Mobile Computing

IV B.TECH - I SEMESTER (JNTUH-R15)

Mr. C Raghavendra, Assistant Professor, CSE
Ms. K Mayuri, Assistant Professor, CSE
Ms. Geetha Yadav, Assistant Professor, CSE
Ms. K Radhika, Assistant Professor, CSE

COMPUTER SCIENCE AND ENGINEERING
INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
DUNDIGAL, HYDERABAD-500 043

UNIT – I
CONTENTS

- Mobile Communication
- Mobile Computing
- Novel Applications
- Location dependent services
- Effects of device portability
- Wireless n/w’s in comparison with fixed n/w’s
- Limitations
- Architecture (simple reference model)

Goal of Mobile Computing

- People and their machines should be able to access information and communicate with each other easily and securely, in any medium or combination of media

  – voice, data, image, video, or multimedia any time, anywhere, in a timely, cost-effective way.

Mobile Computing

- Mobile computing refers to computing in a distributed system in which some processes or processors *can move*.
  - Moving processes ⇒ logical mobility, realized by agents.
  - Moving processors ⇒ physical mobility, realized by moving devices.
- Mobile computing extends a distributed computing environment with a *new dimension of mobility*.
  - Most existing mobile computing systems are based on client-server computing systems.
  - Recent mobile computing solutions consider general distributed computing, namely, peer-to-peer computing environments.
- Many mobile computing techniques have their root in *distributed systems*.

Distributed system

- Definition:
  - A distributed system consists of a collection of autonomous computers, connected through a network and distribution middleware, which enables computers to coordinate their activities and to share the resources of the system, so that users perceive the system as a single, integrated computing facility.
- Multiple autonomous components
- Components are not shared by all users
- Resources may not be accessible
- Software runs in concurrent processes on different processors

Mobile Computing contd..,

- The process of computation on a mobile device
- In mobile computing, a set of distributed computing systems or service provider servers participate, connect, and synchronize through mobile communication protocols
- Mobile computing as a generic term describing ability to use the technology to wirelessly connect to and use centrally located information and/or application software through the application of small, portable, and wireless computing and communication devices

Mobile Computing contd..,

- Provides decentralized (distributed) computations on diversified devices, systems, and networks, which are mobile, synchronized, and interconnected via mobile communication standards and protocols.
- Mobile device does not restrict itself to just one application, such as, voice communication
- Offers mobility with computing power
- Facilitates a large number of applications on a single device

Introduction
Mobile Communication

- User Mobility:
  - Refers to a user who has access to the same or similar telecommunication services at different places.
  - user mobility: users communicate (wireless) “anytime, anywhere, with anyone”
- Device Portability:
  - The communication device moves with or without the user
  - device portability: devices can be connected anytime, anywhere to the network
- A communication device can the exhibit one of the following characteristics:
  - Fixed and wired e.g. Typical desktops computer
  - Mobile and wired e.g. some laptops
  - Fixed and wireless e.g. WIRELESS LANS
  - Mobile and wireless e.g. Mobile phones

Mobility Issues

- Bandwidth restrictions and variability
- Location-aware network operation
  - User may wake up in a new environment
  - Dynamic replication of data
- Querying wireless data & location-based responses
- Busty network activity during connections & handling disconnections
- Disconnection
  - OS and File System Issues - allow for disconnected operation
  - Database System Issues - when disconnected, based on local data

Portability Issues

- Battery power restrictions
- Risks to data

  - Physical damage, loss, theft

  - Unauthorized access

  - encrypt data stored on mobiles

  - Backup critical data to fixed (reliable) hosts
- Small user interface

  - Small displays due to battery power and aspect ratio constraints

  - Cannot open too many windows

  - Difficult to click on miniature icons

  - Input - Graffiti, (Dictionary-based) Expectation

  - Gesture or handwriting recognition with Stylus Pen Voice matching or voice recognition

Evolutions of the Mobile Systems

- 1G Analog systems
- 2G systems
  - voice communication
  - Circuit Switched
- 2.5G systems
  - Circuit switching for voice
  - Packet switching for data
- 3G systems
  - Packet switching for voice and data
  - High speed
  - Compatible with different access technologies

Evolutions of the Mobile Systems

- 3.5G systems
  - Evolved Radio Interface
  - IP based core Network
  - Compatible with different access technologies
- 4G systems
  - New Air Interface
  - Very High bit rate services
  - Convergence of Wireline, Wireless, and IP worlds

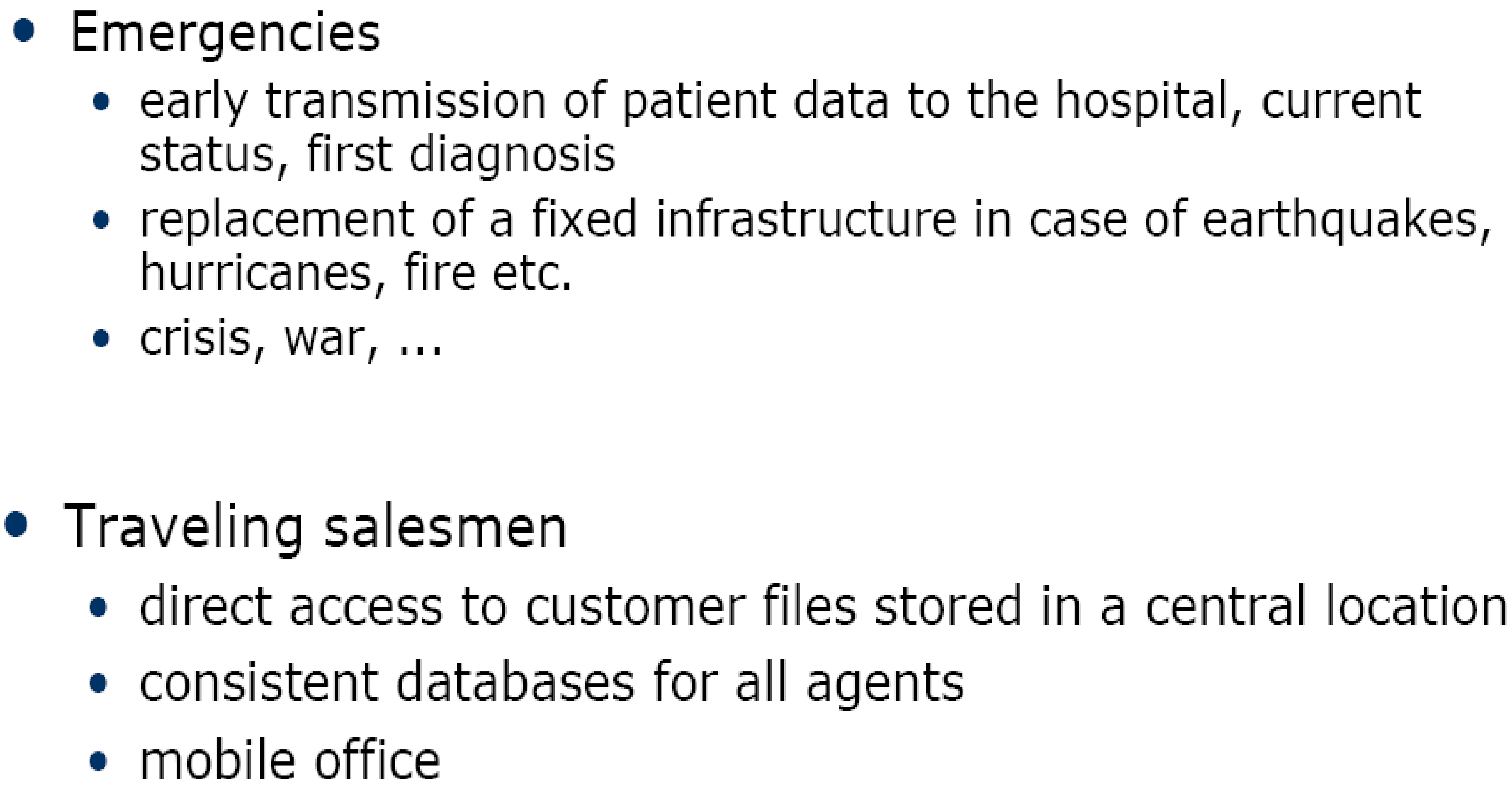
Evolutions of the Mobile Systems

	2G (2000)	EARLY 3G (2001)	LATE 3G (2003)	4G (2005)
AIR INTERFACE VOICE	CIRCUIT	CIRCUIT	CIRCUIT	PACKET
DATA	CIRCUIT	PACKET	PACKET	
ACCESS NETWK VOICE	CIRCUIT	CIRCUIT	PACKET	PACKET
DATA	CIRCUIT	CIRCUIT		
CORE NETWK VOICE	CIRCUIT	PACKET	PACKET	PACKET
DATA	OVERLAY PACKET			

Novel Applications

- Vehicles
  - transmission of news, road condition, weather, music via DAB/DVB-T
  - personal communication using GSM/UMTS
  - position via GPS
  - local ad-hoc network with vehicles close-by to prevent accidents, guidance system, redundancy
  - vehicle data (e.g., from busses, high-speed trains) can be transmitted in advance for maintenance

Novel Applications [2]



- Emergencies
  - early transmission of patient data to the hospital, current status, first diagnosis
  - replacement of a fixed infrastructure in case of earthquakes, hurricanes, fire etc.
  - crisis, war, ...

- Traveling salesmen
  - direct access to customer files stored in a central location
  - consistent databases for all agents
  - mobile office

Novel Applications [3]

- Replacement of fixed networks
  - remote sensors, e.g., weather, earth activities
  - flexibility for trade shows
  - LANs in historic buildings

- Entertainment, education, ...
  - outdoor Internet access
  - intelligent travel guide with up-to-date location dependent information
  - ad-hoc networks for multi user games

Location Dependent Services

- Location aware services
  - what services, e.g., printer, fax, phone, server etc. exist in the local environment
- Follow-on services
  - automatic call-forwarding, transmission of the actual workspace to the current location
- Information services
  - "push": e.g., current special offers in the supermarket
  - "pull": e.g., where is the Black Forrest Cheese Cake?
- Support services
  - caches, intermediate results, state information etc. "follow" the mobile device through the fixed network
- Privacy
  - who should gain knowledge about the location

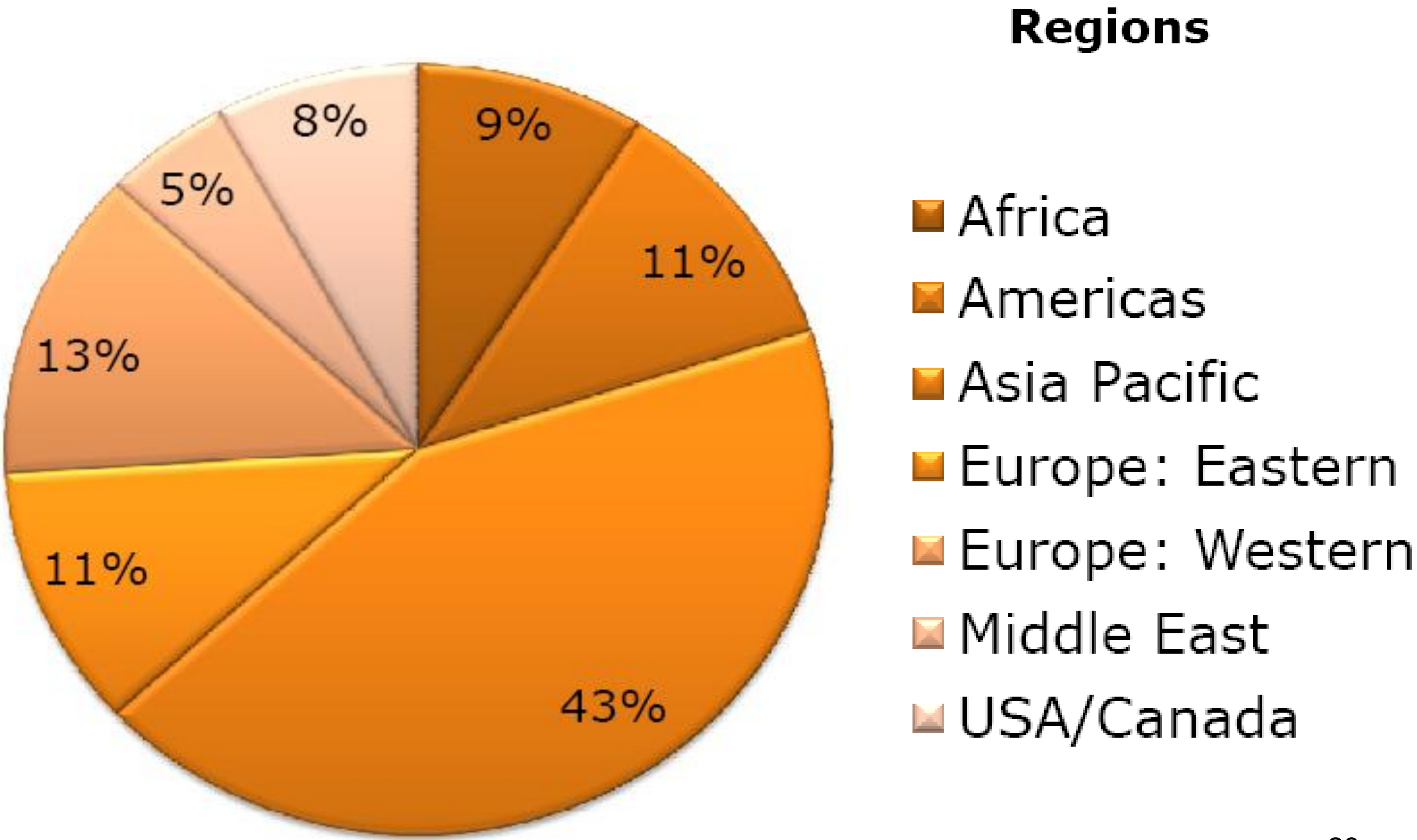
Effects of Device Portability

- Power consumption
  - limited computing power, low quality displays, small disks due to limited battery capacity
  - CPU: **power consumption ~ CV^2f**
    - C: internal capacity, reduced by integration
    - V: supply voltage, can be reduced to a certain limit
    - f: clock frequency, can be reduced temporally
- Loss of data
  - higher probability, has to be included in advance into the design (e.g., defects, theft)
- Limited user interfaces
  - compromise between size of fingers and portability
  - integration of character/voice recognition, abstract symbols
- Limited memory
  - limited usage of mass memories with moving parts
  - flash-memory or ? as alternative

Wireless Networks in Comparison with fixed networks

- Higher loss-rates due to interference
  - emissions of, e.g., engines, lightning
- Restrictive regulations of frequencies
  - frequencies have to be coordinated, useful frequencies are almost all occupied
- Low transmission rates
  - local some Mbit/s, regional currently, e.g., 53kbit/s with GSM/GPRS or about 150 kbit/s using EDGE
- Higher delays, higher jitter
  - connection setup time with GSM in the second range, several hundred milliseconds for other wireless systems
- Lower security, simpler active attacking
  - radio interface accessible for everyone, base station can be simulated, thus attracting calls from mobile phones
- Always shared medium
  - secure access mechanisms important

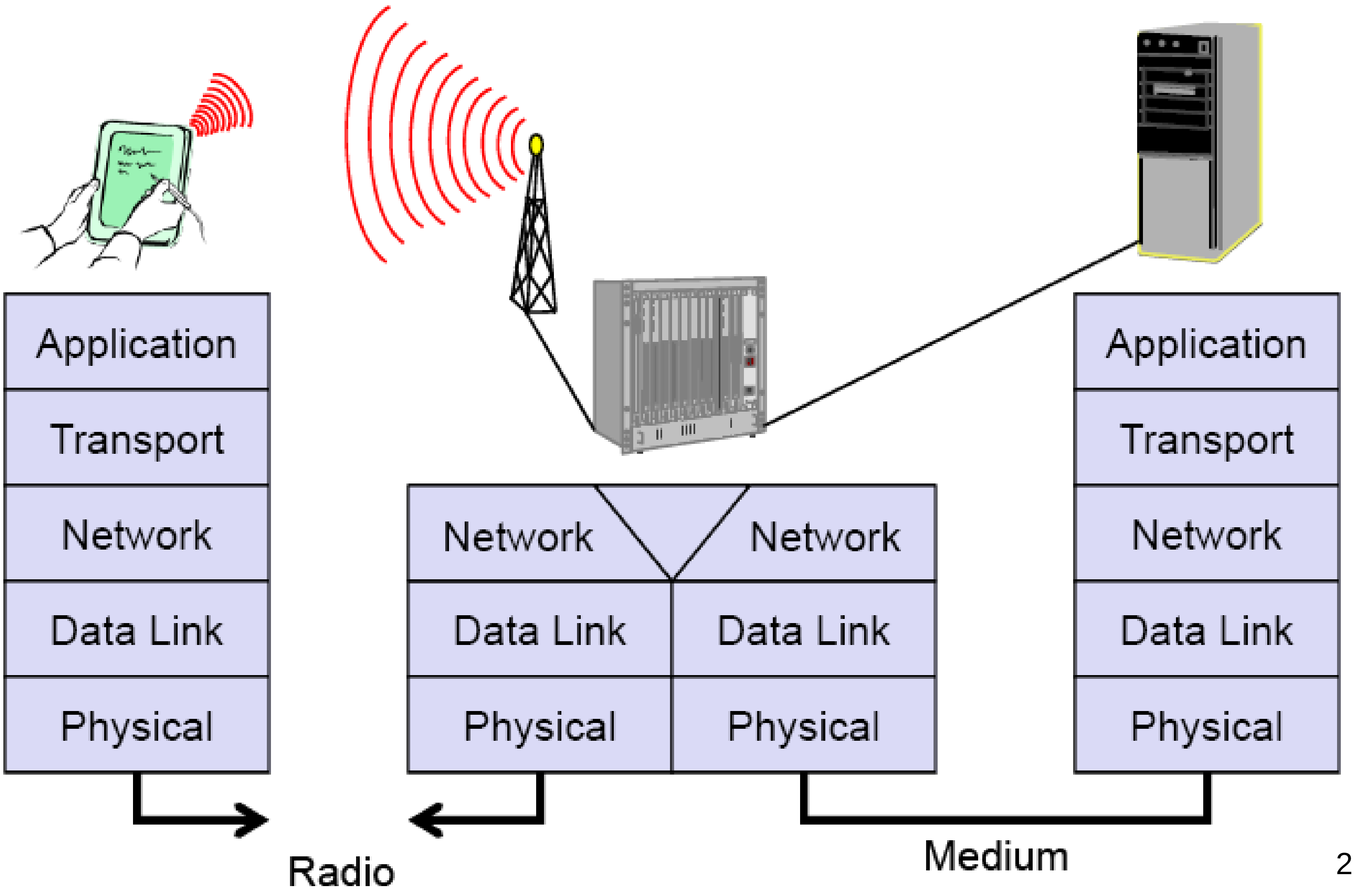
Cellular Subscriber (Sept-2008)



Limitations of the Mobile Environment

- Limitations of the Wireless Network
  - heterogeneity of fragmented networks
  - frequent disconnections
  - limited communication bandwidth
  - Interference: the quality of service (QoS)
  - Bandwidth: connection latency
- Limitations Imposed by Mobility
  - lack of mobility awareness by system/applications
  - route breakages
  - Dynamic changes in communication environment: variations in signal power within a region, thus link delays and connection losses
- Limitations of the Mobile Computer
  - short battery lifetime
  - limited capacities

Simple Reference Model



Application

Transport

Network

Data Link

Physical

Network | Network

Data Link | Data Link

Physical | Physical

Application

Transport

Network

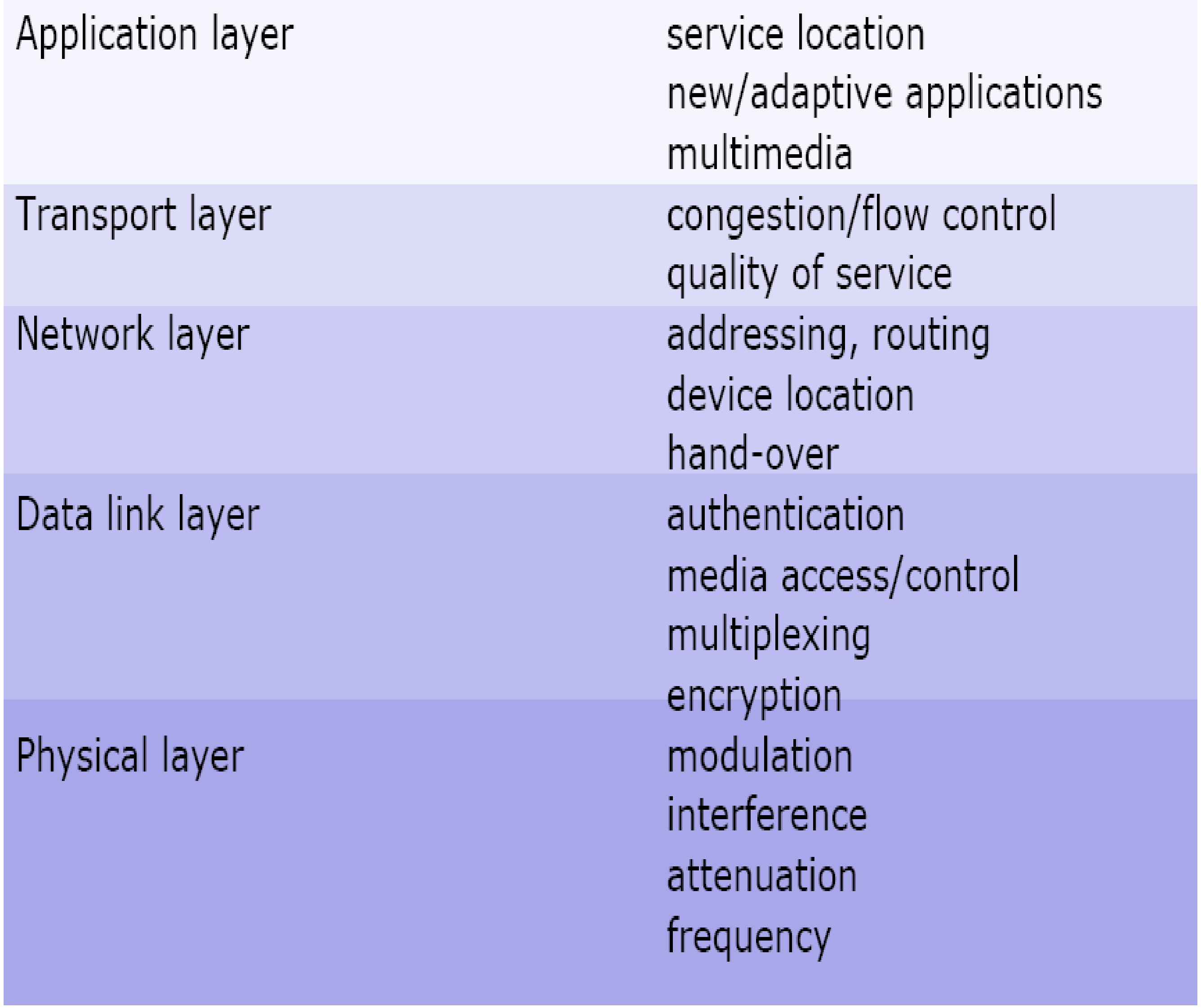
Data Link

Physical

Radio

Medium

Layer functionality



Application layer	service location new/adaptive applications multimedia
Transport layer	congestion/flow control quality of service
Network layer	addressing, routing device location hand-over
Data link layer	authentication media access/control multiplexing encryption
Physical layer	modulation interference attenuation frequency

PART-II

Global System for Mobile Communication [GSM]

CONTENTS

- Introduction
- Mobile Services
- GSM Architecture
- Radio Interface
- Protocols
- Localization and Calling
- Handover
- Security
- New Data Services

Introduction to GSM

Global System for Mobile (GSM) is a second generation cellular standard developed to cater voice services and data delivery using digital modulation

Introduction to GSM [2]

- Developed by Group Special Mobile (founded 1982) which was an initiative of CEPT (Conference of European Post and Telecom)
- **Aim** : to replace the incompatible analog system
- Presently the responsibility of GSM standardization resides with special mobile group under ETSI (European telecom Standards Institute)
- **GSM have 124 duplex channels, each 200Khz wide, are used for FDMA.**
- GSM offer data rates of up to 9.6Kbps and up to a speed of 250Km/hr
- Under ETSI, GSM is named as " **G**lobal **S**ystem for **M**obile communication "
- Today many providers all over the world use GSM (more than 190

  countries in Asia, Africa, Europe, Australia, America)
- More than 1300 million subscribers in world & 45 million subscriber in India.

GSM Developments

GSM 900

Mobile to BTS (uplink): 890-915 Mhz

BTS to Mobile(downlink):935-960 Mhz

GSM 1800 (DCS – Digital Cellular System)

Mobile to BTS (uplink): 1710-1785 Mhz

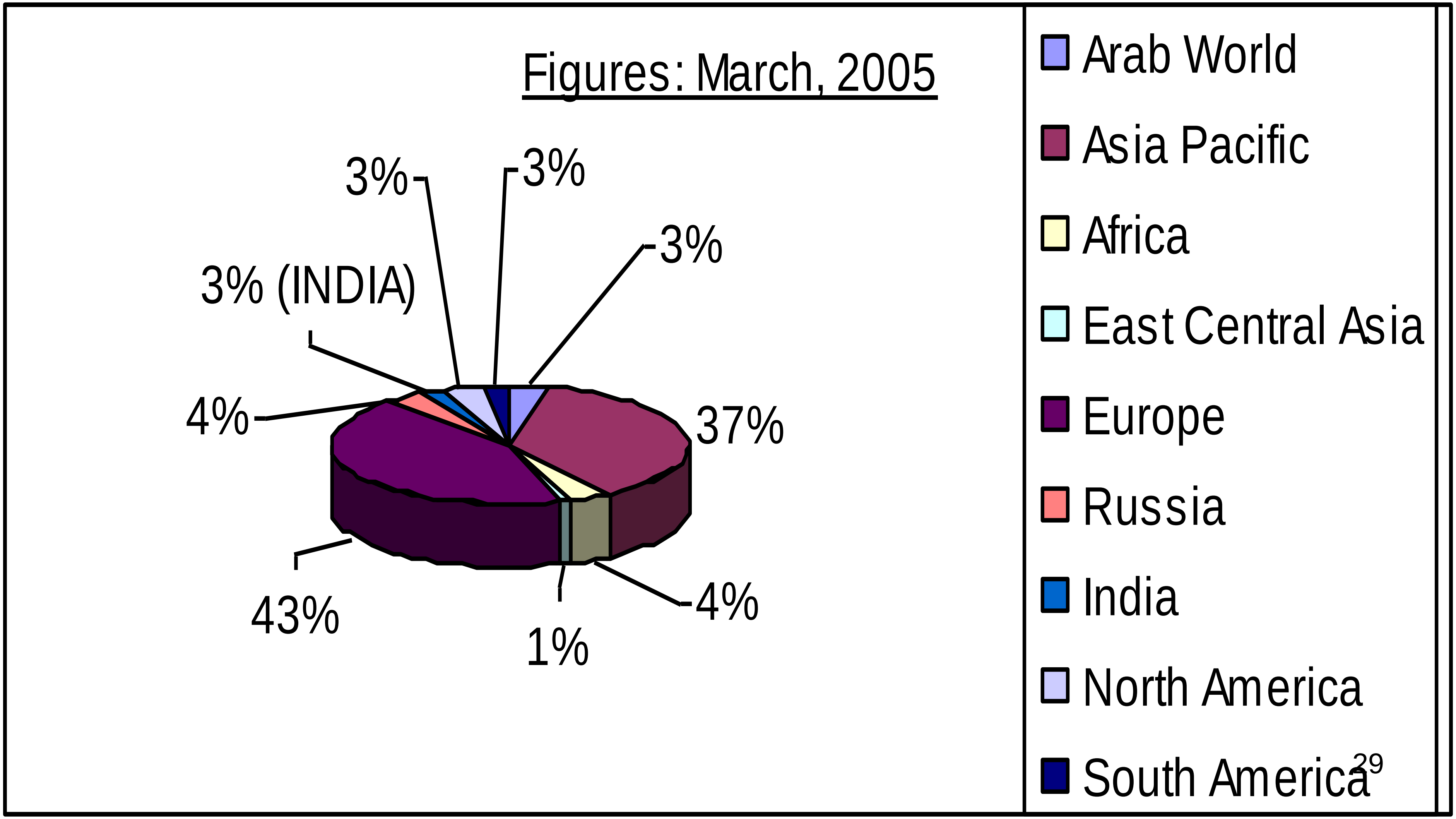
BTS to Mobile(downlink) 1805-1880 Mhz

GSM 1900 (PCS – Personal Communication Service)

Mobile to BTS (uplink): 1850-1910 Mhz

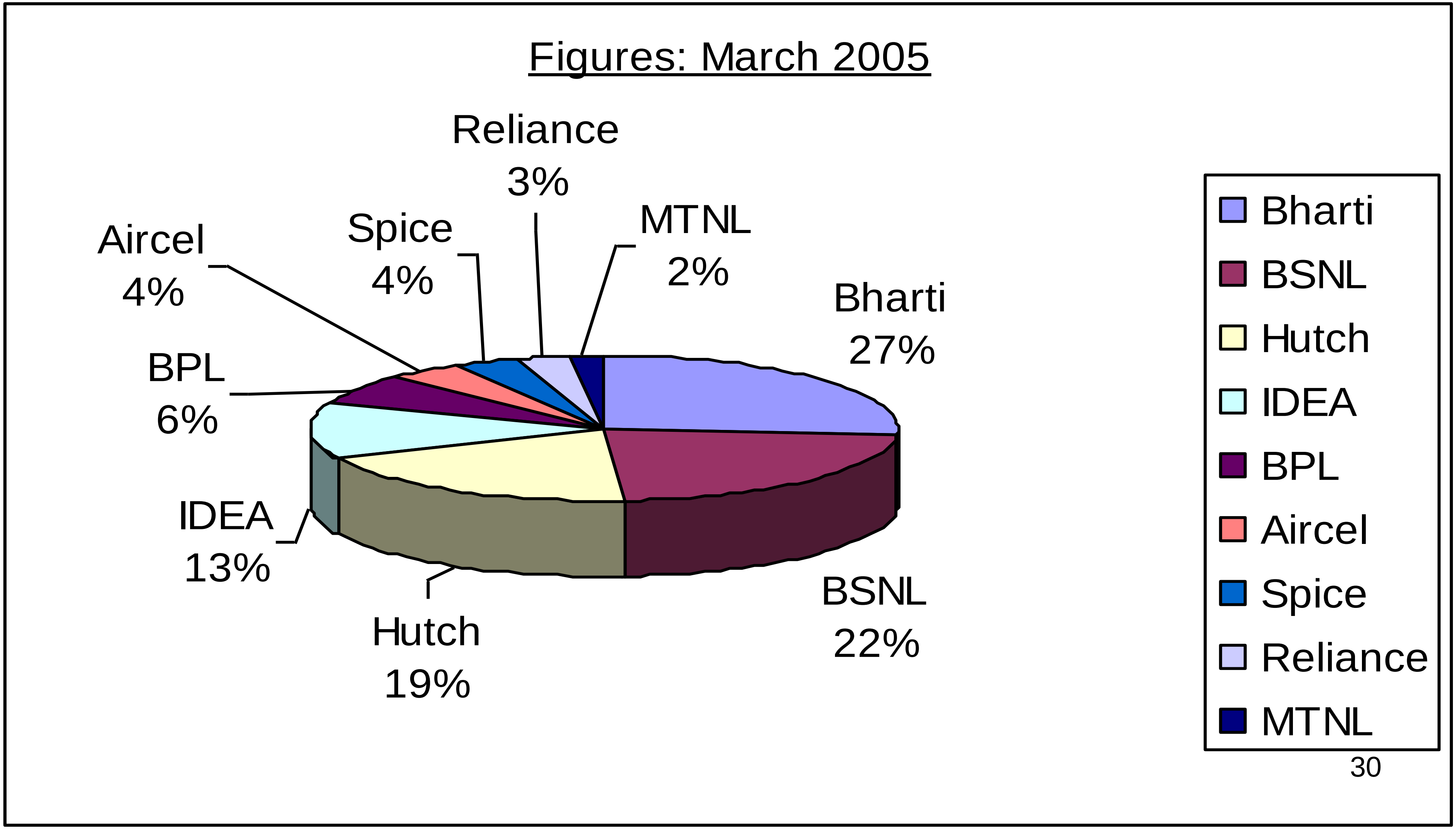
BTS to Mobile(downlink) 1930-1990 Mhz

GSM in World



Figures: March, 2005
3%
3%
3%
3% (INDIA)
4%
37%
43%
1%
4%
Arab World
Asia Pacific
Africa
East Central Asia
Europe
Russia
India
North America
South America

GSM in India


Figures: March 2005
Reliance 3%
MTNL 2%
Spice 4%
Aircel 4%
Bharti 27%
BPL 6%
IDEA 13%
Hutch 19%
BSNL 22%
Bharti
BSNL
Hutch
IDEA
BPL
Aircel
Spice
Reliance
MTNL

Mobile Services

- Tele-services
- Bearer or Data Services
- Supplementary services

Tele Services

• Telecommunication services that enable voice communication

via mobile phones

• Offered services

- Mobile telephony

- Emergency calling

Bearer Services

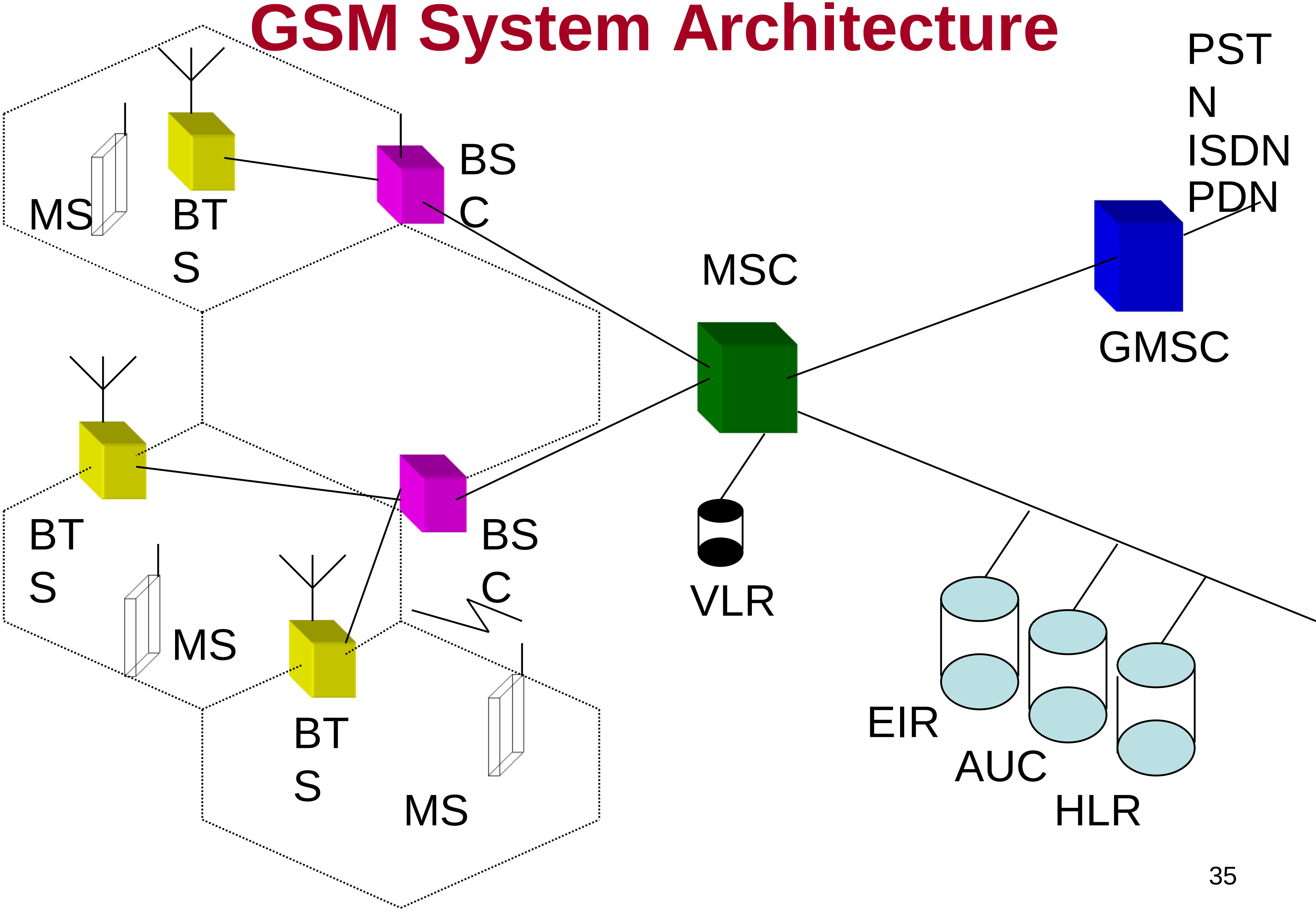
- Include various data services for information transfer between GSM and other networks like PSTN, ISDN etc at rates from 300 to 9600 bps
- Short Message Service (SMS)
  - –up to 160 character alphanumeric data transmission to/from the mobile terminal
- Unified Messaging Services(UMS)
- Group 3 fax
- Voice mailbox
- Electronic mail

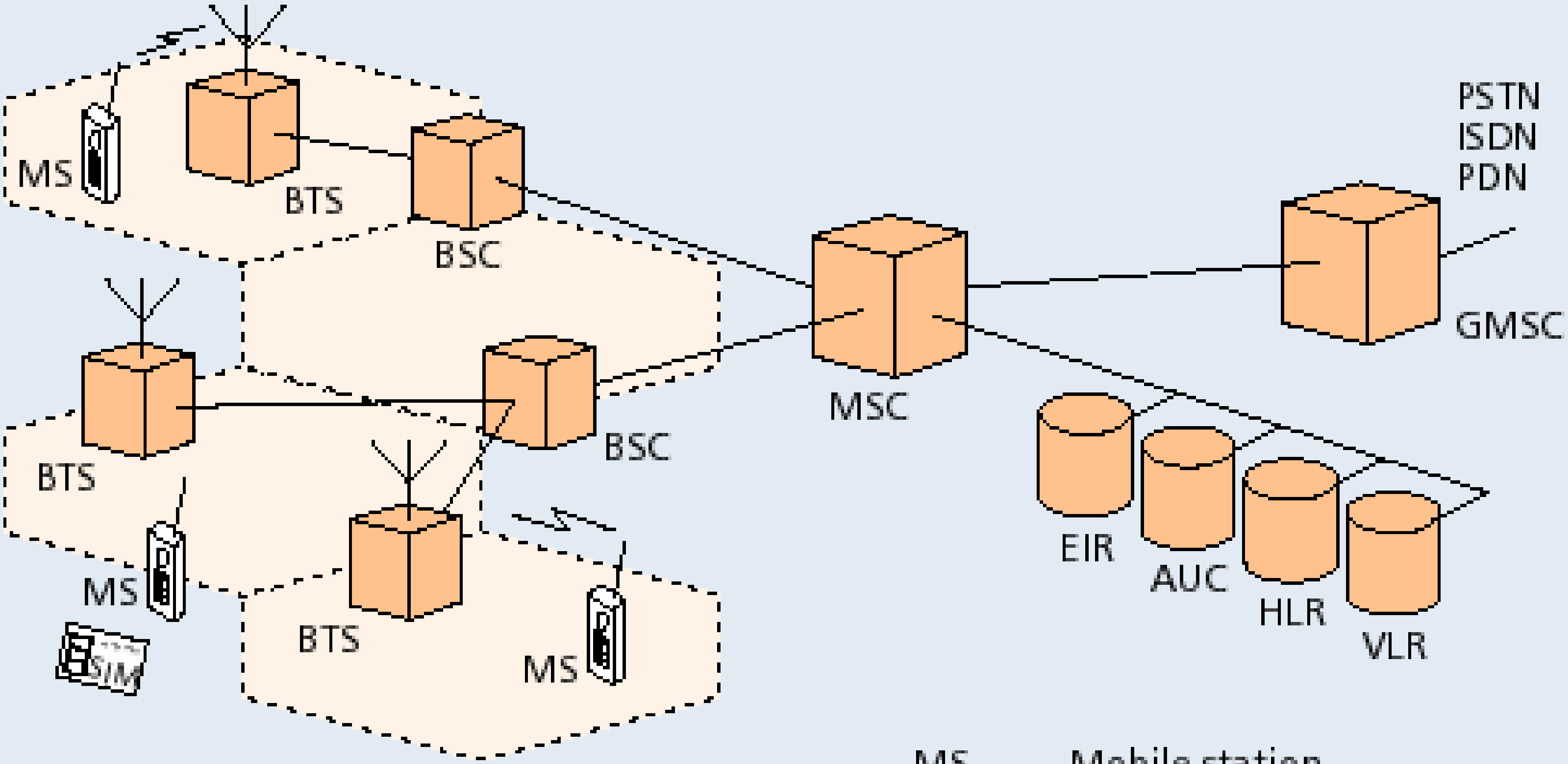
Supplementary Services

Call related services :

- Call Waiting- Notification of an incoming call while on the handset
- Call Hold- Put a caller on hold to take another call
- Call Barring- All calls, outgoing calls, or incoming calls
- Call Forwarding- Calls can be sent to various numbers defined by the user
- Multi Party Call Conferencing - Link multiple calls together
- CLIP – Caller line identification presentation
- CLIR – Caller line identification restriction
- CUG – Closed user group

GSM System Architecture





BTS	Base transceiver station
BSC	Base station controller
BSS	Base station subsystem (BTS+BSC)
MSC	Mobile switching center
GMSC	Gateway MSC
MS	Mobile station
HLR	Home location register
VLR	Visited location register
EIR	Equiment identity register
AUC	Authentication center

Components

- **Mobile Station (MS)**
  - Mobile Equipment (ME)
  - Subscriber Identity Module (SIM)
- **Base Station Subsystem (BSS)**
  - Base Transceiver Station (BTS)
  - Base Station Controller (BSC)
- **Network Switching Subsystem(NSS)**
  - Mobile Switching Center (MSC)
  - Home Location Register (HLR)
  - Visitor Location Register (VLR)
  - Authentication Center (AUC)
  - Equipment Identity Register (EIR)

Mobile Station (MS)

The Mobile Station is made up of two entities:

1. Mobile Equipment (ME)
2. Subscriber Identity Module (SIM)

Mobile Equipment

- Portable, vehicle mounted, hand held device
- Uniquely identified by an **IMEI** (International Mobile Equipment Identity)
- Voice and data transmission
- Monitoring power and signal quality of surrounding cells for optimum handover
- Power level : 0.8W – 20 W
- 160 character long SMS.

Subscriber Identity Module (SIM)

- Smart card contains the International Mobile Subscriber Identity (**IMSI**)
- Allows user to send and receive calls and receive other subscribed services
- Encoded network identification details

  - Key Ki, Kc and A3,A5 and A8 algorithms
- Protected by a password or PIN
- Can be moved from phone to phone – contains key information to activate the phone

Base Station Subsystem (BSS)

Base Station Subsystem is composed of two parts that communicate across the standardized **Abis** interface allowing operation between components made by different suppliers

1. Base Transceiver Station (**BTS**)
2. Base Station Controller (**BSC**)

1. Base Transceiver Station (BTS):

- Encodes, encrypts,multiplexes,modulates and feeds the RF signals to the antenna.
- Frequency hopping
- Communicates with Mobile station and BSC
- Consists of Transceivers (TRX) units

2. Base Station Controller (BSC)

- Manages Radio resources for BTS
- Assigns Frequency and time slots for all MS’s in its area
- Handles call set up
- Transcoding and rate adaptation functionality
- Handover for each MS
- Radio Power control
- It communicates with MSC and BTS

Network Switching Subsystem(NSS)

Mobile Switching Center (MSC)

- Heart of the network
- Manages communication between GSM and other networks
- Call setup function and basic switching
- Call routing
- Billing information and collection
- Mobility management

  - Registration

  - Location Updating

  - Inter BSS and inter MSC call handoff
- MSC does gateway function while its customer roams to other network by using HLR/VLR.

Home Location Registers (HLR)

- permanent database about mobile subscribers in a large service area(generally one per GSM network operator)
- database contains IMSI, MSISDN, prepaid/postpaid, roaming restrictions, supplementary services.

