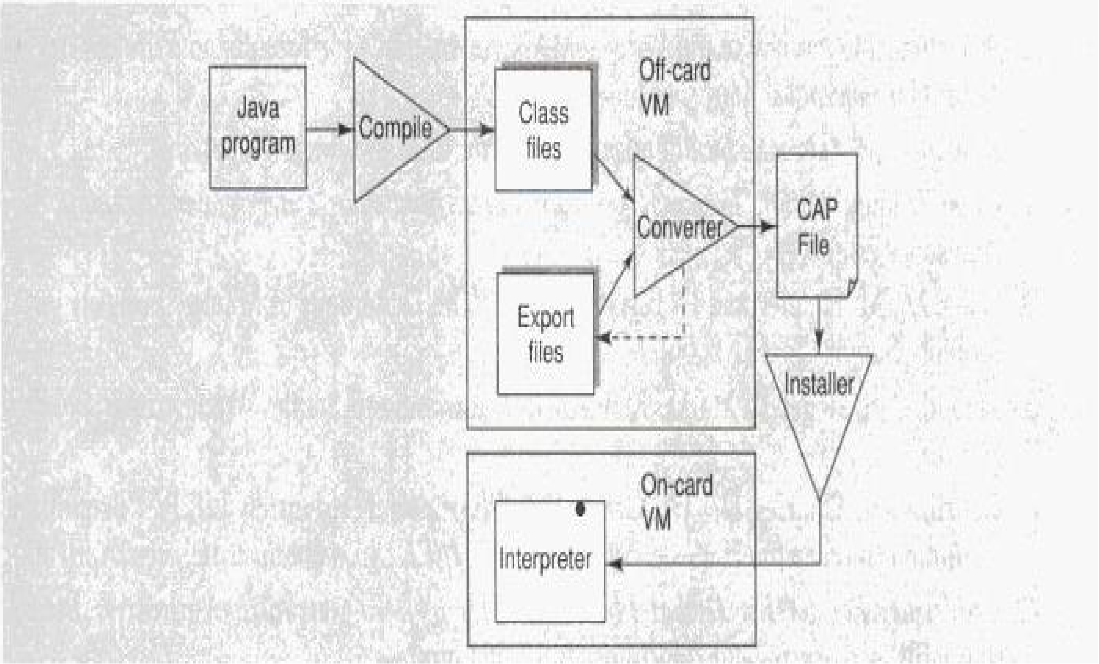
Java Card :

- Java Card is a smart card that is capable of running programs written in Java.
- For this a new Java platform, Sun's JavaSoft division has made available the Java Card 2.0 API specification, and several licensees are now implementing this API on smart cards.

- Java Card refers to a software technology that allows Java-based applications (applets) to be run securely on smart cards and similar small memory footprint devices.
- Java Card is the tiniest of Java platforms targeted for embedded devices.



Java card Architecture:



The development framework in Java card is different from that on a desktop computer. The major challenge of Java Card technology on smart card is to fit Java system software in a resource constraint smart card while conserving enough space for applications. Java Card supports a subset of the features of Java language available on desktop computers. The Java Card virtual machine on a smart card is split into two parts (Fig. 4.15): one that runs off-card and the other that runs on-card. Many processing tasks that are not constrained to execute at runtime, such as class loading, bytecode verification, resolution and linking, and optimization, are dedicated to the virtual machine that is running off-card where resources are usually not a concern. The on-card components of Java Card include components like the Java Card virtual machine (JCVM), the Java Card Runtime Environment (JCRE), and the Java API. Task of the compiler is to convert a Java source into Java class files. The converter will convert class files into a format downloadable into the smart card. Converter ensures the byte code validity before the application is installed into the card. The converter checks the classes off-card for,

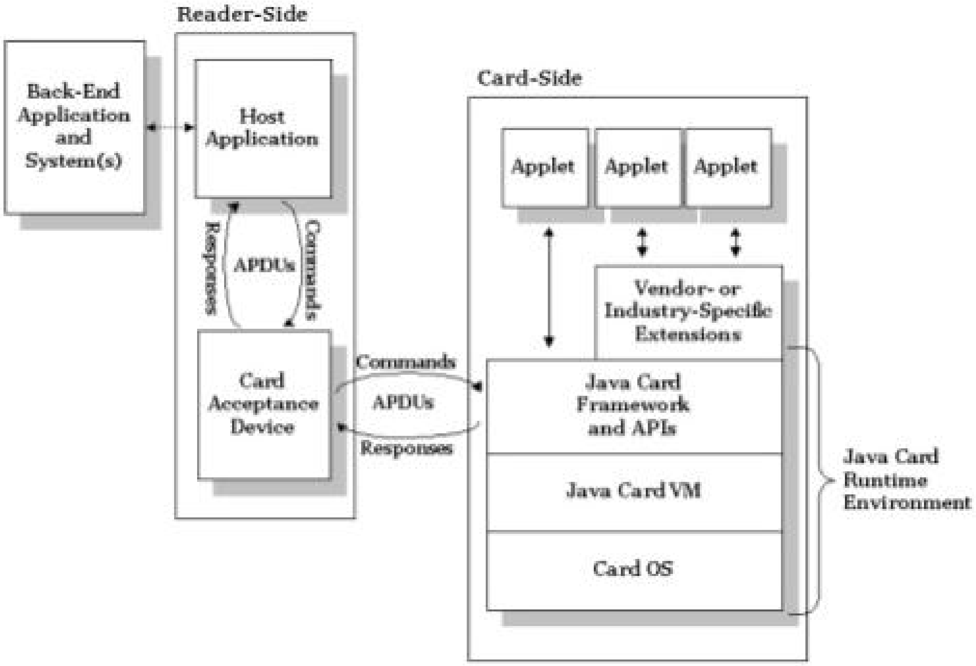
- How well it is formed?
- Java Card subset violations.
- Static variable initialization.
- Reference resolution.
- Byte code optimization.
- Storage allocation.
- The Java Card interpreter.
- Applet execution.
- Controls run time resources.
- Enforces runtime security.