Visitor Location Registers (VLR)

- Temporary database which updates whenever new MS enters its area, by HLR database
- Controls those mobiles roaming in its area
- Reduces number of queries to HLR
- Database contains IMSI, TMSI, MSISDN, MSRN, Location area, authentication key

Authentication Center (AUC)

- Protects against intruders in air interface
- Maintains authentication keys and algorithms and provides security triplets (RAND,SRES,Kc)
- Generally associated with HLR

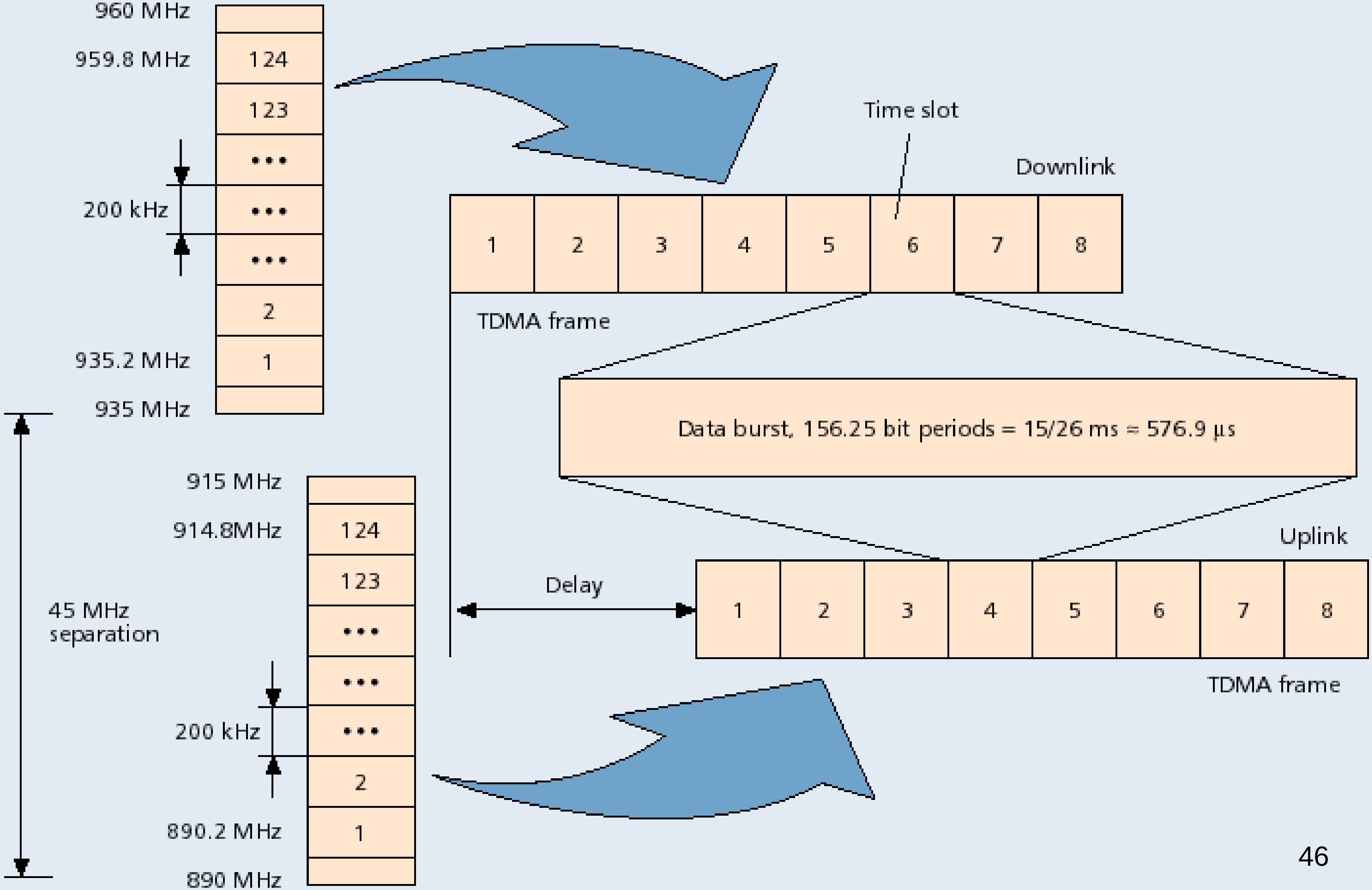
Equipment Identity Register (EIR)

- Database that is used to track handsets using the IMEI (International Mobile Equipment Identity)
- Made up of three sub-classes: The White List, The Black List and the Gray List
- Only one EIR per PLMN

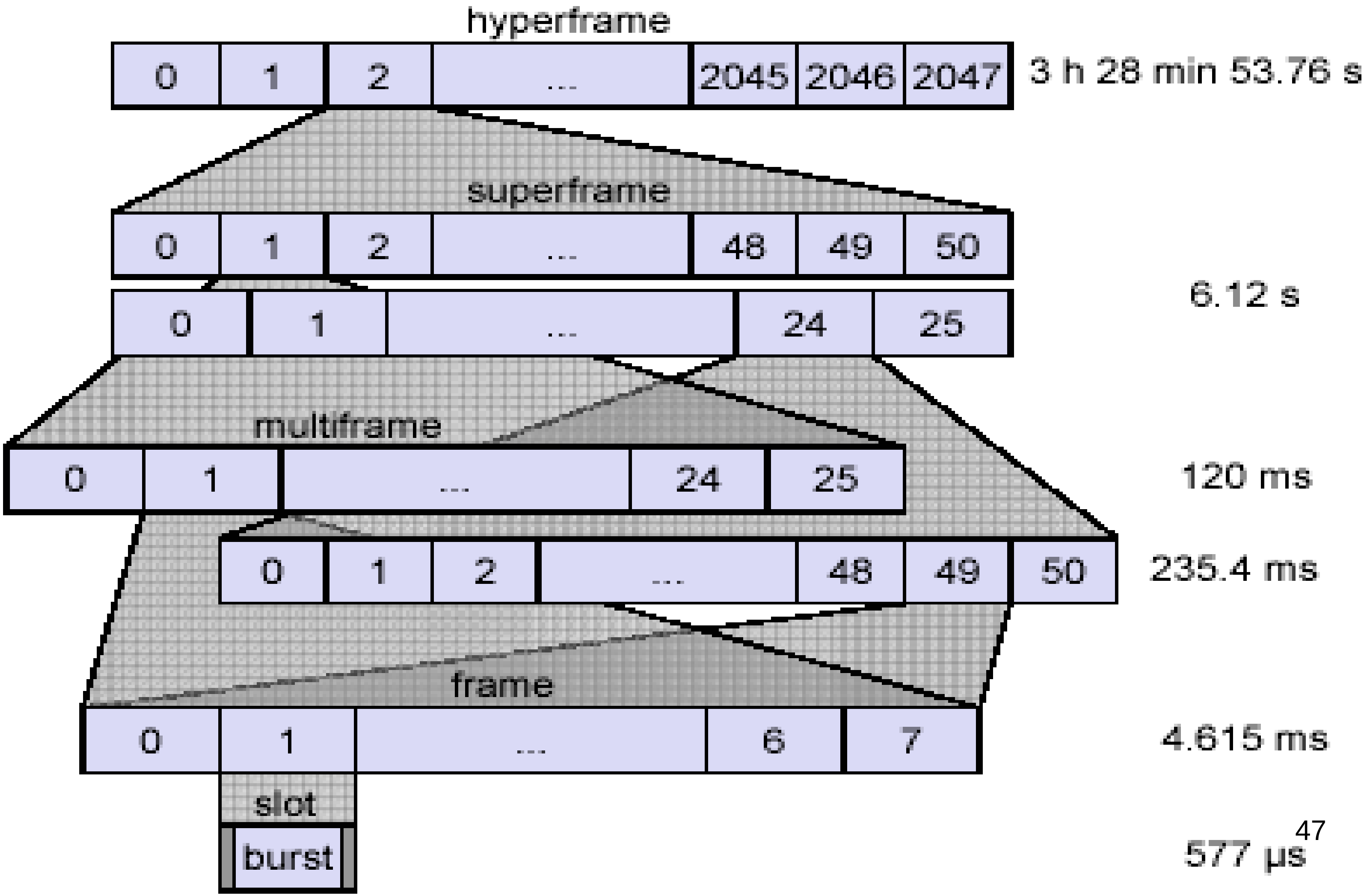
Radio Interface (Um)

- Air Interface: MS to BTS
- Uplink/Downlink of 25MHz
  - 890 - 915 MHz for Up link
  - 935 - 960 MHz for Down link
- Combination of frequency division and time division multiplexing
  - FDMA - 124 channels of 200 kHz
  - TDMA - Burst
- Modulation Method : GMSK (Gaussian Minimum Shift Keying) @ 270.833 Kbps

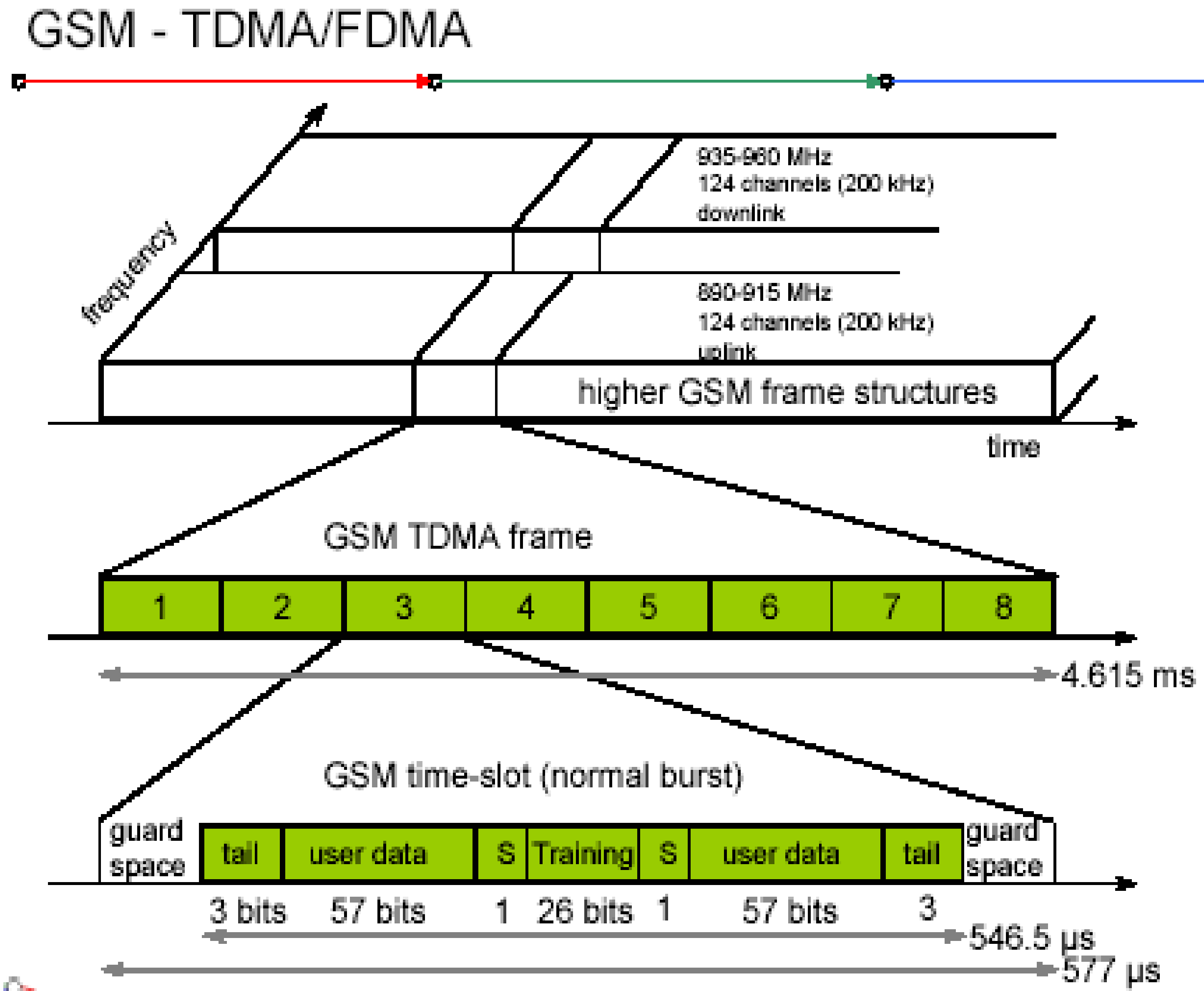
FDMA/TDMA


960 MHz
959.8 MHz
124
123
•••
200 kHz
•••
•••
2
935.2 MHz
1
935 MHz
Time slot
Downlink
1
2
3
4
5
6
7
8
TDMA frame
Data burst, 156.25 bit periods = 15/26 ms ≈ 576.9 µs
45 MHz separation
915 MHz
914.8MHz
124
123
•••
•••
200 kHz
•••
2
890.2 MHz
1
890 MHz
Uplink
Delay
1
2
3
4
5
6
7
8
TDMA frame

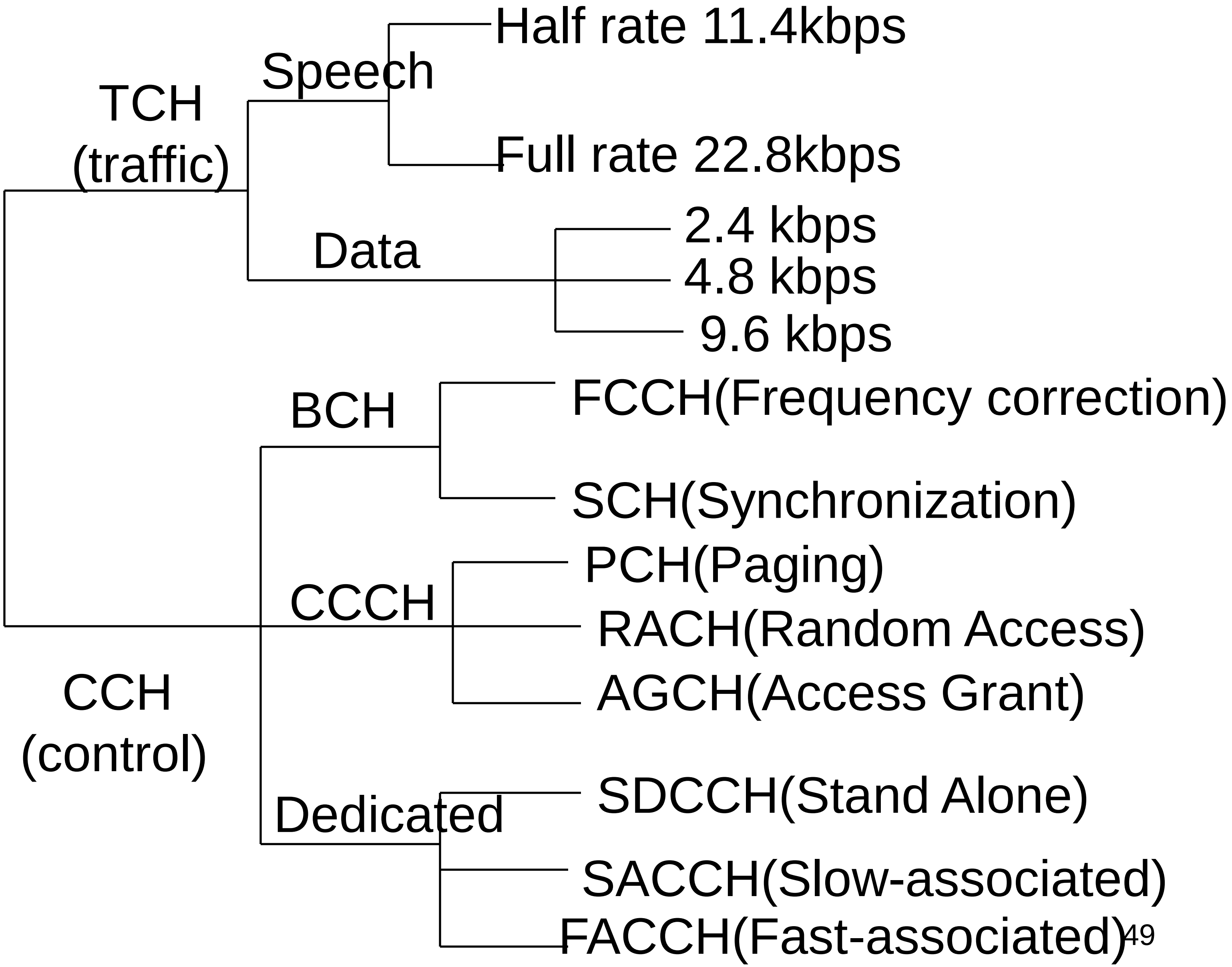
Physical Channel



GSM-Frame Structure



Logical Channels



TCH (traffic)

Speech

Half rate 11.4kbps

Full rate 22.8kbps

Data

2.4 kbps

4.8 kbps

9.6 kbps

CCH (control)

BCH

FCCH(Frequency correction)

SCH(Synchronization)

CCCH

PCH(Paging)

RACH(Random Access)

AGCH(Access Grant)

Dedicated

SDCCH(Stand Alone)

SACCH(Slow-associated)

FACCH(Fast-associated)

BCCH

- BTS to MS
  - send cell identities, organization info about common control channels, cell service available, etc
- Radio channel configuration
    - Current cell + Neighbouring cells
- Synchronizing information
    - Frequencies + frame numbering
- Registration Identifiers
    - LA + Cell Identification (CI) + Base Station Identity Code

BCCH Sub-Channels

- Frequency Correction Channel
    - send a frequency correction data burst containing all zeros to effect a constant frequency shift of RF carrier
        - Mobile station knows which frequency to use
  - Repeated broadcast of Frequency Bursts

- Synchronization Channel
    - send TDMA frame number and base station identity code to synchronize MSs
        - MS knows which timeslot to use
  - Repeated broadcast of Synchronization Bursts

CCC

- Access Grant Channel (AGCH)
  - BTS to MS
  - Used to assign an SDCCH/TCH to MS
- Paging Channel (PCH)
  - BTS to MS
  - Page MS
- Random Access Channel (RACH)
  - MS => BTS
  - Slotted Aloha
  - Request for dedicated SDCCH

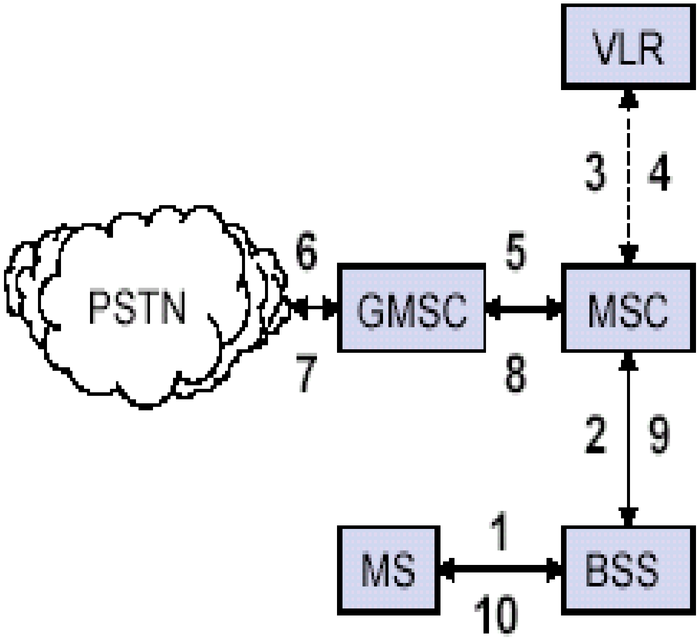
DCCH

- bidirectional point-to-point -- main signaling channels
- SDCCH (stand-alone dedicated control channel): for service request, subscriber authentication, equipment validation, assignment to a traffic channel
- SACCH (slow associated control channel): for out-of-band signaling associated with a traffic channel, eg, signal strength measurements
- FACCH (fast associated control channel): for preemptive signaling on a traffic channel, eg, for handoff messages
  - Uses timeslots which are otherwise used by the TCH

Localization and Calling

- Localization means same phone number is valid worldwide
- Periodic location updates
- VLR informs the HLR about MS location changes
- Changing VLRs with uninterrupted availability of all services called Roaming
- To locate and address to the MS, GSM needs
  - MSISDN, MSRN
  - IMSI, TMSI
- Two ways of calling:
  - Call Originating from MS (MOC)
  - Call termination to MS (MTC)

Outgoing Call (MOC)



1. MS sends dialled number to BSS
2. BSS sends dialled number to MSC

3,4 MSC checks VLR if MS is allowed the requested service.If so,MSC asks BSS to allocate resources for call.

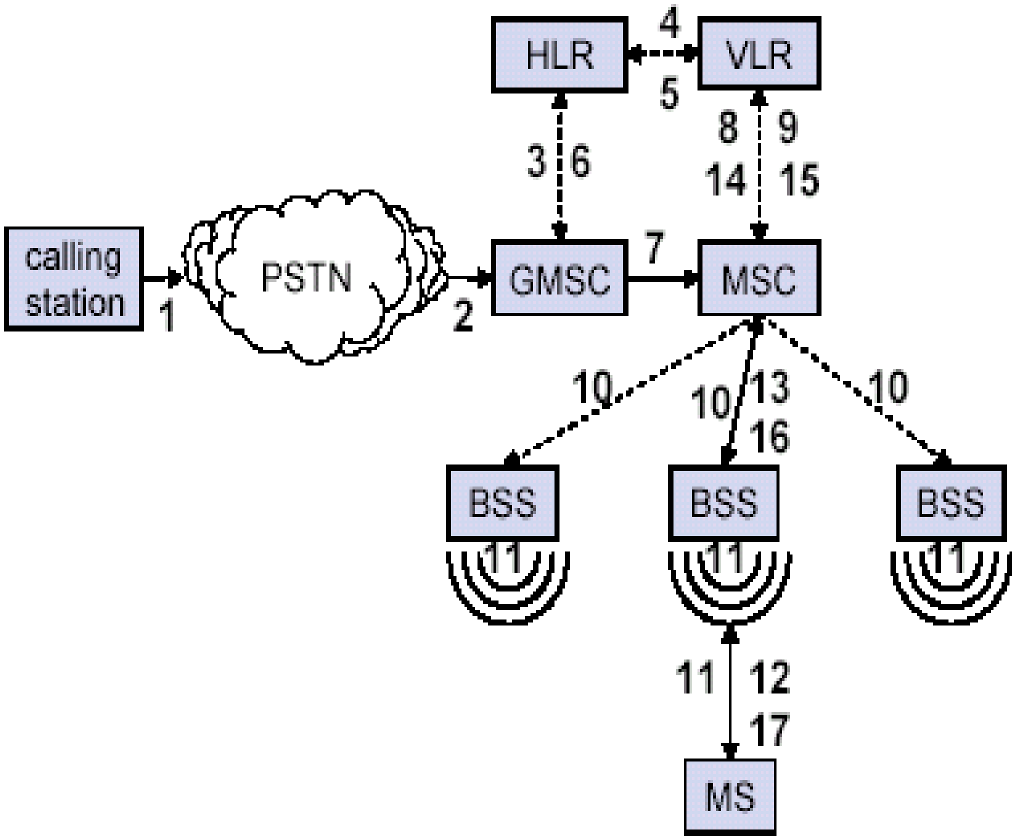
5 MSC routes the call to GMSC

6 GMSC routes the call to local exchange of called user

7, 8,

9,10 Answer back(ring back) tone is routed from called user to MS via GMSC,MSC,BSS

Incoming Call (MTC)



1. Calling a GSM subscribers
2. Forwarding call to GSMC
3. Signal Setup to HLR
4. 5. Request MSRN from VLR
6. Forward responsible MSC to GMSC
7. Forward Call to current MSC
8. 9. Get current status of MS

10.11. Paging of MS

12.13. MS answers

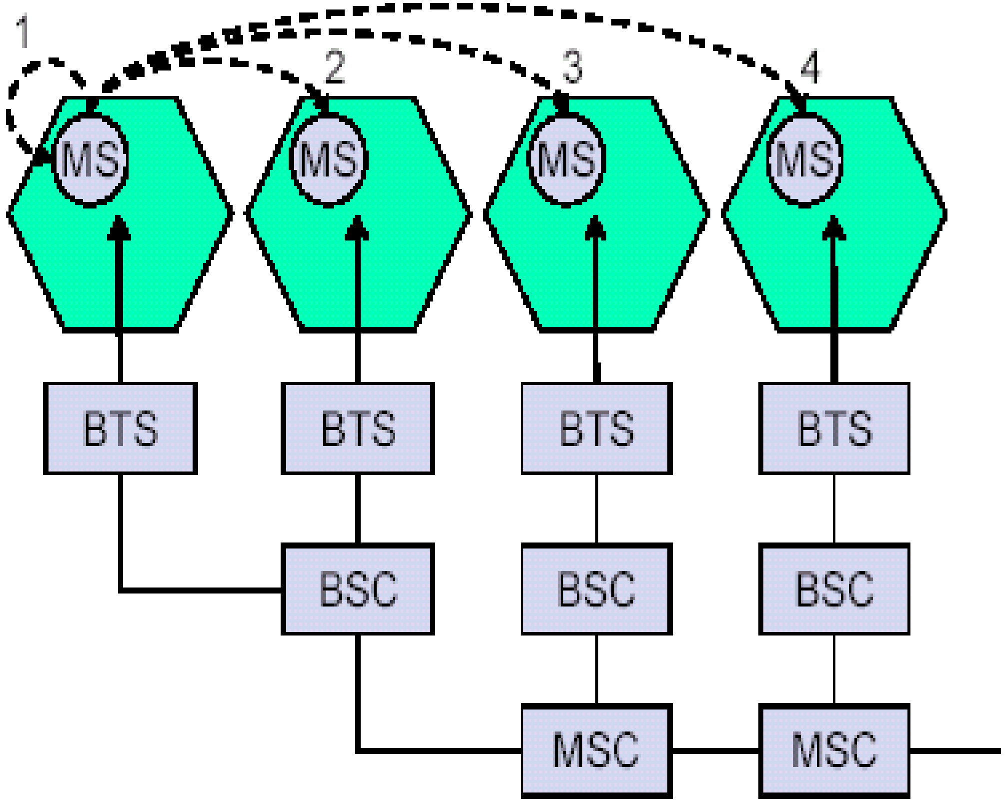
14.15. Security checks

16.17. Set up connection

HANDOVER

- Single cell do not cover the whole service area
- Smaller the cell size, faster the movement of MS, but more handovers
- However, a handover should not cause a cut-off
- Two reasons to use handover
  - Maintain same quality of radio link at reciever
  - Load balancing

Handover Scenario



- Within 1 – Intra Cell
- Between 1 and 2 – Inter BTS / Intra BSC
- Between 1 and 3 – Inter BSC/ Intra MSC
- Between 1 and 4 – Inter MSC

GSM networks various security features

- A wireless radio based network system quite sensitive to the unauthorized use of resources
- GSM employ various security features designed to :

  →Designed to protect subscriber privacy

  →Secured network against misuse of resources by unregistered users

GSM networks various security features

- Controlled access to the network by Mobile station
- Required to use a PIN before it can access the network through Um interface

Security in GSM

- Security Services:
  - Access Control and Authentication
  - Confidentiality : all user data encrypted
  - Anonymity: not disclosed user identity
- GSM Uses the information stored in AuC and SIM
- SIM protected data with PIN against unauthorized use.
- 3 algorithms are specified :

  - A3 algorithm for authentication

  - A5 algorithm for encryption

  - A8 algorithm for key generation

Authentication

- An AuC (authentication centre) for the operation and maintenance subsystem of the GSM network
- Authentication of the Mobile station
- The AuC first authenticates the subscriber Mobile station and only then does the MSC provide the switching service to another terminal TE

Authentication algorithm

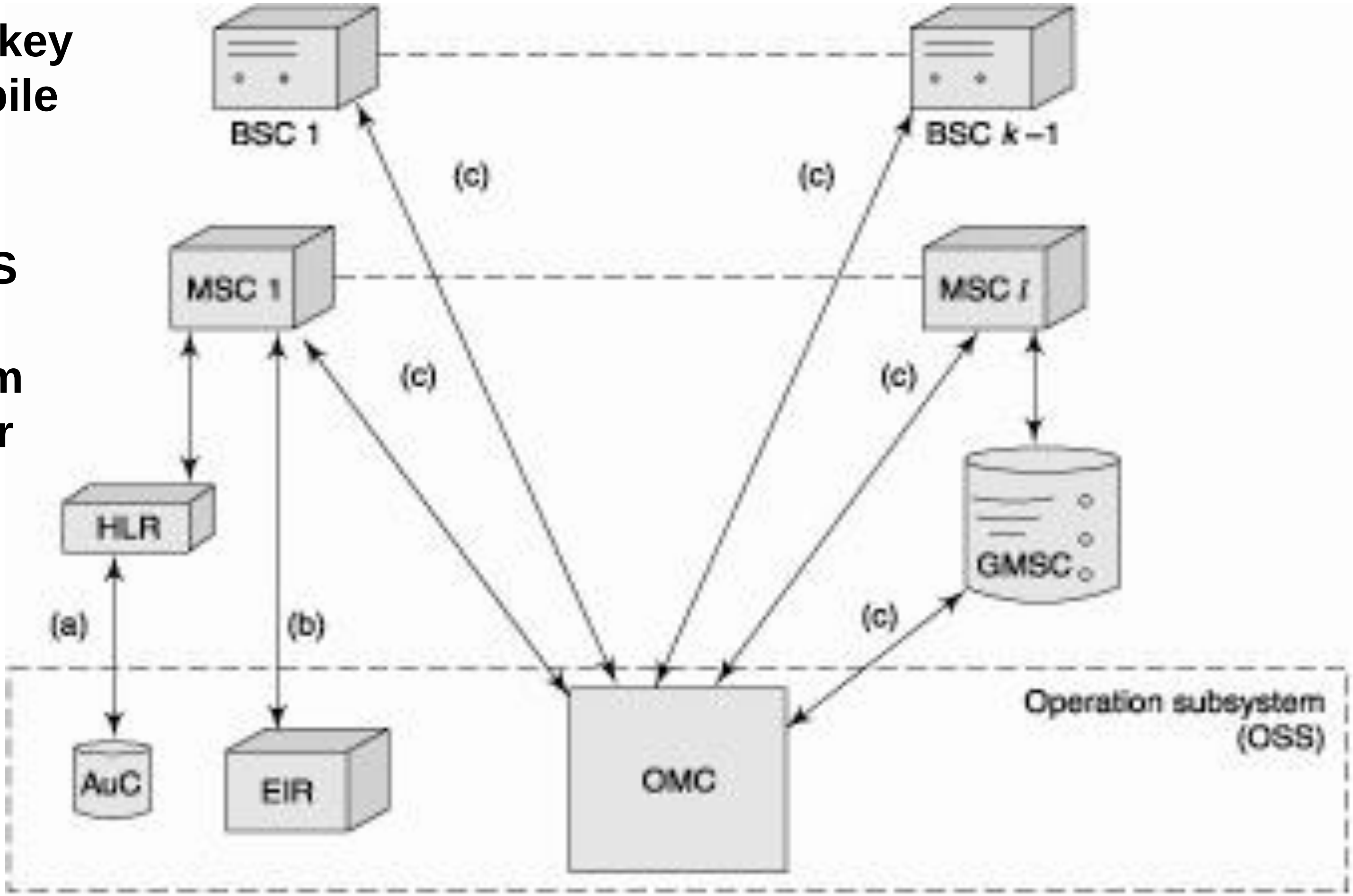
- Use a random number sent by the AuC during the connection set up
- An authentication key which is already saved in the SIM
- Authentication algorithm used differs for different mobile service provider

AuC sending random number for BTS and BTS sending cipher key for encryption

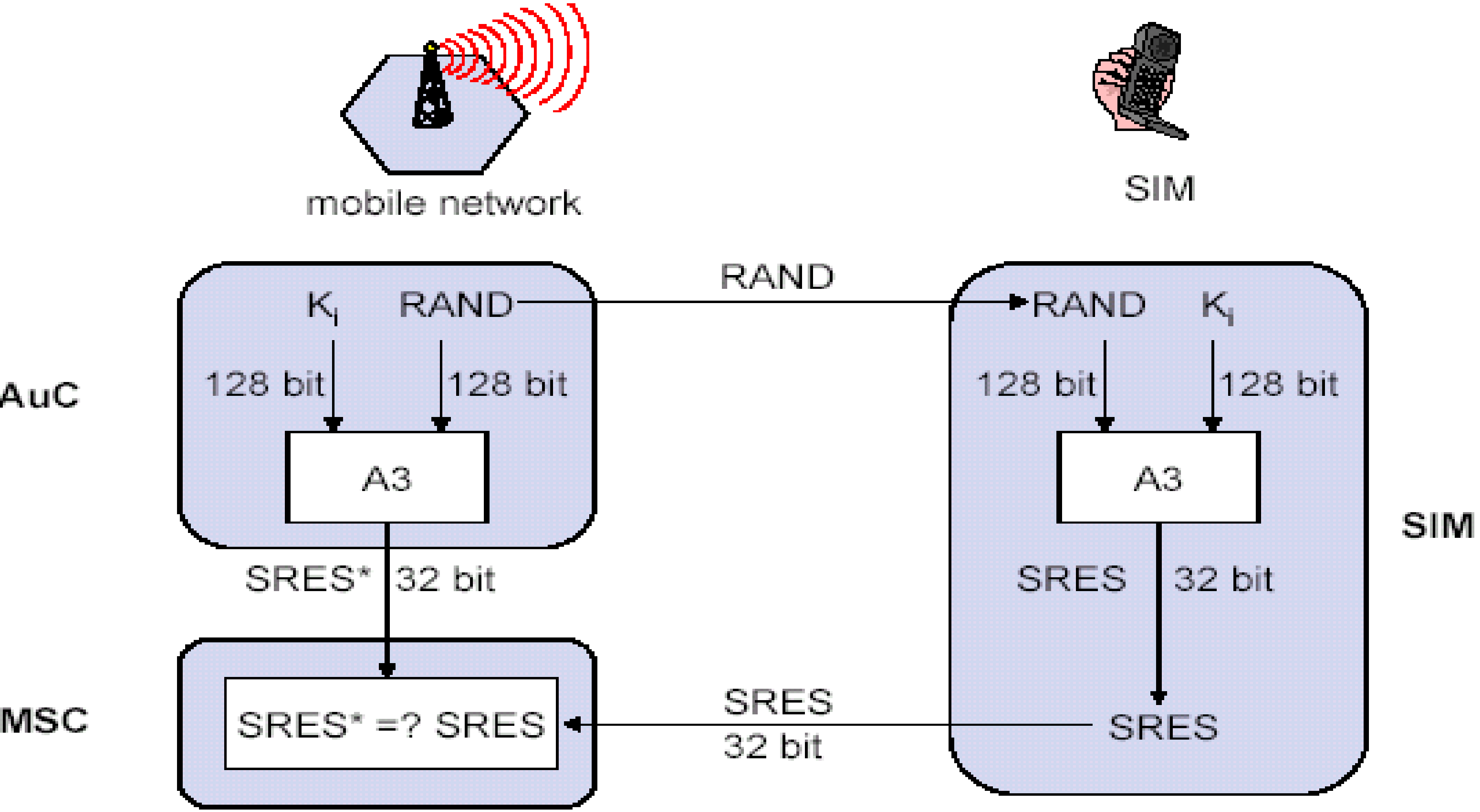
pher key
r Mobile
ation

r BTS

ndom
mber



Authentication in GSM



K_i: individual subscriber authentication key SRES: signed response

IMSI and TMSI of the Mobile station

- Its public identity
- TMSI is the identity granted on moving to a particular location
- When a Mobile station moves to a new location area, the VLR (visitor location register) assigns a TMSI which is stored in the SIM of the Mobile station

TMSI

- The identification of the subscriber during communication done not using the IMSI but the TMSI
- The VLR assigned TMSI generates that ID
- This protects the Mobile station against eavesdropping from external sources

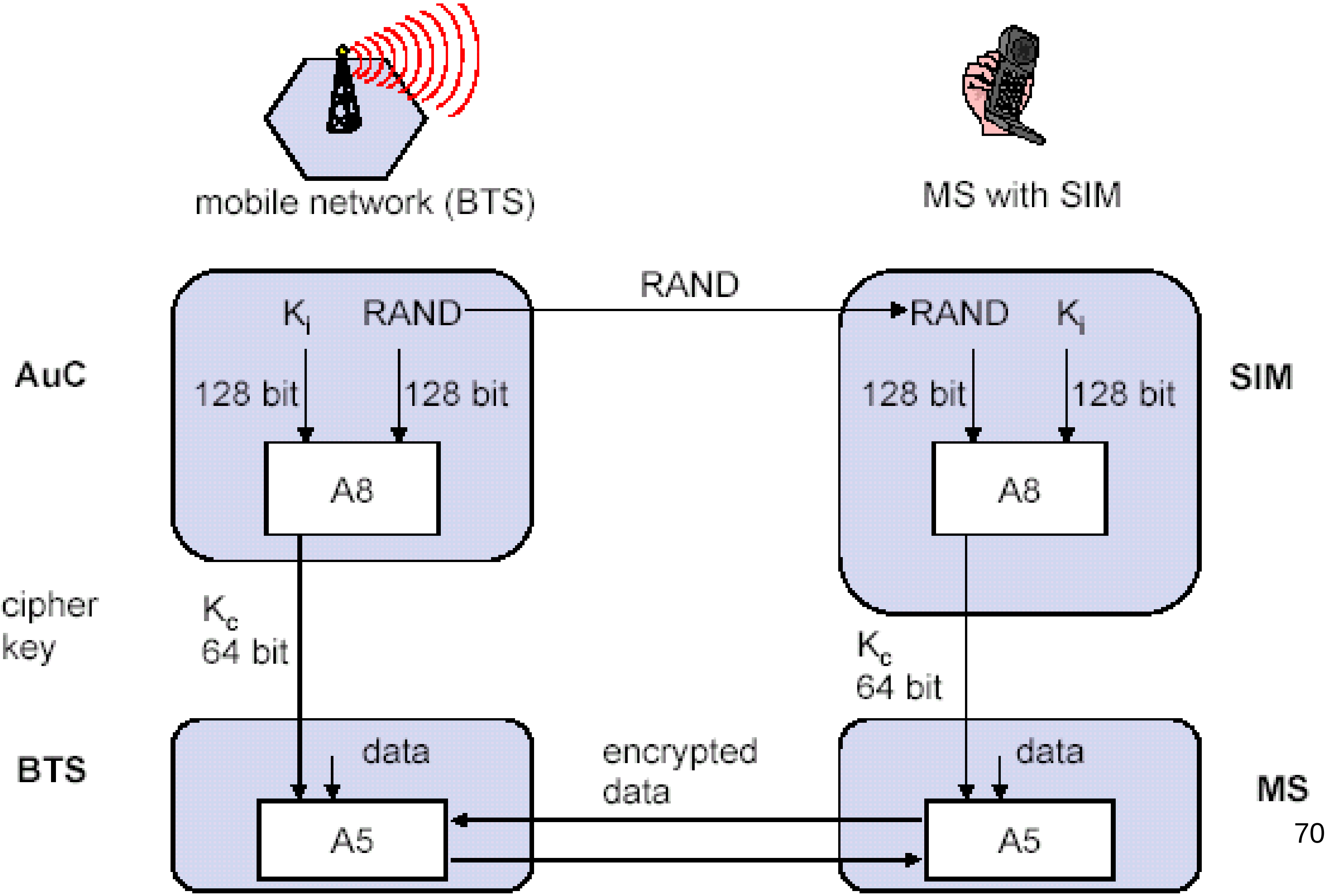
Encryption

- The BTS and the Mobile station perform ciphering before call initiation or before connecting for receiving a call
- The Mobile station uses a cipher (encryption key) for encryption

The cipher

- A result of performing mathematical operations on (a) the cipher key saved in the SIM and (b) the cipher number received from the BTS when the call setup is initiated
- The BTS transmits the cipher number before a call is set up or transmitted

Encryption


mobile network (BTS)
MS with SIM
AuC
K_i
RAND
RAND
RAND
K_i
SIM
128 bit
128 bit
128 bit
128 bit
A8
A8
cipher key
K_c 64 bit
K_c 64 bit
BTS
data
encrypted data
data
MS
A5
A5

Advantages of GSM over Analog system

- Capacity increases
- Reduced RF transmission power and longer battery life.
- International roaming capability.
- Better security against fraud (through terminal validation and user authentication).
- Encryption capability for information security and privacy.
- Compatibility with ISDN,leading to wider range of services

New Data Services: HSCSD

- High Speed Circuit Switched Data (HSCSD)
  - Combined several GSM 9.6 Kbps channels to increase bandwidth
  - It allocates several TDMA slots within a TDMA frame
  - In theory, an MS could use all 8 slots within a TDMA frame to achieve an Air Interface User Rate (AIUR).
  - Only requires software upgrades in an MS and MSC
- HSCSD exhibits some major limitations
  - Still uses connection-oriented mechanism
  - Not efficient for bursty and asymmetrical traffic
  - Charged based on channels allocated

New Data Services: GPRS

- General Packet Radio Service (GPRS) features:
  - It is a 2.5G system, is poised to take off this year in popularity.
  - Building on the GSM network, it will provide the much needed packet data services to most areas of the world.
  - Useful for frequent small volume or infrequent small/medium volume of data
  - Time slots are not allocated in a fixed or pre-determined manner, but on demand
  - Allow broadcast, multicast and unicast service
  - "Always On", no connection has to be setup prior to data transfer
  - Resources are reserved only when needed and charged accordingly

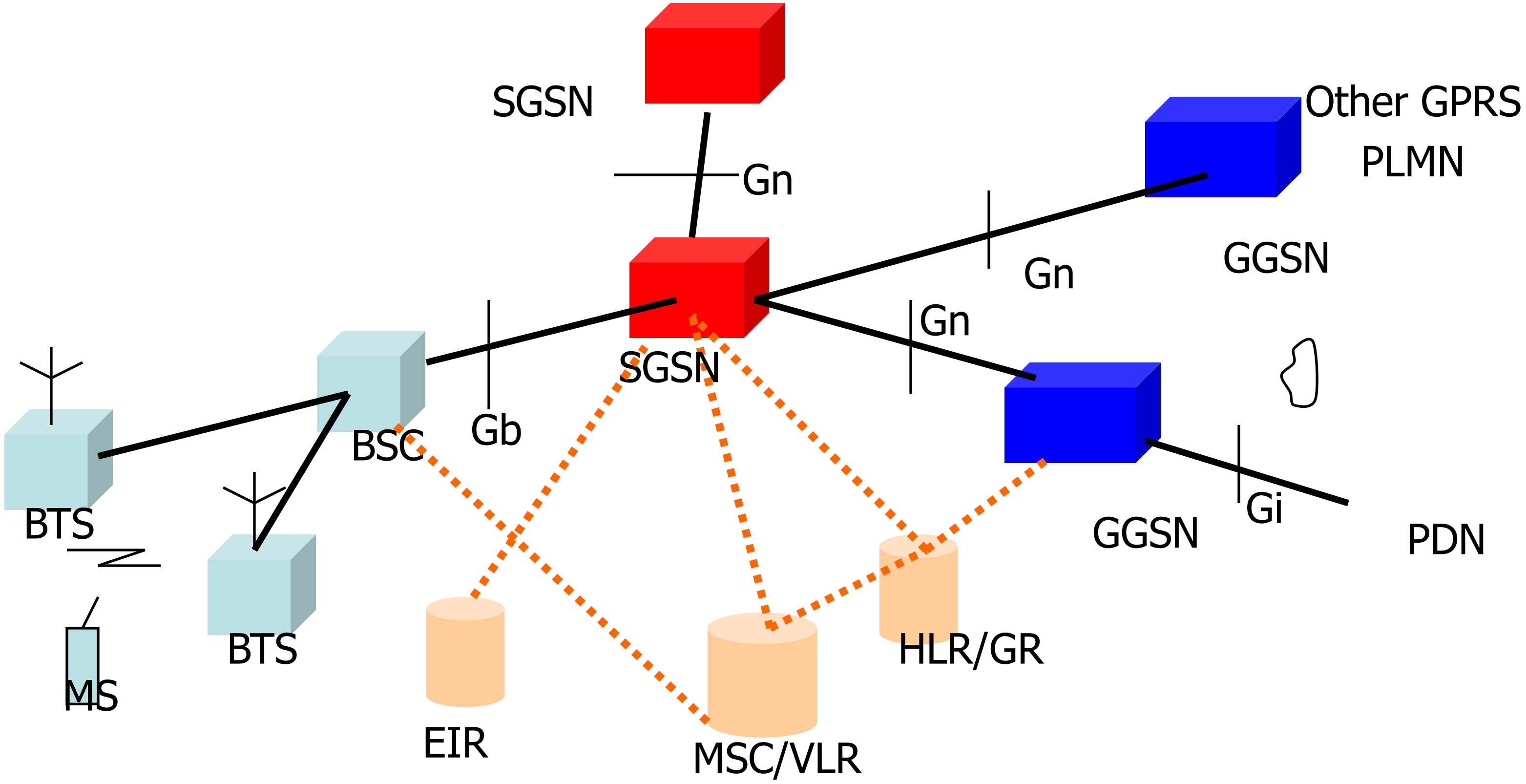
GPRS [2]

- It offers point-to-point packet transfer in 2 versions
  - PTP connection-oriented service (PTP-CONS)
  - PTP connectionless service (PTP-CNLS)

- It also offers Multicasting, called Point-to-Multipoint (PTM) service

- User specify QoS profile
  - Service precedence (high, normal, low)
  - Reliability class
  - Delay class
  - Peak throughput class for data
  - Mean throughput class for data

GPRS Architecture

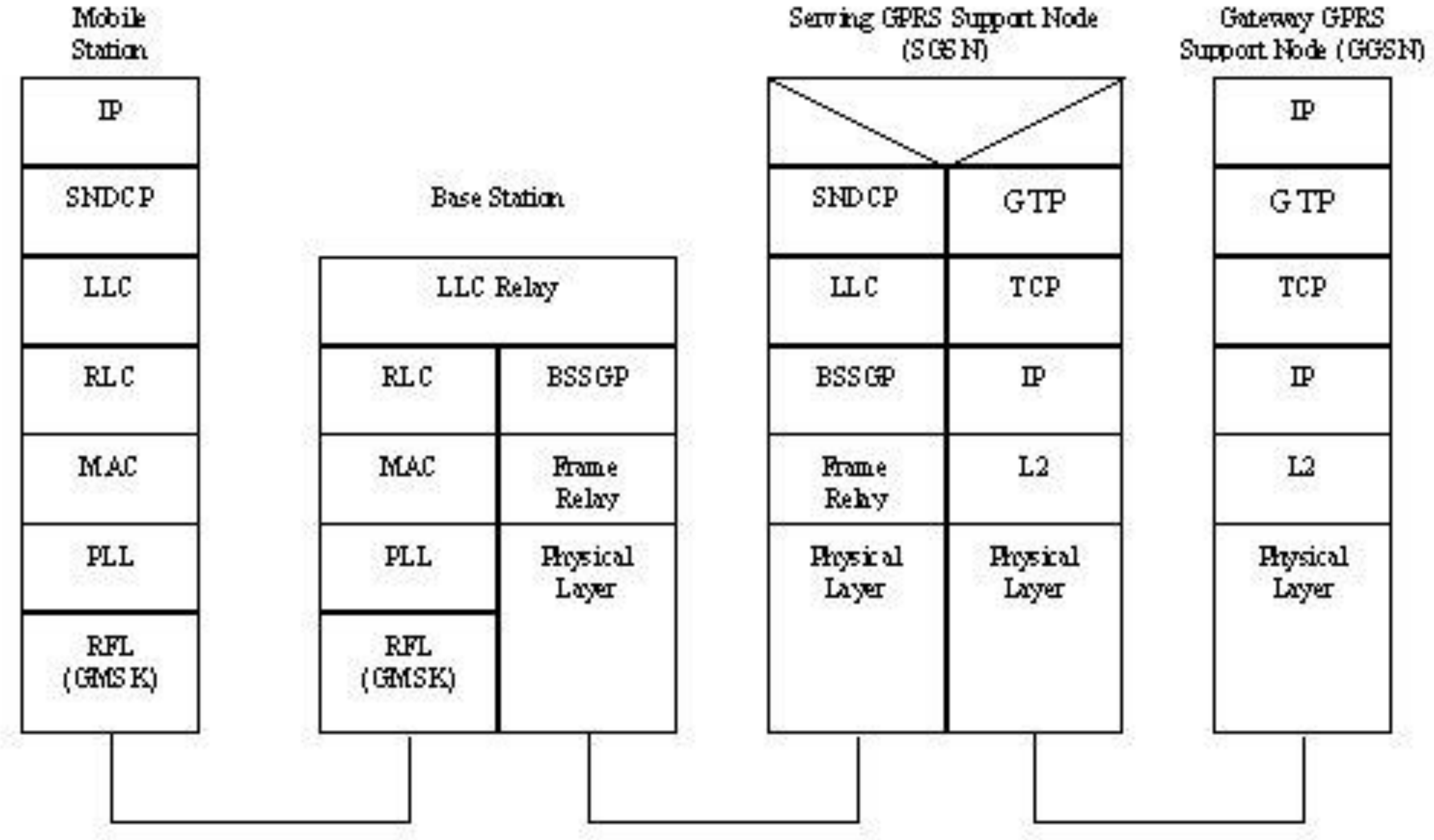


PLMN : GSM Public Land Mobile N/w
PDN: Public Data Network

Entities

- The Serving GPRS Support Node (SGSN)
  - Mobility Management
  - Authentication
  - Requests user information from the GPRS Register (GR)
  - Gathers Charging Information
- Gateway GPRS Support Node (GGSN)
  - Gateway between UMTS Core Network and external networks
  - Address allocation for MS
  - Gathers Charging Information
  - Filtering
- Base Station Subsystem (BSS) : BSC, BTS

GPRS Protocol Stack



Physical Layer = L1
PLL+RFL = Radio
GMSK= Gaussian minimum shift keying
Top Layer = IP/X.25

Layer Functionalities

- All data within GPRS backbone (b/w GSNs), is transferred using the GTP (GPRS Tunneling Protocol).
- GTP can uses two transport protocol TCP and UDP.
- The N/w protocol for the GPRS backbone is IP.
- The SubNetwork Dependent Convergence Protocol (SNDCP) used to adapt different characteristics of the underlying n/ws b/w an MS and SGSN.
- On top of SNDCP and GTP, user data packet is tunneled from the MS to the SGSN and vice versa.
- To achieve a high reliability of packet transfer b/w SGSN and MS, a special LLC is used, which compute ARQ and FEC.
- A Base Station Subsystem GPRS Protocol (BSSGP) is used to convey routing and QoS related info b/w BSS and SGSN.
- BSSGP does not perform error correction and works on the top of frame relay (FR),
- The Radio Link Control (RLC) provides a reliable link, while MAC controls access with signaling procedure for radio channel.