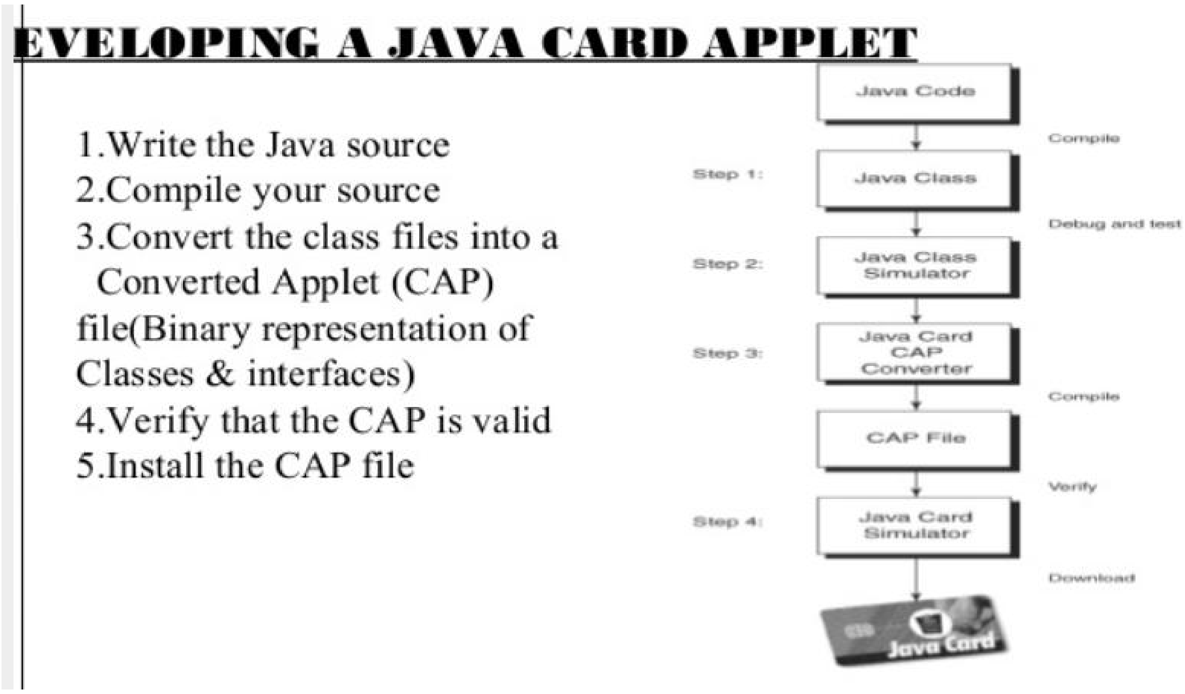
Following conversion by the off-card VM into CAP (Converted APlet) format, the applet is transferred into the card using the installer. The applet is selected for execution by the JCRE. JCRE is made up of the on-card virtual machine and the Java Card API classes. JCRE performs additional runtime security checks through the applet firewall. Applet firewall partitions the objects stored into separate protected object spaces, called contexts. Applet firewall controls the access to shareable interfaces of these objects. The JCVM is a scaled down version of standard JVM (Java Virtual Machine). Elements of standard Java not supported in JCVM are,

- Security manager.
- Dynamic class loading.
- Bytecode verifier.
- Threads.
- Garbage collection.
- Multi-dimensional arrays.
- Char and strings.
- Floating point operation.
- Object serialization.
- Object cloning.

As mentioned above, Java applications for Java Cards are called Applets. Java Card applets should not be confused with Java applets on the Internet. A Java Card applet is not intended to run within an Internet browser environment. The reason for choosing the name applet is that Java Card applets can be loaded into the Java Card runtime environment after the card has been manufactured. That is, unlike applications in many embedded systems, Java Card applets do not need to be burned into the ROM during manufacture.

Working :





EVELOPING A JAVA CARD APPLET

1.Write the Java source
2.Compile your source
3.Convert the class files into a Converted Applet (CAP) file(Binary representation of Classes & interfaces)
4.Verify that the CAP is valid
5.Install the CAP file

Java Code
Compile
Step 1:
Java Class
Debug and test
Step 2:
Java Class Simulator
Step 3:
Java Card CAP Converter
Compile
CAP File
Verify
Step 4:
Java Card Simulator
Download
Java Card

Advantages :

- Interoperable
- Secure
- Multi-Application capable
- Dynamic

- Compatible with existing standards
- Hardware Independence

PalmOS :

- **Palm OS** (also known as **Garnet OS**) is a discontinued mobile operating systeminitially developed by Palm, Inc., for personal digital assistants
- Palm OS was designed for ease of use with a touchscreen-based graphical user interface. It is provided with a suite of basic applications for personal information management. Later versions of the OS have been extended to support smartphones.

• An OS for handheld devices

• Designed for highly efficient running of small productivity programs for devices with a few application tasks

• Offers high performance due to a special feature that it supports only one process which controls all computations by the event handlers

• No multi-processing or multi-tasking

• Simplifies the kernel of the OS─ there is an infinite waiting loop in the only process that kernel runs

• The loop polls for an event

PalmOS Features

• Single process (no multi-processing and multi-threading)

• Compiled for a specific set of hardware, performance very finely tuned

• Memory space partitioned into program memory and multiple storage heaps for data and applications

• A file in format of a database

• IP-based network connectivity and WiFi (in later version only)

• Integration to cellular GSM/CDMA phone

The key features of the current Palm OS Garnet are:

- Simple, single-tasking environment to allow launching of full screen applications with a basic, common GUI set
- Monochrome or color screens with resolutions up to 480x320 pixel

- Handwriting recognition input system called Graffiti 2
- HotSync technology for data synchronization with desktop computers
- Sound playback and record capabilities
- Simple security model: Device can be locked by password, arbitrary application records can be made private
- TCP/IP network access
- Serial port/USB, infrared, Bluetooth and Wi-Fi connections
- Expansion memory card support
- Defined standard data format for personal information management applications to store calendar, address, task and note entries, accessible by third-party applications.

Linux for mobile devices :

Linux for mobile devices, is about the use of Linux kernel-based operating systems on all sorts of mobile devices, whose primary or only Human interface device (HID) is a touchscreen.

This mainly comprises smartphones and tablet computers, but also some mobile phones, personal digital assistants (PDAs) portable media players that come with a touchscreen separately.

This is a list of many Linux kernel-based operating systems used on mobile devices. They differ from one another in parts of the middleware or the entire middleware, and in that they employ individual UIs.

- Android (operating system)
- Replicant (operating system)
- AsteroidOS
- Plasma Mobile
- postmarketOS
- Sailfish OS
- SHR (operating system)
- Tizen
- Ubuntu Touch
- webOS
- PureOS (announced for 2018)
- Firefox OS (discontinued)
- Openmoko Linux (discontinued)
- OpenZaurus (discontinued)
- Bada (discontinued)
- Ubuntu Mobile (discontinued in favor of Ubuntu Touch)
- Maemo (discontinued)
- Moblin (discontinued)
- MeeGo (discontinued)

Middlewares

- BusyBox – small footprint alternative to GNU Core Utilities, under GNU GPLv2

- Toybox – BSD licenseed alternative to BusyBox
- mer
- Smart Common Input Method
- Maliit
- Intelligent Input Bus
- Uim
- Fcitx

UI

- KDE Plasma Workspaces
- Unity
- GPE Palmtop Environment
- OPIE user interface

MODULE 6

LTE :

LTE = Long Term Evolution
LTE = 4thGeneration Wireless Technology(3.9G according to the experts)
Standardized by 3GPP and included in ITU-R M1457

LTE Architecture
LTE encompasses the evolution of:

- the radio access through the E-UTRAN
- the non-radio aspects under the term **System Architecture Evolution (SAE)**
- Entire system composed of both LTE and SAE is called the **Evolved Packet System (EPS)**

At a high-level, the network is comprised of:

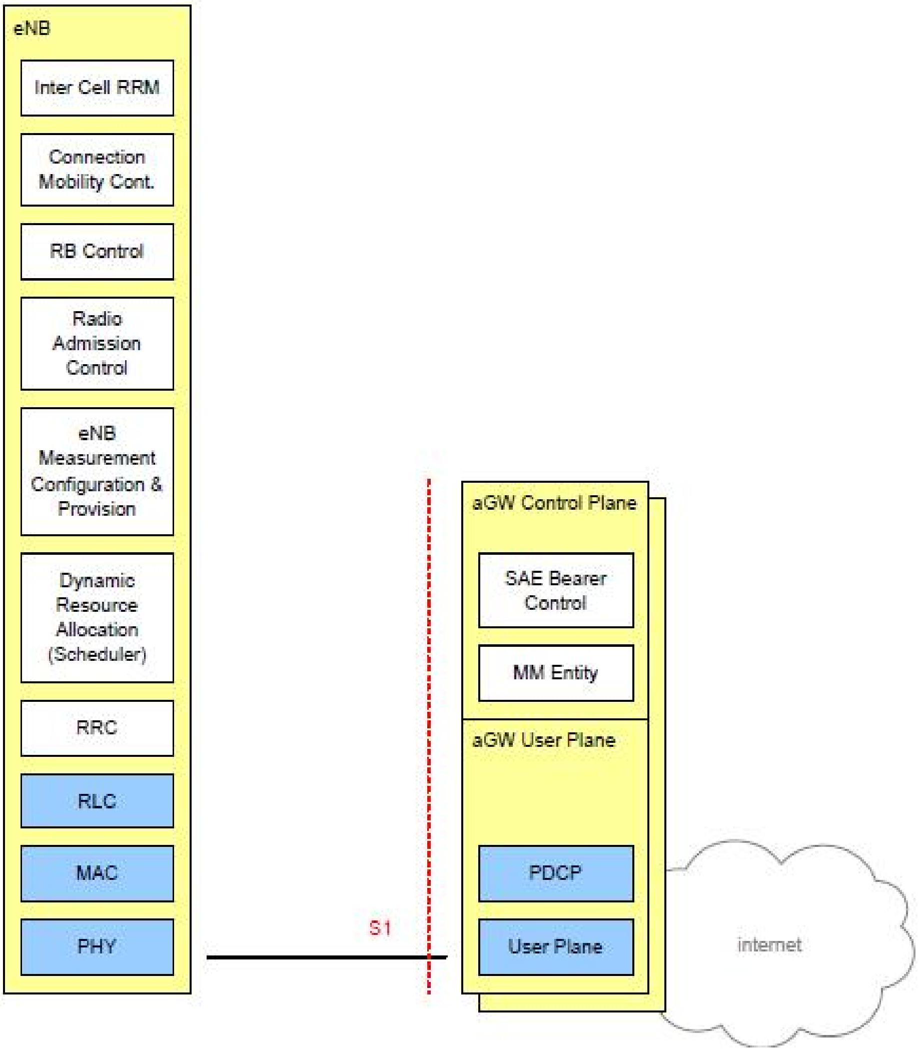
- Core Network (CN), called **Evolved Packet Core (EPC)** in SAE access network (E-UTRAN)
- A bearer is an IP packet flow with a defined QoS between the gateway and the User Terminal (UE)
- CN is responsible for overall control of UE and establishment of the bearers

LTE is a system with larger bandwidths (up to 20 MHz), low latency and Packet optimized radio access technology having peak data rates of 100 Mbps in downlink and 50 Mbps in the uplink . Radio access technology for LTE is OFDM (Orthogonal frequency division multiplexing) which provides higher spectral efficiency and more robustness against mulitpath and fading, as compared to CDMA (Code division multiple access). In order to offer the operators increased flexibility in network deployment, the LTE system supports bandwidth scalability and both FDD and TDD duplexing methods. The system also supports both unicast and multicast traffic – in cell sizes from local area or micro cells (hundreds of meters) up to large macro cells (>10 km in radius).

key requirements and capability targets for the long-term evolution are :

· Low latency : for both user plane and control plane, with a 5MHz spectrum allocation the latency target is below 5 ms
· Bandwidth Scalability : different bandwidths can be used depending upon the requirements (1.25 to 20 MHz)
· Peak Data Rates : 100 Mbps for DL , 50 Mbps for UL
· 2 to 3 times capacity over existing Release 6 scenarios with HSUPA
· 2 to 4 times capacity over existing Release 6 scenarios with HSDPA
· Only Packet Switched Domain support
· Improved Cell edge performance
· Inter-working with the existing 2G and 3G systems and non-3GPP systems
· Optimized for low mobile speed but also support high mobile speeds
· Reduction of complexity in both system and terminals
· Ease of migration from existing networks

· Simplification and minimization of the number of interfaces Network Architecture LTE architecture is characterised by three special requirements: support for PS domain only, low latency and reduced cost. To achieve the above objectives and to overcome the complexities of the previous network architectures, LTE must be designed to contain fewer network nodes. This is important because smaller number of network nodes reduces overall amount of protocol-related processing, cost of testing and number of interfaces. It also translates into ease of optimizing radio interface protocols. It can be done by merging some control protocols and using shorter signaling sequences resulting into rapid session setups. LTE uses two-node architecture. Figure gives an overview of the E-UTRAN architecture where yellow-shaded boxes depict the logical nodes, white boxes depict the functional entities of the C-plane, and blue boxes depict the functional entities of the U-plane.



The E-UTRAN consists of:

· eNB (Enhanced Node B)

· aGW (access Gateway)

eNB is the basic access network element covering a single cell or installed on one site. It provides the E-UTRA user plane (PDCP/RLC/MAC/PHY) and control plane (RRC) protocol terminations towards the UE(User Equipment). Two eNBs are connected with each other through X2 interface. LTE is designed to give eNBs a greater degree of intelligence to reduce the overhead. As a result, functions for Radio Resource

Management are provided by eNB. This includes Radio Bearer Control, Radio Admission Control, Connection Mobility Control, Dynamic allocation of resources to UEs in both uplink and downlink. eNB is involved in security services by encryption of user data stream and routing of user plane data towards serving gateway.Moreover, it also carries out scheduling and transmission of paging messages and BCCH information.

aGW is one level above eNB. A aGW can be connected to one or more eNBs depending upon the network design. aGW performs many different function, together with paging origination, ciphering of user plane data and SAE bearer control. aGW is functionally divided into two parts, MME (Mobility Management Entity) and UPE (User Plane Entity). MME is the control plane part of aGW. Its functionalities include management and storage of temporary user IDs, Termination of U-plane packets for paging reasons and management and NAS security. On the other hand, UPE is responsible for tasks related to user plane. It is accountable for Packet routing and forwarding, allocation of local IP address for mobility, charging for roaming and anchoring for inter eNB mobility, charging of paging messages to eNBs and inter-3GPP access Mobility.