Future Of GSM

- 2nd Generation
  - GSM -9.6 Kbps (data rate)

- 2.5 Generation (Future of GSM)
  - HSCSD (High Speed ckt Switched data)
    - Data rate : 76.8 Kbps (9.6 x 8 kbps)
  - GPRS (General Packet Radio service)
    - Data rate: 14.4 - 171.2 Kbps
  - EDGE (Enhanced data rate for GSM Evolution)
    - Data rate: 547.2 Kbps (max)
- 3 Generation
  - WCDMA(Wide band CDMA)
    - Data rate : 0.348 – 2.0 Mbps

UNIT-II
(Wireless) Medium Access Control

Contents

1. Motivation for Specialized MAC
   - Hidden and Exposed Terminals
   - Far and Near Terminals
2. Access Methods
   - SDMA
   - FDMA
   - TDMA
   - CDMA

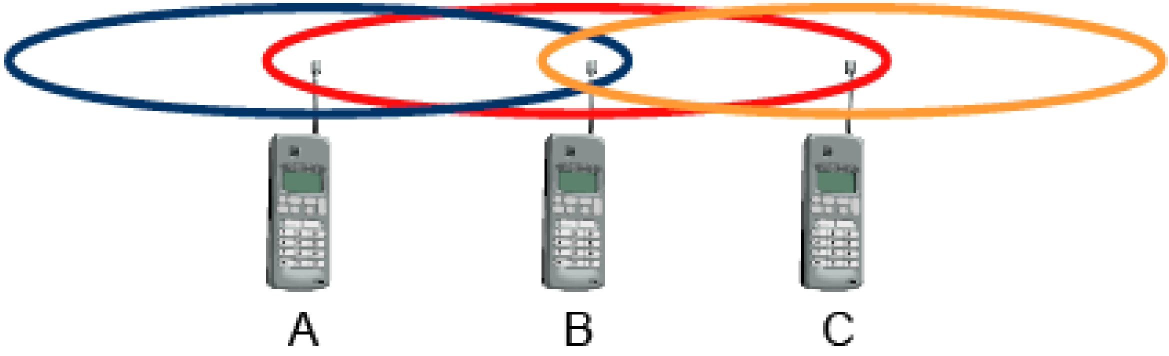
1. Motivation for Specialized MAC

- Can we apply media access methods from fixed networks?

- Example CSMA/CD
  - **C**arrier **S**ense **M**ultiple **A**ccess with **C**ollision **D**etection
  - send as soon as the medium is free, listen into the medium if a collision occurs (legacy method in IEEE 802.3)
- Problems in wireless networks
  - signal strength decreases proportional to the square of the distance
  - the sender would apply CS and CD, but the collisions happen at the receiver
  - it might be the case that a sender cannot "hear" the collision, i.e., CD does not work
  - furthermore, CS might not work if, e.g., a terminal is "hidden"

Motivation - hidden and exposed terminals

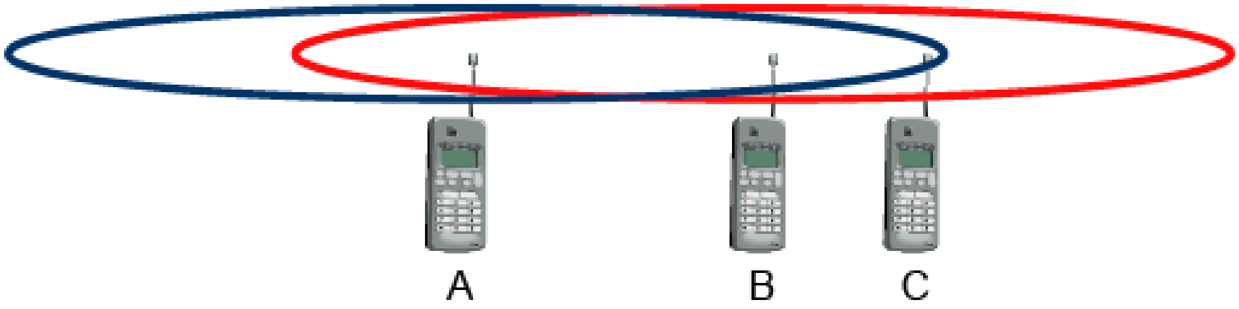
- Hidden terminals
  - A sends to B, C cannot receive A
  - C wants to send to B, C senses a "free" medium (CS fails)
  - collision at B, A cannot receive the collision (CD fails)
  - A is "hidden" for C



- Exposed terminals
  - B sends to A, C wants to send to another terminal (not A or B)
  - C has to wait, CS signals a medium in use
  - but A is outside the radio range of C, therefore waiting is not necessary
  - C is "exposed" to B

Motivation - near and far terminals

- Terminals A and B send, C receives
  - signal strength decreases proportional to the square of the distance
  - the signal of terminal B therefore drowns out A’s signal
  - C cannot receive A

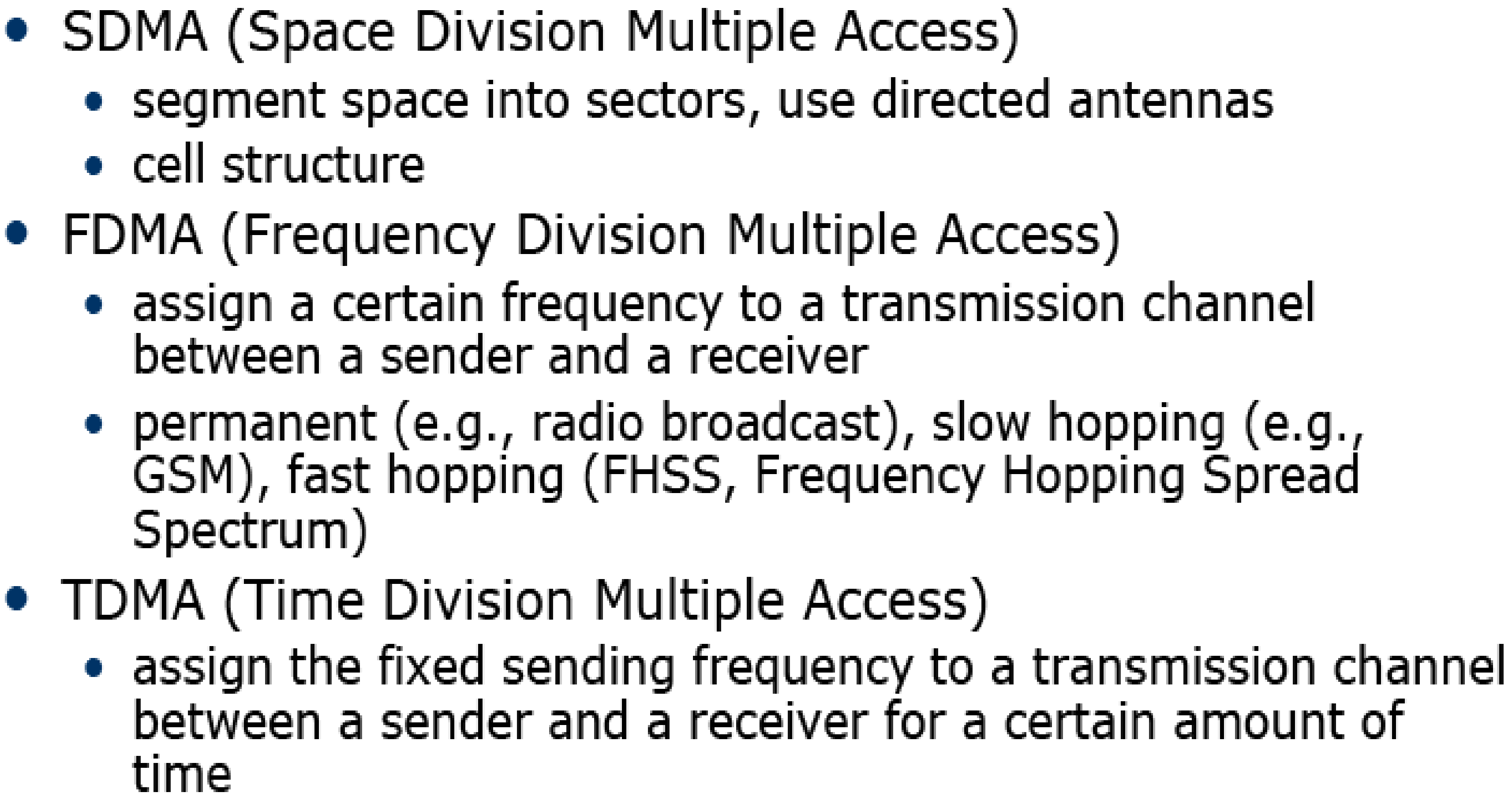


- If C for example was an arbiter for sending rights, terminal B would drown out terminal A already on the physical layer
- Also severe problem for CDMA-networks - precise power control needed!

MULTIPLEXING

- A fundamental mechanism in communication system and networks
- Enables multiple users to share a medium
- For wireless communication, multiplexing can be carried out in four dimensions: **space, time, frequency and code**

Access Methods

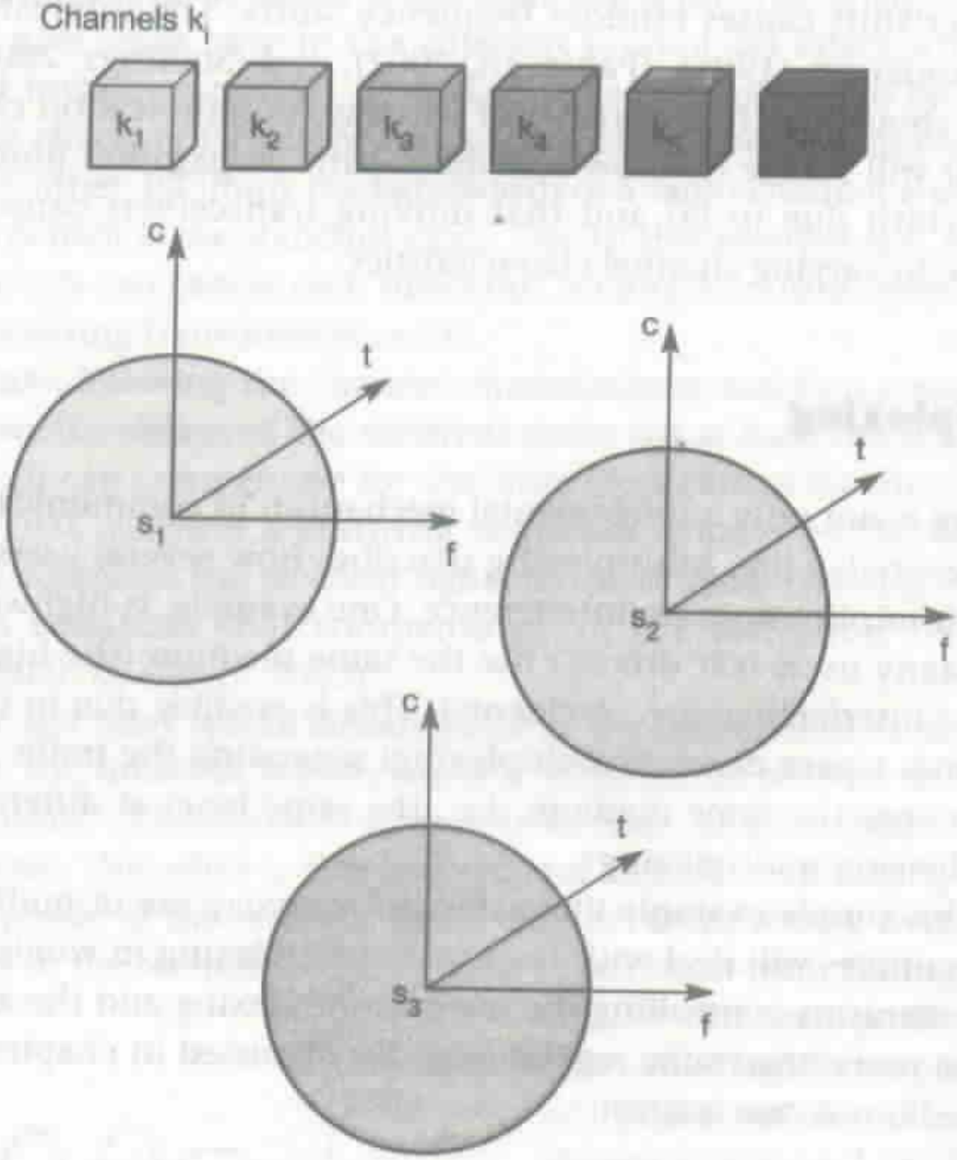


- SDMA (Space Division Multiple Access)
  - segment space into sectors, use directed antennas
  - cell structure
- FDMA (Frequency Division Multiple Access)
  - assign a certain frequency to a transmission channel between a sender and a receiver
  - permanent (e.g., radio broadcast), slow hopping (e.g., GSM), fast hopping (FHSS, Frequency Hopping Spread Spectrum)
- TDMA (Time Division Multiple Access)
  - assign the fixed sending frequency to a transmission channel between a sender and a receiver for a certain amount of time

2. Space Division Multiple Access (SDMA)

- Channels are assigned on the basis of “space” (but operate on same frequency)
- The assignment makes sure that the transmission do not interfere with each (with a guard band in between)

Figure 2.16 Space division multiplexing (SDM)



3. FDMA: Frequency Division Multiple Access

1. Frequency domain is subdivided into several non-overlapping frequency bands
2. Each channel is assigned its own frequency band (with guard spaces in between)

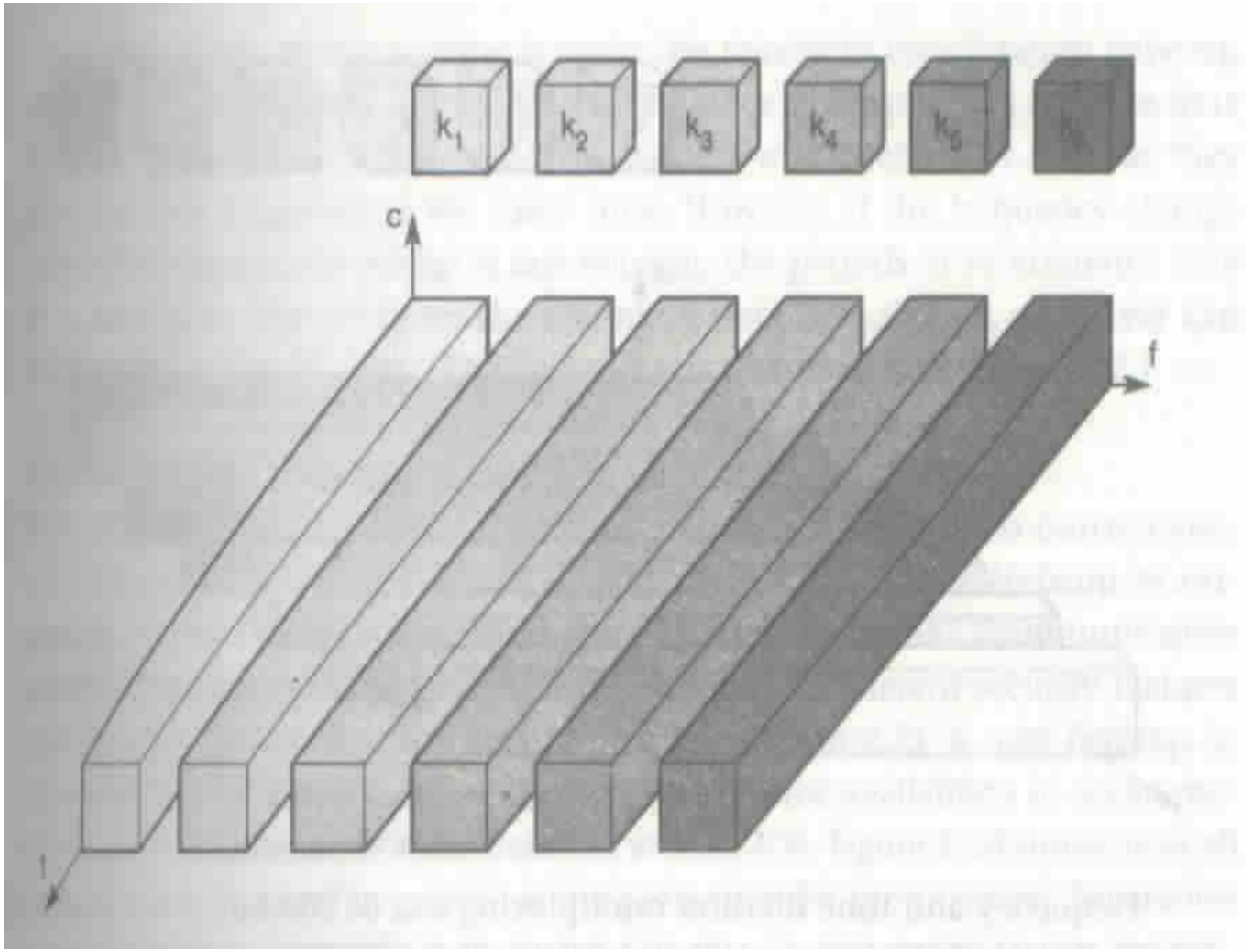
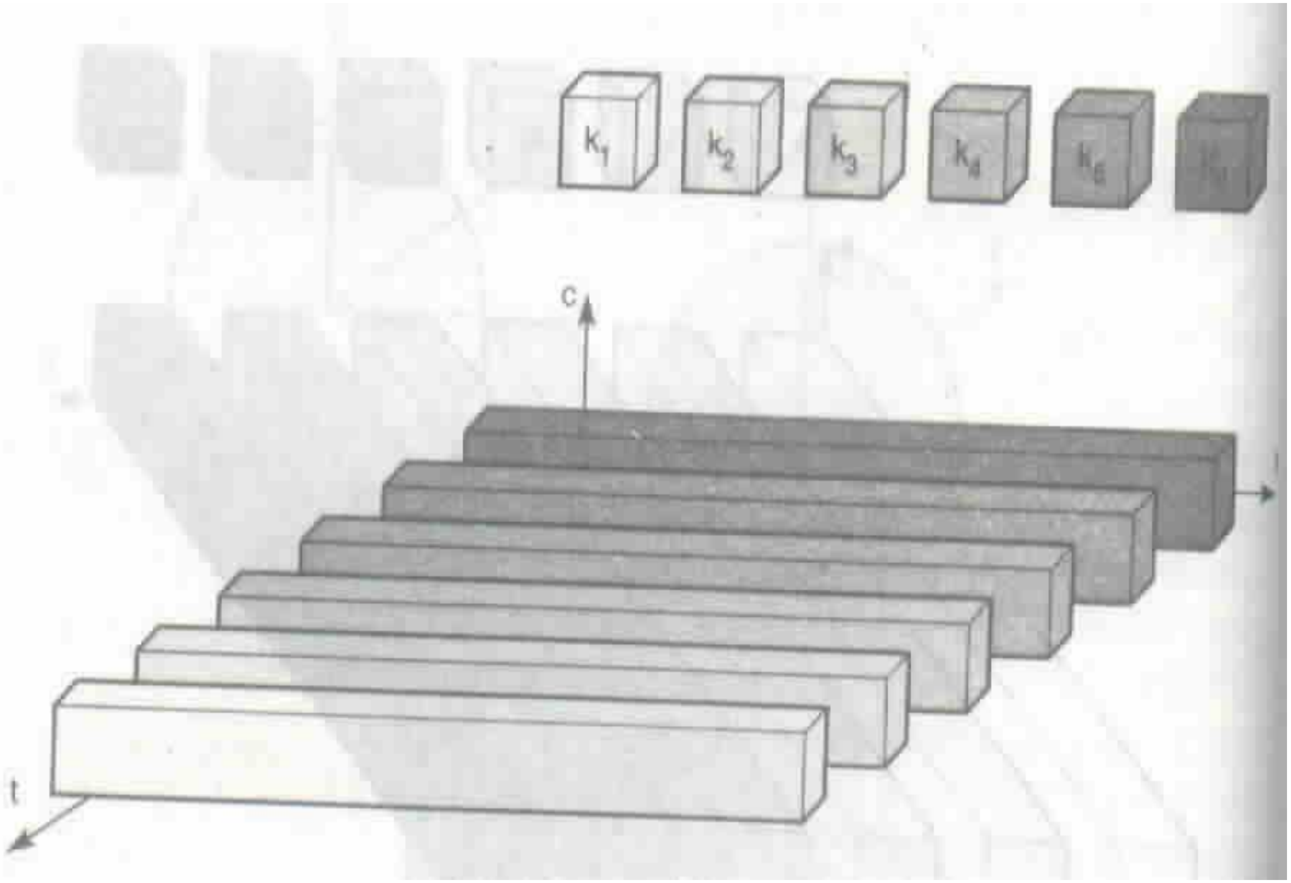


Figure 2.17 Frequency division multiplexing (FDM)

3. TDMA : Time Division Medium Access

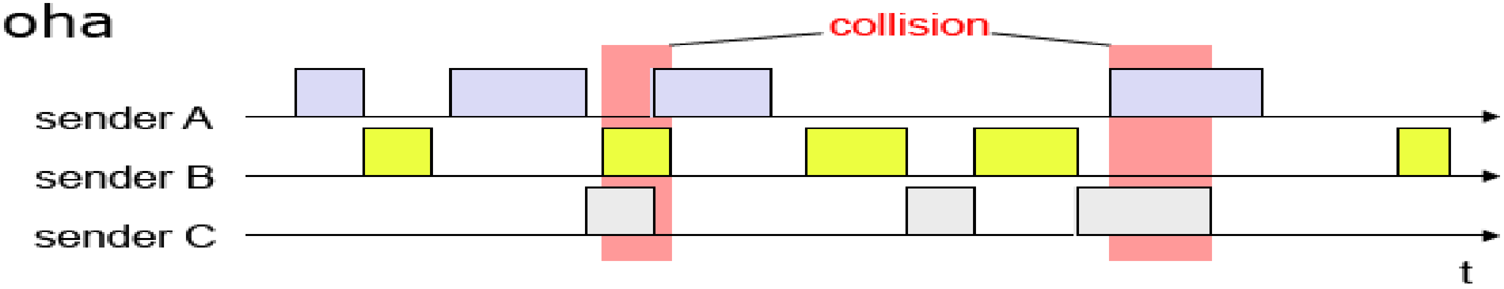
- A channel is given the whole bandwidth for a certain amount of time
  - All senders use the same frequency, but at different point of time
- Synchronization can be done by using
  - Fixed Allocation Scheme or
  - Dynamic Allocation Scheme

Figure 2.18
Time division
multiplexing (TDM)

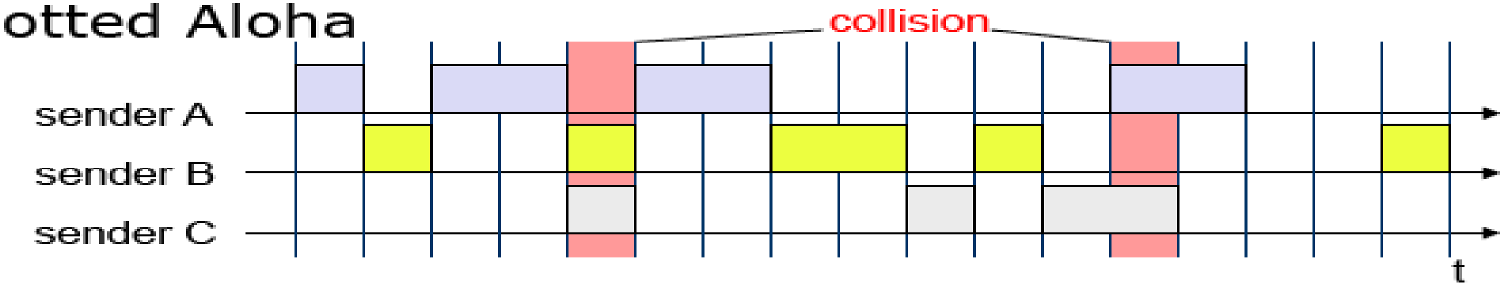


Aloha/slotted aloha

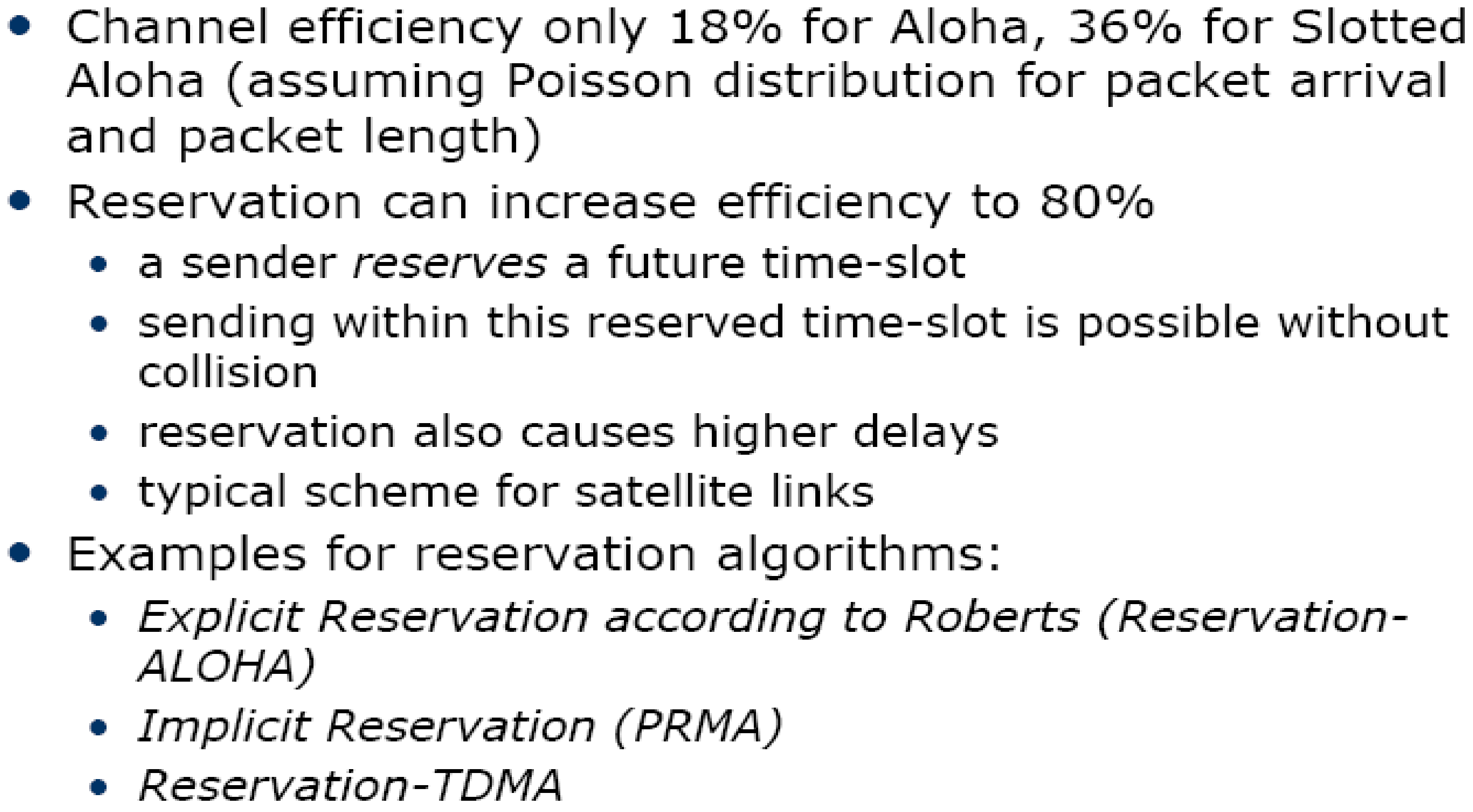
- Mechanism
  - random, distributed (no central arbiter), time-multiplex
  - Slotted Aloha additionally uses time-slots, sending must always start at slot boundaries
- Aloha



- Slotted Aloha



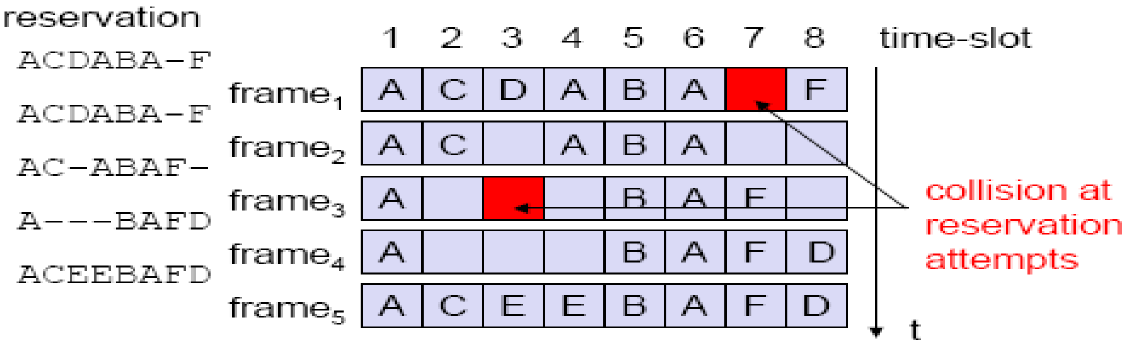
Reservation Algorithms



- Channel efficiency only 18% for Aloha, 36% for Slotted Aloha (assuming Poisson distribution for packet arrival and packet length)
- Reservation can increase efficiency to 80%
  - a sender *reserves* a future time-slot
  - sending within this reserved time-slot is possible without collision
  - reservation also causes higher delays
  - typical scheme for satellite links
- Examples for reservation algorithms:
  - *Explicit Reservation according to Roberts (Reservation-ALOHA)*
  - *Implicit Reservation (PRMA)*
  - *Reservation-TDMA*

Packet Reservation Multiple Access (PRMA) / Implicit Reservation

- Implicit reservation (PRMA - Packet Reservation MA):
    - a certain number of slots form a frame, frames are repeated
    - stations compete for empty slots according to the slotted aloha principle
    - once a station reserves a slot successfully, this slot is automatically assigned to this station in all following frames as long as the station has data to send
    - competition for this slots starts again as soon as the slot was empty in the last frame

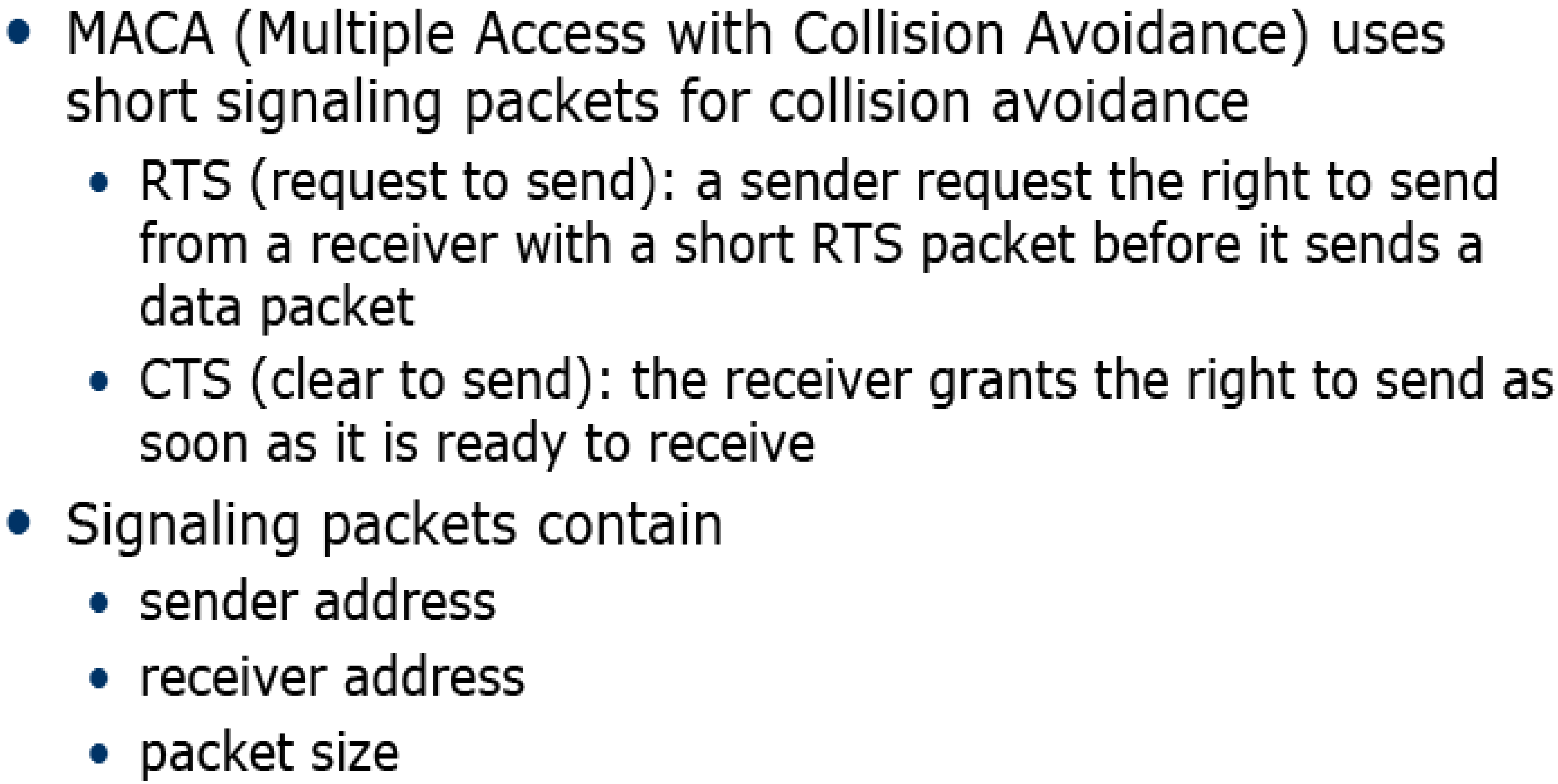


reservation		1	2	3	4	5	6	7	8
ACDABA-F	frame_1	A	C	D	A	B	A	🟥	F
ACDABA-F	frame_2	A	C		A	B	A		
AC-ABAF-	frame_3	A		🟥		B	A	F	
A---BAFD	frame_4	A				B	A	F	D
ACEEBAFD	frame_5	A	C	E	E	B	A	F	D

Carrier Sense Multiple Access Protocols (CSMA)

- In this each terminal on the network is able to monitor the status of the channel before transmitting information
- Variations:
  - 1-persistent CSMA
  - non-persistent CSMA -
  - p-persistent CSMA
  - CSMA/CA
  - Elimination yield – non-preemptive multiple access (EY-NMPA)

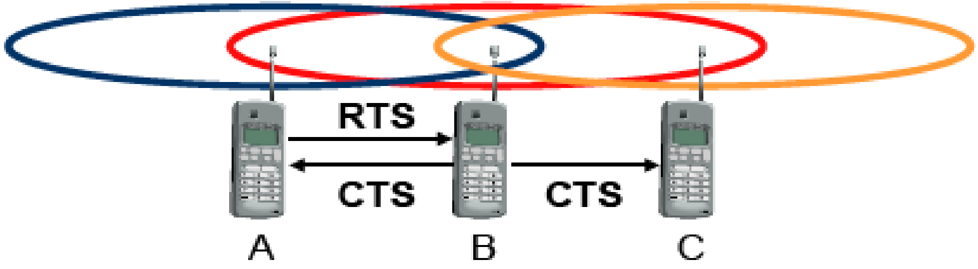
Multiple Access with Collision Avoidance (MACA)



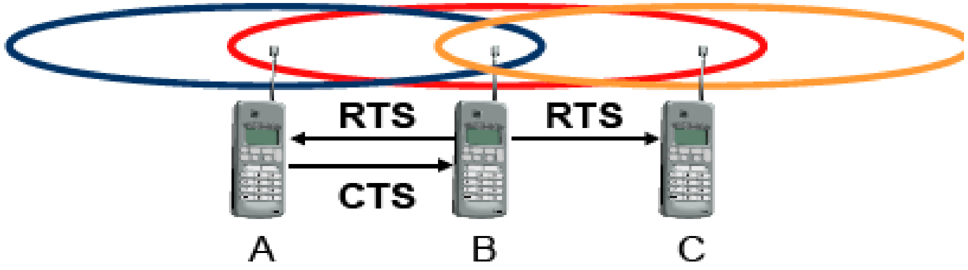
- MACA (Multiple Access with Collision Avoidance) uses short signaling packets for collision avoidance
    - RTS (request to send): a sender request the right to send from a receiver with a short RTS packet before it sends a data packet
    - CTS (clear to send): the receiver grants the right to send as soon as it is ready to receive
- Signaling packets contain
    - sender address
    - receiver address
    - packet size

MACA examples

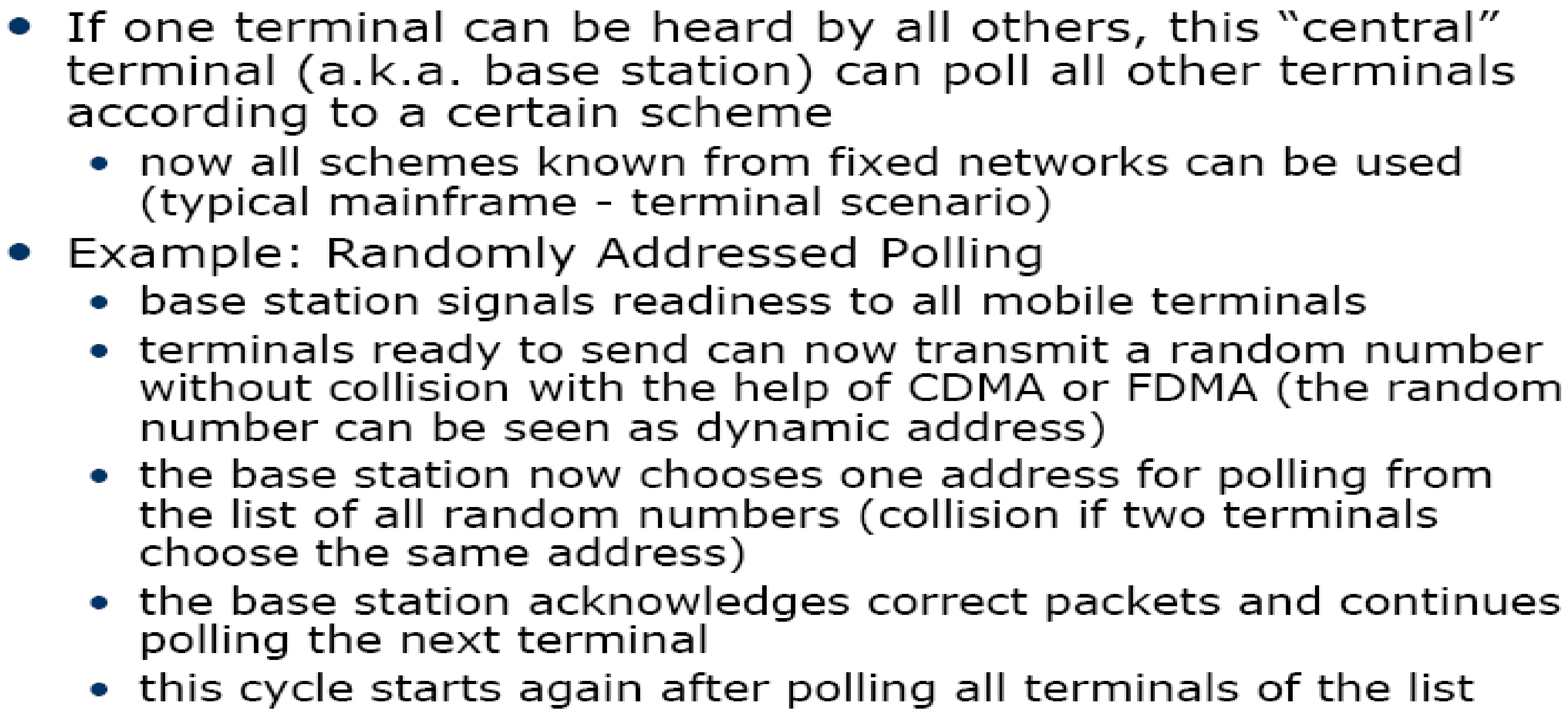
- MACA avoids the problem of hidden terminals
  - A and C want to send to B
  - A sends RTS first
  - C waits after receiving CTS from B



- MACA avoids the problem of exposed terminals
  - B wants to send to A, C to another terminal
  - now C does not have to wait for it cannot receive CTS from A



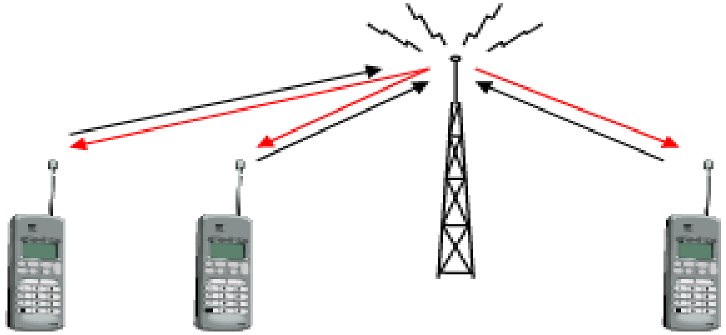
POLLING



- If one terminal can be heard by all others, this “central” terminal (a.k.a. base station) can poll all other terminals according to a certain scheme
  - now all schemes known from fixed networks can be used (typical mainframe - terminal scenario)
- Example: Randomly Addressed Polling
  - base station signals readiness to all mobile terminals
  - terminals ready to send can now transmit a random number without collision with the help of CDMA or FDMA (the random number can be seen as dynamic address)
  - the base station now chooses one address for polling from the list of all random numbers (collision if two terminals choose the same address)
  - the base station acknowledges correct packets and continues polling the next terminal
  - this cycle starts again after polling all terminals of the list

Inhibit Sense Multiple Access (ISMA)

- Current state of the medium is signaled via a “busy tone”
  - the base station signals on the downlink (base station to terminals) if the medium is free or not
  - terminals must not send if the medium is busy
  - terminals can access the medium as soon as the busy tone stops
  - the base station signals collisions and successful transmissions via the busy tone and acknowledgements, respectively (media access is not coordinated within this approach)
  - mechanism used, e.g., for CDPD (USA, integrated into AMPS)

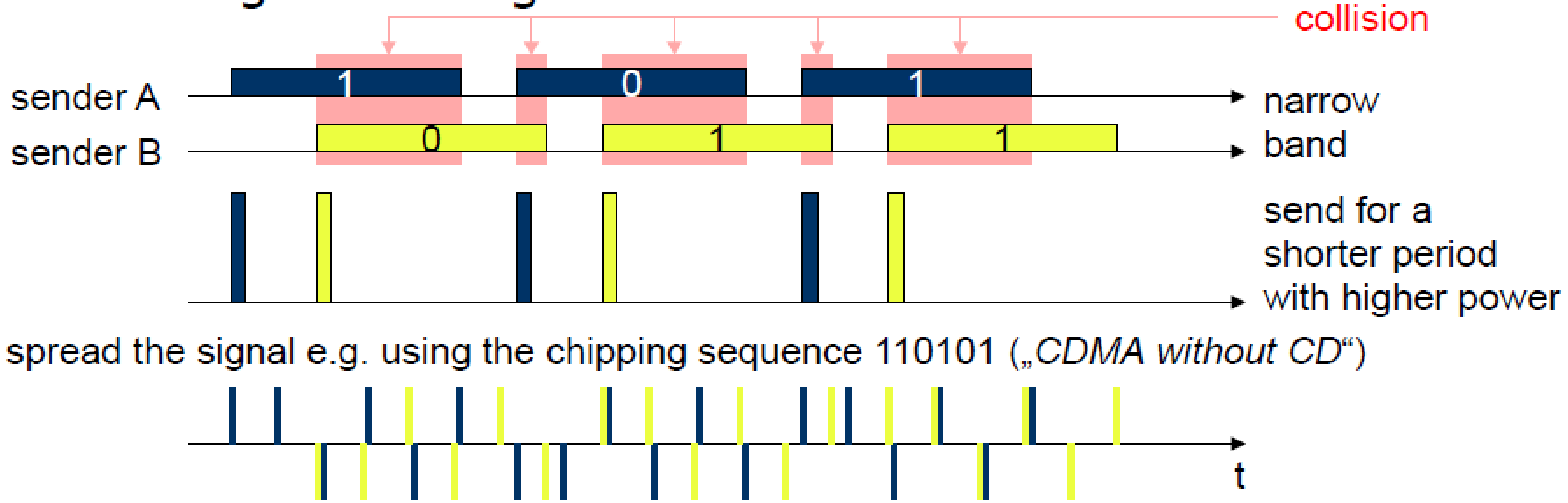


4. CDMA: Code Division Multiple Access

- separation of channels achieved by assigning each channel its own *code*
- guard spaces are realized by having *distance* in code space (e.g. orthogonal codes)
- transmitter can transmit in the same frequency band at the same time, but have to use different code
- Provides good protection against interference and tapping
- but the receivers have relatively high complexity.