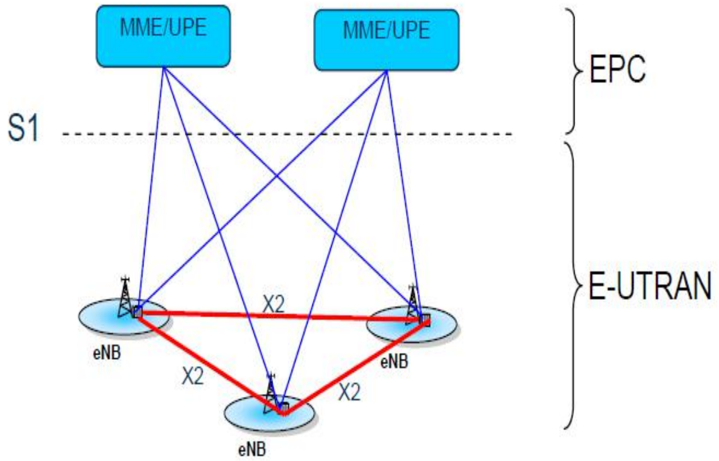
E-UTRAN Interfaces

One of the objectives of EUTRAN is to simplify and reduce the number of interfaces between different network elements. Interfaces between different network elements are S1 (eNodeBaGW) and X2 (inter ENodeB) as shown in figure

S1 is the interface between eNB and UPE. This interface can be subdivided into two parts :

· C-plane: S1-C is the interface between eNB and MME function in EPC

· U-plane: S1-U is the interface between eNB and UPE function in EPC



From the S1 perspective, the EUTRAN access point is an eNB and the EPC access point is either the control plane MME node or the user plane SAE gateway logical node. S1 access point shall independently fulfill the requirements of the relevant S1 specifications. S1 interface supports many functions which include initial context setup, UE context management and mobility functions. Initial context setup function supports the establishment of overall initial UE context plus SAE bearer context, security context, roaming restriction, UE capability information, etc. in the eNB to enable idle-to-Active transition. S1 interface also establishes and releases the UE contexts in eNB and in EPC to support user signaling. Moreover, S1 also provide mobility functions for handover. This can be intra-LTE handover or inter-3GPP handover (with a system other than LTE) .

X2 interface allows the interconnection between eNBs. X2 has the status of an open interface. It supports the signal information exchange between two eNBs, along with the forwarding of PDUs to their destination. In terms of logical point of view, X2 is a point-to-point interface within E-UTRAN. Therefore, it is possible to create an X2 interface between two eNBs even if there is no physical and direct connection between them .

X2 facilitates the interconnection between eNBs of different vendors and offers a continuation of the services offered via S1 interface for a seamless network. In addition, it makes possible the introduction of new future technologies by clearly separating radio network and transport network functionalities. With significant improvements in the radio interface and other components, enabling a lower data access cost per megabyte, as well as several potentially important new services, 3G Long- Term Evolution (LTE) will bring substantial technological improvements. These efforts are expected to deliver economic benefits to operators, and therefore provide a decisive advantage over alternative wireless technologies, keeping the mobile cellular systems competitive during the next decade.

Radio Planning :

Coverage planning includes radio link budget and coverage analysis. RLB computes the power received by the user (receiver) given a specific transmitted power (from the transmitter or base station). RLB comprises of all the gains and losses in the path of signal from transmitter to the receiver. This includes transmitter and receiver gains as well as losses and the effect of the wireless medium between them. Free space propagation loss, fast fading and slow fading is taken into account. Additionally, parameters that are particular to some systems are also considered. Frequency hopping and antenna diversity margins are two examples.

Coverage Planning is the first step in the process of dimensioning. It gives an estimatie of the resources needed to provide service in the deployment area with the given system parameters, without any capacity concern. Therefore, it gives an assessment of the resources needed to cover the area under consideration, so that the transmitters and receivers can listen to each other. In other words, there are no QoS concerns involved in this process.

Coverage planning consists of evaluation of DL and UL radio link budgets. The maximum path loss is calculated based on the required SINR level at the receiver, taking into account the extent of the interference caused by traffic. The minimum of the maximum path losses in UL and DL directions is converted into cell radius, by using a propagation model appropriate to the deployment area. Radio Link Budget is the most prominent component of coverage planning exercise.

Radio Link Budget (RLB) is calculated in order to estimate the allowed path loss. Transmission powers, antenna gains, system losses, diversity gains, fading margins, etc. are taken into account in a RLB. RLB gives the maximum allowed path loss, from which cell size is calculated using a suitable propagation model.

For LTE, the basic RLB equation can be written as follows (in units of dB):

PathLos = TxPower + TxGains - TxLosses - Re*quiredSINR + RxGains - RxLosses - RxNoise*

Where,
Path Loss = Total path loss encountered by the signal from transmitter to receiver (W)
TxPowerdB = Power transmitted by the transmitter antenna (dBm)
TxGainsdB = Gain of transmitter antenna (dB)
TxLossesdB = Transmitter losses (dB)
RequiredSINRdB = Minimum required SINR for the signal to be received at the receiver with the required quality or strength (dB)
RxGainsdB = Gain of receiver antenna (dB)
RxLossesdB = Receiver losses (dB)
RxNoisedB = Receiver Noise (dBm)

$$PathLoss = \frac{TxPower \bullet TxGains \bullet RxGains}{TxLosses \bullet \mathrm{Re}\,quiredSINR \bullet RxLosses \bullet RxNoise}$$

Where,
Path Loss = Total path loss encountered by the signal from transmitter to receiver (W)
TxPower = Power transmitted by the transmitter antenna (W)
TxGains = Gain of transmitter antenna TxLosses = Transmitter losses (W)
RequiredSINR = Minimum required SINR for the signal to be received at the receiver with the required quality or strength
RxGains = Gain of receiver antenna
RxLosses = Receiver losses (W)
RxNoise = Receiver Noise (W)
In LTE, the basic performance indicator is 'Required SINR'. Maximum allowed path loss is calculated according to the condition:

$$\begin{cases} SINR \geq RequiredSINR \\ SINR = \dfrac{AveRxPower}{Interference + RxNoise} = \dfrac{AveRxPower}{OwnCellInterference + OtherCellInterference + RxNoise} \end{cases}$$

Where,
SINR = Signal to interference and noise ratio
AveRxPower = Average received power (W)
Interference = Interference power (W)
OwnCellInterference = Power due to own cell interference (W)
OtherCellInterference = Power received for neighboring cells (W)
In downlink, assuming the maximum available transmission power is equally divided over the
cell bandwidth, the average received power (AveRxPowerDL) in the bandwidth allocated to the
user is derived as follows:

$$AveRxPowerDL = \frac{AveTxPower}{LinkLossDL} = \frac{MaxNodeBTxPower}{CellBandwidth} \cdot \frac{AllocatedBandwidth}{LinkLossDL}$$

Where,

SINR = Signal to interference and noise ratio

AveTxPower = Average transmitted power (W)

LinkLossDL = Total link loss in downlink (W)

MaxNodeBTxPower = Maximum Power transmitted from NodeB (W)

CellBandwidth = Allocated bandwidth of LTE network cell (MHz)

AllocatedBandwidth = Bandwidth of channel over which the signal is transmitted (MHz)

The MaxNodeBTxPower in LTE depends on the cell bandwidth, which can range from 1.25 to 20 MHz. Specifically, MaxNodeTxPower is 20 Watt (43 dBm) up to 5 MHz and 40 Watt (46 dBm) above this limit .

In uplink, assuming no power control, the average received power (AveRxPowerUL) is:

$$AveRxPowerUL = \frac{MaxUETxPower}{LinkLossUL}$$

Where,

MaxUETxPower= Max transmission power of user equipment (W)

LinkLossUL = Total link loss in uplink (W)

The MaxUETxPower can be either 0.125 W or 0.25 W (21 or 24 dBm). The LinklossUL includes the distance-dependent Pathloss and all other gains and losses at the transmitter and the receiver.

<u>5 G Technology :</u>

5G is the name given to the next generation of mobile data connectivity. It will definitely provide great speeds between 10Gbps to 100Gbps and it will have enough capacity. But the thing that separated 5G from 4G is latency; the latency provided by 4G is between 40ms to 60ms, whereas in 5G it will provide ultra latency between 1ms to 10ms.Then in future we can actually watch a cricket or football or any conference actually live without any delay. 5G is a technology that will appear to be invisible; it will be just there like electricity.