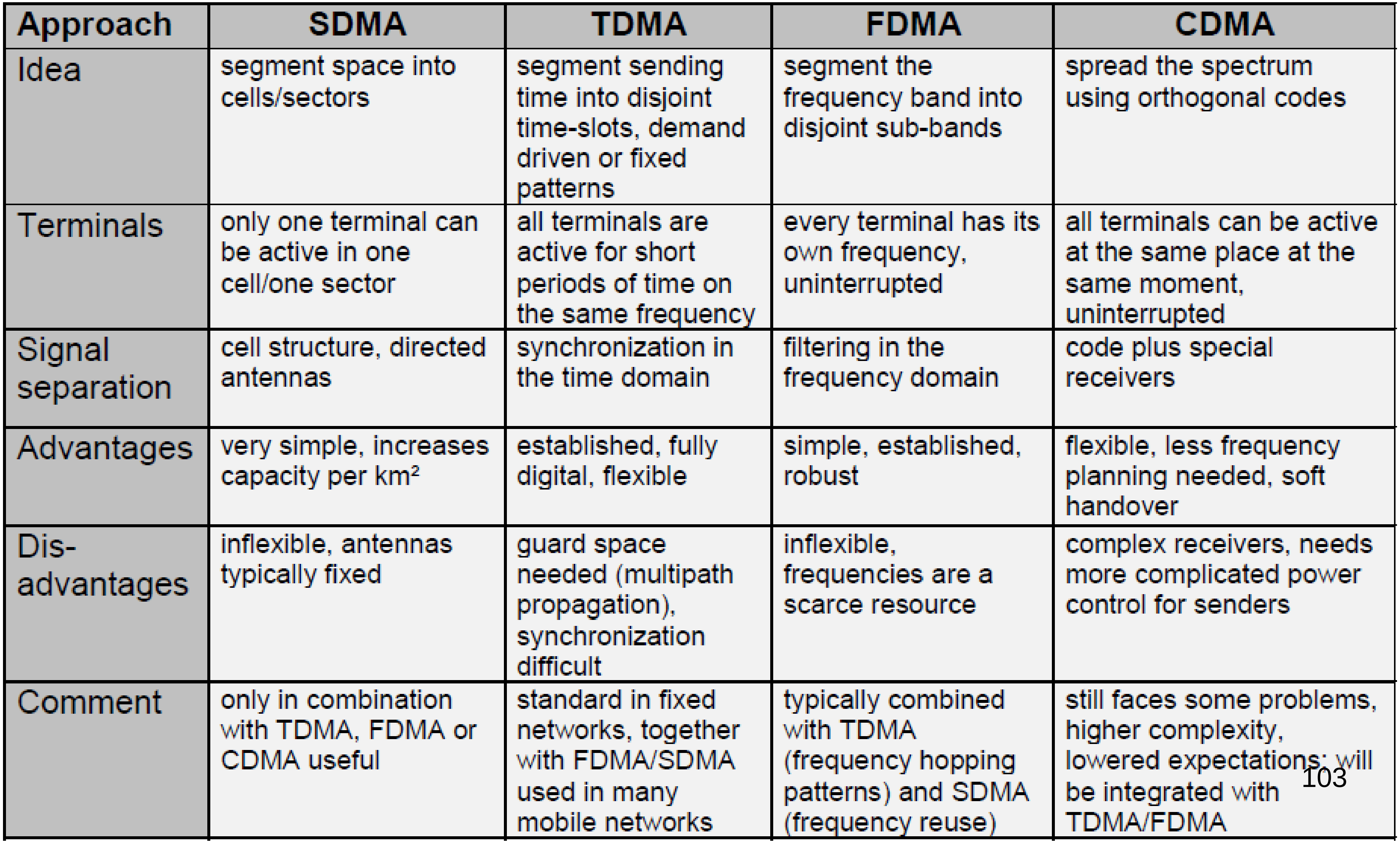
Spread Aloha Multiple Access (SAMA)

- Aloha has only a very low efficiency, CDMA needs complex receivers to be able to receive different senders with individual codes at the same time
- Idea: use spread spectrum with only one single code (chipping sequence) for spreading for all senders accessing according to aloha



spread the signal e.g. using the chipping sequence 110101 („*CDMA without CD*“)

Comparisons of S/T/F/CDMA



Approach	SDMA	TDMA	FDMA	CDMA
Idea	segment space into cells/sectors	segment sending time into disjoint time-slots, demand driven or fixed patterns	segment the frequency band into disjoint sub-bands	spread the spectrum using orthogonal codes
Terminals	only one terminal can be active in one cell/one sector	all terminals are active for short periods of time on the same frequency	every terminal has its own frequency, uninterrupted	all terminals can be active at the same place at the same moment, uninterrupted
Signal separation	cell structure, directed antennas	synchronization in the time domain	filtering in the frequency domain	code plus special receivers
Advantages	very simple, increases capacity per km²	established, fully digital, flexible	simple, established, robust	flexible, less frequency planning needed, soft handover
Dis-advantages	inflexible, antennas typically fixed	guard space needed (multipath propagation), synchronization difficult	inflexible, frequencies are a scarce resource	complex receivers, needs more complicated power control for senders
Comment	only in combination with TDMA, FDMA or CDMA useful	standard in fixed networks, together with FDMA/SDMA used in many mobile networks	typically combined with TDMA (frequency hopping patterns) and SDMA (frequency reuse)	still faces some problems, higher complexity, lowered expectations; will be integrated with TDMA/FDMA

UNIT-III

Mobile Network Layer

CONTENTS

1. **Mobile IP**
   - i) Goals, Assumptions & Requirements
   - ii) Entities and Terminology
   - iii) IP Packet Delivery
   - iv) Agent Advertisement & Discovery
   - v) Registration
   - vi) Tunneling & Encapsulation
   - vii) Optimizations

2. **DHCP**
   - i) Basic Configuration
   - ii) Client Initialization

Goal of Network Layer

- Goal of Routing Protocols
  - decrease routing-related overhead
  - find short routes
  - find “stable” routes (despite mobility)

- Goal of Mobile IP
  - Supporting end-system mobility while maintaining scalability, efficiency and compatibility in all respects with existing systems.

Mobile IP : Motivation

- Traditional routing
  - based on IP address; network prefix determines the subnet
  - change of physical subnet implies
    - change of IP address (conform to new subnet), or
    - special routing table entries to forward packets to new subnet

Quick Solution

- Changing of IP address
  - – Use DHCP to have a new IP address when mobile device moves to a new subnet, but then the new address may not be known to anyone
  - – Take help of DNS to update the entry, but DNS updates take long time
  - – TCP connections break
  - – security problems
- Changing entries in routing tables
  - – change routing table entries as the Mobile Node moves from one network to another does not scale with the number of mobile hosts and frequent changes in the location

Requirements

- Compatibility
  - support of the same layer 2 protocols as IP
  - no changes to current end-systems and routers required
  - mobile end-systems can communicate with fixed systems
- Transparency
  - mobile end-systems keep their IP address
  - continuation of communication after interruption of link possible
  - point of connection to the fixed network can be changed
- Efficiency and scalability
  - only little additional messages to the mobile system required (connection typically via a low bandwidth radio link)
  - world-wide support of a large number of mobile systems
- Security
  - authentication of all registration messages

Terminology

- Mobile Node (MN) - is an end-system that can change the point of connection to the network without changing its IP address.

- Home Network (HN) – is the subnet the MN belongs to with respect to its IP address.

- Foreign Network (FA) – is the current subnet the MN visits.

- Correspondent Node (CN) – is a fixed or Mobile Node act as partner for communication with MN.

- Care-of Address (COA)
  - address of the current tunnel end-point for the MN (at FA or MN)
  - actual location of the MN from an IP point of view
  - can be chosen, e.g., via DHCP

- Home Agent (HA)
  - Is a system (or router) located in the home network of the MN,
  - registers the location of the MN, then tunnels IP datagrams to the COA

- Foreign Agent (FA)
  - system in the current foreign network of the MN, typically a router
  - typically the default router for the MN

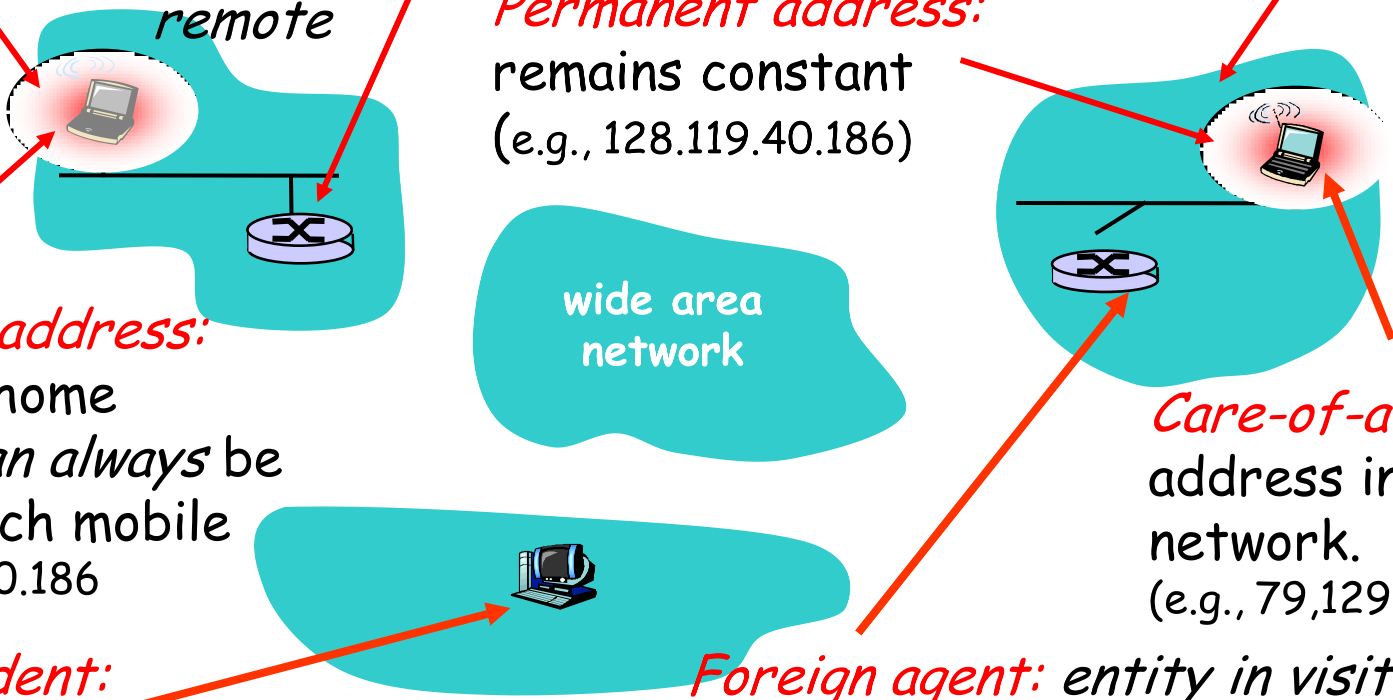
Mobility: Vocabulary

home network: permanent “home” of mobile (e.g., 128.119.40/24)

home agent: entity that will perform mobility functions on behalf of mobile, when mobile is remote

visited network: network in which mobile currently resides (e.g., 79.129.13/24)

Permanent address: remains constant (e.g., 128.119.40.186)



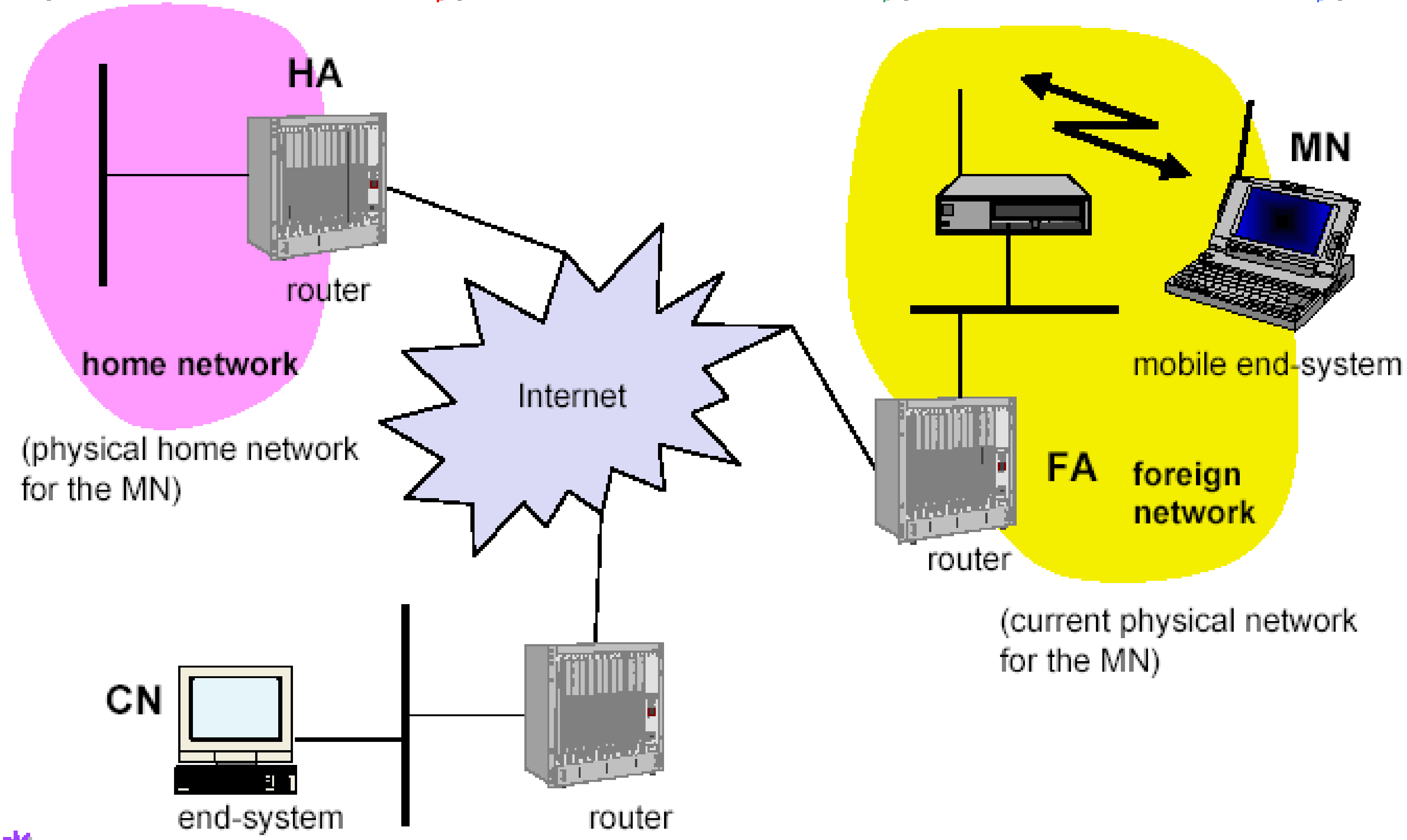
Permanent address: address in home network, *can always* be used to reach mobile e.g., 128.119.40.186

Care-of-address: address in visited network. (e.g., 79,129.13.2)

correspondent: wants to communicate with mobile

Foreign agent: entity in visited network that performs mobility functions on behalf of mobile.

Example network



HA

router

home network

(physical home network
for the MN)

Internet

MN

mobile end-system

FA

foreign
network

router

(current physical network
for the MN)

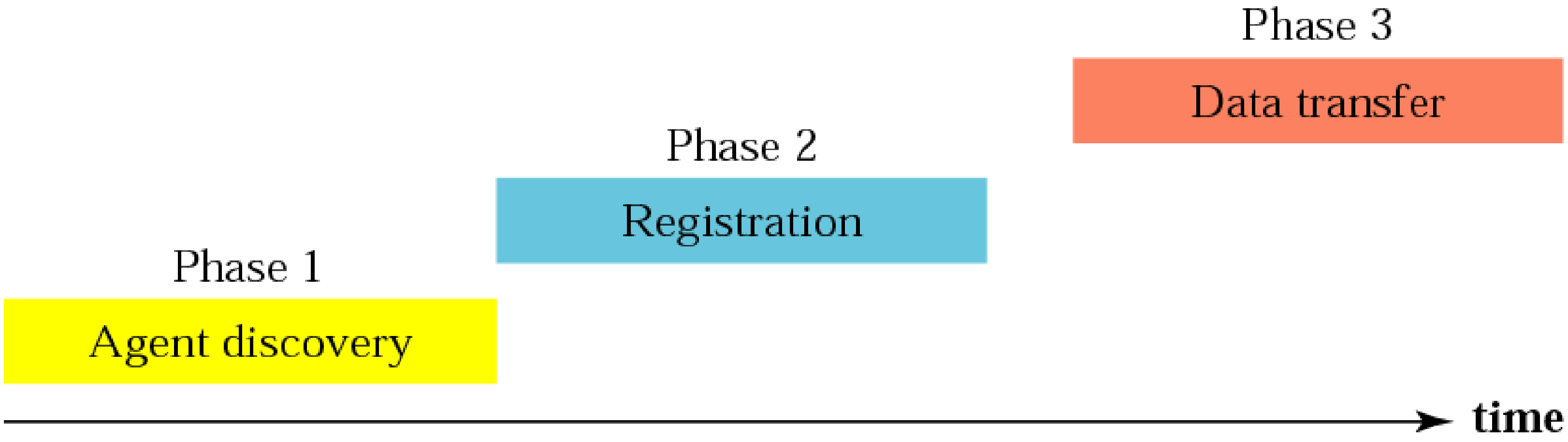
CN

end-system

router

Three Phases

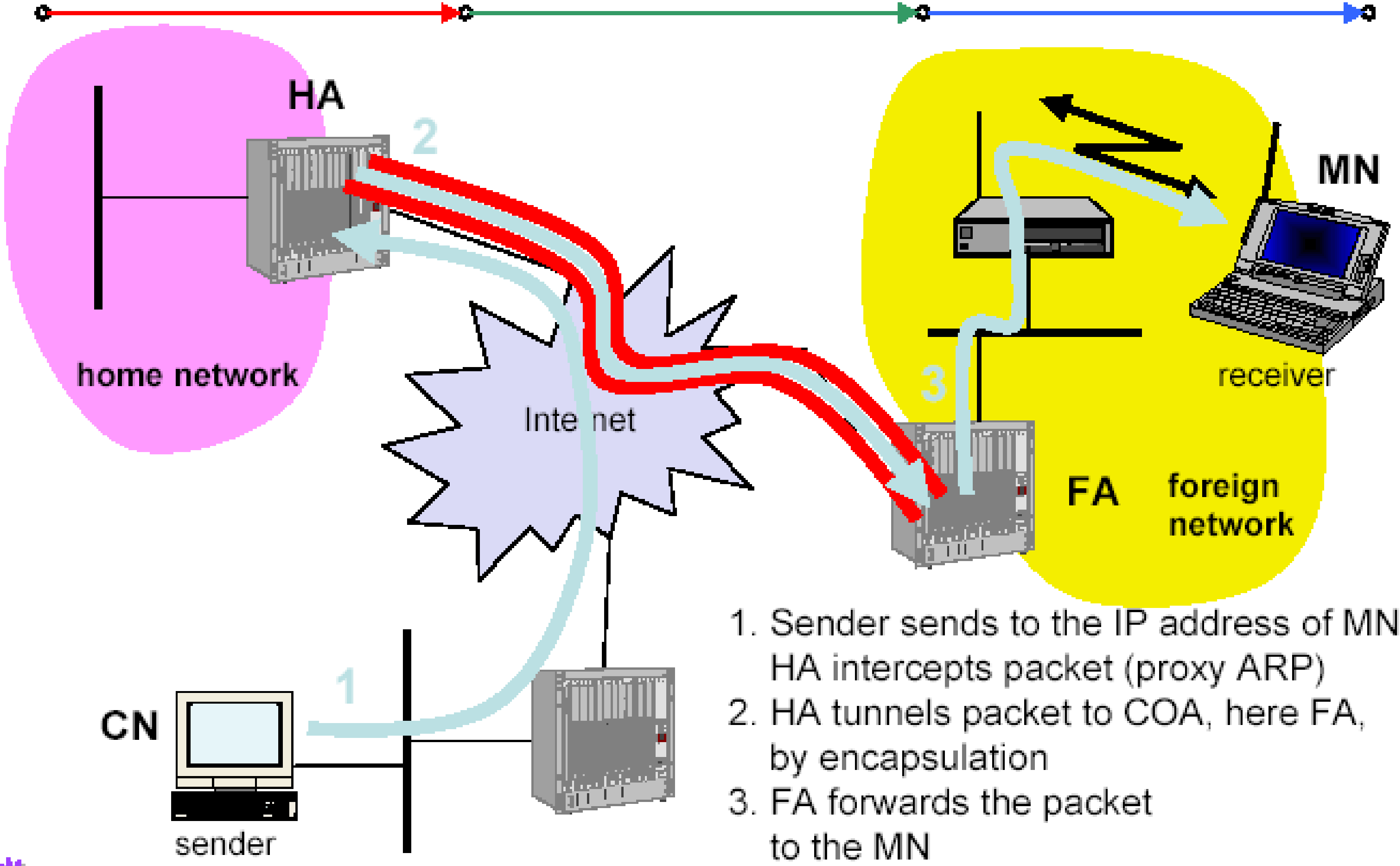
To communicate with a remote host, a mobile host goes through three phases: agent discovery, registration, and data transfer.



Phase-III : Data Transfer or IP Packet Delivery

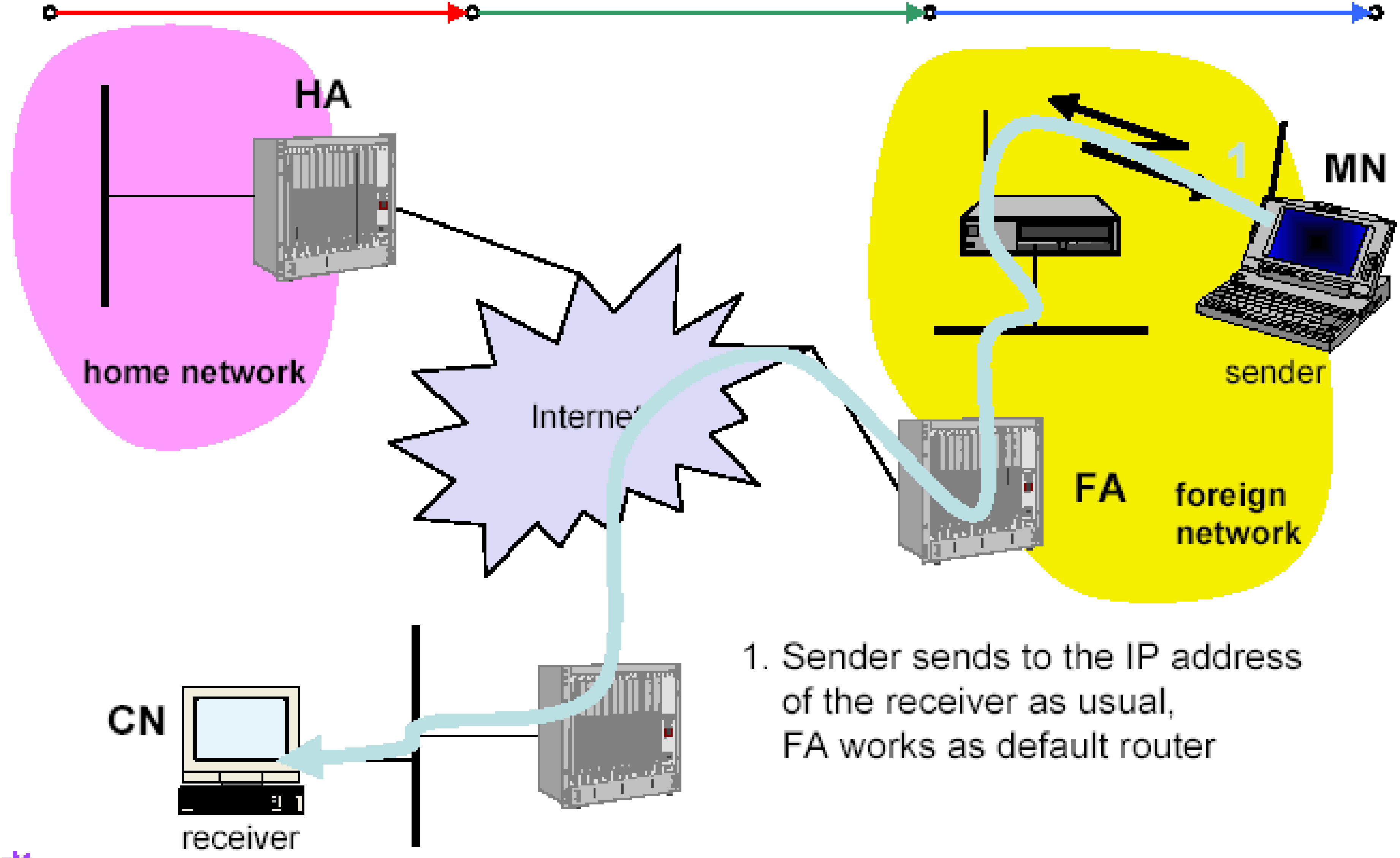
- Step 1: CN send the packet as usual to the IP address of MN.
- Step 2: The HA intercepts the packet and the forwarded into the subnet as usual, but encapsulated and tunneled to the COA.
- Step 3: The FA now decapsulates the packet and forwards the original packet with CN as source and MN as destination to the MN
- Step 4: The MN sends the packet as usual with its own fixed IP address as source and CN's address as the destination.

Data transfer to the mobile system



1. Sender sends to the IP address of MN, HA intercepts packet (proxy ARP)
2. HA tunnels packet to COA, here FA, by encapsulation
3. FA forwards the packet to the MN

Data transfer from the mobile system



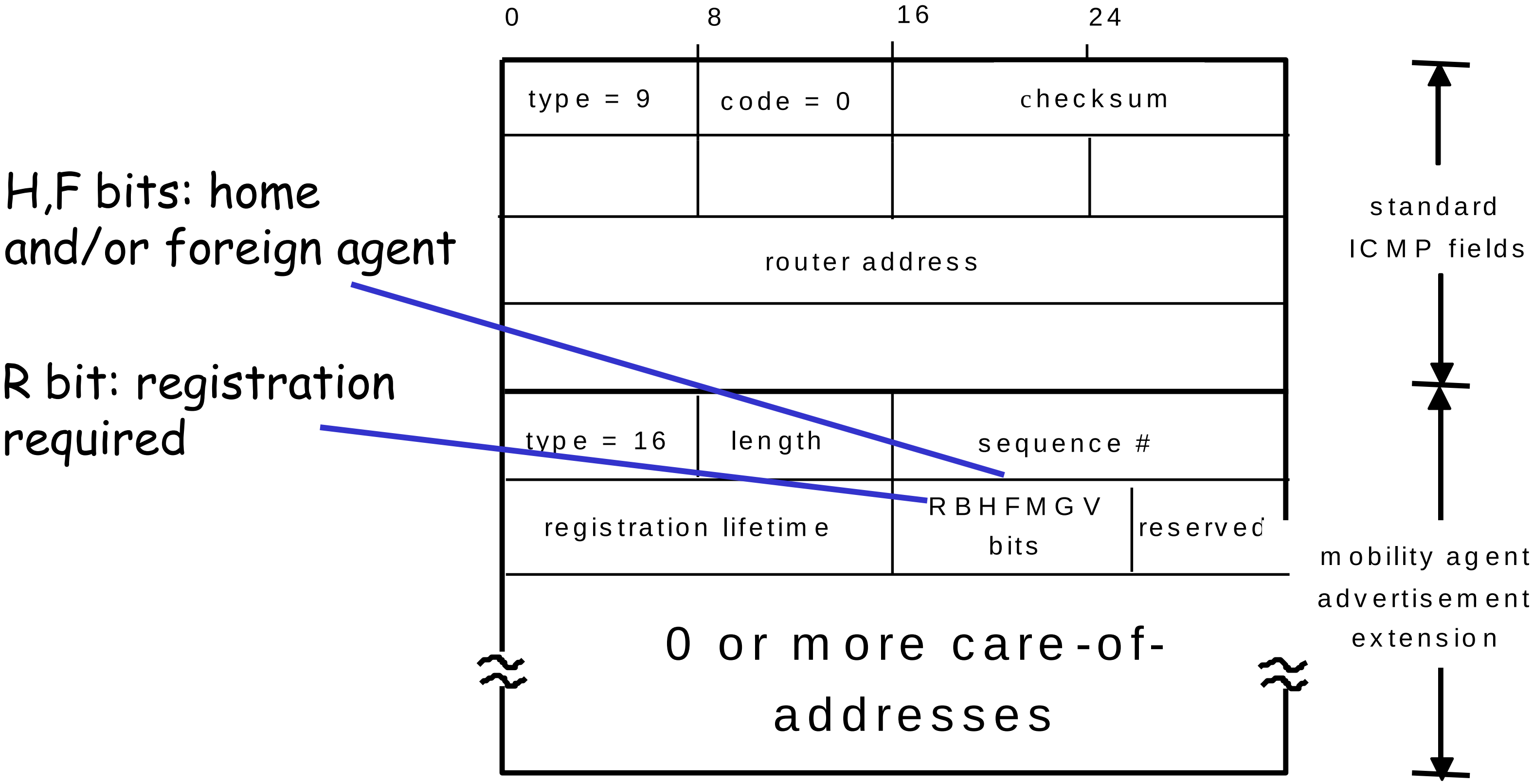
1. Sender sends to the IP address of the receiver as usual, FA works as default router

Phase – I : Agent Discovery

- HA and FA periodically send Advertisement messages into air
- MN listens to these messages and detects, if it is in the home or a foreign network
  - If MN does not wish to wait for the periodic advertisement, it can send out Agent Solicitation messages that will be responded by HA or FS
- MN reads a COA from the FA advertisement messages

Mobile IP: agent discovery

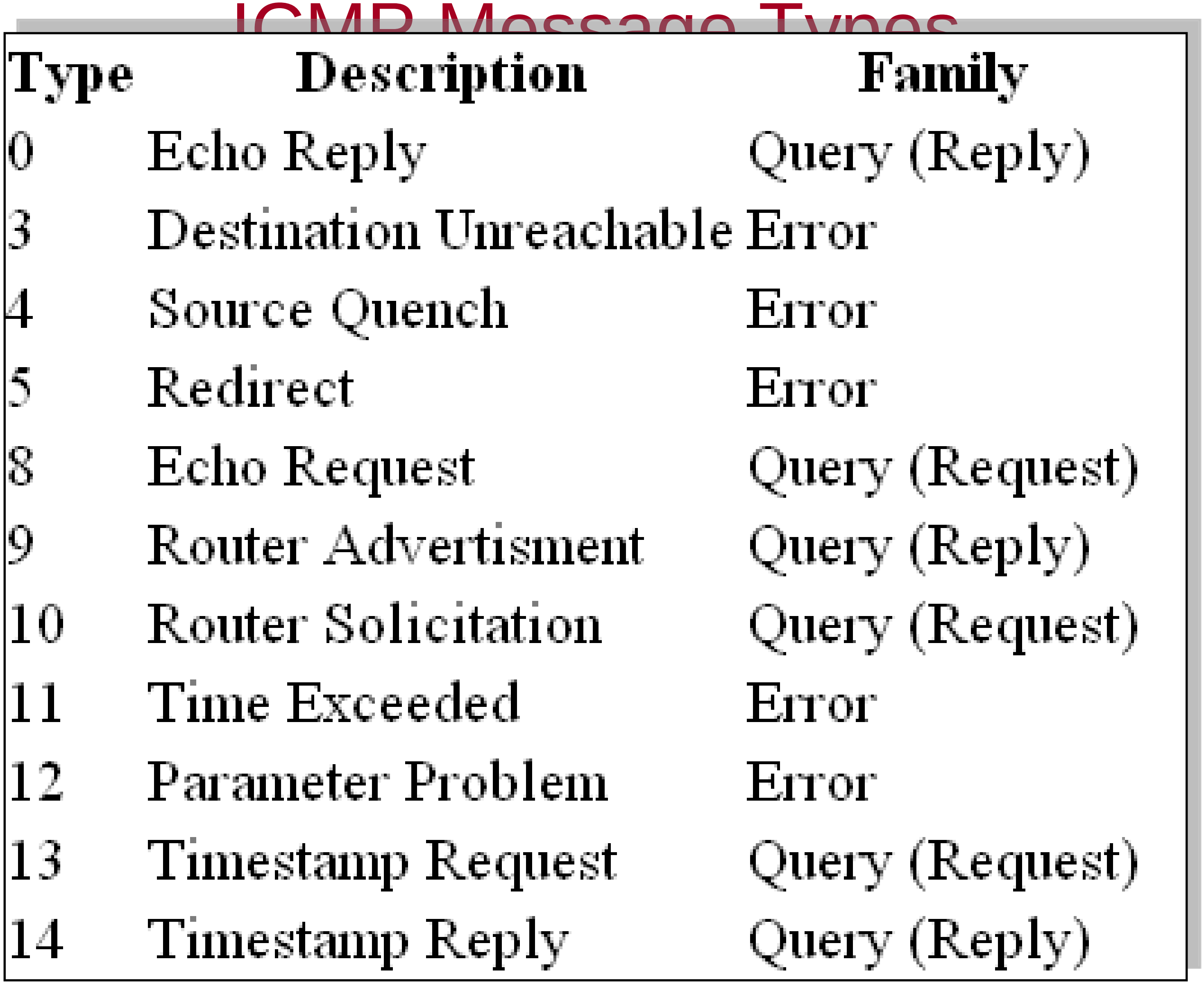
- agent advertisement: foreign/home agents advertise service by broadcasting ICMP messages (typefield = 9)



Field Description

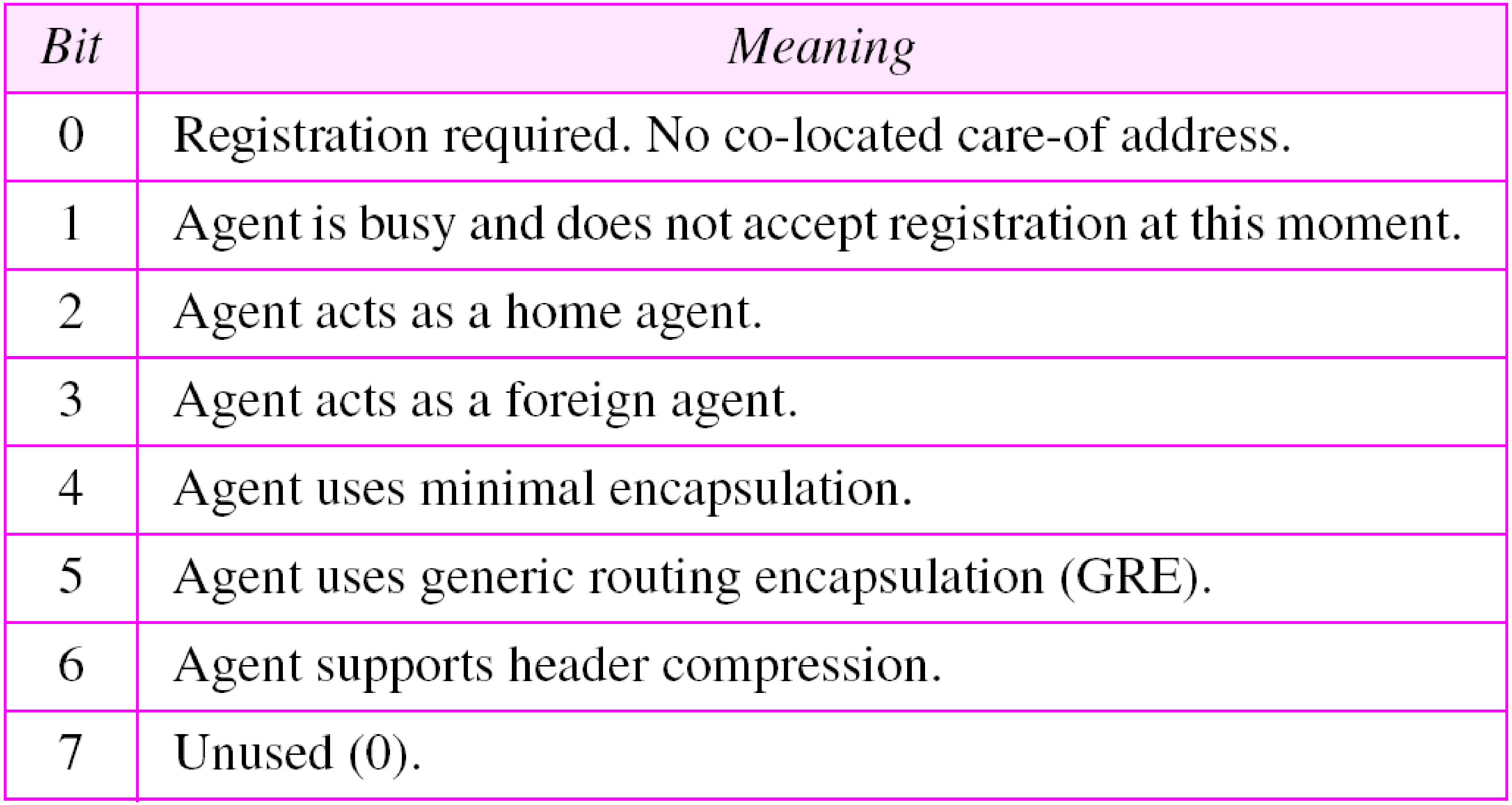
- Type – ICMP packet type (9 – Router Advt)
- Code – 0 - agent routes traffic for mobile/non- mobile node – 16 – only for mobile traffic
- #address – number of router addresses
- Life time – length of time this advertise valid
- Preference – choose the router, most eager one to get a new node.

ICMP Message Types



Type	Description	Family
0	Echo Reply	Query (Reply)
3	Destination Unreachable	Error
4	Source Quench	Error
5	Redirect	Error
8	Echo Request	Query (Request)
9	Router Advertisment	Query (Reply)
10	Router Solicitation	Query (Request)
11	Time Exceeded	Error
12	Parameter Problem	Error
13	Timestamp Request	Query (Request)
14	Timestamp Reply	Query (Reply)

Code Bits



Bit	*Meaning*
0	Registration required. No co-located care-of address.
1	Agent is busy and does not accept registration at this moment.
2	Agent acts as a home agent.
3	Agent acts as a foreign agent.
4	Agent uses minimal encapsulation.
5	Agent uses generic routing encapsulation (GRE).
6	Agent supports header compression.
7	Unused (0).