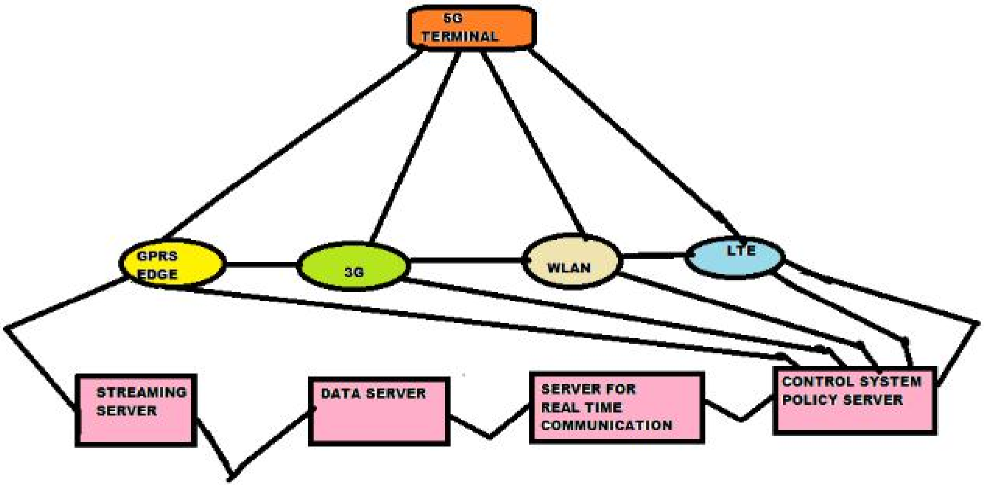
Some of the key technologies to be used in 5G are massive MIMO, device-to-device communication, millimetre wave communication and some multiple access techniques like beam division multiple access (BDMA).

5G ARCHITECTURE

In general a research have shown that a mobile subscriber stays inside for approximately 80 percentage of time and outside for approximately 20 percentage of time. From this scenario for a subscriber inside will receive a call when signal penetration through the walls, then that signal will undergo many losses and hence efficiency will be less, bit rate will be low and low energy efficiency. This is happening because there is only one base station at the middle of the cell site that handles all these. When it comes to 5G ar-chitecture it has different models for outside and inside. By doing so some of the penetration losses can be reduced. This will be implemented using massive MIMO technology by deploying hundreds of antennas.

Normally in MIMO system we utilized two or four antennas, by using massive MIMO we are increasing number of transmitter and receiver antennas approximately between ten to hundred, by doing so we are increasing the capacity gain [1]. In massive MIMO network two things are setup for establishing a reliable network. First, a base station will be installed in a cell site with multiple antennas on it or in the area of cell; these are connected with the base station using optical fibre cables. When a subscriber is outside he is connected to the base station directly or connected via multiple hops from the antennas creating virtual massive MIMO network. Secondly an antenna array will be installed in every building; these antennas will be in line of sight with the base station. The communication inside is done using technologies like Wi-Fi, visible light communication, millimetre wave communication etc .

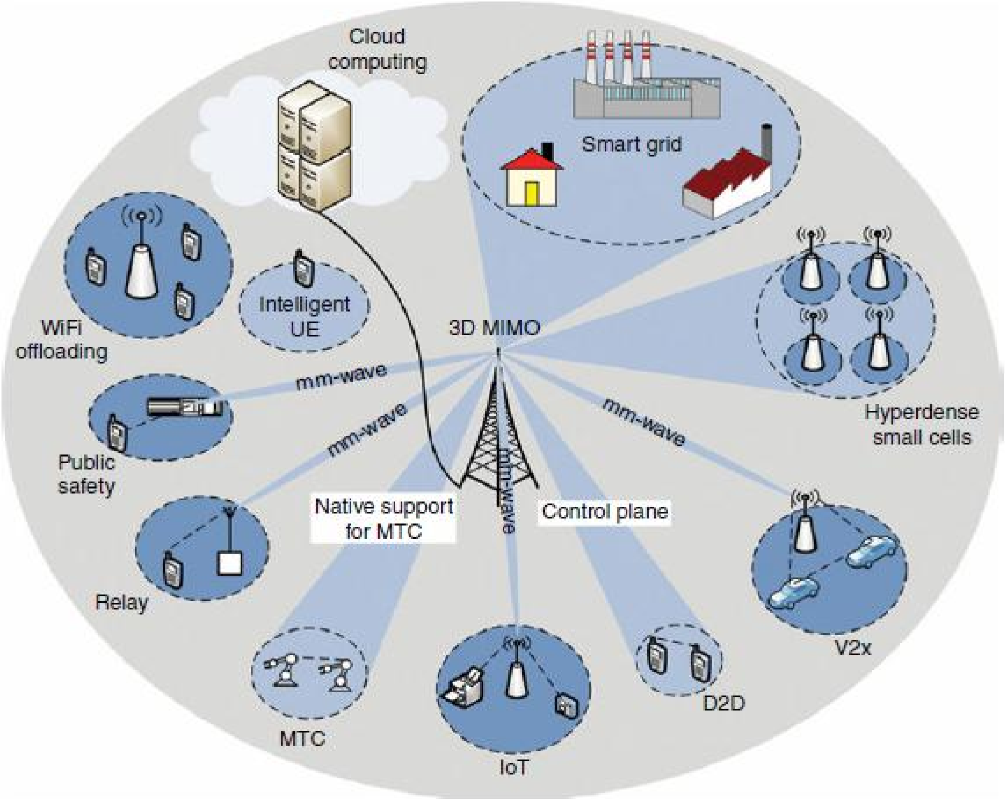
The proposed 5th generation cellular architecture is shown in figure below.The architecture shows that it is an IP based model for wireless and mobile networks interoperability.User terminal plays a very important role in this system.


5G TERMINAL
GPRS EDGE
3G
WLAN
LTE
STREAMING SERVER
DATA SERVER
SERVER FOR REAL TIME COMMUNICATION
CONTROL SYSTEM POLICY SERVER

Radio access technology (RAT) means a physical connection method for a radio based communication network. It is shown in the figure that, with in each of the terminals each RATs are the IP link to the outside world and also in mobile terminal there should be different radio interface for each RATs. If someone wants to access four different radio access technologies than four different accesses specific interfaces should be there in the mobile terminal and all of them should be activated at the same time in order to function the architecture.

It is called as the IP based model and the main purpose is to ensure control data for proper routing of IP packets belonging to a certain application connections. Routing of packets is carried out by the users with some specific policies and rules. Application connections are established between clients and servers in the internet i.e. via sockets. These Internet sockets are end points for data communication flows and each socket of the web is an unique combination of local IP address and appropriate local transport communications port, target IP address and target suitable communication port, and type of transport protocol.

In case of interoperability between heterogeneous networks and for the vertical handover between the respective radio technologies, the local Internet Protocol address and destination Internet Protocol address should be fixed and unchanged. Fixing of these two parameters should certain handover transparency to the Internet connection end-to-end, when there is a mobile telephone user at least on one end of such connection. In order to preserve the proper design of the packets and to reduce or prevent packets losses, drive off to the target destination and vice versa should be unique and it should be using the same path.