Phase-II: Registration

- MN signals COA to the HA via the FA, HA acknowledges via FA to MN
  - – these actions have to be secured by a

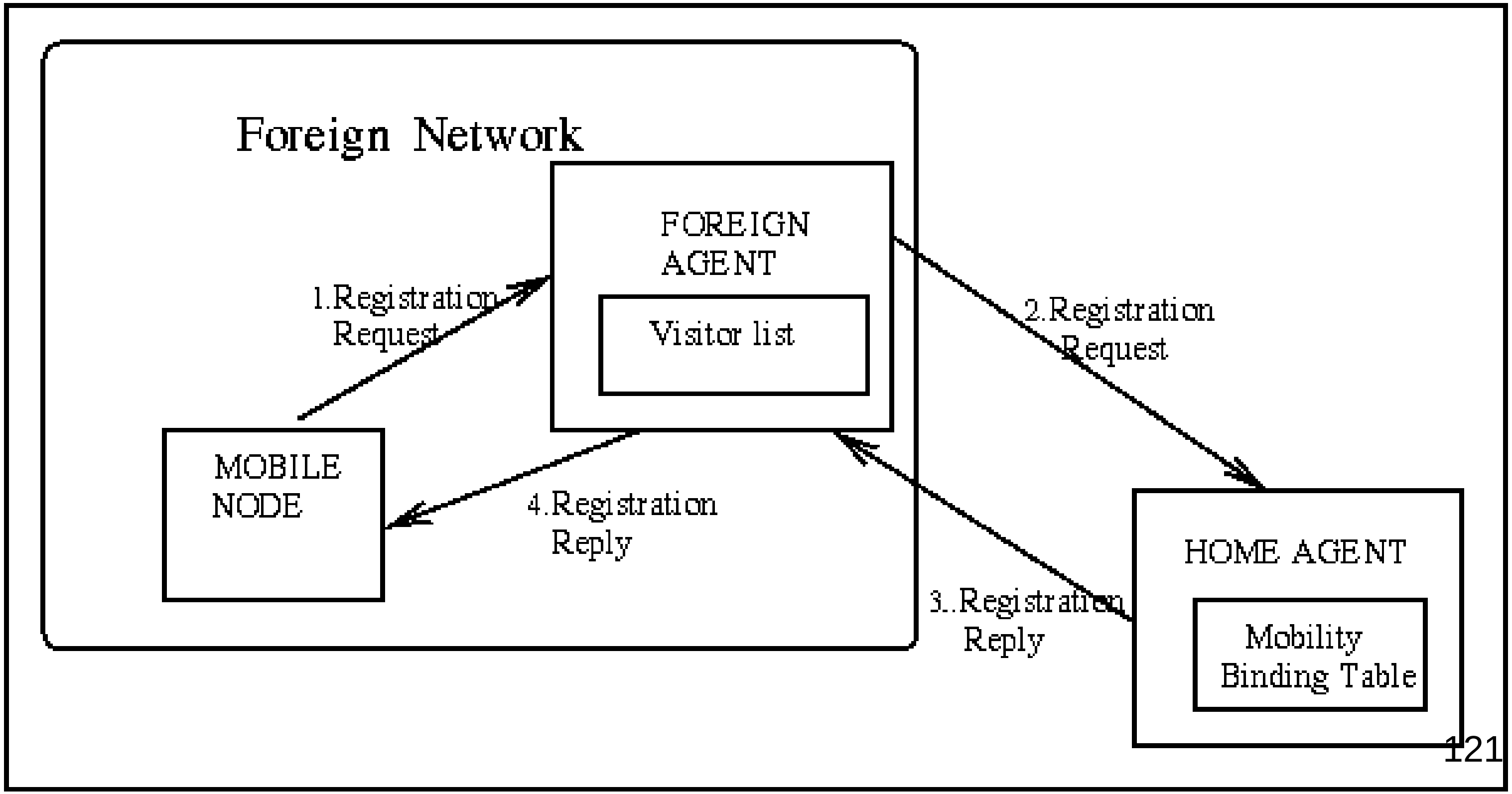
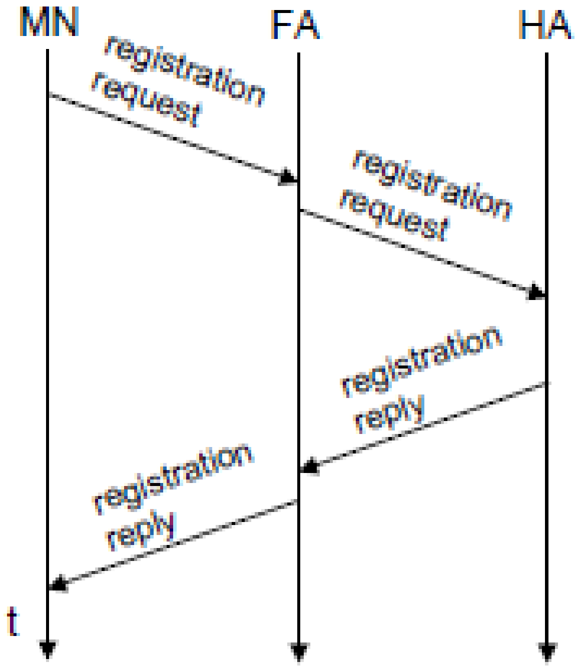


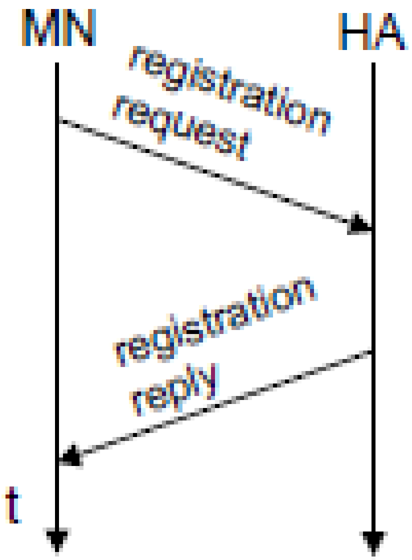
Figure 3: Registration process in Mobile IP

Registration

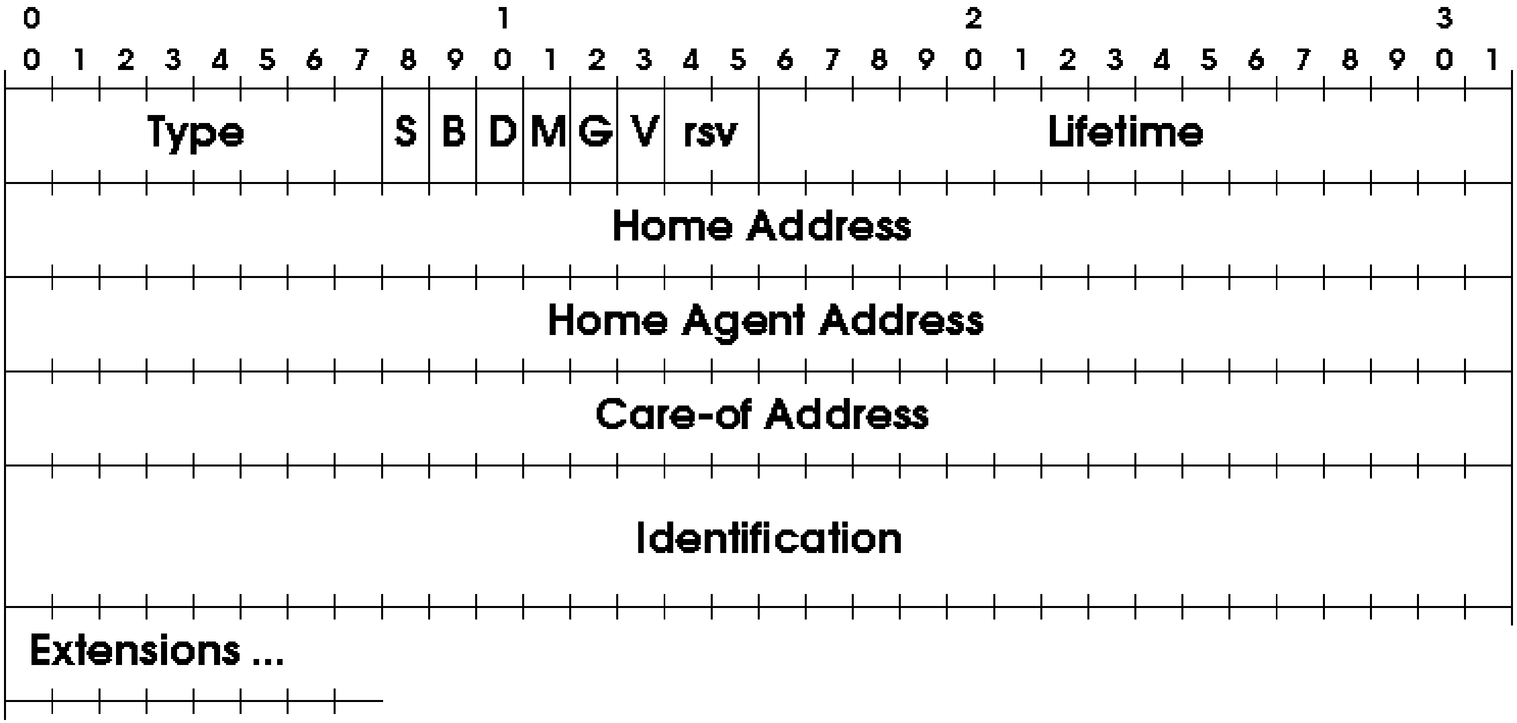
Two Ways : 1. Via FA

2. directly with HA



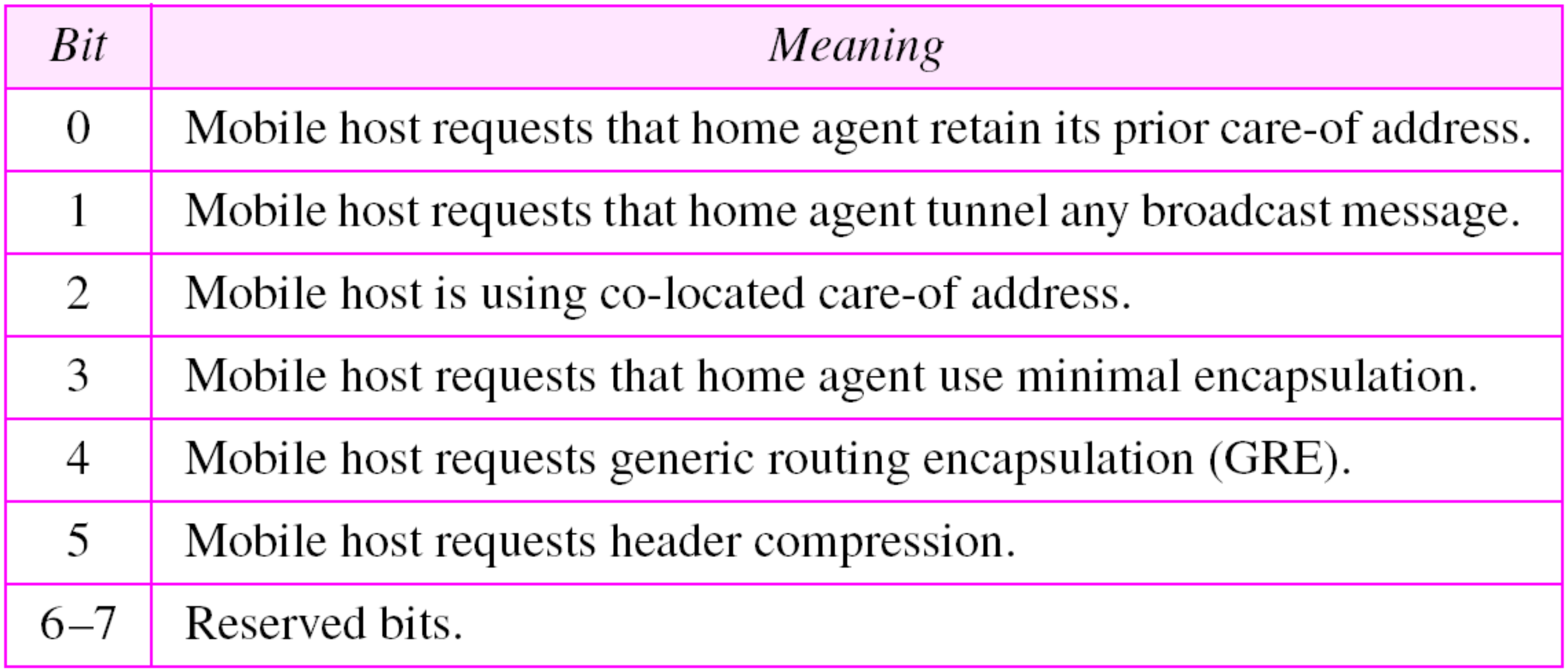


Registration Request



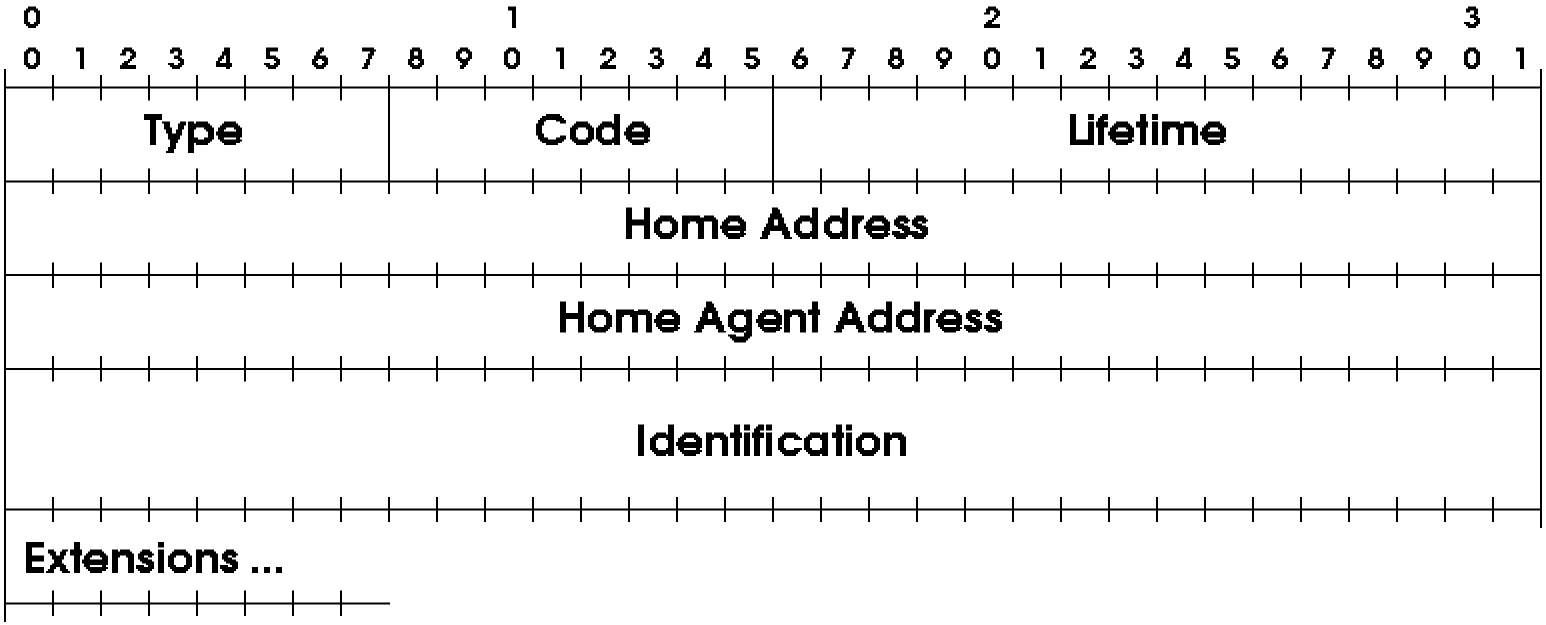
- *Type* is set to 1 for registration request
- *Flags* – S – simultaneous bindings, B – receive Broadcast pkts, D – de-capsulation, M & G – Minimul & generic encapsulation, T – reverse tunnel (replaced with V)
- Identification used for replay protection
- Uses UDP messages

Registration request flag field bits



Bit	*Meaning*
0	Mobile host requests that home agent retain its prior care-of address.
1	Mobile host requests that home agent tunnel any broadcast message.
2	Mobile host is using co-located care-of address.
3	Mobile host requests that home agent use minimal encapsulation.
4	Mobile host requests generic routing encapsulation (GRE).
5	Mobile host requests header compression.
6–7	Reserved bits.

Registration Reply



Type

Code

Lifetime

Home Address

Home Agent Address

Identification

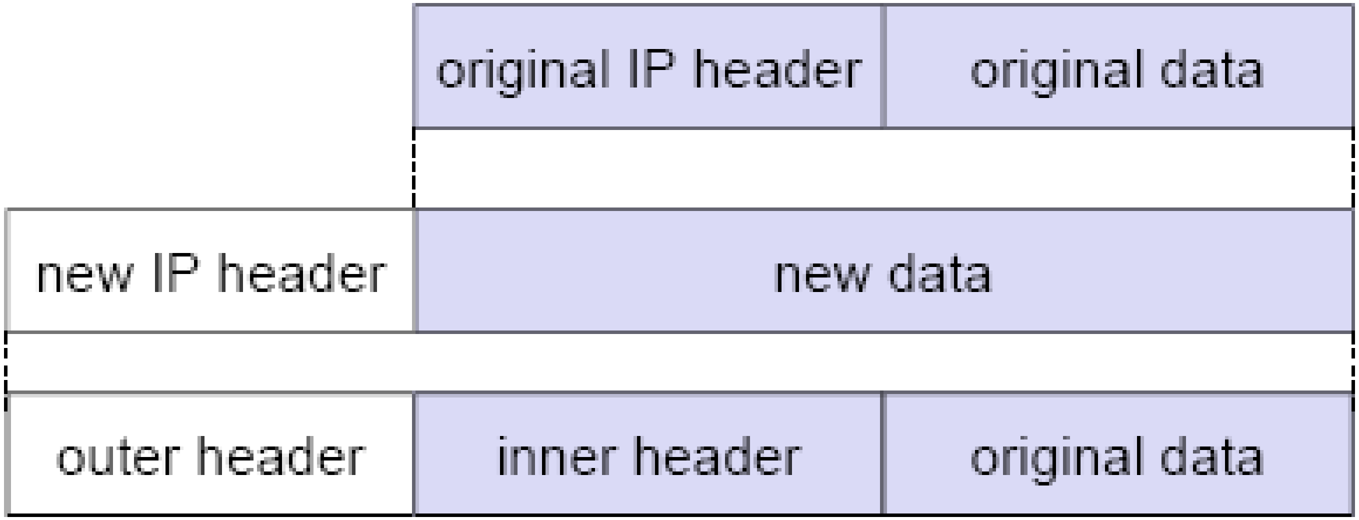
Extensions ...

- ❑ *Type* field set to 3
- ❑ Extensions must at least contain parameter for authentication
- ❑ *Code* field describes status information, e.g. why the registration failed. These include
  - authentication failed, ID mismatch , unknown HA

Tunneling and Encapsulation

- It describes the mechanism used for forwarding packets between the HA and COA.
- A tunnel establishes the virtual pipe for data packets between a tunnel entry and tunnel end point.
- Encapsulation is the mechanism of taking a packet consisting of packet header and data and put into a data part of another packet.
- Outer Header – Header of New Packet
- Inner Header – identical to header of Original Packet or can be computed during encapsulation.
- There are different ways of performing the encapsulation – IP-in-IP (mandatory), minimal and generic routing.

IP Encapsulation



IP-in-IP Encapsulation

- Tunneling between HA and COA

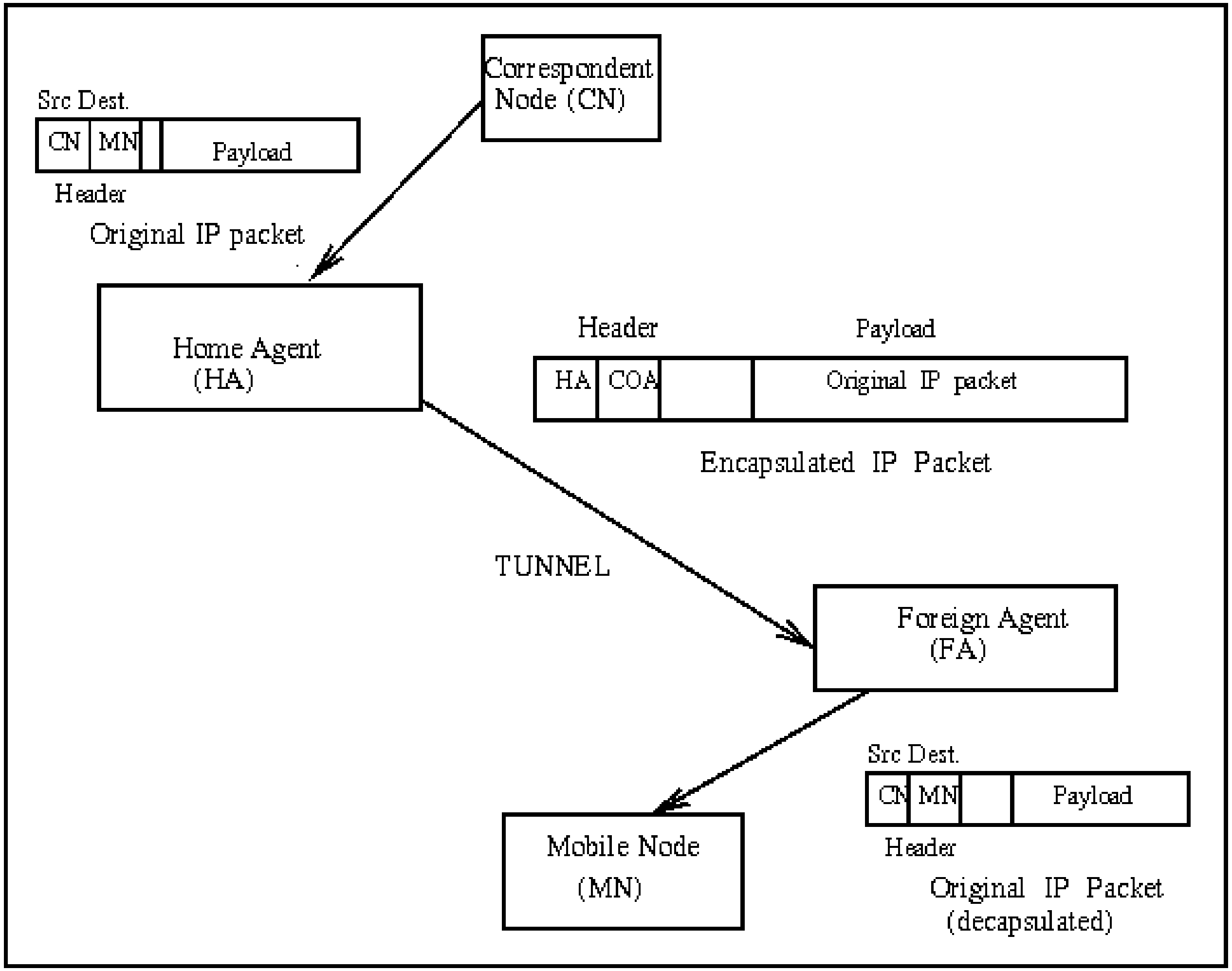
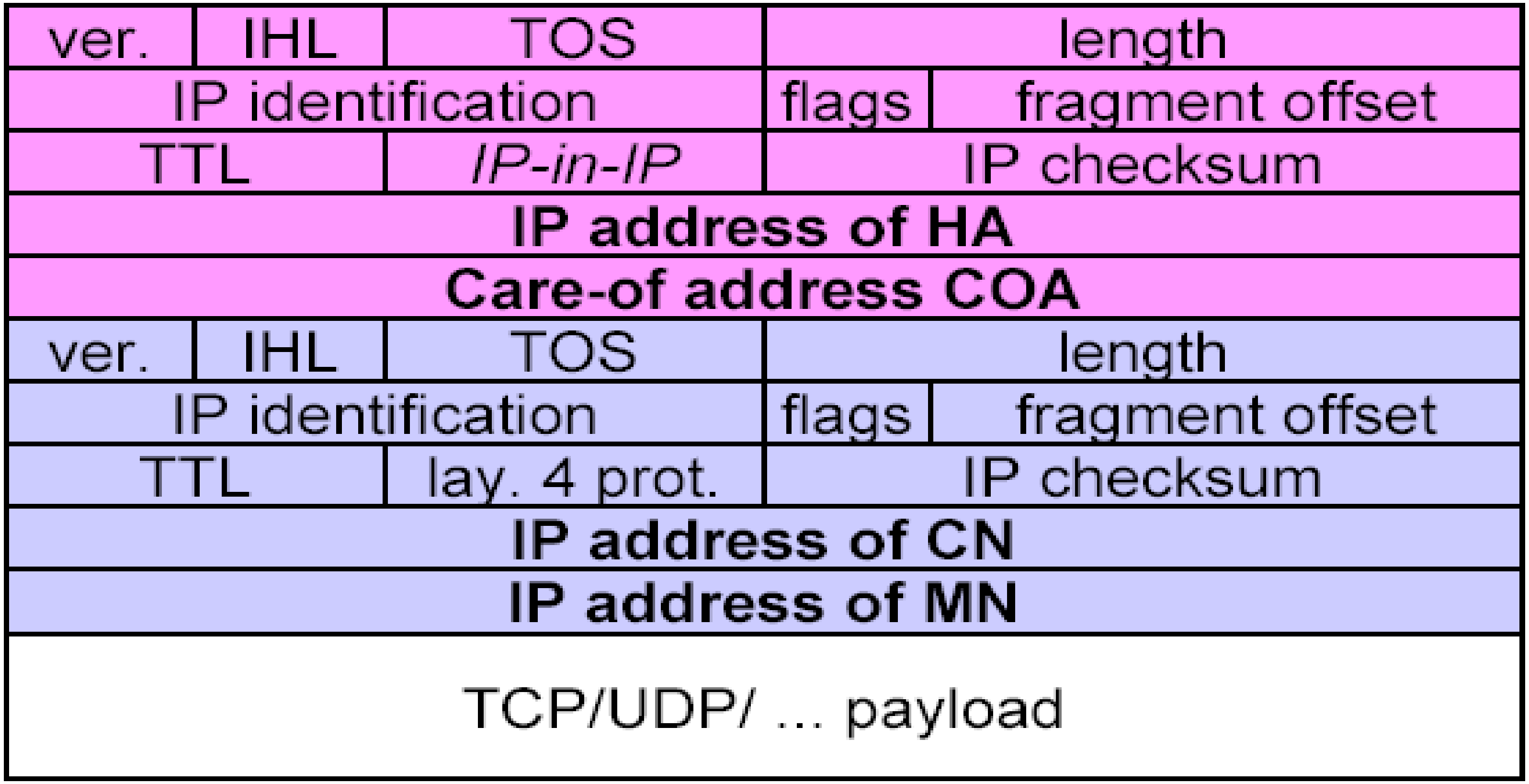


Figure 4: Tunneling operation in Mobile IP

IP-in-IP Encapsulation



<table>
<tr><td>ver.</td><td>IHL</td><td>TOS</td><td colspan="2">length</td></tr>
<tr><td colspan="3">IP identification</td><td>flags</td><td>fragment offset</td></tr>
<tr><td colspan="2">TTL</td><td>*IP-in-IP*</td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="5">**IP address of HA**</td></tr>
<tr><td colspan="5">**Care-of address COA**</td></tr>
<tr><td>ver.</td><td>IHL</td><td>TOS</td><td colspan="2">length</td></tr>
<tr><td colspan="3">IP identification</td><td>flags</td><td>fragment offset</td></tr>
<tr><td colspan="2">TTL</td><td>lay. 4 prot.</td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="5">**IP address of CN**</td></tr>
<tr><td colspan="5">**IP address of MN**</td></tr>
<tr><td colspan="5">TCP/UDP/ ... payload</td></tr>
</table>

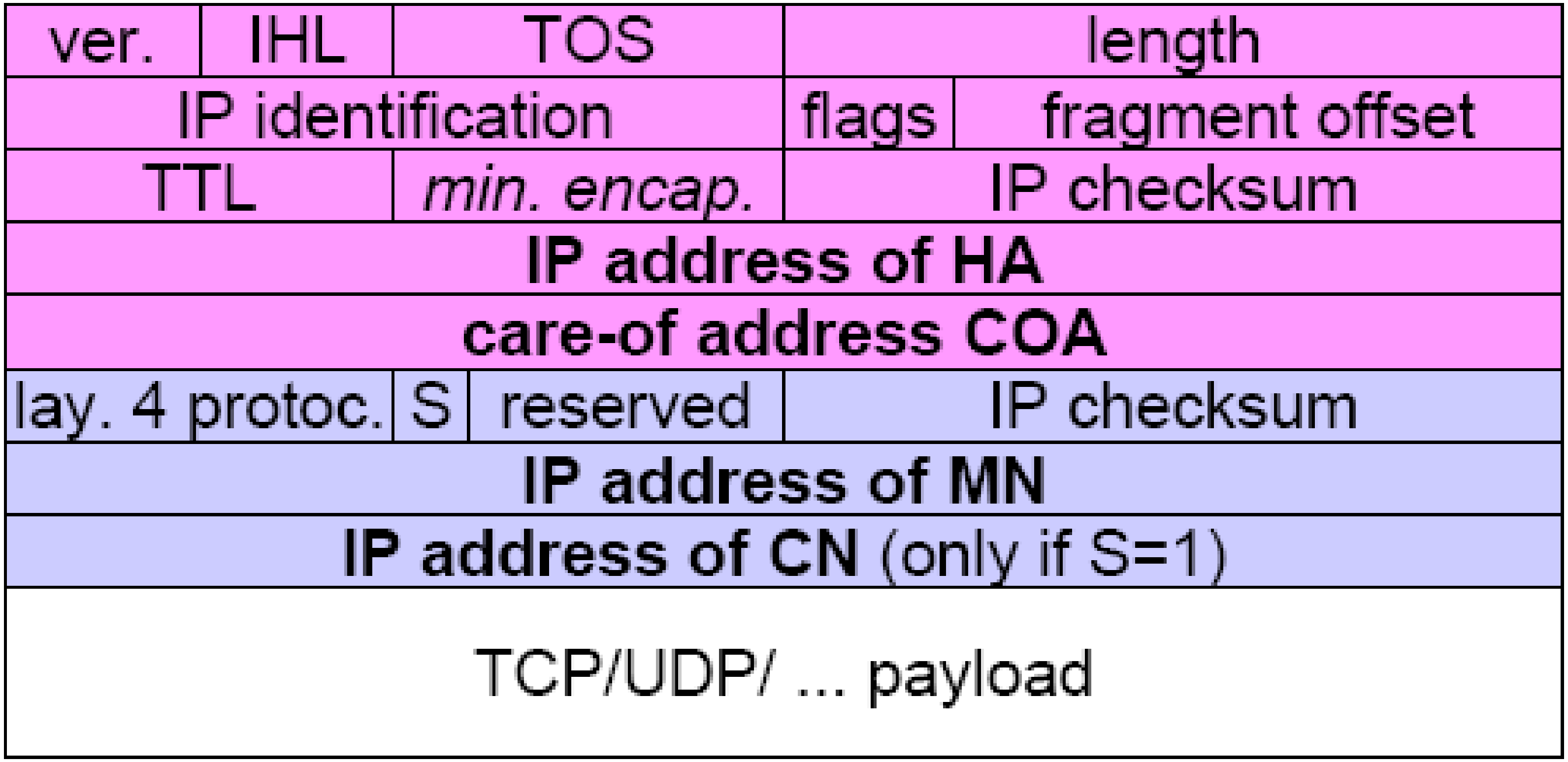
Field Description

- The outer IP header source & destination address identify the *tunnel endpoints* (i.e. HA & FA).
- Ver – is ‘4’ for IP protocol.
- TOS – now it is DS in the context of Differentiated Services.
- The inner IP header source and destination address identify the original sender & recipient
- Other headers for *authentication* might be added to outer header.
- Some outer IP header fields are copied from the inner IP fields (TOS), most are re-computed (checksum, length) based on new datagram.
- The outer TTL must be high enough so that packet can reach the tunnel end-point. The inner TTL decremented by one only, that is, whole tunnel to be considered as single hop.

Minimal Encapsulation

- We can save space by recognizing that much of the inner header can be derived from the outer header
  - Copy inner header
  - Modify protocol field to be 55, for the minimal protocol
  - Destination address replaced by tunnel exit
  - If encapsulator isn't originator of message, replace source address with address of encapsulator
  - Increment total length by the size of the additional header (either 12 or 8 octets)
  - Recompute checksum
- Sender sends all packets via HA to MN
- Higher latency and network load

Minimal Encapsulation Header



<table>
<tr><td>ver.</td><td>IHL</td><td colspan="2">TOS</td><td colspan="2">length</td></tr>
<tr><td colspan="4">IP identification</td><td>flags</td><td>fragment offset</td></tr>
<tr><td colspan="2">TTL</td><td colspan="2"><i>min. encap.</i></td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="6"><b>IP address of HA</b></td></tr>
<tr><td colspan="6"><b>care-of address COA</b></td></tr>
<tr><td colspan="2">lay. 4 protoc.</td><td>S</td><td>reserved</td><td colspan="2">IP checksum</td></tr>
<tr><td colspan="6"><b>IP address of MN</b></td></tr>
<tr><td colspan="6"><b>IP address of CN</b> (only if S=1)</td></tr>
<tr><td colspan="6">TCP/UDP/ ... payload</td></tr>
</table>

Generic Routing Encapsulation

It allows the encapsulation of packets of one protocol suite in to the payload portion of a packet of another protocol suite.

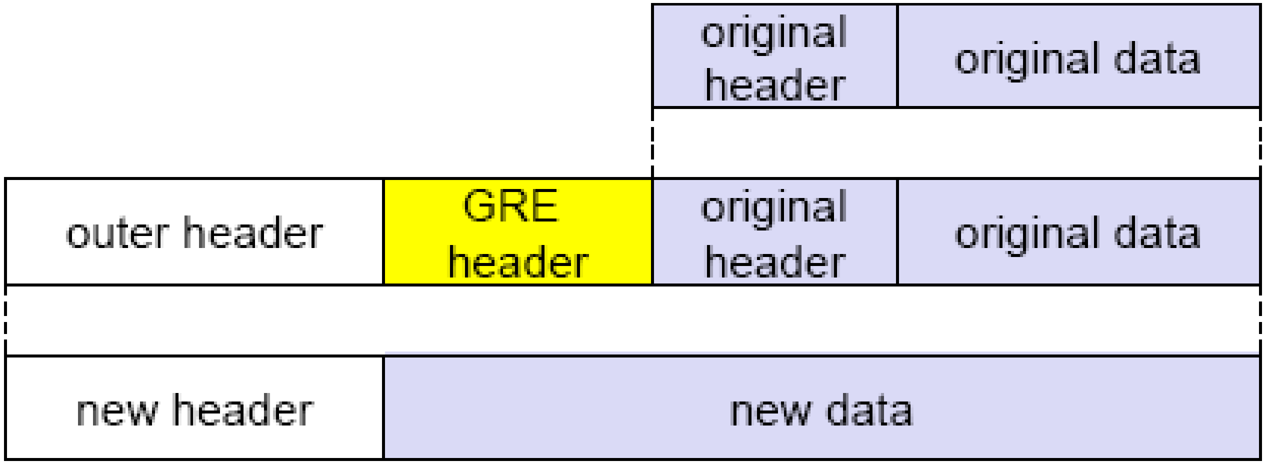
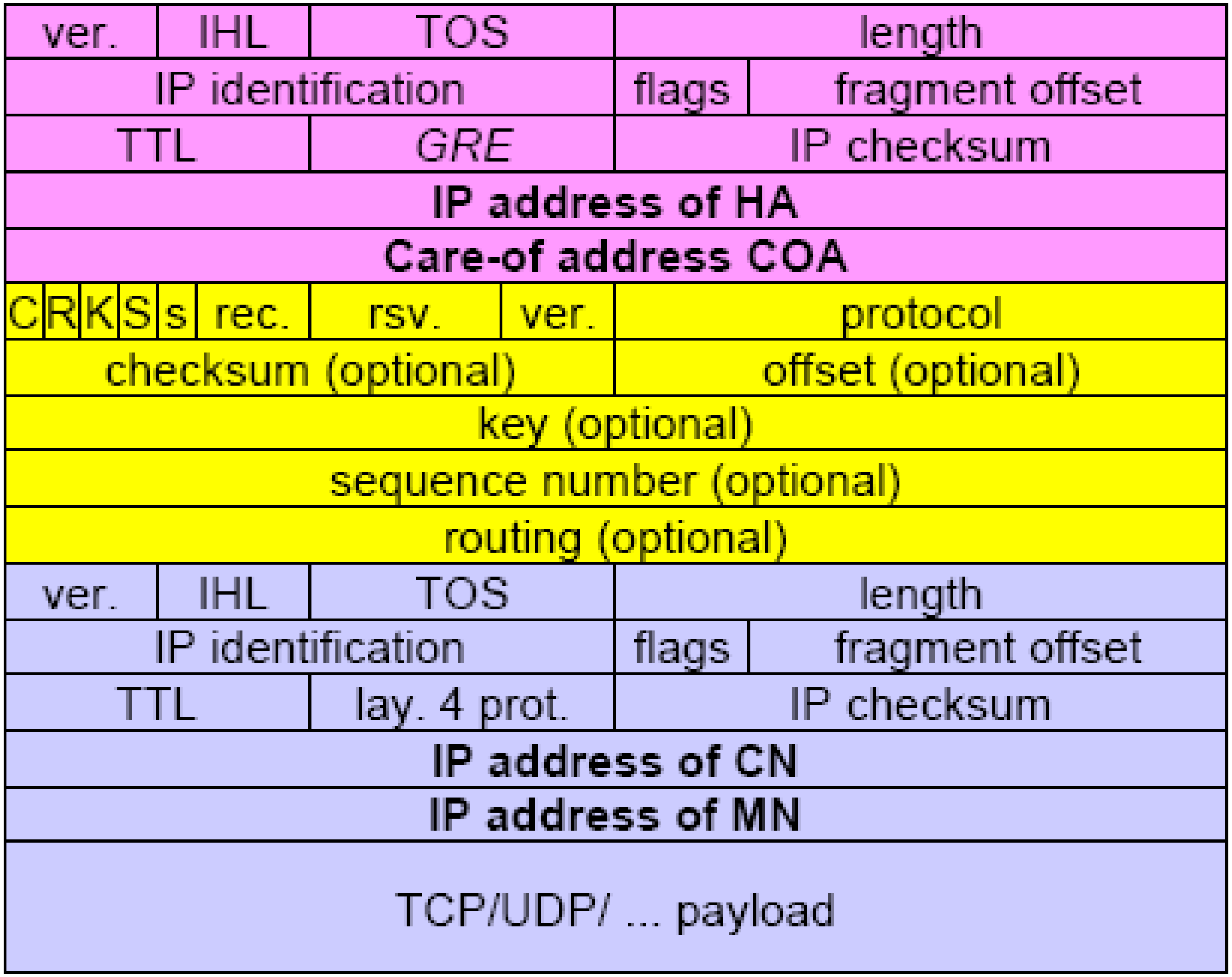


Fig: GRE

Protocol fields for GRE



ver. | IHL | TOS | length

IP identification | flags | fragment offset

TTL | *GRE* | IP checksum

IP address of HA

Care-of address COA

C | R | K | S | s | rec. | rsv. | ver. | protocol

checksum (optional) | offset (optional)

key (optional)

sequence number (optional)

routing (optional)

ver. | IHL | TOS | length

IP identification | flags | fragment offset

TTL | lay. 4 prot. | IP checksum

IP address of CN

IP address of MN

TCP/UDP/ ... payload

Optimizations

- The inefficient behavior of a non-optimized mobile IP is called Triangular Routing.
- In this all packets to through the HA. This cause unnecessary overhead between CN & HA and HA & COA/FA .
- One way to optimize the route is inform the CN of the current location of MN.
- The optimized mobile IP needs 4 additional messages.

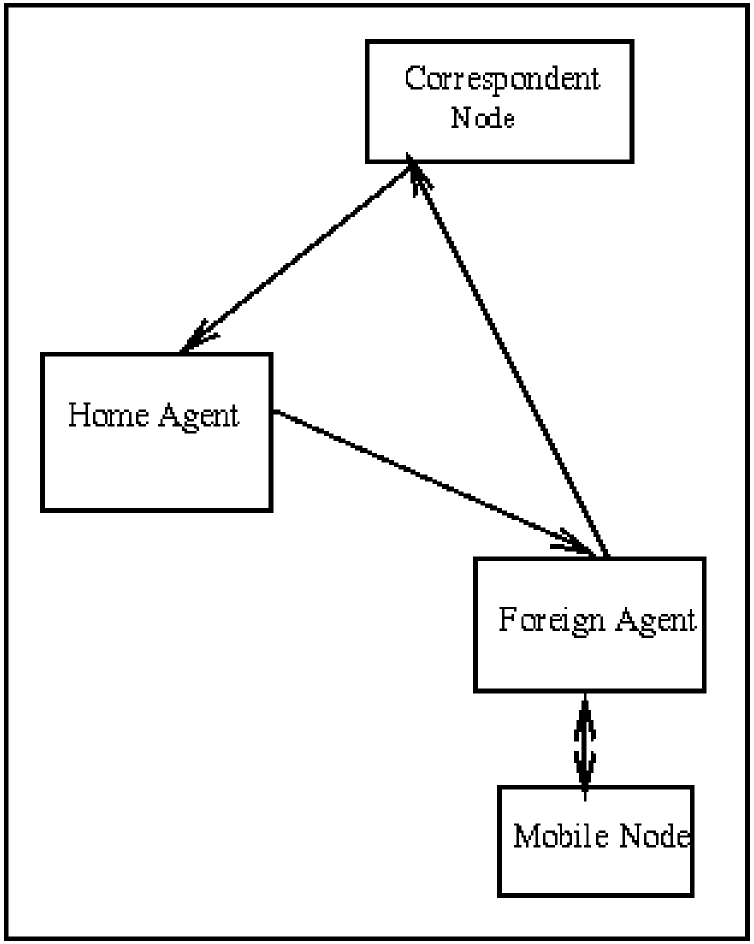
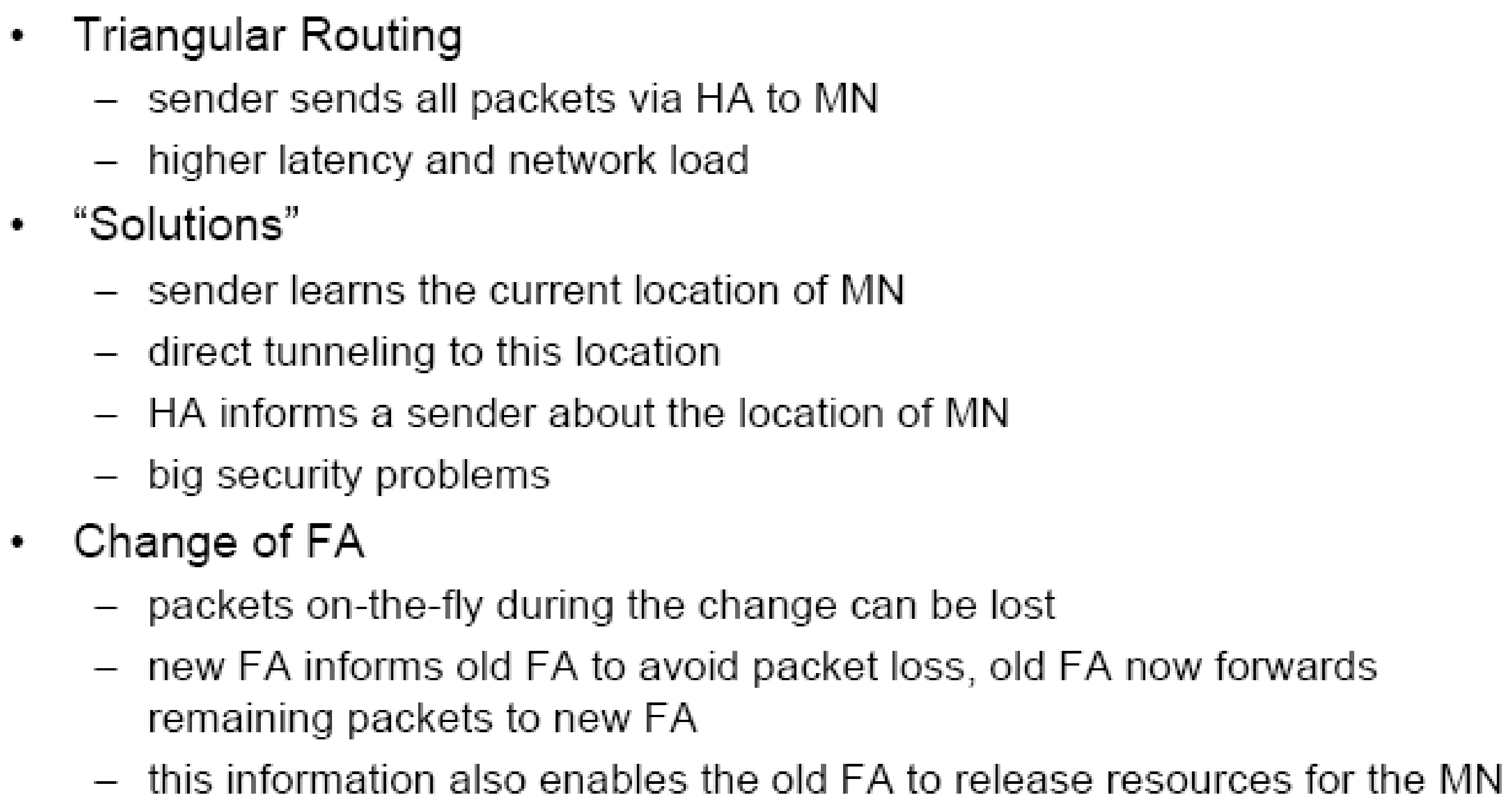


Figure 5: Triangle Routing

Four Messages

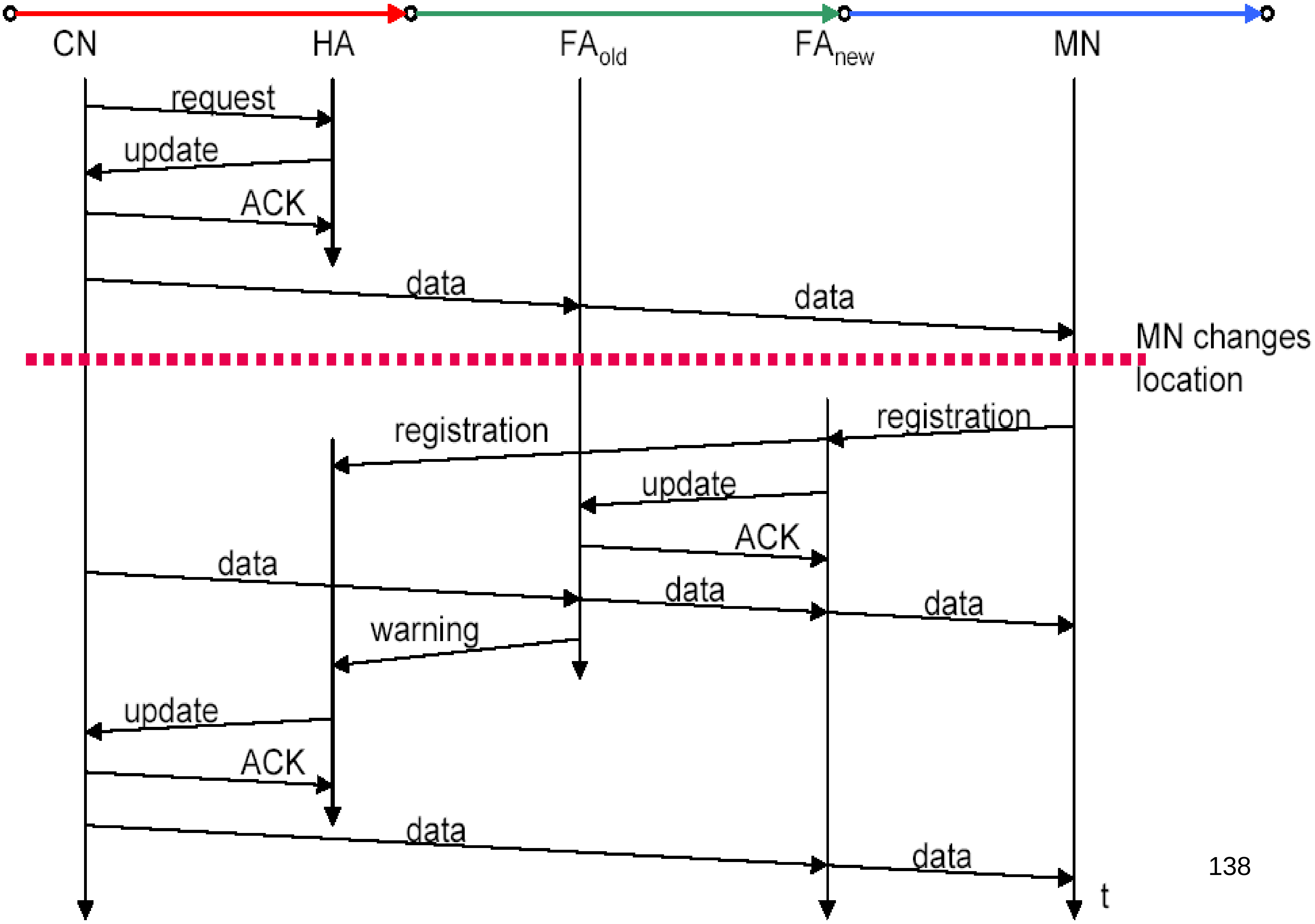
- Binding Request (BR) – Any node wants to know the current location of MN can send BR to the HA.
- Binding Update (BU) – It is sent by the HA to CN reveals the current location of MN.
- Binding ACK (BA) – If requested, a node returns this ACK after receiving a BU.
- Binding Warning (BW) – If a node decapsulates a packet for an MN, but it is not the current FA for this MN, this nods send this BW.