5G System architecture

Cognitive radio technology is one of the key concepts of 5G. Cognitive radio technology, also known as smart-radio: allowing different radio technologies to share the same spectrum efficiently by adaptively finding unused spectrum and adapting the transmission scheme to the requirements of the technologies currently sharing the spectrum. This powerful radio resource management is achieved in a distributed fashion, and relies on software defined ratio.

Small cells by strict definition are low-power wireless access points that operate in licensed spectrum are operator-managed and provide improved cellular coverage, capacity and applications for homes and enterprises as well as metropolitan and rural public spaces.

MASSIVE MIMO

MIMO stands for Multiple Input and Multiple Output that means we use multiple antennas at the trans-mitter and receiver, this is called spatial diversity. spatial diversity was often limited to systems that switched between two antennas. If we use multiple antennas at the transmitter we call it as transmitter diversity and at receiver we call it as receiver diversity. By doing so we are increasing the channel capacity and reliability of the wireless network. During the start of this technology point-to-point MIMO were used were both transmitter and receiver have multiple antennas, soon it was over taken by multi-user MIMO where there were multiple antennas at the base station which communicated with the single antenna receiver. Due to this cost of the whole system was reduced because now costly antennas were only needed at the base stations, cheap antennas can be used at the single antenna end

One advantage of this technology is that we can increase the capacity and reliability, the other is that we can reduce the error rate. If we can transmit multiple versions of our message through different chan-nels the probability all the signals will be affected same will be less. At the receiver these multiple copies are received and processed to get our original message. Hence Diversity also helps to stabilise a com-munication link, improves its performance, and reduces error rate. Due to all these advantages MIMO technology is deployed as a part of communication standards such as 802.11 (WiFi), 802.16 (WiMAX), and LTE

The communication in MIMO take place in two formats called spatial diversity and spatial multiplex-ing. In spatial diversity, the same data is transmitted through different paths; the data is received at the multiple antennas and processed. By spatial multiplexing we can improve the reliability of the link. The other technique is spatial multiplexing, where the data is divided into small parts and different part is transmitted through different path, by doing so we are increasing the speed, but reliability is less.

A MIMO system consists of a number of transmitter and receiver antennas and a fading channel through which the data will be sent. Let us consider we have *M,* number of transmitter antennas and *N,* number of receiver antenna i.e. we form a matrix for transmitter and receiver antennas having t number of rows in transition matrix similarly r number of rows in receiver matrix.

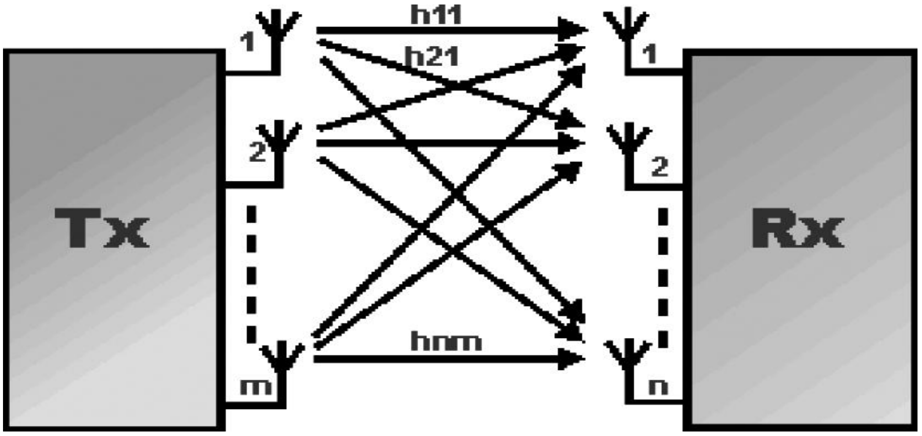
The basic equation for MIMO system is given by Y= H.X + W

Where, Y = N x 1 Receiver matrix

H = N x M Channel matrix

X = M x 1 Transition matrix

W = Noise



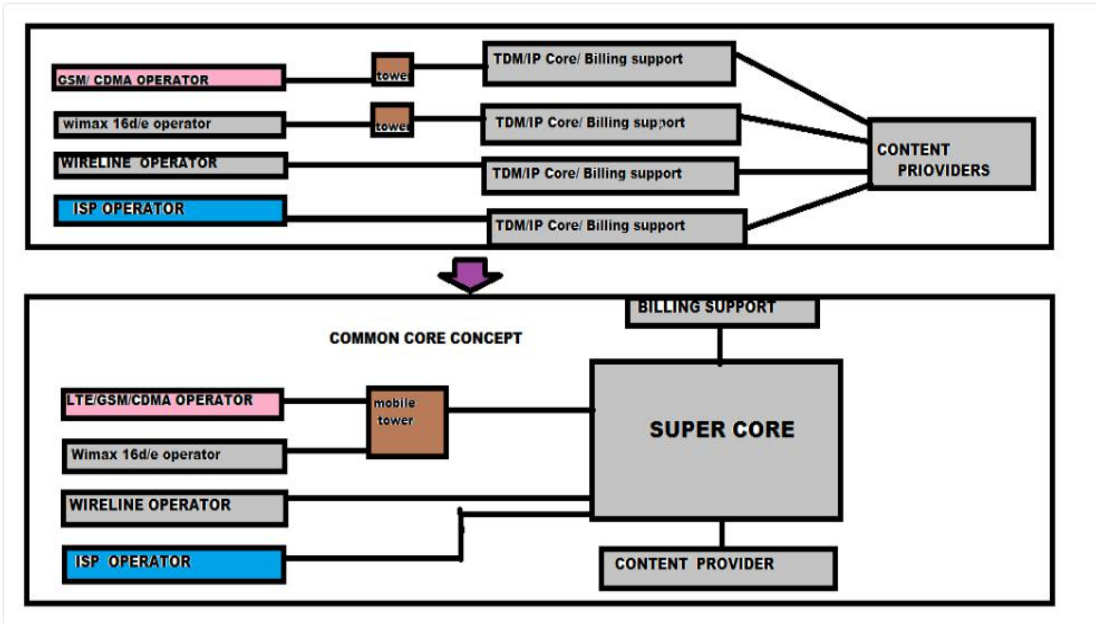
The general MIMO system consists of not more than 10 an-tennas, whereas in massive MIMO there could be 100 or more antennas

Super Core Concept :

The 5G will have the following three main technologies:

- Nanotechnology.
- Cloud Computing.
- All Flat IP Platform.

The existing telecom networks are arranged in an ordered way, where subscriber traffic is aggregated at aggregation point (BSC/RNC) and then drives off to gateways. Flat Internet Protocol architecture will lessen burden on aggregation point and traffic will directly move from Base station to Media gateways. When transition from legacy (TDM, ATM) platforms to IP will be concluded, a common ALL IP platform will be appeared. All network operators (GSM, CDMA, Wi-max, Wire line) can be connected to one Super core with enormous capacity. The super core concept will roughly calculate all interconnecting charges, which is now days network operator is facing.



Applications of 5G Technology

1. One can be able to feel her kid's stroke when he/she is in her mother's womb.

2. One can be able to perceive his/her sugar level with his/her mobile.

3. One can be able to charge his/her mobile with his/her own heartbeat.

4. One can be able to view his/her residence in his/her mobile when someone enters.

5. The mobile will ring according to our mood.

6. One can be able to pay all bills in a single payment with his/her mobile.

7. One can get the live share value.

8. One can be able to navigate the train for which he/she might be waiting.

9. One can be able to vote from his/her mobile.

10. One can be able to know the exact time of his/her child birth that too in nanoseconds.

11. One can be able to sense tsunami/earthquake before it occurs.

12. Our mobile can share our work load.

13. One can get an alert in his/her mobile when someone opens his/ her intelligent car.

14. One can be ale to lock his/her car or bike with his/her mobile when he/she forgets to do so.

15. We can be able to expand our coverage using our mobile phone.

16. Our mobile can perform radio resource management.

Features Of 5G :

- Very High speed, high capacity, and low cost per bit.
- It supports interactive multimedia, voice, video, Internet, and other broadband services, more effective and more attractive, and have Bi-directional, accurate traffic statistics.

- 5G technology offers Global access and service portability.
- It offers the high quality services due to high error tolerance.
- It is providing large broadcasting capacity up to Gigabit which supporting almost 65,000 connections at a time.

- More applications combined with artificial intelligent (AI) as human life will be surrounded by artificial sensors which could be communicating with mobile phones .
- 5G technology use remote management that user can get better and fast solution.
- The uploading and downloading speed of 5G technology is very high.
- 5G technology offer high resolution for crazy cell phone user and bi-directional large bandwidth shaping .
- 5G technology offer transporter class gateway with unparalleled consistency

Every mobile phone in a 5G wireless system will have an IP address. The technology is expected to support virtual private networks and advanced billing interfaces. The remote diagnostics also a great feature of 5G. The uploading and also downloading speed of 5G technology will be very high. The traffic statistics will be more accurate by using 5G technology. 5G technology provides large broadcasting of data in gigabits which supports almost 65000 connections.

LiFi :

Li-Fi can be thought of as a light-based Wi-Fi. That is, it uses light instead of radio waves to transmit information. And instead of Wi-Fi modems, Li-Fi would use transceiver-fitted LED lamps that can light a room as well as transmit and receive information. Since simple light bulbs are used, there can technically be any number of access points.

This technology uses a part of the electromagnetic spectrum that is still not greatly utilized- The Visible Spectrum. Light is in fact very much part of our lives for millions and millions of years and does not have any major ill effect. Moreover there is 10,000 times more space available in this spectrum and just counting on the bulbs in use, it also multiplies to 10,000 times more availability as an infrastructure, globally.

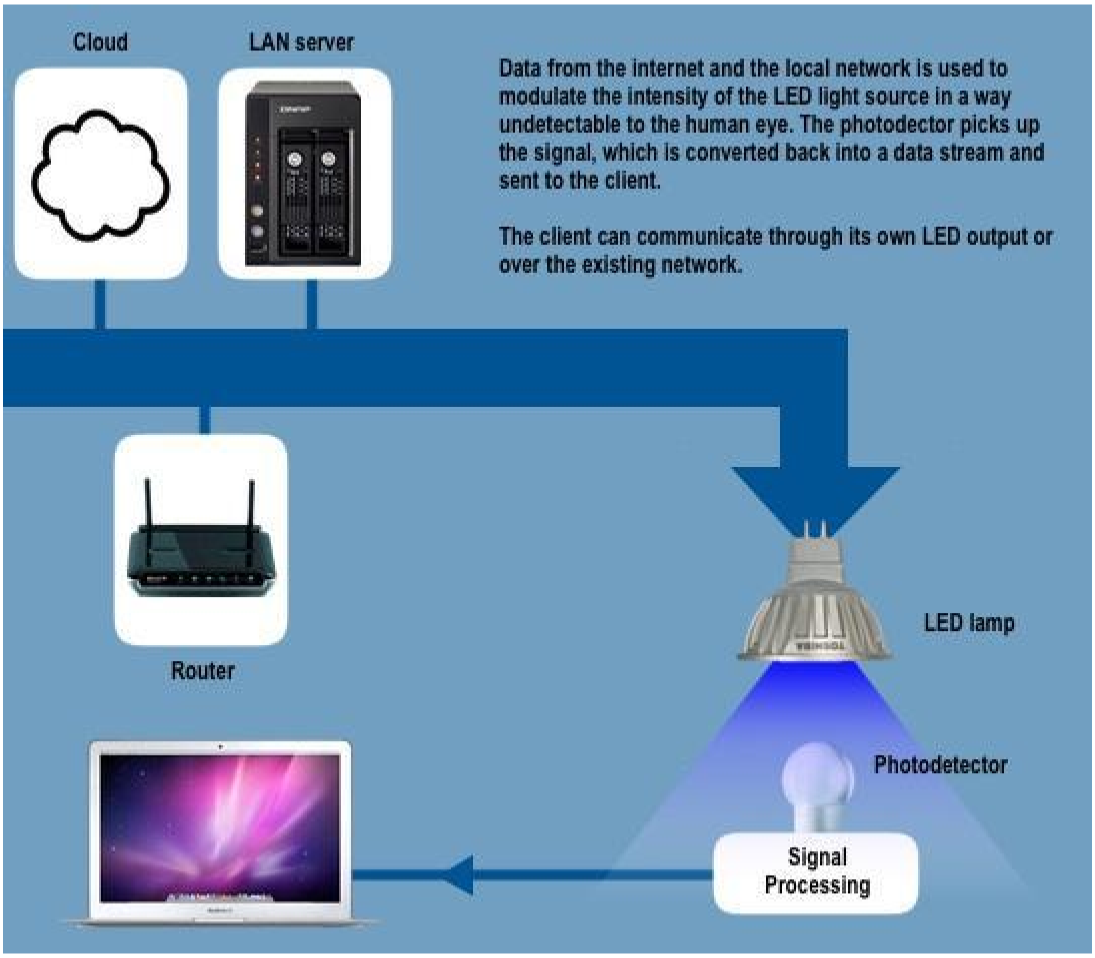
It is possible to encode data in the light by varying the rate at which the LEDs flicker on and off to give different strings of 1s and 0s. The LED intensity is modulated so rapidly that human eyes cannot notice, so the output appears constant.

Working :

Li-Fi is typically implemented using white LED light bulbs at the downlink transmitter. These devices are normally used for illumination only by applying a constant current. However, by fast and subtle variations of the current, the optical output can be made to vary at extremely high speeds.

This very property of optical current is used in Li-Fi setup. The operational procedure is very simple-, if the LED is on, you transmit a digital 1, if it’s off you transmit a 0. The LEDs can be switched on and off very quickly, which gives nice opportunities for transmitting data. Hence all that is required is some LEDs and a controller that code data into those LEDs. All one has to do is to vary the rate at which the LED’s flicker depending upon the data we want to encode.

Further enhancements can be made in this method, like using an array of LEDs for parallel data transmission, or using mixtures of red, green and blue LEDs to alter the light’s frequency with each frequency encoding a different data channel. Such advancements promise a theoretical speed of 10 Gbps – meaning one can download a full high-definition film in just 30 seconds.