Change of FA with an Optimized mobile IP



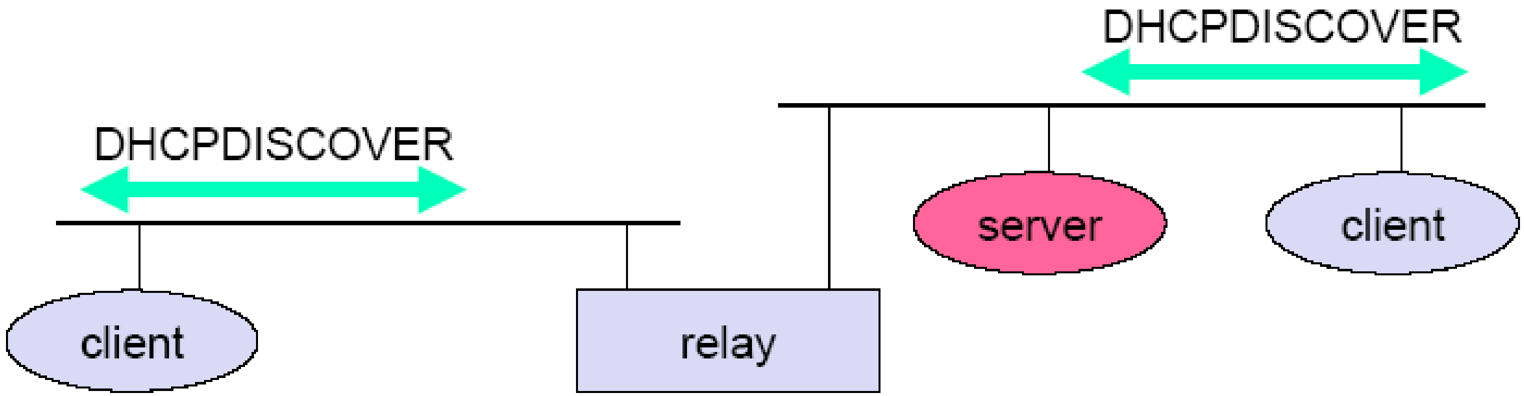
- Triangular Routing
  - sender sends all packets via HA to MN
  - higher latency and network load
- “Solutions”
  - sender learns the current location of MN
  - direct tunneling to this location
  - HA informs a sender about the location of MN
  - big security problems
- Change of FA
  - packets on-the-fly during the change can be lost
  - new FA informs old FA to avoid packet loss, old FA now forwards remaining packets to new FA
  - this information also enables the old FA to release resources for the MN

Change of foreign agent

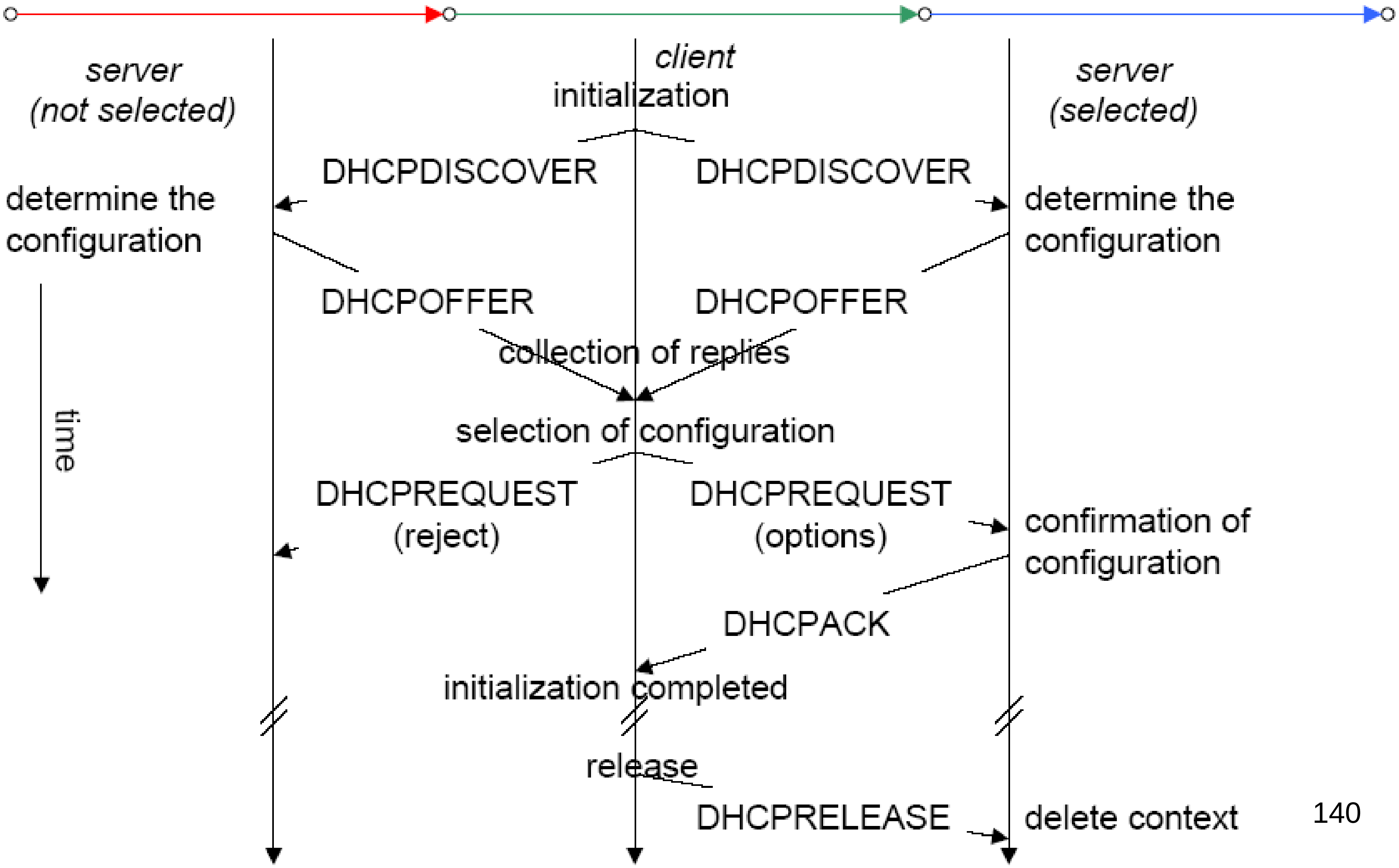


DHCP: Dynamic Host Configuration Protocol

- Application
  - simplification of installation and maintenance of networked computers
  - supplies systems with all necessary information, such as IP address, DNS server address, domain name, subnet mask, default router etc.
  - enables automatic integration of systems into an Intranet or the Internet, can be used to acquire a COA for Mobile IP
- Client/Server-Model
  - the client sends via a MAC broadcast a request to the DHCP server (might be via a DHCP relay)



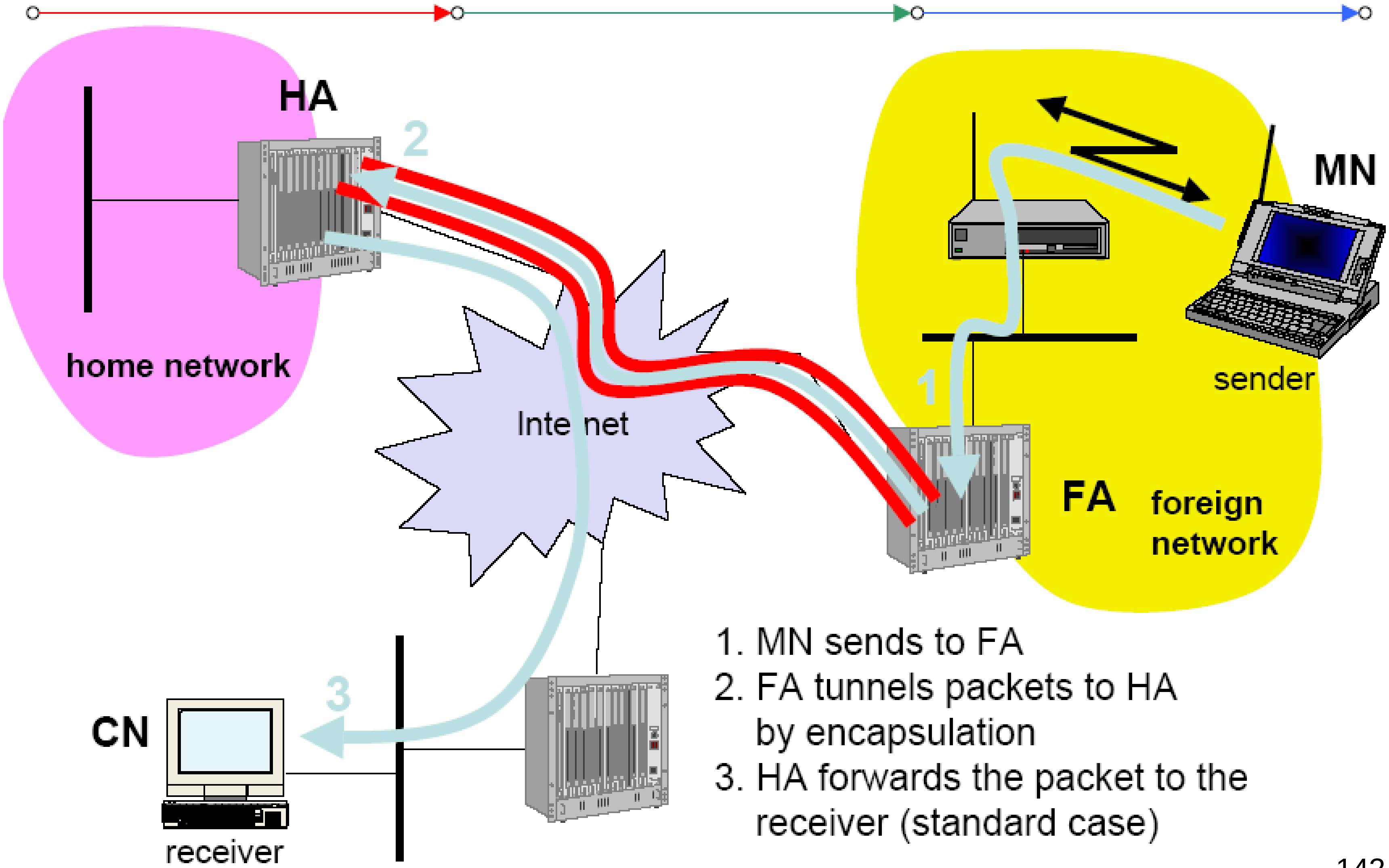
DHCP - protocol mechanisms



Mobile IP with reverse tunneling

- Router accept often only “topologically correct“ addresses (firewall!)
  - a packet from the MN encapsulated by the FA is now topologically correct
  - furthermore multicast and TTL problems solved (TTL in the home network correct, but MN is too far away from the receiver)
- Reverse tunneling does not solve
  - problems with *firewalls*, the reverse tunnel can be abused to circumvent security mechanisms (tunnel hijacking)
  - optimization of data paths, i.e. packets will be forwarded through the tunnel via the HA to a sender (double triangular routing)
- Reverse tunneling is backwards compatible
  - the extensions can be implemented easily and cooperate with current implementations without these extensions

Reverse tunneling



1. MN sends to FA
2. FA tunnels packets to HA by encapsulation
3. HA forwards the packet to the receiver (standard case)

Mobile IP and IPv6

- Mobile IP was developed for IPv4, but IPv6 simplifies the protocols
  - security is integrated and not an add-on, authentication of registration is included
  - COA can be assigned via auto-configuration (DHCPv6 is one candidate), every node has address auto-configuration
  - no need for a separate FA, **all** routers perform router advertisement which can be used instead of the special agent advertisement
  - MN can signal a sender directly the COA, sending via HA not needed in this case (automatic path optimization)
  - „soft“ hand-over, i.e. without packet loss, between two subnets is supported
    - MN sends the new COA to its old router
    - the old router encapsulates all incoming packets for the MN and forwards them to the new COA
    - authentication is always granted

UNIT-III
Mobile Transport Layer

CONTENTS

- ❑ Traditional TCP
- ❑ Indirect TCP
- ❑ Snooping TCP
- ❑ Mobile TCP

- ❑ Fast retransmit/recovery
- ❑ Transmission freezing
- ❑ Selective retransmission
- ❑ Transaction oriented TCP

- ❑ Provide mobility support to the applications at transport Layer.
- ❑ Keep original concepts while doing TCP improvements for mobile and wireless networks.
- ❑ Increase TCP’s performance.

Introduction

- ❑ Supporting mobility up to the network layer is not enough to support mobility support for applications.
- ❑ Most application rely on a transport layer, such as TCP in case of internet.
- ❑ Two functions of the transport layer in the internet are
  - Checksumming user data
  - Mux/demux of data to/from application
- ❑ While network layer only addresses a host, ports in UDP/TCP allow dedicated services to be addressed.
- ❑ The connectionless UDP does not offer much more than TCP, so, this unit concentrates only on TCP.
- ❑ Mobility Support in IP (like Mobile IP) is already enough for UDP to work.
- ❑ TCP has built-in mechanism to behave in a ‘network friendly’ manner.

Traditional TCP

Transport protocols typically designed for

- Fixed end-systems
- Fixed, wired networks

Research activities

- Performance
- Congestion control
- Efficient retransmissions

Congestion control

- packet loss in fixed networks typically due to (temporary) overload situations
- router have to discard packets as soon as the buffers are full
- TCP recognizes congestion only indirect via missing acknowledgements, retransmissions unwise, they would only contribute to the congestion and make it even worse
- slow-start algorithm as reaction

Traditional TCP (2)

Slow-start

- sender calculates a congestion window for a receiver
- start with a congestion window size equal to one segment
- exponential increase of the congestion window up to the congestion threshold, then linear increase
- missing acknowledgement causes the reduction of the congestion threshold to one half of the current congestion window
- congestion window starts again with one segment

Fast retransmit/fast recovery

- TCP sends an acknowledgement only after receiving a packet
- if a sender receives several acknowledgements for the same packet, this is due to a gap in received packets at the receiver
- however, the receiver got all packets up to the gap and is actually receiving packets
- therefore, packet loss is not due to congestion, continue with current congestion window (do not use slow-start)

Implications on mobility

TCP assumes congestion if packets are dropped

- typically wrong in wireless networks, here we often have packet loss due to *transmission errors*
- furthermore, *mobility* itself can cause packet loss, if e.g. a mobile node roams from one access point (e.g. foreign agent in Mobile IP) to another while there are still packets in transit to the wrong access point and forwarding is not possible

The performance of an unchanged TCP degrades severely

- however, TCP cannot be changed fundamentally due to the large base of installation in the fixed network, TCP for mobility has to remain compatible
- the basic TCP mechanisms keep the whole Internet together

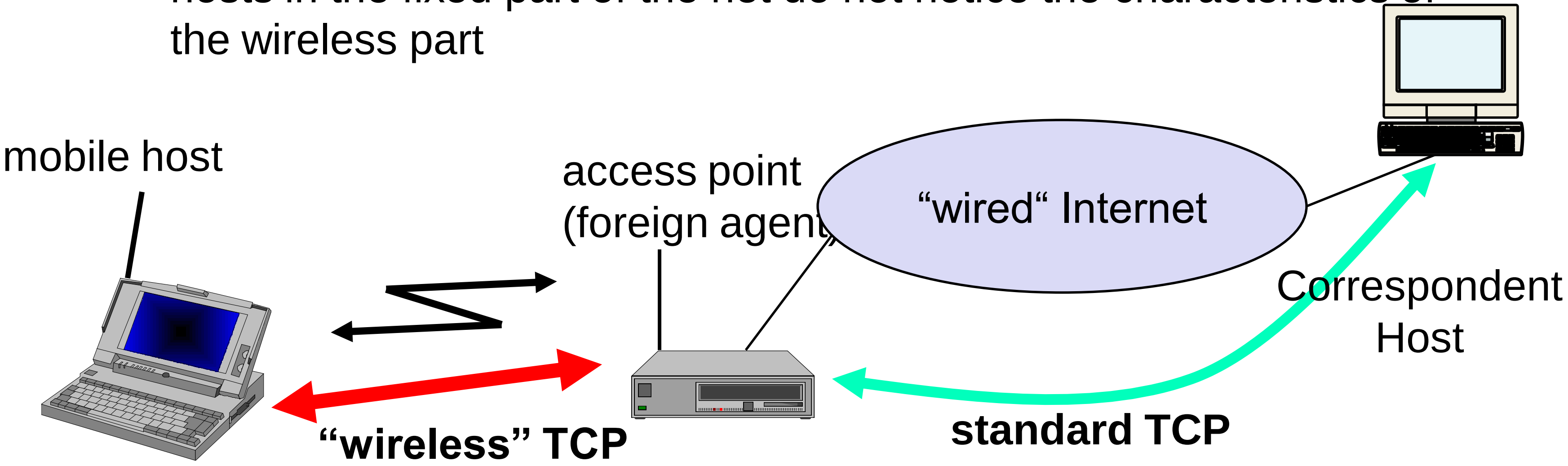
Classical TCP Improvements

- Indirect TCP (I-TCP)
- Snooping TCP
- Mobile TCP (M-TCP)
- Fast retransmit/fast recovery
- Transmission/time-out freezing
- Selective retransmission
- Transaction-oriented TCP (T/TCP)

Indirect TCP (I-TCP)

Indirect TCP segments the connection

- no changes to the TCP protocol for hosts connected to the wired Internet, millions of computers use (variants of) this protocol
- optimized TCP protocol for mobile hosts
- splitting of the TCP connection at, e.g., the foreign agent into 2 TCP connections, no real end-to-end connection any longer
- hosts in the fixed part of the net do not notice the characteristics of the wireless part



Indirect TCP (I-TCP) [2]

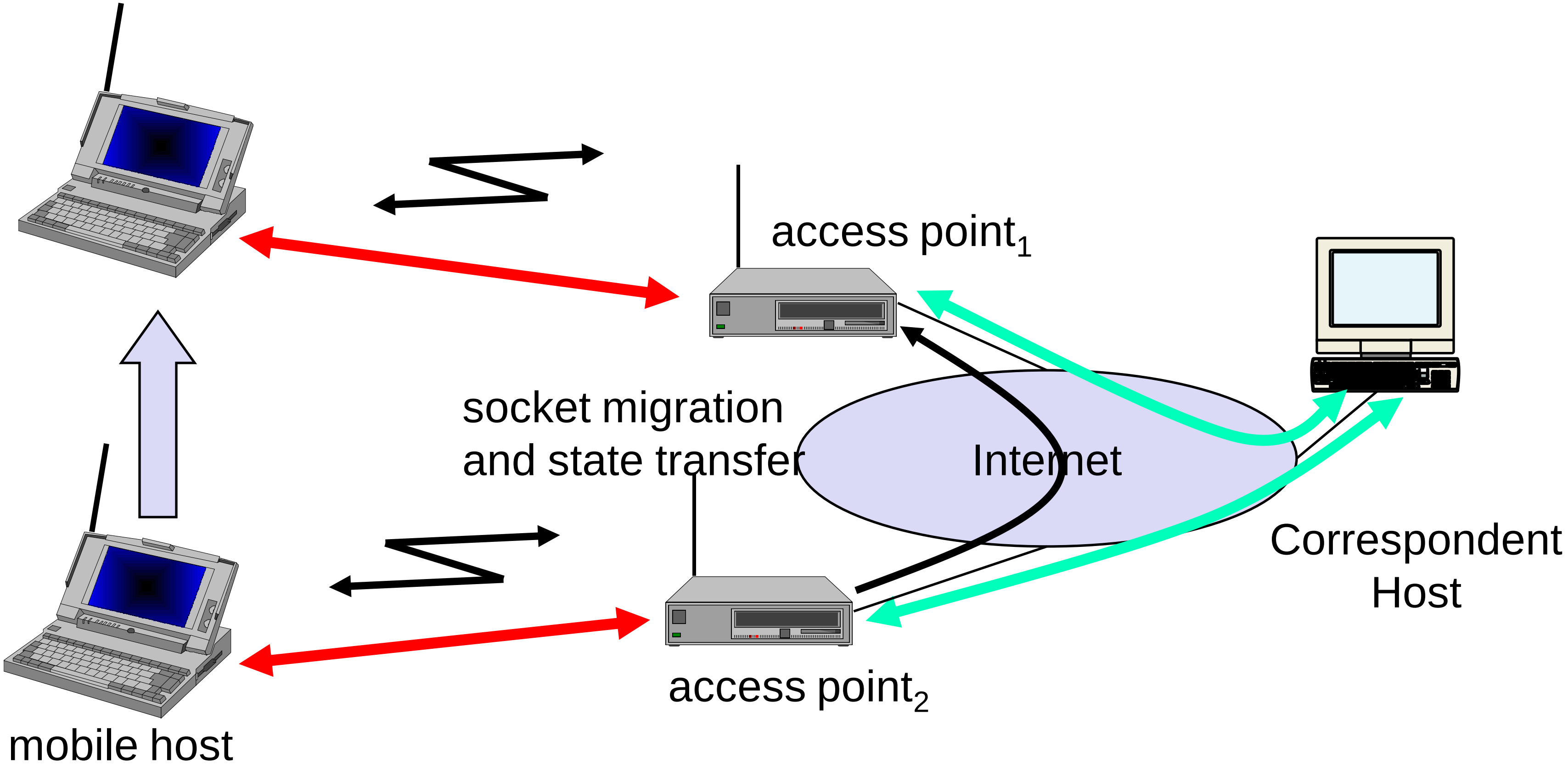


Figure: I-TCP socket and state migration

Indirect TCP or I-TCP [3]

Advantages

- no changes in the fixed network necessary, no changes for the hosts (TCP protocol) necessary, all current optimizations to TCP still work
- transmission errors on the wireless link do not propagate into the fixed network
- simple to control, mobile TCP is used only for one hop between, e.g., a foreign agent and mobile host
- therefore, a very fast retransmission of packets is possible, the short delay on the mobile hop is known

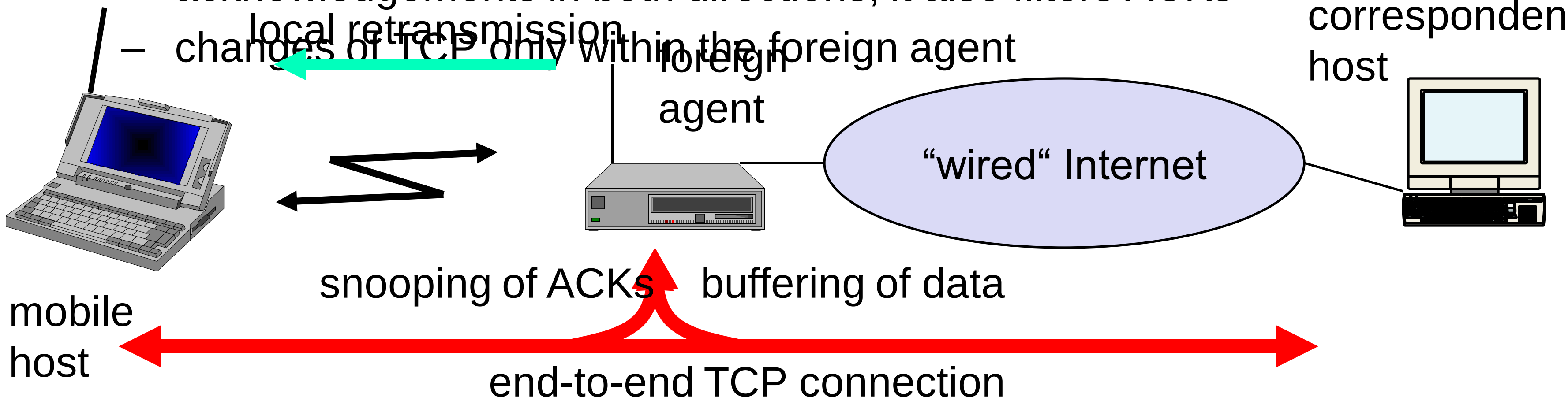
Disadvantages

- loss of end-to-end semantics, an acknowledgement to a sender does now not any longer mean that a receiver really got a packet, foreign agents might crash
- higher latency possible due to buffering of data within the foreign agent and forwarding to a new foreign agent

Snooping TCP

“Transparent” extension of TCP within the foreign agent

- buffering of packets sent to the mobile host
- lost packets on the wireless link (both directions!) will be retransmitted immediately by the mobile host or foreign agent, respectively (so called “local” retransmission)
- the foreign agent therefore “snoops” the packet flow and recognizes acknowledgements in both directions, it also filters ACKs
- changes of TCP only within the foreign agent



Snooping TCP [2]

Data transfer to the mobile host

- FA buffers data until it receives ACK of the MH, FA detects packet loss via duplicated ACKs or time-out
- fast retransmission possible, transparent for the fixed network

Data transfer from the mobile host

- FA detects packet loss on the wireless link via sequence numbers, FA answers directly with a NACK to the MH
- MH can now retransmit data with only a very short delay

Integration of the MAC layer

- MAC layer often has similar mechanisms to those of TCP
- thus, the MAC layer can already detect duplicated packets due to retransmissions and discard them

Problems

- snooping TCP does not isolate the wireless link as good as I-TCP
- snooping might be useless depending on encryption schemes

Mobile TCP (M-TCP)

- Special handling of lengthy and/or frequent disconnections
- M-TCP splits as I-TCP does
  - unmodified TCP fixed network to supervisory host (SH)
  - optimized TCP SH to MH
- Supervisory host
  - no caching, no retransmission
  - monitors all packets, if disconnection detected
    - set sender window size to 0
    - sender automatically goes into persistent mode
  - old or new SH reopen the window
- Advantages
  - maintains semantics, supports disconnection, no buffer forwarding
- Disadvantages

Fast retransmit/fast recovery

Change of foreign agent often results in packet loss

- TCP reacts with slow-start although there is no congestion

Forced fast retransmit

- as soon as the mobile host has registered with a new foreign agent, the MH sends duplicated acknowledgements on purpose
- this forces the fast retransmit mode at the communication partners
- additionally, the TCP on the MH is forced to continue sending with the actual window size and not to go into slow-start after registration

Advantage

- simple changes result in significant higher performance

Disadvantage

- further mix of IP and TCP, no transparent approach

Transmission/time-out freezing

Mobile hosts can be disconnected for a longer time

- no packet exchange possible, e.g., in a tunnel, disconnection due to overloaded cells or multiplexing with higher priority traffic
- TCP disconnects after time-out completely

TCP freezing

- MAC layer is often able to detect interruption in advance
- MAC can inform TCP layer of upcoming loss of connection
- TCP stops sending, but does now not assume a congested link
- MAC layer signals again if reconnected

Advantage

- scheme is independent of data

Disadvantage

- TCP on mobile host has to be changed, mechanism depends on MAC layer

Selective retransmission

TCP acknowledgements are often cumulative

- ACK n acknowledges correct and in-sequence receipt of packets up to n
- if single packets are missing quite often a whole packet sequence beginning at the gap has to be retransmitted (go-back-n), thus wasting bandwidth

Selective retransmission as one solution

- RFC2018 allows for acknowledgements of single packets, not only acknowledgements of in-sequence packet streams without gaps
- sender can now retransmit only the missing packets

Advantage

- much higher efficiency

Disadvantage

- more complex software in a receiver, more buffer needed at the

Transaction oriented TCP (T/TCP)

TCP phases

- connection setup, data transmission, connection release
- using 3-way-handshake needs 3 packets for setup and release, respectively
- thus, even short messages (one byte) need a minimum of 7 packets!

Transaction oriented TCP

- RFC1644, T-TCP, describes a TCP version to avoid this overhead
- connection setup, data transfer and connection release can be combined
- thus, only 2 or 3 packets are needed

Advantage

- efficiency

Disadvantage

- requires changed TCP
- mobility not longer transparent

Transaction oriented TCP [2]

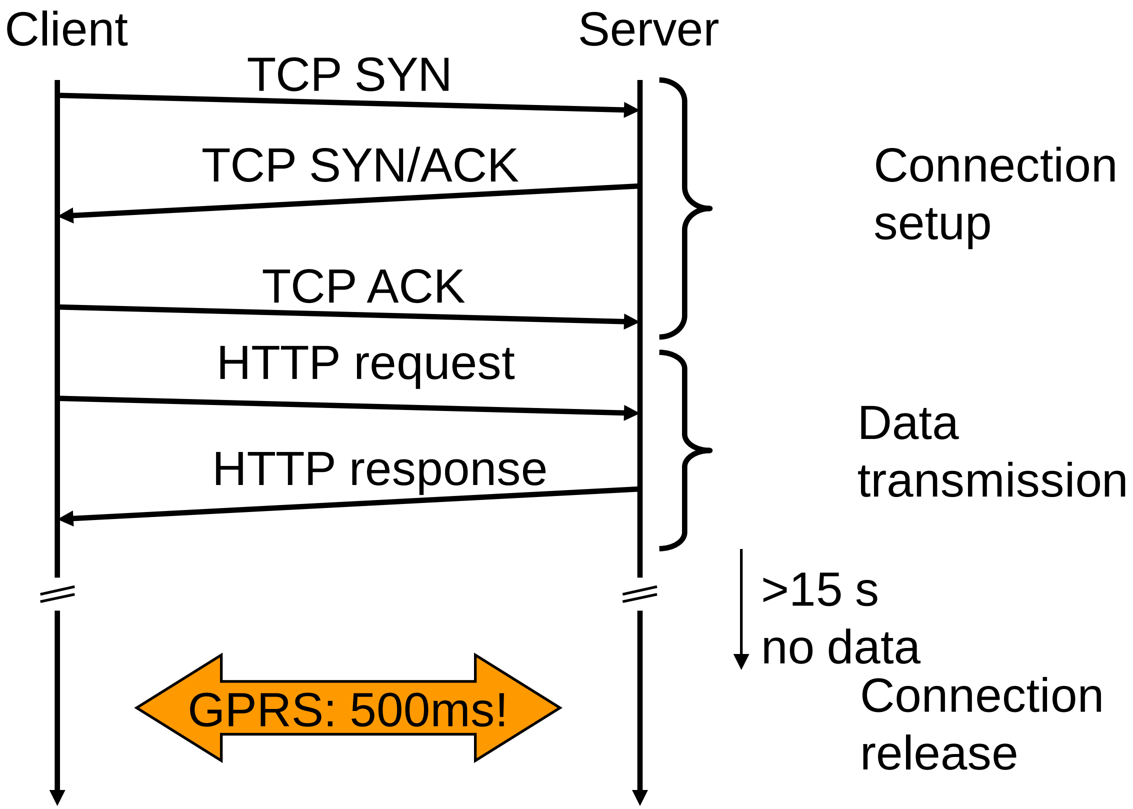


Figure: Example TCP Connection setup overhead

Comparison of different approaches for a "mobile" TCP

Approach	Mechanism	Advantages	Disadvantages
Indirect TCP	splits TCP connection into two connections	isolation of wireless link, simple	loss of TCP semantics, higher latency at handover
Snooping TCP	"snoops" data and acknowledgements, local retransmission	transparent for end-to-end connection, MAC integration possible	problematic with encryption, bad isolation of wireless link
M-TCP	splits TCP connection, chokes sender via window size	Maintains end-to-end semantics, handles long term and frequent disconnections	Bad isolation of wireless link, processing overhead due to bandwidth management
Fast retransmit/ fast recovery	avoids slow-start after roaming	simple and efficient	mixed layers, not transparent
Transmission/ time-out freezing	freezes TCP state at disconnect, resumes after reconnection	independent of content or encryption, works for longer interrupts	changes in TCP required, MAC dependant
Selective retransmission	retransmit only lost data	very efficient	slightly more complex receiver software, more buffer needed
Transaction oriented TCP	combine connection setup/release and data transmission	Efficient for certain applications	changes in TCP required, not transparent

Database Hoarding Techniques

A mobile device cannot store a large database due to memory constraints.

The large databases are available on the servers, remote computing systems, or networks.

Retrieving the required data from a database server during every computation- impractical due to time constraints

A mobile device is not always connected to the server or network; neither does the device retrieve data from a server or a network for each computation.

Database Hoarding Techniques

Rather, the device caches some specific data, which may be required for future computations, during the interval in which the device is connected to the server or network.

Cache is a list or database consisting of saved items or records at the device, which the device saves for faster access at a later time, rather than reselecting or re-tuning or re-fetching the data when required.

The cached items are the ones which the device fetches by request , demand or subscription to the server and saves in the list or database.

Caching entails saving a copy of select data or a part of a database from a connected system with a large database.

The cached data is hoarded in the mobile device database as and when the device connects to the server or network

Hoarding of the cached data in the database ensure that even when the device is not connected to a network or server(disconnected mode), the data required from the database is available instantly for computing

Example-caching and hoarding:

Consider the train schedules in a railway timetable.

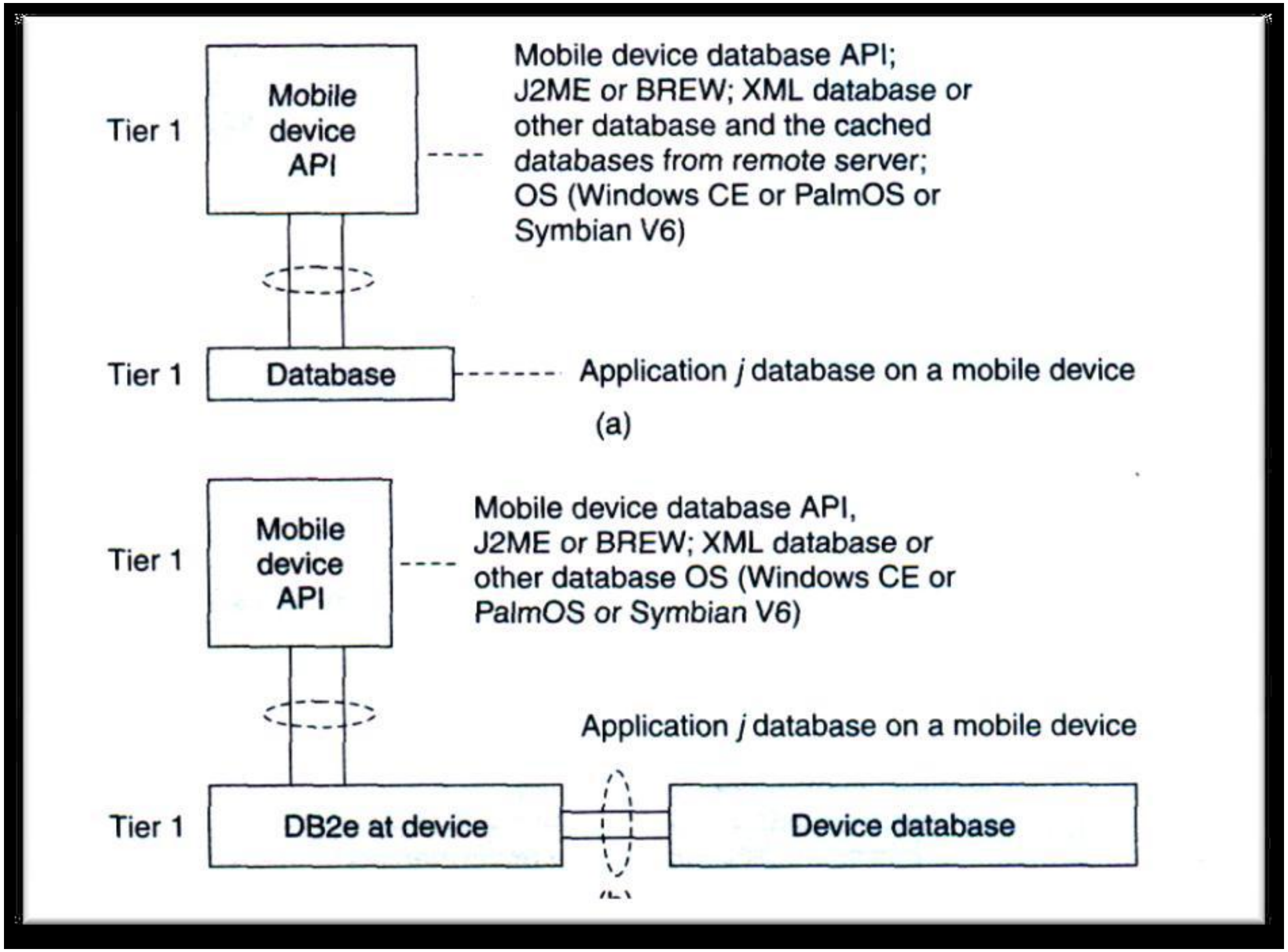
The schedule of specific trains to and from the device user location is stored in the huge databases of the servers and networked of the company operating the trains.

The user device caches and hoards some specific information from the database and the hoarded device-database is used during computations from retrieving the data for a specific set of train schedules

Database Hoarding

Database hoarding may be done at the application tier itself. The following figure shows a simple architecture in which a mobile device API directly retrieves the data from a database.

It also shows another simple architecture in which a mobile device API directly retrieves the data from a database through a program, for ex: IBM DB2 Everyplace (DB2e).


Tier 1
Mobile device API
Mobile device database API; J2ME or BREW; XML database or other database and the cached databases from remote server; OS (Windows CE or PalmOS or Symbian V6)
Tier 1
Database
Application j database on a mobile device
(a)
Tier 1
Mobile device API
Mobile device database API, J2ME or BREW; XML database or other database OS (Windows CE or PalmOS or Symbian V6)
Application j database on a mobile device
Tier 1
DB2e at device
Device database

(a) API at mobile device sending queries and retrieving data from local database (Tier 1) (b) API at mobile device retrieving data from database using DB2e (Tier 1)

Database Hoarding

Both the two architectures belong to the class of one-tier database architecture.

Some examples are downloaded ringtones, music etc. **IBM DB2 Everyplace (DB2e) is a relational database engine which has been designed to reside at the device.**

It supports J2ME and most mobile device operating systems.

Hoarding multi-tier database

In two-tier or multi-tier databases the databases reside in the remote servers and the copies of these databases are cached at the client tiers.

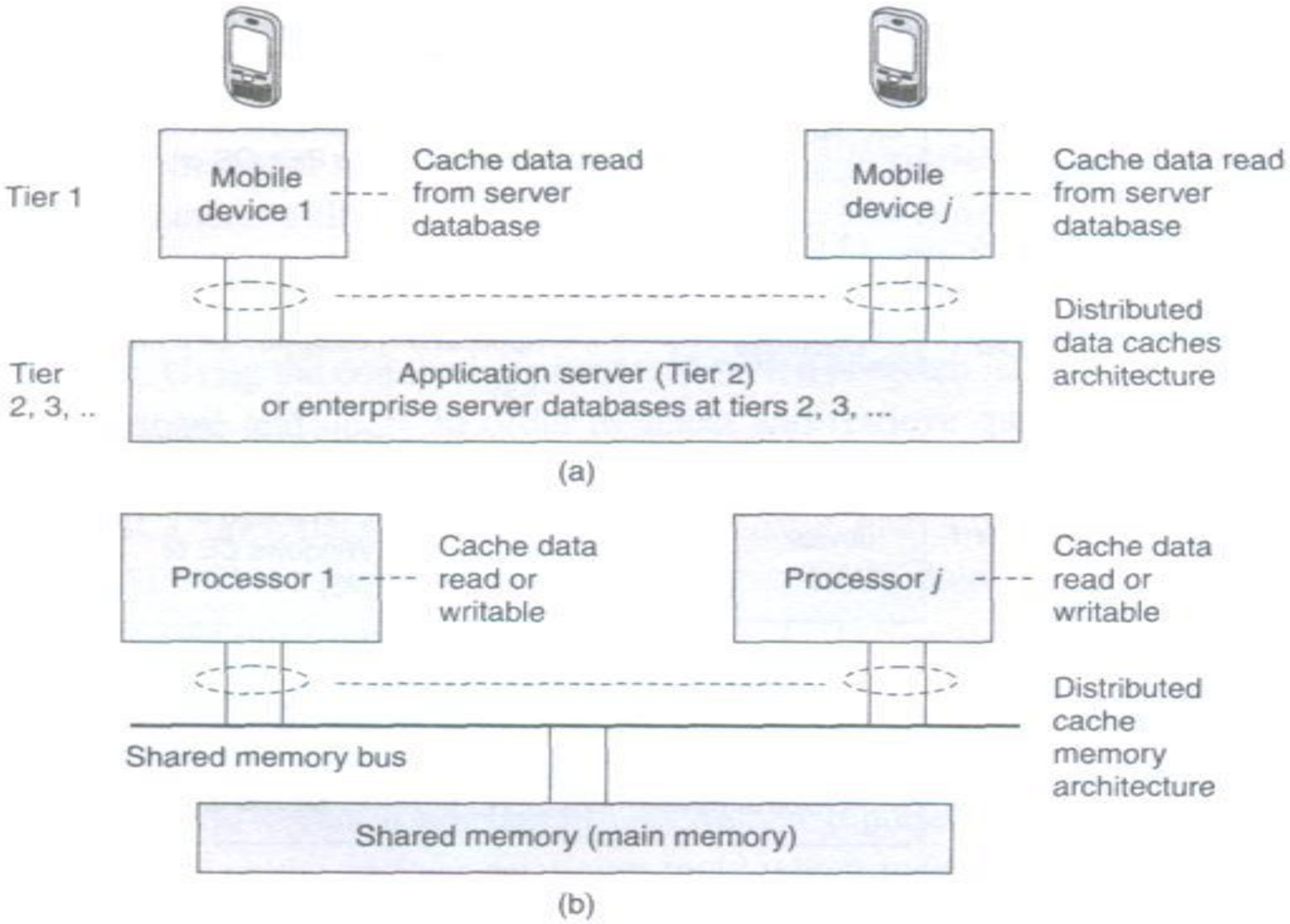
This type of architecture is known as client-server computing architecture.

Cache is a list or database consisting of saved items or records at the device.

The device may select and save the item from a set of records broadcasted or pushed by a server or the device may access (by request ,demand, or subscription) the item at the server and and save it to the list or database

Databases are hoarded at the application or enterprise tier, where the database server uses business logic and connectivity for retrieving the data and then transmitting it to the device.

The server provides and updates local copies of the database at each mobile device connected to it


Tier 1
Mobile device 1
Cache data read from server database
Mobile device *j*
Cache data read from server database
Distributed data caches architecture
Tier 2, 3, ..
Application server (Tier 2) or enterprise server databases at tiers 2, 3, ...
(a)
Processor 1
Cache data read or writable
Processor *j*
Cache data read or writable
Distributed cache memory architecture
Shared memory bus
Shared memory (main memory)
(b)

(a) Distributed data caches in mobile devices (b) Similar architecture for a distributed cache memory in multiprocessor systems

The computing API at the mobile device (first tier) uses the cached local copy.

At first tier (tier 1), the API uses the cached data records using the computing architecture.

From tier 2 or tier 3, the server retrieves and transmits the data records to tier 1 using business logic and synchronizes the local copies at the device.

Advantage of hoarding

The advantage of hoarding is that there is no access latency (delay in retrieving the queried record from the server over wireless mobile networks).

The client device API has instantaneous data access to hoarded or cached data.

After a device caches the data distributed by the server, the data is hoarded at the device.

The disadvantage of hoarding is that the consistency of the cached data with the database at the server needs to be maintained.

Data Caching

Hoarded copies of the databases at the servers are distributed or transmitted to the mobile devices from the enterprise servers or application databases.

The copies cached at the devices are equivalent to the cache memories at the processors in a multiprocessor system with a shared main memory and copies of the main memory data stored at different locations.

Data Caching cotd..

Cache Access Protocols: A client device caches the pushed (disseminated) data records from a server.

Caching of the pushed data leads to a reduced access interval as compared to the pull (on-demand) mode of data fetching.

Caching of data records can be-based on pushed 'hot records' (the most needed database records at the client device).

Also, caching can be based on the ratio of two parameters—access probability (at the device) and pushing rates (from the server) for each record.

This method is called cost-based data replacement or caching

Data Caching cotd..

Pre-fetching: Pre-fetching is another alternative to caching of disseminated data.

The process of pre-fetching entails requesting for and pulling records that may be required later.

The client device can pre-fetch instead of caching from the pushed records keeping future needs in view.

Pre-fetching reduces server load. Further, the cost of cache-misses can thus be reduced.

The term 'cost of cache-misses' refers to the time taken in accessing a record at the server in case that record is not found in the device database when required by the device API.

Caching Invalidation Mechanisms

A cached record at the client device may be invalidated.

This may be due to expiry or modification of the record at the database server.

Cache invalidation is a process by which a cached data item or record becomes invalid and thus unusable because of modification, expiry, or invalidation at another computing system or server.

Cache invalidation mechanisms are used to synchronize the data at other processors whenever the cache-data is written (modified) by a processor in a multiprocessor system.

Caching Invalidation Mechanisms contd…

A cache consists of several records. Each record is called a cache-line, copies of which can be stored at other devices or servers.

The cache at the mobile devices or server databases at any given time can be assigned one of four possible tags indicating its state

—modified (after rewriting), exclusive, shared, and invalidated (after expiry or when new data becomes available) at any given instance.

These four states are indicated by the letters M, E, S, and I, respectively (MESI).

The states indicated by the various tags are as follows:

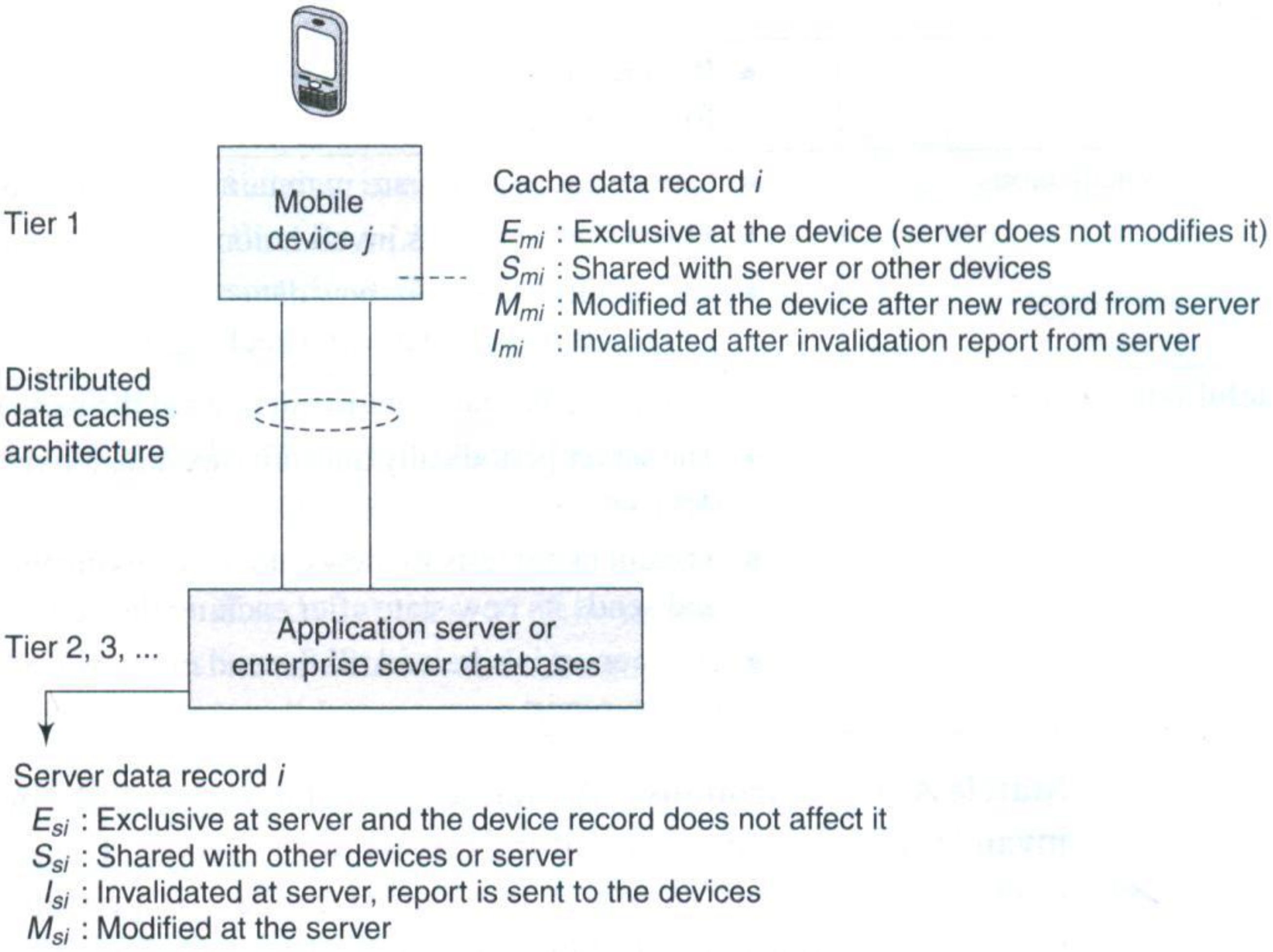
a) The E tag indicates the exclusive state which means that the data record is for internal use and cannot be used by any other device or server.

b) The S tag indicates the shared state which indicates that the data record can be used by others.

c) The M tag indicates the modified state which means that the device cache

d) The I tag indicates the invalidated state which means that the server database no longer has a copy of the record which was shared and used for computations earlier.

The following figure shows the four possible states of a data record i at any instant in the server database and its copy at the cache of the mobile device j.



Tier 1
Mobile device *j*
Distributed data caches architecture
Tier 2, 3, ...
Application server or enterprise sever databases

Cache data record *i*

E_{mi} : Exclusive at the device (server does not modifies it)
S_{mi} : Shared with server or other devices
M_{mi} : Modified at the device after new record from server
I_{mi} : Invalidated after invalidation report from server

Server data record *i*

E_{si} : Exclusive at server and the device record does not affect it
S_{si} : Shared with other devices or server
I_{si} : Invalidated at server, report is sent to the devices
M_{si} : Modified at the server

Four possible states (M, E, S, or /) of a data record /at any instance at the server database and device j cache

Cache consistency

Another important factor for cache maintenance in a mobile environment is cache consistency (also called cache coherence).

This requires a mechanism to ensure that a database record is identical at the server as well as at the device caches and that only the valid cache records are used for computations.

Data Cache Maintenance in Mobile Environment

Device needs a data-record while running an application.

A request must be sent to the server for the data record(this mechanism is called pulling)

The time taken for the application software to access a particular record is known as access latency.

Caching and hoarding the record at the device reduces access latency to zero.

Data cache maintenance is necessary in a mobile environment to overcome access latency.

Data cache inconsistency means that data records cached for applications are not invalidated at the device when modified at the server but not modified at the device

Data cache consistency can be maintained by the three methods:

Cache invalidation mechanism(server-initiated case)

Polling mechanism(client-initiated case)

Time-to-live mechanism(client-initiated case)

Cache invalidation mechanisms

Cache invalidation mechanisms in mobile devices are triggered or initiated by the server.

There are four possible invalidation mechanisms –

Stateless asynchronous,

stateless synchronous,

stateful asynchronous and

stateful synchronous.

Stateless Asynchronous:

Stateless Asynchronous: A stateless mechanism entails broadcasting of the invalidation of the cache to all the clients of the server.

The server does not keep track of the records stored at the device caches.

It just uniformly broadcasts invalidation reports to all clients irrespective of whether the device cache holds that particular record or not.

The term 'asynchronous' indicates that the invalidation information for an item is sent as soon as its value changes.

The server advertises the invalidation information only. The client can either request for a modified copy of the record or cache the relevant record when data is pushed from the server.

The server advertises as and when the corresponding data-record at the server is invalidated and modified (deleted or replaced).

The advantage of the asynchronous approach is that there are no frequent, unnecessary transfers of data reports, thus making the mechanism more bandwidth efficient.

The disadvantages of this approach are—

(a) every client device gets an invalidation report, whether that client requires that copy or not and

(b) client devices presume that as long as there is no invalidation report, the copy is valid for use in computations.

Therefore, even when there is link failure, the devices may be using the invalidated data and the server is unaware of state changes at the clients after it sends the invalidation report.

Stateless Synchronous

Stateless Synchronous: This is also a stateless mode, i.e., the server has no information regarding the present state of data records at the device caches and broadcasts to all client devices.

However, unlike the asynchronous mechanism, here the server advertises invalidation information at periodic intervals as well as whenever the corresponding data-record at server is invalidated or modified.

This method ensures synchronization because even if the in-between period report is not detected by the device due to a link failure, the device expects the period-end report of invalidation and if that is not received at the end of the period, then the device sends a request for the same (deleted or replaced).

In case the client device does not get the periodic report due to link failure, it requests the server to send the report.

Polling mechanism(client-initiated case)

Polling means checking the state of the data record, from the server and determining whether the record is in the valid, invalid, modified, or exclusive state.

Each cached record copy is polled whenever required by the application software during computation

The device connects to the server and finds out whether the cached data record copy at the device has become invalid or has been modified at the server

If the record is found to be modified or invalidated, then the device requests for the modified data and replaces the earlier cached record copy.

Time-to-live mechanism(client-initiated case)

Each cached record is assigned a TTL value

The TTL assignment is adaptive for previous update intervals of that record.

The client device requests the server to check whether the cached data record is invalid or modified at the end of the TTL time.

If it is modified , then the device requests the server to replace the invalid cached record with the modified data.

when TTL is set 0, the TTL mechanism is equivalent to the polling mechanism

Client-server computing for mobile computing and adaptation

The network architecture can be designed so that a node is either client or server

A client node runs application software which depends on server nodes resources(files, databases, web pages, processor power or other devices or computers connected or networked to it)

The server node has larger resources and computing power than the client nodes.

This architecture is known as client-server computing architecture

It is different from peer-to-peer architecture, where each node on the network has similar resources and the various nodes depend on each others resources

Client-server architecture is used for mobile computing

Mobile devices function as client nodes due to their resources constraints.

Two Network Based Computing Architectures

Client-Server:

Client nodes dépends on server resources

A client requests the server for data or responses which the client then uses in computations

The client can either access the data records at the server or cache these records through broadcasts or distribution from the server

Client-server Computing

- An N-tier architecture (N = 1, 2, …)
- On the same computing system (not on a network), then the number of tiers, N = 1
- When the client and the server are on different computing systems on the network, then N = 2

Server networks or connecting to other computing systems

- if server connecting to other systems provide additional resources to the server for the client then N > 2
- N > 1 means that the client device at tier 1 connects to the server at tier 2 which, in turn, may connected to other

Application server in two-tier client–server computing architecture

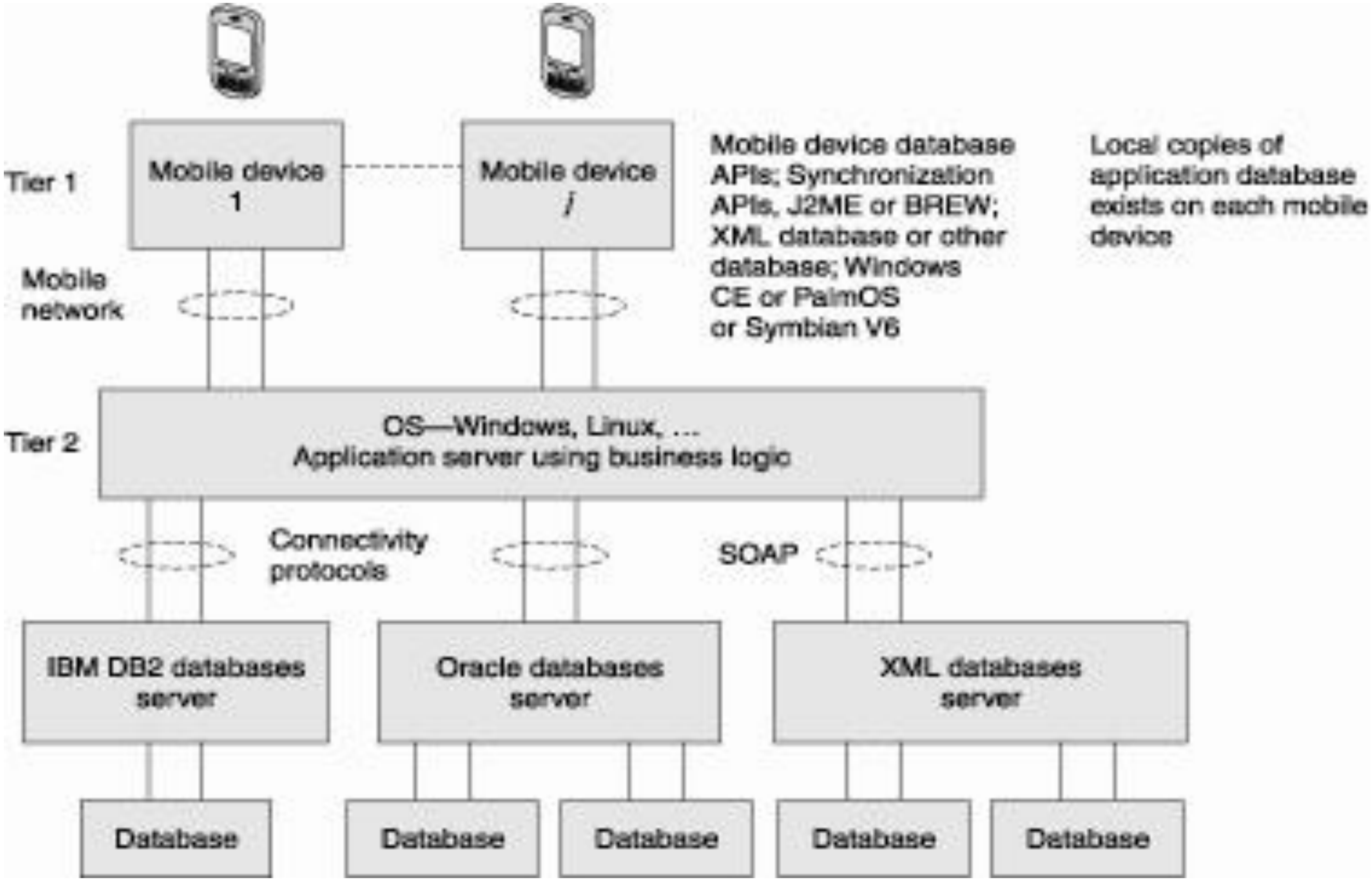
A client is a program(API) used to retrieve records from databases.

A server is a program that connects to database and sends the outputs(response) to the client

A server is defines as a computing system, which responds to request from one or more clients

A client is defined as a computing system, which request the server for a resources or for executing a task.

Two-tier Client–Server Architecture


Tier 1
Mobile device 1
Mobile device *i*
Mobile device database APIs; Synchronization APIs, J2ME or BREW; XML database or other database; Windows CE or PalmOS or Symbian V6
Local copies of application database exists on each mobile device
Mobile network
Tier 2
OS—Windows, Linux, ...
Application server using business logic
Connectivity protocols
SOAP
IBM DB2 databases server
Oracle databases server
XML databases server
Database
Database
Database
Database
Database

APIs and Synchronization API

- Various APIs synchronization with each other
- Synchronization— means that when copies at the server-end modifies, the cached copies accordingly modified
- The APIs designed independent of hardware and software platforms as far as possible as different devices may have different platforms

Three-tier Client–Server Architecture

- The application interface, the functional logic, and the database are maintained at three different layers
- The database is associated with the enterprise server tier (tier 3)
- Only local copies of the database exist at mobile devices

Three-tier Client–Server Architecture

Data records at tier 3 are sent to tier 1 through synchronization-cum-application server at tier 2.

The synchronization-cum application server has synchronization server programs, which retrieves data records from the enterprise tier (tier 3) using business logic,

The enterprise tier connects to the databases using a connectivity protocol and sends the database records as per the business logic query to tier 2

There is an in-between server, called synchronization which sends and synchronizes the copies at the multiple devices

UNIT-IV

Data Dissemination

- Ongoing advances in communications including the proliferation of internet, development of mobile and wireless networks, high bandwidth availability to homes have led to development of a wide range of new-information centered applications.
- Many of these applications involve data dissemination, i.e. delivery of data from a set of producers to a larger set of consumers.

- Data dissemination entails distributing and pushing data generated by a set of computing systems or broadcasting data from audio, video, and data services.
- The output data is sent to the mobile devices. A mobile device can select, tune and cache the required data items, which can be used for application programs.

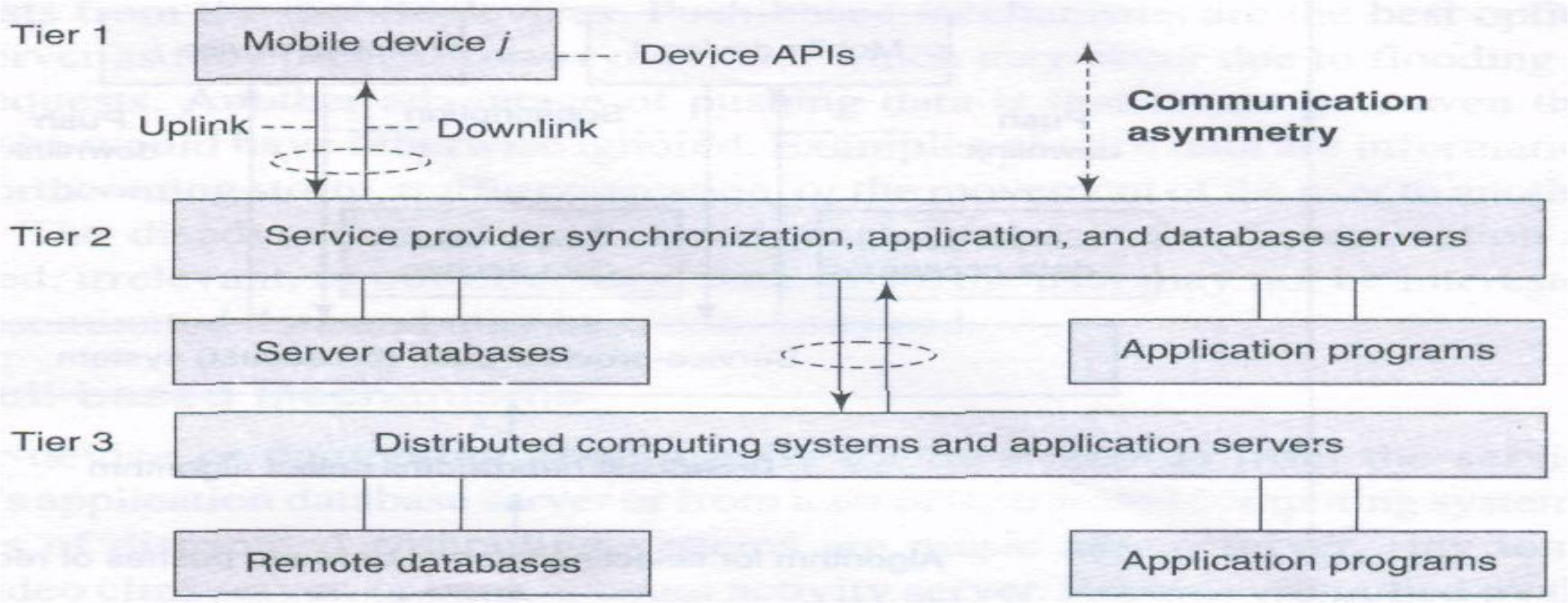
- Efficient utilization of wireless bandwidth and battery power are two of the most important problems facing software designed for mobile computing.
- Broadcast channels are attractive in tackling these two problems in wireless data dissemination.
- Data disseminated through broadcast channels can be simultaneously accessed by an arbitrary number of mobile users, thus increasing the efficiency of bandwidth usage.

Communications Asymmetry

- One key aspect of dissemination-based applications is their inherent communications asymmetry.
- That is, the communication capacity or data volume in the downstream direction (from servers-to-clients) is much greater than that in the upstream direction (from clients-to-servers).
- Content delivery is an asymmetric process regardless of whether it is performed over a symmetric channel such as the internet or over an asymmetric one, such as cable television (CATV) network.
- Techniques and system architectures that can efficiently support asymmetric applications will therefore be a requirement for future use.

- Mobile communication between a mobile device and a static computer system is intrinsically asymmetric. A device is allocated a limited bandwidth.
- This is because a large number of devices access the network. Bandwidth in the downstream from the server to the device is much larger than the one in the upstream from the device to the server.
- This is because mobile devices have limited power resources and also due to the fact that faster data transmission rates for long intervals of time need greater power dissipation from the devices.
- In GSM networks data transmission rates go up to a maximum of 14.4 kbps for both uplink and downlink.
- The communication is symmetric and this symmetry can be maintained because GSM is only used for voice communication.

Data Dissemination



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

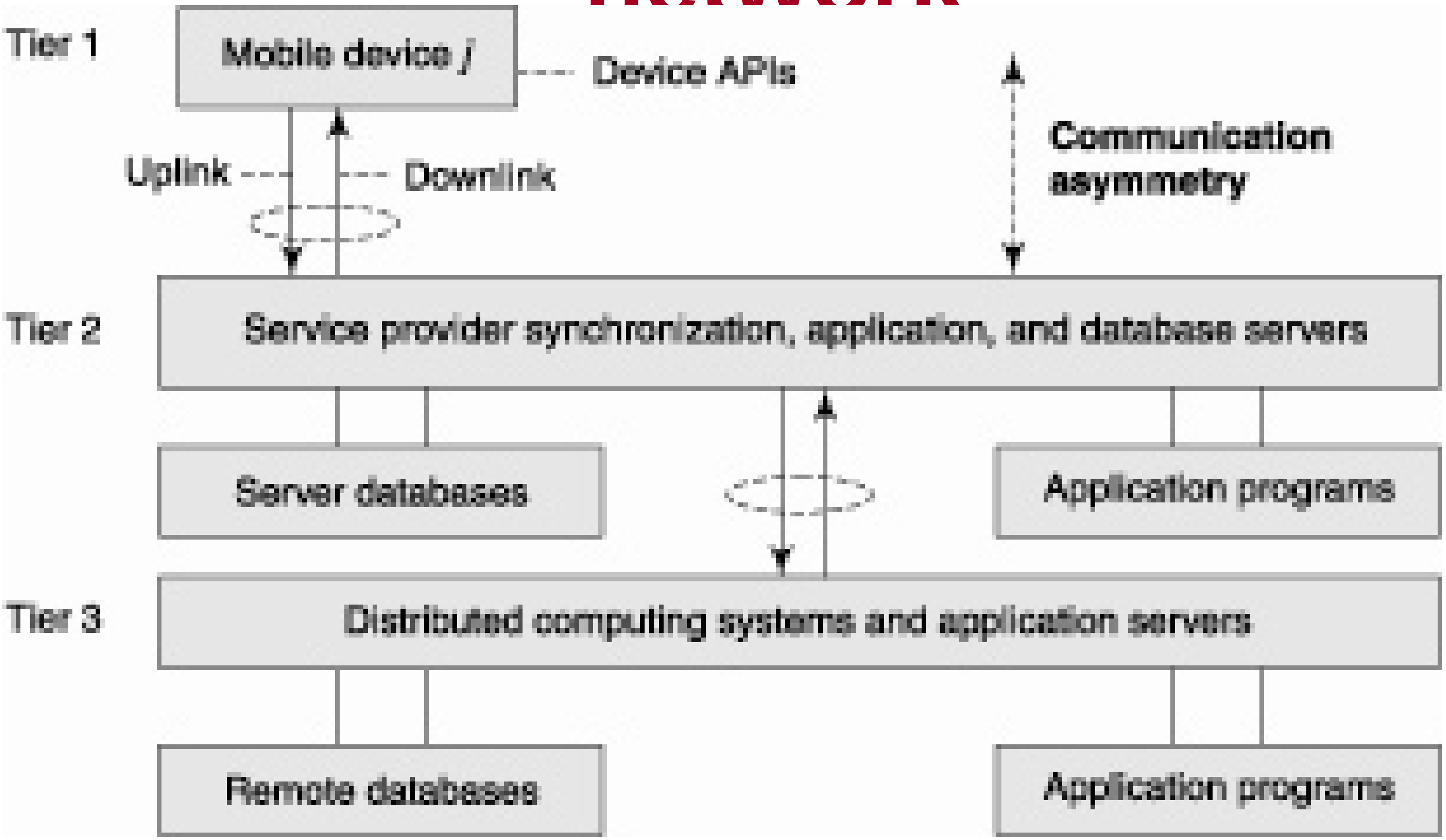
The above figure shows communication asymmetry in uplink and downlink in a mobile network. The participation of device APIs and distributed computing systems in the running of an application is also shown.

Communication Asymmetry

- • Intrinsically asymmetric Mobile
- communication between the mobile
- device and static computer system
- • Device allocated a limited bandwidth
- • Because of a large number of devices

- • Bandwidth in the downstream from the
- server to device much larger than the
- one in the upstream from the device to
- server
- • Because mobile devices have limited
- power resources
- • Faster data transmission rates for long
- intervals of time need greater power
- dissipation from the devices

Communication asymmetry in uplink and downlink in a mobile network



GSM networks data transmission

- • Rates go up to a maximum of 14.4 kbps
- for both uplink and downlink
- • Symmetric communication
- • Only used for voice communication

i-mode for many applications

- • Used for voice, multimedia transmission,
- Internet access, voice communication
- • Base station provides downlink 384 kbps
- • Uplink from the devices restricted to 64
- kbps
- • Asymmetric communication

The characteristics in wireless signals

- • Interference and time-dispersion
- • Signal distortion and transmission errors
- at the receiver end
- • Lead to path loss and signal fading,
- which cause data loss
- • Greater access latency compared to
- wired networks

The characteristics in wireless signals

- • Data loss has to be taken care of by
- repeat transmissions
- • Transmission errors have to be corrected
- • Taken care of by appending additional
- bits, such as the forward error correction
- bits

The characteristics in Mobile communication

- • Mobile devices also have low storage
- capacity (memory)
- • Cannot hoard large databases
- • Accessing the data online not only has a
- latency period (is not instantaneous) but
- also dissipates bandwidth resources of
- the device

Broadcasting

- • Corresponds to unidirectional (downlink
- from the server to the devices)
- • Unicast communication— Unicast means
- the transmission of data packets in a
- computer network such that a single
- destination receives the packets

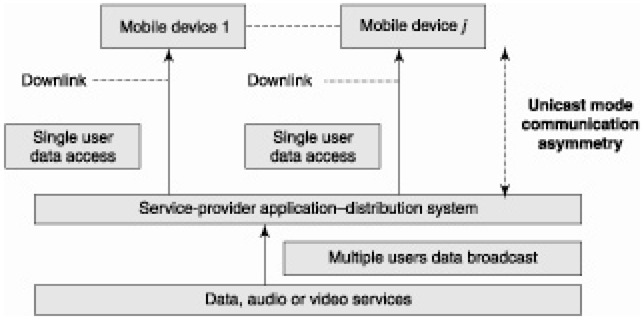
Broadcasting or application distribution service

- • This destination generally the one which
- has subscribed to the service
- • Mobile TV— an example of unidirectional
- unicast mode of broadcasting
- • Each device receives broadcast data
- packets from the service provider's
- application– distribution system

Broadcasting or application distribution service

- • Application–distribution system
- broadcasts data of text, audio, or video
- services

A broadcasting architecture



Summary

•

- • GSM symmetric and voice only
- • Mobile communication asymmetric in
- general
- • Limited device capability
- • Device memory, energy and uplink and
- downlink bandwidths
- • Broadcast architecture

Classification of Data-Delivery Mechanisms

- There are two fundamental information delivery methods for wireless data applications: Point-to-Point access and Broadcast.
- Compared with Point-to-Point access, broadcast is a more attractive method. A single broadcast of a data item can satisfy all the outstanding requests for that item simultaneously.
- As such, broadcast can scale up to an arbitrary number of users. There are three kinds of broadcast models, namely *push-based broadcast, On-demand (or pull-based) broadcast, and hybrid broadcast.*
- *In push based broadcast, the server disseminates information using a periodic/aperiodic broadcast program (generally without any intervention of clients).*
-

- *In on demand broadcast, the server disseminates information based on the outstanding requests submitted by clients; In hybrid broadcast, push based broadcast and on demand data deliveries are combined to complement each other.*
- *In addition, mobile computers consume less battery power on monitoring broadcast channels to receive data than accessing data through point-to-point communications.*
- Data-delivery mechanisms can be classified into three categories, namely, push-based mechanisms (publish-subscribe mode), pull-based mechanisms (on-demand mode), and hybrid mechanisms (hybrid mode).

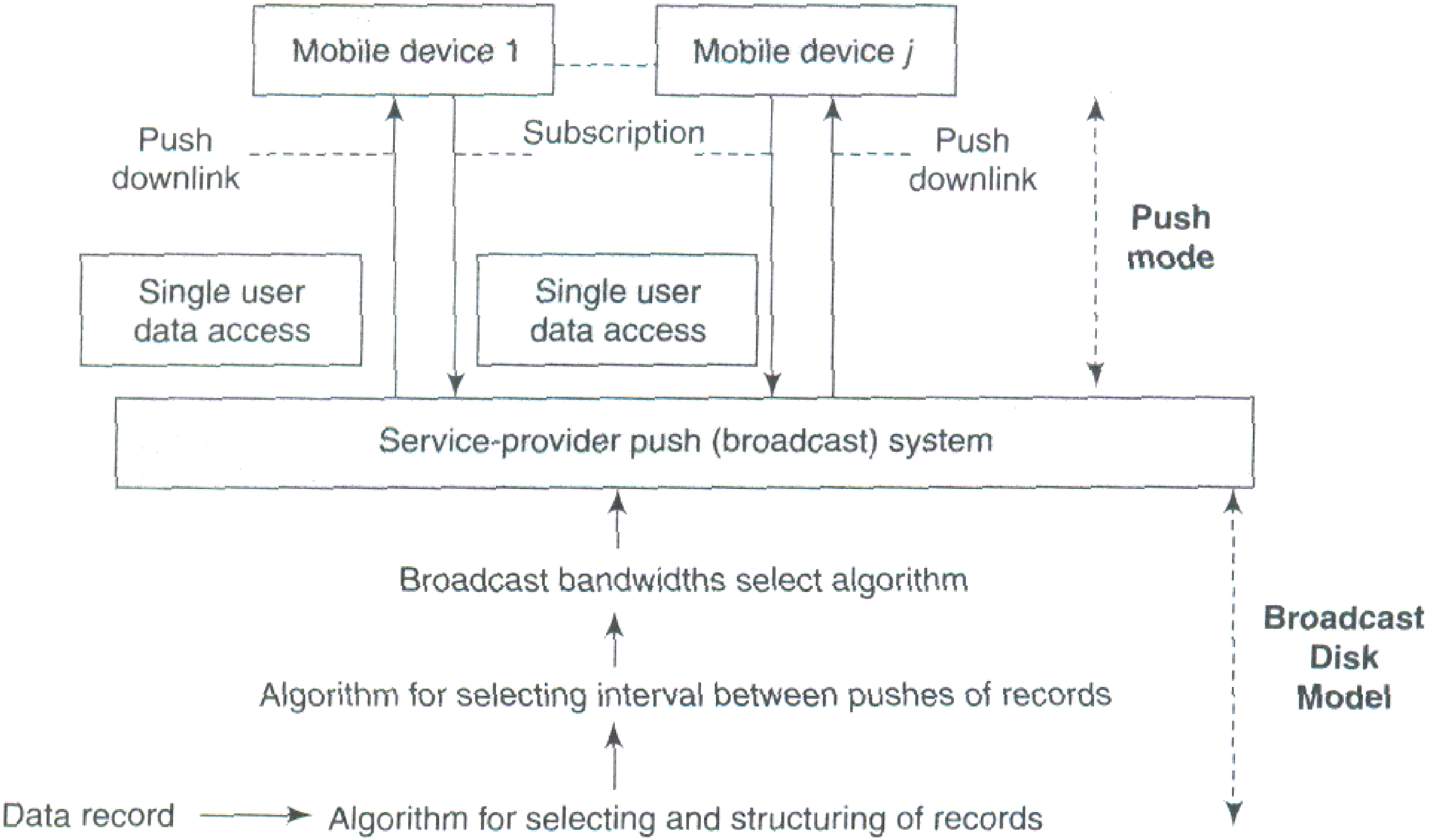
Classification of Data-Delivery Mechanisms

- • Push-based mechanisms (publish–
- subscribe mode)
- • Pull-based mechanisms (on-demand
- mode)
- • Hybrid mechanisms (hybrid mode)

Push-based Mechanisms

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

Push-based data-delivery mechanism


Mobile device 1
Mobile device *j*
Push downlink
Subscription
Push downlink
Push mode
Single user data access
Single user data access
Service-provider push (broadcast) system
Broadcast bandwidths select algorithm
Broadcast Disk Model
Algorithm for selecting interval between pushes of records
Data record
Algorithm for selecting and structuring of records

Push-based mechanisms function in the following manner:

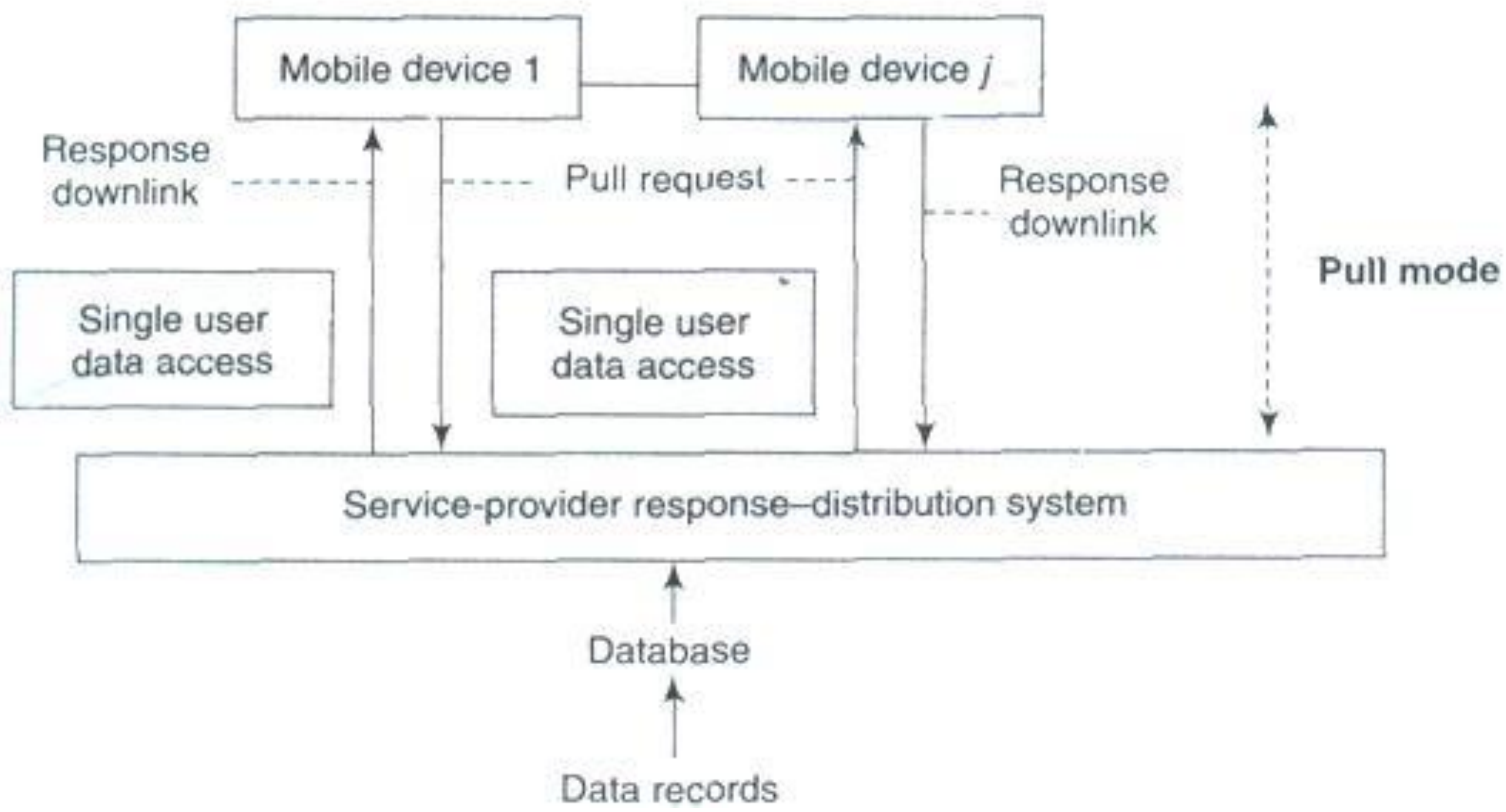
- 1. A structure of data records to be pushed is selected. An algorithm provides an adaptable multi-level mechanism that permits data items to be pushed uniformly or non-uniformly after structuring them according to their relative importance.
- 2. Data is pushed at selected time intervals using an adaptive algorithm. Pushing only once saves bandwidth. However, pushing at periodic intervals is important because it provides the devices that were disconnected at the time of previous push with a chance to cache the data when it is pushed again.
- 3. Bandwidths are adapted for downlink (for pushes) using an algorithm. Usually higher bandwidth is allocated to records having higher number of subscribers or to those with higher access probabilities.

- 4. A mechanism is also adopted to stop pushes when a device is handed over to another cell.

- ***Advantages of Push based mechanisms:***
  - – □ Push-based mechanisms enable broadcast of data services to multiple devices.
  - – □ The server is not interrupted frequently by requests from mobile devices.
  - – □ These mechanisms also prevent server overload, which might be caused by flooding of device requests
  - – □ Also, the user even gets the data he would have otherwise ignored such as traffic congestion, forthcoming weather reports etc
- ***Disadvantages:***
  - – □ Push-based mechanisms disseminate of unsolicited, irrelevant, or out-of-context data, which may cause inconvenience to the user.

Pull based Mechanisms

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


Mobile device 1
Mobile device *j*
Response downlink
Pull request
Response downlink
Pull mode
Single user data access
Single user data access
Service-provider response–distribution system
Database
Data records

Pull-based mechanisms function in the following manner:

  - 1. The bandwidth used for the uplink channel depends upon the number of pull requests.
  - 2. A pull threshold is selected. This threshold limits the number of pull requests in a given period of time. This controls the number of server interruptions.
  - 3. A mechanism is adopted to prevent the device from pulling from a cell, which has handed over the concerned device to another cell. On device handoff, the subscription is cancelled or passed on to the new service provider cell

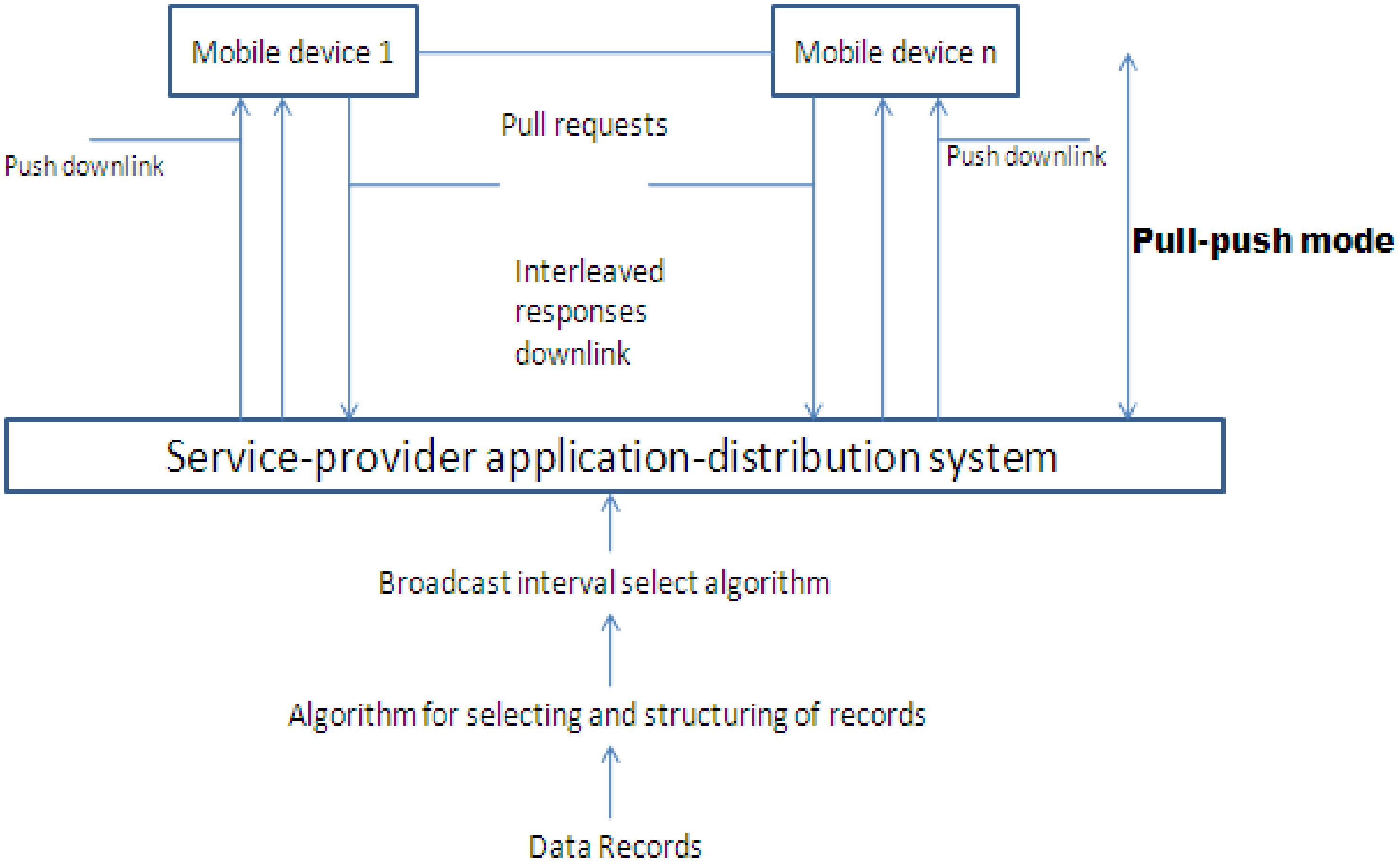
- In pull-based mechanisms the user-device receives data records sent by server on demand only.

- ***Advantages of Pull based mechanisms:***
  - □ With pull-based mechanisms, no unsolicited or irrelevant data arrives at the device and the relevant data is disseminated only when the user asks for it.
  - □ Pull-based mechanisms are the best option when the server has very little contention and is able to respond to many device requests within expected time intervals.

- ***Disadvantages:***
  - □ The server faces frequent interruptions and queues of requests at the server may cause congestion in cases of sudden rise in demand for certain data record.
  - □ In on-demand mode, another disadvantage is the energy and bandwidth required for sending the requests for hot items and temporal records

Hybrid Mechanisms

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


Mobile device 1
Mobile device n
Pull requests
Push downlink
Push downlink
Interleaved responses downlink
Pull-push mode
Service-provider application-distribution system
Broadcast interval select algorithm
Algorithm for selecting and structuring of records
Data Records

Hybrid interleaved push-pull-based data-delivery mechanism

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

Hybrid mechanisms function in the following manner:

- 1. There are two channels, one for pushes by front channel and the other for pulls by back channel.
- 2. Bandwidth is shared and adapted between the two channels depending upon the number of active devices receiving data from the server and the number of devices requesting data pulls from the server.
- 3. An algorithm can adaptively chop the slowest level of the scheduled pushes successively The data records at lower level where the records are assigned lower priorities can have long push intervals in a broadcasting model.
- ***Advantages of Hybrid mechanisms:***
  - □ The number of server interruptions and queued requests are significantly reduced.
- ***Disadvantages:***
  - □ IPP does not eliminate the typical server problems of too many interruptions and queued requests.
  - □ Another disadvantage is that adaptive chopping of the slowest level of scheduled pushes.

Selective Tuning and Indexing Techniques

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

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

Directory Method

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

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

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

Hash-Based Method

- Hash is a result of operations on a pair of key and record.
- Advantage of broadcasting a hash is that it contains a fewer bits compared to key and record separately.
- The operations are done by a hashing function. From the server end the hash is broadcasted and from the device end a key is extracted by computations from the data in the record by operating the data with a function called hash function (algorithm).
- This key is called hash key.
- Hash-based method entails that the hash for the hashing parameter (hash key) is broadcasted.
- Each device receives it and tunes to the record as per the extracted key.
- In this method, the records that are of interest to a device or those required by it are cached from the broadcast cycle by first extracting and identifying the hash key which provides the location of the record.

- This helps in tuning of the device. Hash-based method can be described as follows:
  - 1. A separate directory is not broadcast as overhead with each broadcast cycle.
  - 2. Each broadcast cycle has hash bits for the hash function H, a shift function S, and the data that it holds. The function S specifies the location of the record or remaining part of the record relative to the location of hash and, thus, the time interval for wait before the record can be tuned and cached.
  - 3. Assume that a broadcast cycle pushes the hashing parameters **H*(R$_i$) [H and S] and record R$_i$. The functions H and S help in tuning to the H(R$_i$) and hence to R$_i$ as follows—H gives a key which in turn gives the location of H(R$_i$) in the broadcast data. In case H generates a key that does not provide the location of H(R$_i$) by itself, then the device computes the location from S after the location of H(R$_i$). That location has the sequential records R$_i$ and the devices tunes to the records from these locations.***
  - 4. In case the device misses the record in first cycle, it tunes and caches that in next or some other cycle.

Index-Based Method

- Indexing is another method for selective tuning. Indexes temporarily map the location of the buckets.
- At each location, besides the bits for the bucket in record of interest data, an offset value may also be specified there.
- While an index maps to the absolute location from the beginning of a broadcast cycle, an offset index is a number which maps to the relative location after the end of present bucket of interest.
- Offset means a value to be used by the device along with the present location and calculate the wait period for tuning to the next bucket. All buckets have an offset to the beginning of the next indexed bucket or item.

- Indexing is a technique in which each data bucket, record, or record block of interest is assigned an index at the previous data bucket, record, or record block of interest to enable the device to tune and cache the bucket after the wait as per the offset value.
- The server transmits this index at the beginning of a broadcast cycle as well as with each bucket corresponding to data of interest to the device.
- A disadvantage of using index is that it extends the broadcast cycle and hence increases taccess.

- The index I has several offsets and the bucket type and flag information. A typical index may consist of the following:
  - 1. Ioffset(1) which defines the offset to first bucket of nearest index.
  - 2. Additional information about *Tb, which is the time required for caching the bucket bits in full after the device tunes to and starts caching the bucket. This enables transmission of buckets of variable lengths.*
  - 3. Ioffset (next) which is the index offset of next bucket record of interest.
  - 4. I_{offset}(end) which is the index offset for the end of broadcast cycle and the start of next cycle. This enables the device to look for next index I after the time interval as per I_{offset}(end). This also permits a broadcast cycle to consist of variable number of buckets.
  - 5. I_{type}, which provides the specification of the type of contents of next bucket to be tuned, that is, whether it has an index value or data.
  - 6. A flag called dirty flag which contains the information whether the indexed buckets defined by I_{offset}(1) and I_{offset}(next) are dirty or not. An indexed bucket being dirty means that it has been rewritten at the server with new values. Therefore, the device should invalidate the previous caches of these buckets and update them by tuning to and caching them.