Cloud
LAN server
Data from the internet and the local network is used to modulate the intensity of the LED light source in a way undetectable to the human eye. The photodector picks up the signal, which is converted back into a data stream and sent to the client.
The client can communicate through its own LED output or over the existing network.
Router
LED lamp
Photodetector
Signal Processing

Light is inherently safe and can be used in places where radio frequency communication is often deemed problematic, such as in aircraft cabins or hospitals. So visible light communication not only has the potential to solve the problem of lack of spectrum space, but can also enable novel application. The visible light spectrum is unused, it's not regulated, and can be used for communication at very high speeds.

Application

- **You Might Just Live Longer :** You Might Just Live Longer For a long time, med ical technology has lagged behind the rest of the wireless world. Operating rooms do not allow Wi-Fi over radiation concerns, Li-Fi solves problems: lights are not only allowed in operating rooms
- **Airlines**
- **Smarter Power Plants**

Wi-Fi and many other radiation types are bad for sensitive areas.Li-Fi could offer safe, abundant connectivity for all areas of these sensitive locations.

- **Undersea Awesomeness**

Security Issues in Mobile Computing

The mobile computing is the communication between computing devices without a physical connection between them through wireless networks, which mean there are some of new mobile security issues that are originated from wireless security issues. The security issues and threats of mobile computing can be divided into two categories: security issues that related to transmission of information over wireless networks, and the issues that related to information and data residing on mobile devices.

Confidentiality: Preventing unauthorized users from gaining access to critical information of any particular user.

Integrity: Ensures unauthorized modification, destruction or creation of information cannot take place.

Availability: Ensuring authorized users getting the access they require.

Legitimate: Ensuring that only authorized users have access to services.

E. Accountability: Ensuring that the users are held responsible for their security related activities by arranging the user and his/her activities are linked if and when necessary.

B Wireless Security Issues

The security issues that related of wireless networks are happened by intercepted of their radio signals by hacker, and by non-management of its network entirely by user because most of wireless networks are dependent on other private networks which managed by others, so the user has less control of security procedures. There are some of the main security issues of mobile computing, which introduced by using of wireless networks are:

Denial of Service (DOS) attacks: It's one of common attacks of all kinds of networks and specially in wireless network, which mean the prevent of users from using network services by sending large amounts of unneeded data or connection requests to the communication server by an attacker which cause slow network and therefore the users cannot benefit from the use of its service.

Traffic Analysis: It's identifying and monitoring the communicating between users through listening to traffic flowing in the wireless channel, in order to access to private information of users that can be badly used by attacker.

Eavesdropping: The attacker can be log on to the wireless network and get access to sensitive data, this happens if the wireless a network was not enough secure and also the information was not encrypted. Session Interception and Messages Modification: Its interception the session and modify transmitted data in this session by the attacker through scenario which called: man in the middle which inserts the attacker’s host between sender and receiver host.

Spoofing: The attacker is impersonating an authorized account of another user to access sensitive data and unauthorized services.

Captured and Re transmitted Messages: Its can get some of network services to attacker by get unauthorized access through capture a total message and replay it with some modifications to the same destination or another

C Device Security Issues Mobile devices are vulnerable to new types of security attacks and vulnerable to theft not because of the get these devices itself, but because of get to sensitive data That exists within its devices. Mobile computing, like any computer software may damage by malware such as Virus, Spyware and Trojan. A virus is a real part of malicious software and Spyware is gathering information about the user without his knowledge. Some of main new mobile computing security issues introduced by using mobile devices include:

Pull Attacks: In pull Attack, the attacker controls the device as a source of data by an attacker which obtained data by device itself.

Push Attacks: It's creation a malicious code at mobile device by attacker and he may spread it to affect on other elements of the network.

Forced De-authentication: The attacker convinces the mobile end-point to drop its connection and re-connection to get new signal, then he inserts his device between a mobile device and the network. Multi-protocol Communication: It is the ability of many mobile devices to operate using multiple protocols, e.g. a cellular provider’s network protocol, most of the protocols have a security holes, which help the attacker to exploit this weakness and access to the device.

Mobility: The mobility of users and their data that would introduce security threats determined in the location of a user, so it must be replicate of user profiles at different locations to allow roaming via different places without any concern regarding access to personal and sensitive data in any place and at any time. But the repetition of sensitive data on different sites that increase of security threats.

Disconnections: When the mobile devices cross different places it occurs a frequent disconnections caused by external party resulting hand off.

CONTENT BEYOND SYLLABUS

Current mobile technologies

3G

3G or third generation mobile telecommunications is a generation of standards for mobile phones and mobile telecommunication services fulfilling the International Mobile Telecommunications-2000 (IMT-2000) specifications by the International Telecommunication Union. Application services include wide-area wireless voice telephone, mobile Internet access, video calls and mobile TV, all in a mobile environment.

Global Positioning System (GPS)

The Global Positioning System (GPS) is a space-based satellite navigation system that provides location and time information in all weather, anywhere on or near the Earth, where there is an unobstructed line of sight to four or more GPS satellites. The GPS program provides critical capabilities to military, civil and commercial users around the world. In addition, GPS is the backbone for modernizing the global air traffic system, weather, and location services.

Long Term Evolution (LTE)

LTE is a standard for wireless communication of high-speed data for mobile phones and data terminals. It is based on the GSM/EDGE and UMTS/HSPA network technologies, increasing the capacity and speed using new modulation techniques. It is related with the implementation of fourth Generation (4G) technology.

WiMAX

WiMAX (Worldwide Interoperability for Microwave Access) is a wireless communications standard designed to provide 30 to 40 megabit-per-second data rates, with the latest update providing up to 1 Gbit/s for fixed stations. It is a part of a fourth generation or 4G wireless-communication technology. WiMAX far surpasses the 30-metre wireless range of a conventional Wi-Fi Local Area Network (LAN), offering a metropolitan area network with a signal radius of about 50 km. WiMAX offers data transfer rates that can be superior to conventional cable-modem and DSL connections, however, the bandwidth must be shared among multiple users and thus yields lower speed in practice.

Near Field Communication

Near Field Communication (NFC) is a set of standards for smartphones and similar devices to establish radio communication with each other by touching them together or bringing them into close proximity, usually no more than a few centimeters. Present and anticipated applications include contactless transactions, data exchange, and simplified setup of more complex communications such as Wi-Fi. Communication is also possible between an NFC device and an unpowered NFC chip, called a "tag".

Smartphones

This kind of phone combines the features of a PDA with that of a mobile phone or camera phone. It has a superior edge over other kinds of mobile phones. Smartphones have the capability to run multiple programs concurrently. These phones include high-resolution touch screens, web browsers that can access and properly display standard web pages rather than just mobile-optimized sites, and high-speed data access via Wi-Fi and high speed cellular broadband.

The most common mobile Operating Systems (OS) used by modern smartphones include Google's Android, Apple's iOS, Nokia's Symbian, RIM's BlackBerry OS, Samsung's Bada, Microsoft's Windows Phone, and embedded Linux distributions such as Maemo and MeeGo. Such operating systems can be installed on different phone models, and typically each device can receive multiple OS software updates over its lifetime.

Personal Digital Assistant (PDA)

The main purpose of this device is to act as an electronic organizer or day planner that is portable, easy to use and capable of sharing information with your computer systems.

PDA is an extension of the PC, not a replacement. These systems are capable of sharing information with a computer system through a process or service known as synchronization. Both devices will access each other to check for changes or updates in the individual devices. The use of infrared and Bluetooth connections enables these devices to always be synchronized.