Distributed Index Based Method

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

Flexible Indexing Method

- Assume that a broadcast cycle has number of data segments with each of the segments having a variable set of records. For example, let *n records, Ro to Rn-1, be present in four data segments, R() to Ri-1, Ri to Rj-1 , Rj to Rj-1 and Rk to Rn-1.*
- *Some possible index parameters are (i) Iseg,having just 2 bits for the offset, to specify the location of a segment in a broadcast cycle, (ii) Irec, having just 6 bits for the offset, to specify the location of a record of interest within a segment of the broadcast cycle, (iii) Ib, having just 4 bits for the offset, to specify the location of a bucket of interest within a record present in one of the segments of the broadcast cycle.*
- *Flexible indexing method provides dual use of the parameters (e.g., use of Iseg or Irec in an index segment to tune to the record or buckets of interest) or multi-parameter indexing (e.g., use of Iseg, Irec, or Ib in an index segment to tune to the bucket of interest).*

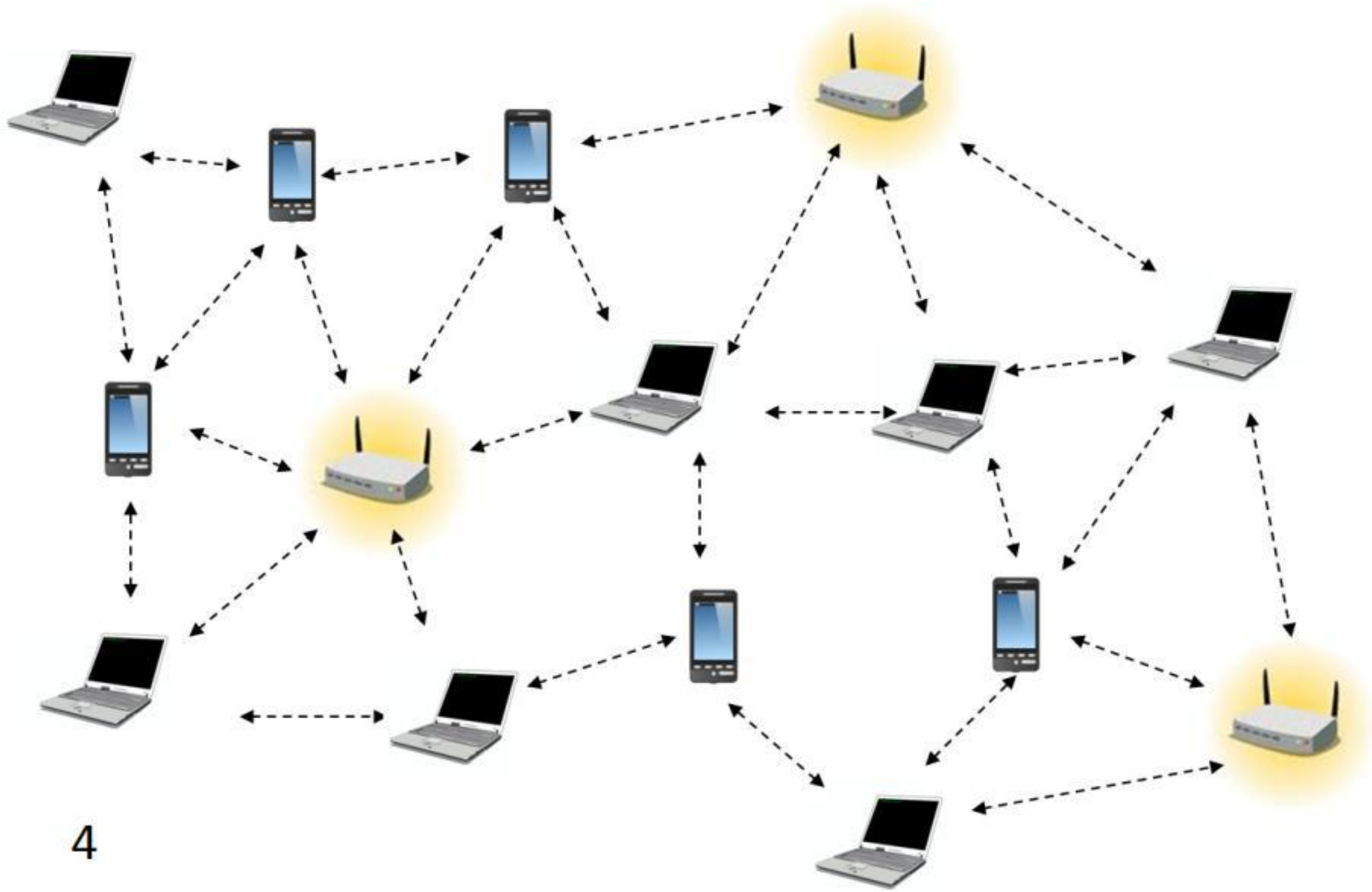
- Assume that broadcast cycle has *m sets of records (called segments). A set of binary bits defines the index parameter Iseg,. A local index is then assigned to the specific record (or bucket). Only local index (Irec or Ib) is used in (Iloc, m) based data tuning which corresponds to the case of flexible indexing method being discussed. The number of bits in a local index is much smaller than that required when each record is assigned an index. Therefore, the flexible indexing method proves to be beneficial.*

UNIT-V

Mobile Ad hoc NETworks (MANETs)

- Mobile Ad hoc NETworks (MANETs) are wireless networks which are characterized by dynamic topologies and no fixed infrastructure.
- Each node in a MANET is a computer that may be required to act as both a host and a router and, as much, may be required to forward packets between nodes which cannot directly communicate with one another.
- Each MANET node has much smaller frequency spectrum requirements that that for a node in a fixed infrastructure network.
- A MANET is an autonomous collection of mobile users that communicate over relatively bandwidth constrained wireless links.
- Since the nodes are mobile, the network topology may change rapidly and unpredictably over time.
- The network is decentralized, where all network activity including discovering the topology and delivering messages must be executed by the nodes themselves, i.e., routing functionality will be incorporated into mobile nodes.

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



MANET- Characteristics

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

Need for Ad Hoc Networks

- Setting up of fixed access points and backbone infrastructure is not always viable
  - Infrastructure may not be present in a disaster area or war zone
  - Infrastructure may not be practical for short-range radios; Bluetooth (range ~ 10m)
- Ad hoc networks:
  - Do not need backbone infrastructure support
  - Are easy to deploy
  - Useful when infrastructure is absent, destroyed or impractical

Properties of MANETs

- □ MANET enables fast establishment of networks. When anew network is to be established, the only requirement is to provide a new set of nodes with limited wireless communication range. A node has limited capability, that is, it can connect only to the nodes which are nearby. Hence it consumes limited power.
- □ A MANET node has the ability to discover a neighboring node and service. Using a service discovery protocol, a node discovers the service of a nearby node and communicates to a remote node in the MANET.
- □ MANET nodes have peer-to-peer connectivity among themselves.
- □ MANET nodes have independent computational, switching (or routing), and communication capabilities.
- □ The wireless connectivity range in MANETs includes only nearest node connectivity.
- □ The failure of an intermediate node results in greater latency in communicating with the remote server.

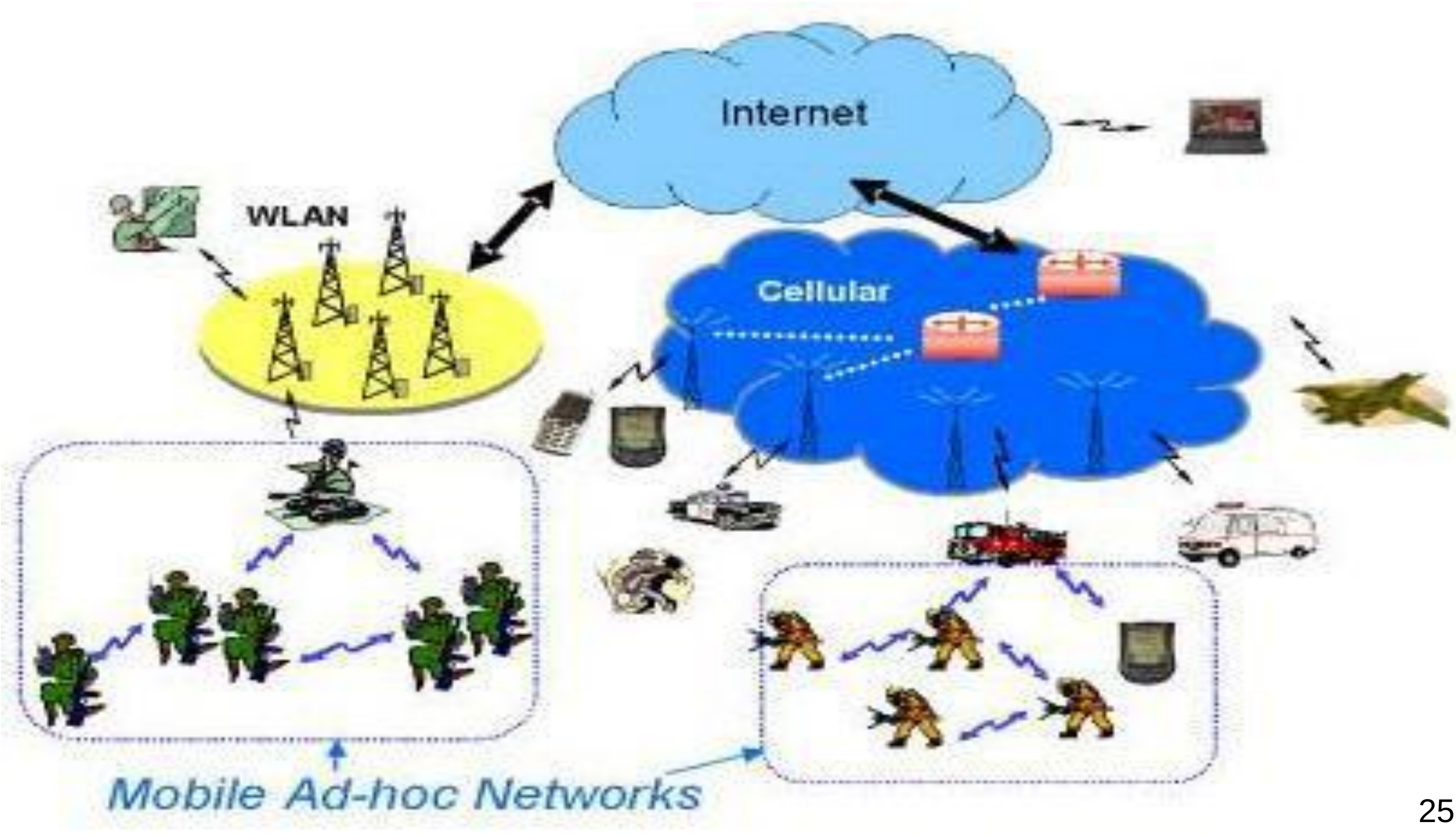
- □ Limited bandwidth available between two intermediate nodes becomes a constraint for the MANET. The node may have limited power and thus computations need to be energy-efficient.
- □ There is no access-point requirement in MANET. Only selected access points are provided for connection to other networks or other MANETs.
- □ MANET nodes can be the iPods, Palm handheld computers, Smartphones, PCs, smart labels, smart sensors, and automobile-embedded systems\
- □ MANET nodes can use different protocols, for example, IrDA, Bluetooth, ZigBee, 802.11, GSM, and TCP/IP.MANET node performs data caching, saving, and aggregation.
- □ MANET mobile device nodes interact seamlessly when they move with the nearby wireless nodes, sensor nodes, and embedded devices in automobiles so that the seamless connectivity is maintained between the devices.

MANET challenges

- To design a good wireless ad hoc network, various challenges have to be taken into account:
- ***Dynamic Topology****: Nodes are free to move in an arbitrary fashion resulting in the topology changing arbitrarily. This characteristic demands dynamic configuration of the network.*
- ***Limited security****: Wireless networks are vulnerable to attack. Mobile ad hoc networks are more vulnerable as by design any node should be able to join or leave the network at any time. This requires flexibility and higher openness.*
- ***Limited Bandwidth:*** *Wireless networks in general are bandwidth limited. In an ad hoc network, it is all the more so because there is no backbone to handle or multiplex higher bandwidth*
- ***Routing****: Routing in a mobile ad hoc network is complex. This depends on many factors, including finding the routing path, selection of routers, topology, protocol etc.*

Applications of MANETS

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


Internet
WLAN
Cellular
Mobile Ad-hoc Networks

Applications of MANETS

- ***Military battlefield****– soldiers, tanks, planes. Ad- hoc networking would allow the military to take advantage of commonplace network technology to maintain an information network between the soldiers, vehicles, and military information headquarters***.**
- ***Sensor networks*** *– to monitor environmental conditions over a large area*
- ***Local level*** *– Ad hoc networks can autonomously link an instant and temporary multimedia network using notebook computers or palmtop computers to spread and share information among participants at e.g. conference or classroom. Another appropriate local level application might be in home networks where devices can communicate directly to exchange information.*

- ***Personal Area Network (PAN)*** *– pervasive computing i.e. to provide flexible connectivity between personal electronic devices or home appliances. Short-range MANET can simplify the intercommunication between various mobile devices (such as a PDA, a laptop, and a cellular phone). Tedious wired cables are replaced with wireless connections. Such an ad hoc network can also extend the access to the Internet or other networks by mechanisms e.g. Wireless LAN (WLAN), GPRS, and UMTS.*
- ***Vehicular Ad hoc Networks*** *– intelligent transportation i.e. to enable real time vehicle monitoring and adaptive traffic control*

- ***Civilian environments*** *– taxi cab network, meeting rooms, sports stadiums, boats, small aircraft*
- ***Emergency operations*** *– search and rescue, policing and fire fighting and to provide connectivity between distant devices where the network infrastructure is unavailable. Ad hoc can be used in emergency/rescue operations for disaster relief efforts, e.g. in fire, flood, or earthquake. Emergency rescue operations must take place where non-existing or damaged communications infrastructure and rapid deployment of a communication network is needed. Information is relayed from one rescue team member to another over a small hand held.*

Routing in MANET’s

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

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

- Asymmetric links: - Routing information collected for one direction is of no use for the other direction. Many routing algorithms for wired networks rely on a symmetric scenario.
- Redundant links: - In wired networks, some redundancy is present to survive link failures and this redundancy is controlled by a network administrator. In ad-hoc networks, nobody controls redundancy resulting in many redundant links up to the extreme of a complete meshed topology.
- Interference: - In wired networks, links exist only where a wire exists, and connections are planned by network administrators. But, in ad-hoc networks links come and go depending on transmission characteristics, one transmission might interfere with another and nodes might overhear the transmission of other nodes.

- Dynamic topology: - The mobile nodes might move in an arbitrary manner or medium characteristics might change. This result in frequent changes in topology, so snapshots are valid only for a very short period of time. So, in ad-hoc networks, routing tables must somehow reflect these frequent changes in topology and routing algorithms have to be adopted.

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

- Traditional routing algorithms known from wired networks will not work efficiently or fail completely. These algorithms have not been designed with a highly dynamic topology, asymmetric links, or interference in mind.
- Routing in wireless ad-hoc networks cannot rely on layer three knowledge alone. Information from lower layers concerning connectivity or interference can help routing algorithms to find a good path.
- Centralized approaches will not really work, because it takes too long to collect the current status and disseminate it again. Within this time the topology has already changed.

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

- 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.

Routing Algorithms

- Always maintain routes:- Little or no delay for route determination
- Consume bandwidth to keep routes up-to-date
- Maintain routes which may never be used
- Advantages: low route latency, State information, QoS guarantee related to connection set-up or other real-time requirements
- Disadvantages: high overhead (periodic updates) and route repair depends on update frequency

- only obtain route information when needed
- Advantages: no overhead from periodic update, scalability as long as there is only light traffic and low mobility.
- Disadvantages: high route latency, route caching can reduce latency

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

Destination sequence distance vector (DSDV)

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

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

- Each node maintains a routing table which stores next hop, cost metric towards each destination and a sequence number that is created by the destination itself.
- Each node periodically forwards routing table to neighbors. Each node increments and appends its sequence number when sending its local routing table.
- Each route is tagged with a sequence number; routes with greater sequence numbers are preferred.
- Each node advertises a monotonically increasing even sequence number for itself.
- When a node decides that a route is broken, it increments the sequence number of the route and advertises it with infinite metric. Destination advertises new sequence number.

Security in MANET's

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

These attacks can be classified into two types:

- 1. **External Attack:**
  - External attacks are carried out by nodes that do not belong to the network.
  - It causes congestion sends false routing information or causes unavailability of services.
- 2. **Internal Attack:**
  - Internal attacks are from compromised nodes that are part of the network. In an internal attack the malicious node from the network gains unauthorized access and impersonates as a genuine node.
  - It can analyze traffic between other nodes and may participate in other network activities.

- **Denial of Service attack**: This attack aims to attack the availability of a node or the entire network. If the attack is successful the services will not be available. The attacker generally uses radio signal jamming and the battery exhaustion method.
- **Impersonation**: If the authentication mechanism is not properly implemented a malicious node can act as a genuine node and monitor the network traffic. It can also send fake routing packets, and gain access to some confidential information.
- **Eavesdropping**: This is a passive attack. The node simply observes the confidential information. This information can be later used by the malicious node. The secret information like location, public key, private key, password etc. can be fetched by eavesdropper.
- **Routing Attacks**: The malicious node makes routing services a target because it’s an important service in MANETs. There are two flavors to this routing attack. One is attack on routing protocol and another is attack on packet forwarding or delivery mechanism.

- **Black hole Attack:** In this attack, an attacker advertises a zero metric for all destinations causing all nodes around it to route packets towards it.[9] A malicious node sends fake routing information, claiming that it has an optimum route and causes other good nodes to route data packets through the malicious one. A malicious node drops all packets that it receives instead of normally forwarding those packets. An attacker listen the requests in a flooding based protocol.
- **Wormhole Attack:** In a wormhole attack, an attacker receives packets at one point in the network, ―tunnels‖ them to another point in the network, and then replays them into the network from that point. Routing can be disrupted when routing control message are tunnelled. This tunnel between two colluding attacks is known as a wormhole.
- **Replay Attack:** An attacker that performs a replay attack are retransmitted the valid data repeatedly to inject the network routing traffic that has been captured previously. This attack usually targets the freshness of routes, but can also be used to undermine poorly designed security solutions.

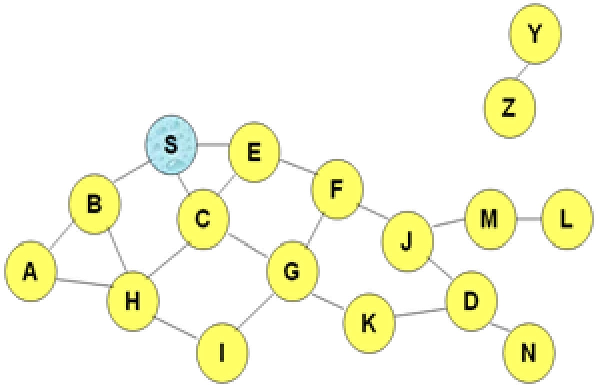
- **Jamming:** In jamming, attacker initially keep monitoring wireless medium in order to determine frequency at which destination node is receiving signal from sender. It then transmit signal on that frequency so that error free receptor is hindered.
- **Man- in- the- middle attack:** An attacker sites between the sender and receiver and sniffs any information being sent between two nodes. In some cases, attacker may impersonate the sender to communicate with receiver or impersonate the receiver to reply to the sender.
- **Gray-hole attack:** This attack is also known as routing misbehavior attack which leads to dropping of messages. Gray-hole attack has two phases. In the first phase the node advertise itself as having a valid route to destination while in second phase, nodes drops intercepted packets with a certain probability.

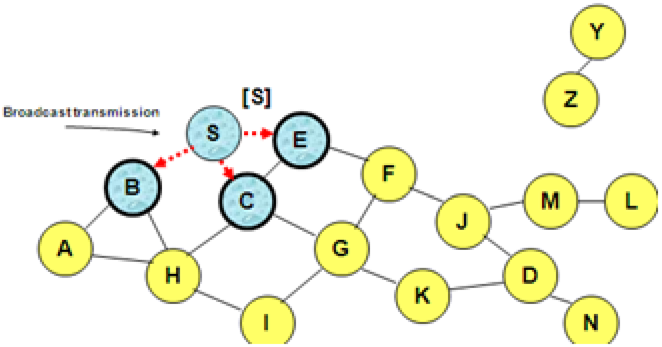
Dynamic Source Routing

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

- *Route discovery. If the source does not have a route to the destination in its route cache, it broadcasts a route request (RREQ) message specifying the destination node for which the route is requested.*
- *The RREQ message includes a route record which specifies the sequence of nodes traversed by the message. When an intermediate node receives a RREQ, it checks to see if it is already in the route record. If it is, it drops the message.*
- *This is done to prevent routing loops. If the intermediate node had received the RREQ before, then it also drops the message. The intermediate node forwards the RREQ to the next hop according to the route specified in the header.*
- *When the destination receives the RREQ, it sends back a route reply message. If the destination has a route to the source in its route cache, then it can send a route response (RREP) message along this route. Otherwise, the RREP message can be sent along the reverse route back to the source.*
- *Intermediate nodes may also use their route cache to reply to RREQs. If an intermediate node has a route to the destination in its cache, then it can append the route to the route record in the RREQ, and send an RREP back to the source containing this route. This can help limit flooding of the RREQ. However, if the cached route is out-of-date, it can result in the source receiving stale routes.*

- *Route maintenance. When a node detects a broken link while trying to forward a packet to the next hop, it sends a route error (RERR) message back to the source containing the link in error. When an RERR message is received, all routes containing the link in error are deleted at that node.*
- As an example, consider the following MANET, where a node S wants to send a packet to D, but does not know the route to D. So, it initiates a route discovery. Source node S floods Route Request (RREQ). Each node appends its own identifier when forwarding RREQ as shown below.

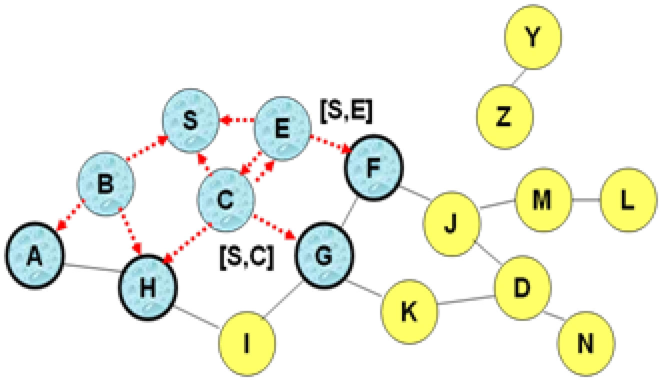




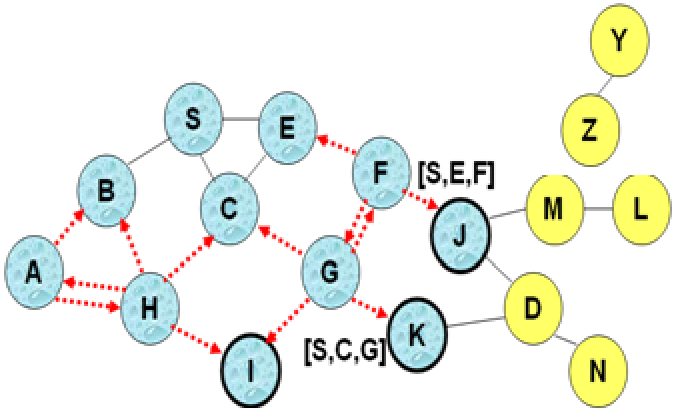
● Represents a node that has received RREQ for D from S

⋯▸ Represents transmission of RREQ

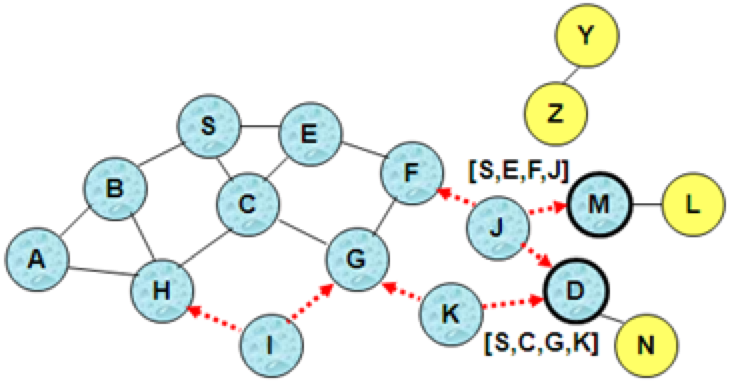
[X,Y] Represents list of identifiers appended to RREQ



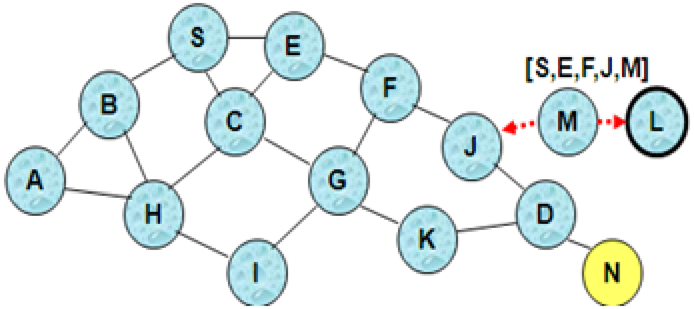
Node H receives packet RREQ from two neighbors: potential for collision



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



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

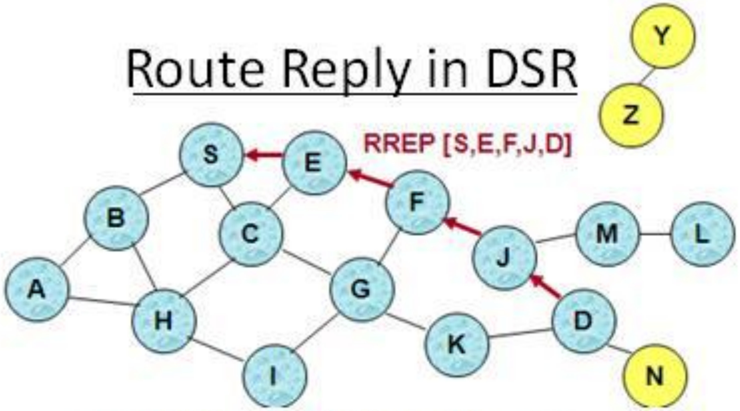


Node D does not forward RREQ, because node D is the intended target of the route discovery

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

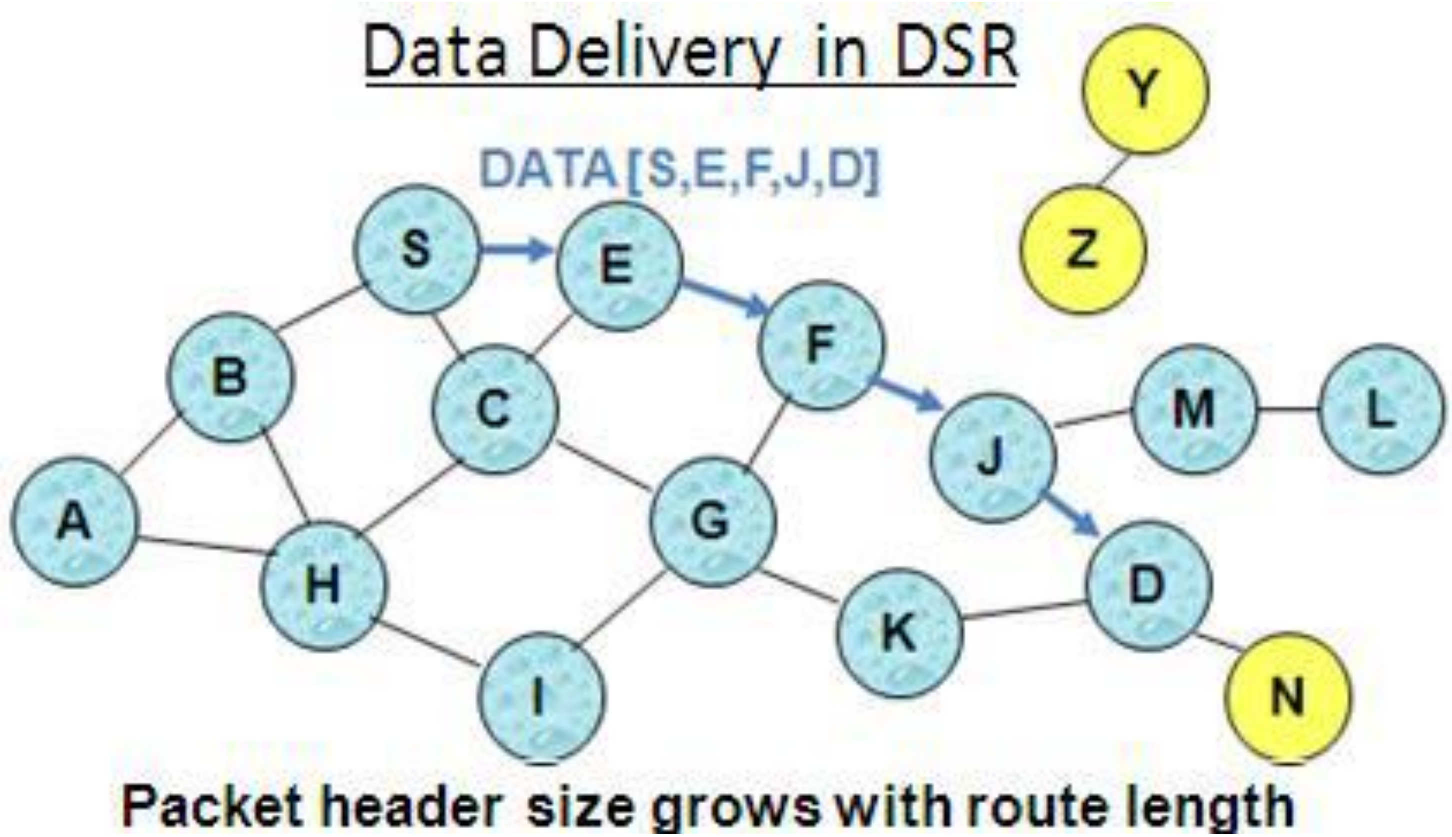
- Route Reply can be sent by reversing the route in Route Request (RREQ) only if links are guaranteed to be bi-directional.
- If Unidirectional (asymmetric) links are allowed, then RREP may need a route discovery from S to D.
- Node S on receiving RREP, caches the route included in the RREP. When node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

Route Reply in DSR



← Represents RREP control message

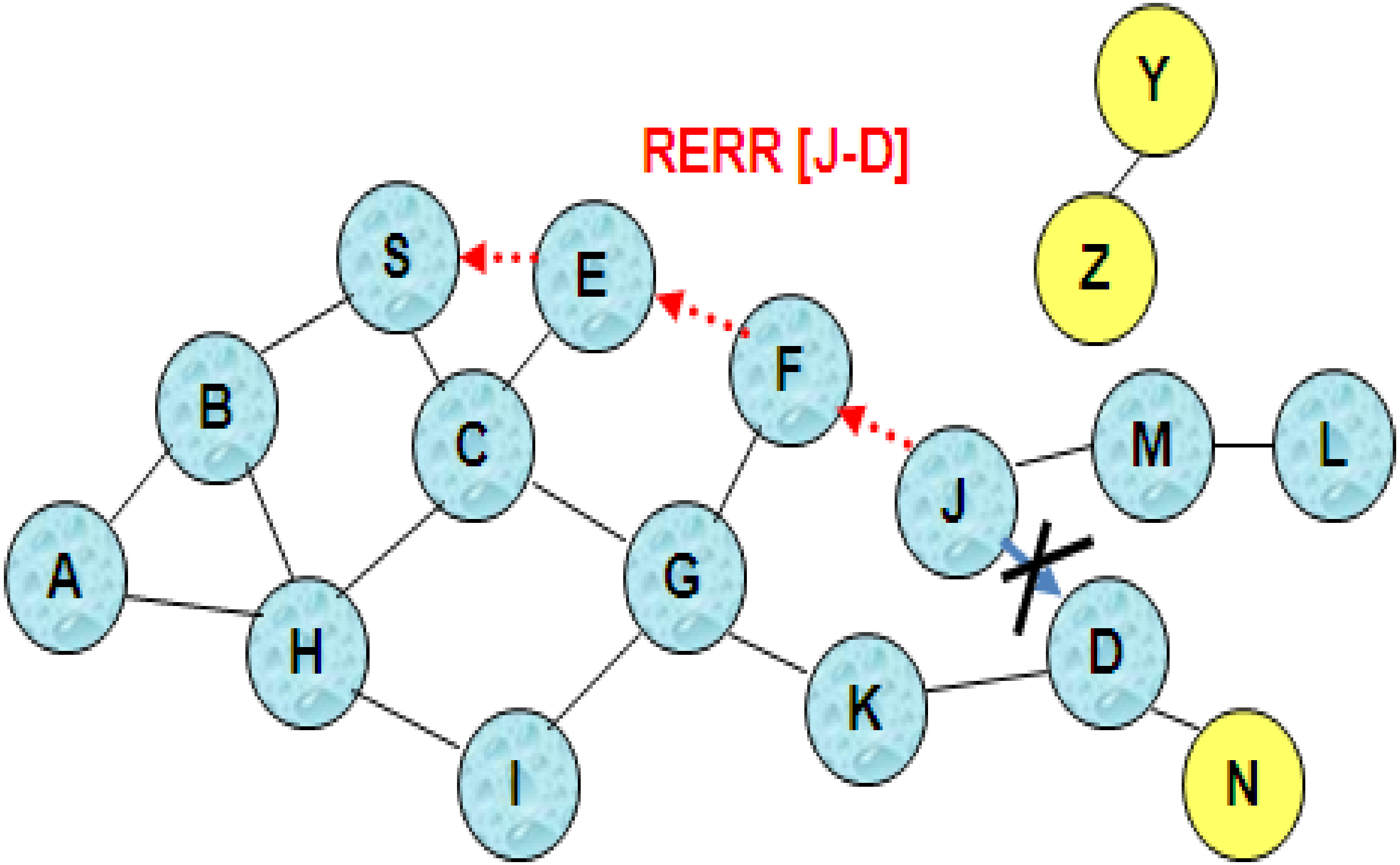
- Route Reply can be sent by reversing the route in Route Request (RREQ) only if links are guaranteed to be bi-directional. If Unidirectional (asymmetric) links are allowed, then RREP may need a route discovery from S to D. Node S on receiving RREP, caches the route included in the RREP. When node S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



Data Delivery in DSR

Packet header size grows with route length

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


RERR [J-D]
Y
Z
S
E
F
B
C
J
M
L
A
G
H
D
K
I
N

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

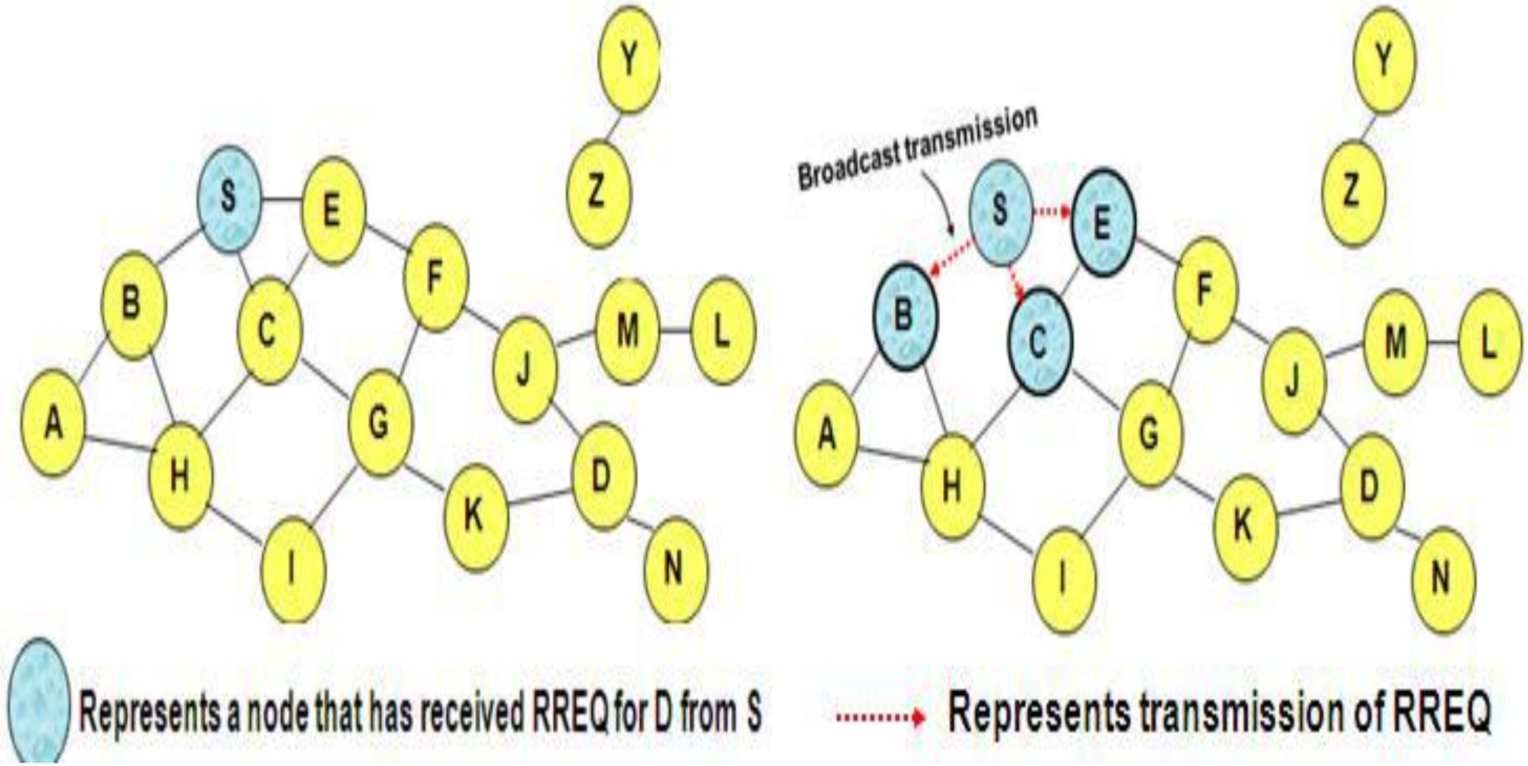
- ***Disadvantages of DSR:***
- Packet header size grows with route length due to source routing
- Flood of route requests may potentially reach all nodes in the network
- Care must be taken to avoid collisions between route requests propagated by neighboring nodes -- insertion of random delays before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

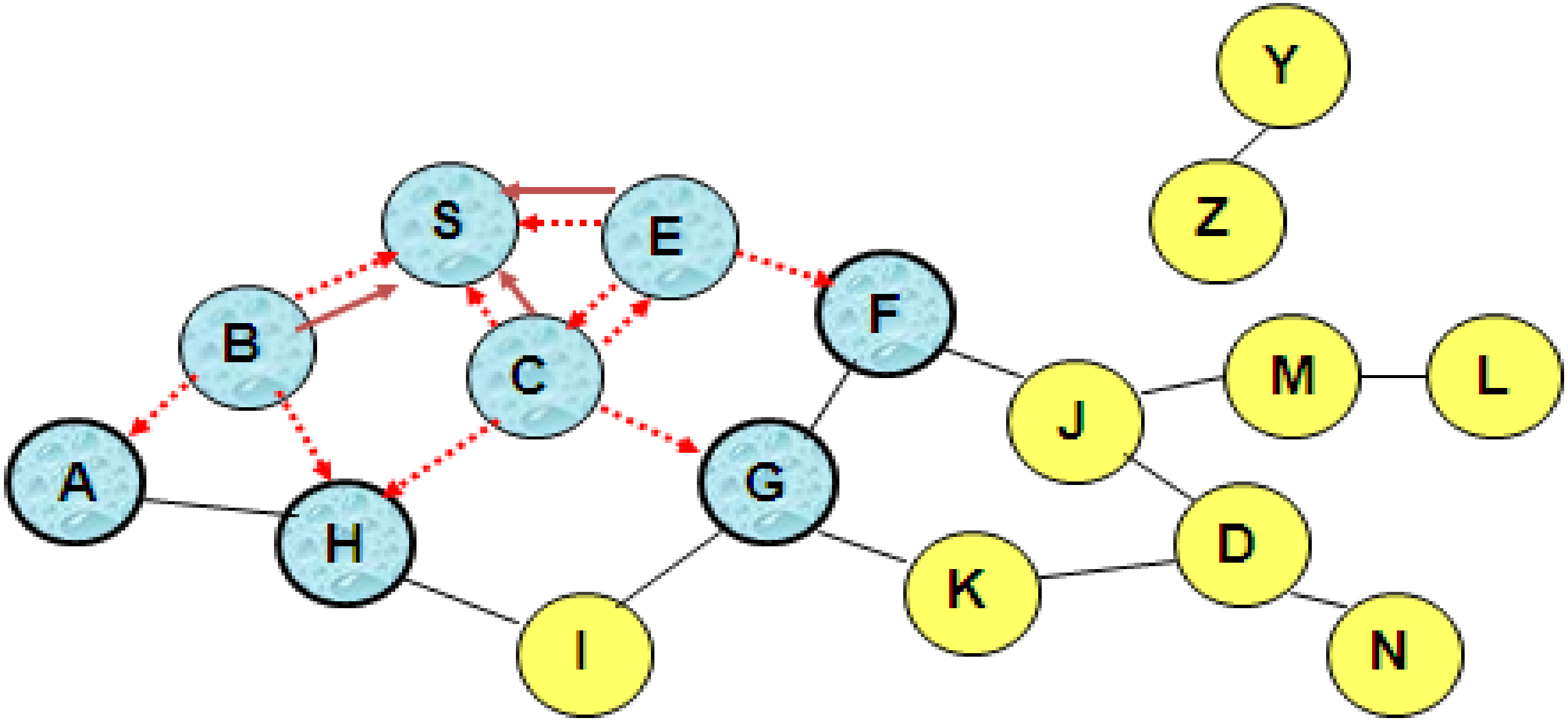
Ad Hoc On-Demand Distance Vector Routing (AODV)

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

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

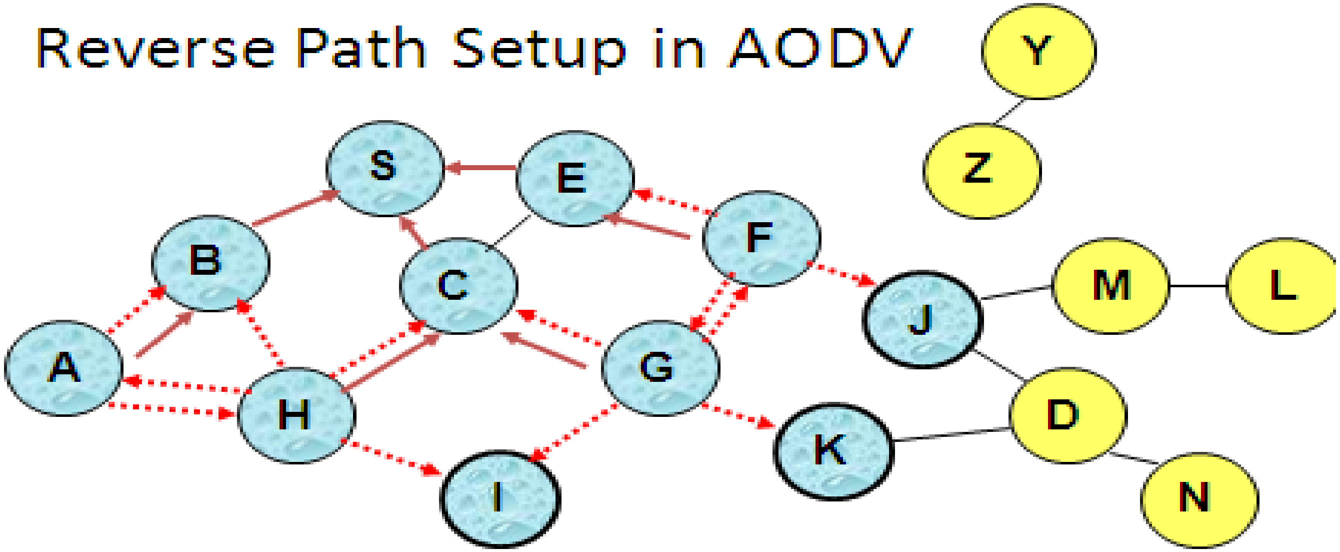
- A larger destination sequence number indicates a more current (or more recent) route. Upon receiving an RREQ or RREP packet, a node updates its routing information to set up the reverse or forward path, respectively, only if the route contained in the RREQ or RREP packet is more current than its own route.
- ***Route Maintenance.*** *When a node detects a broken link while attempting to forward a packet to the next hop, it generates a RERR packet that is sent to all sources using the broken link. The RERR packet erases all routes using the link along the way. If a source receives a RERR packet and a route to the destination is still required, it initiates a new route discovery process. Routes are also deleted from the routing table if they are unused for a certain amount of time.*


Y
Z
S
E
B
C
F
M
L
J
A
G
H
D
K
I
N
Broadcast transmission
Represents a node that has received RREQ for D from S
Represents transmission of RREQ

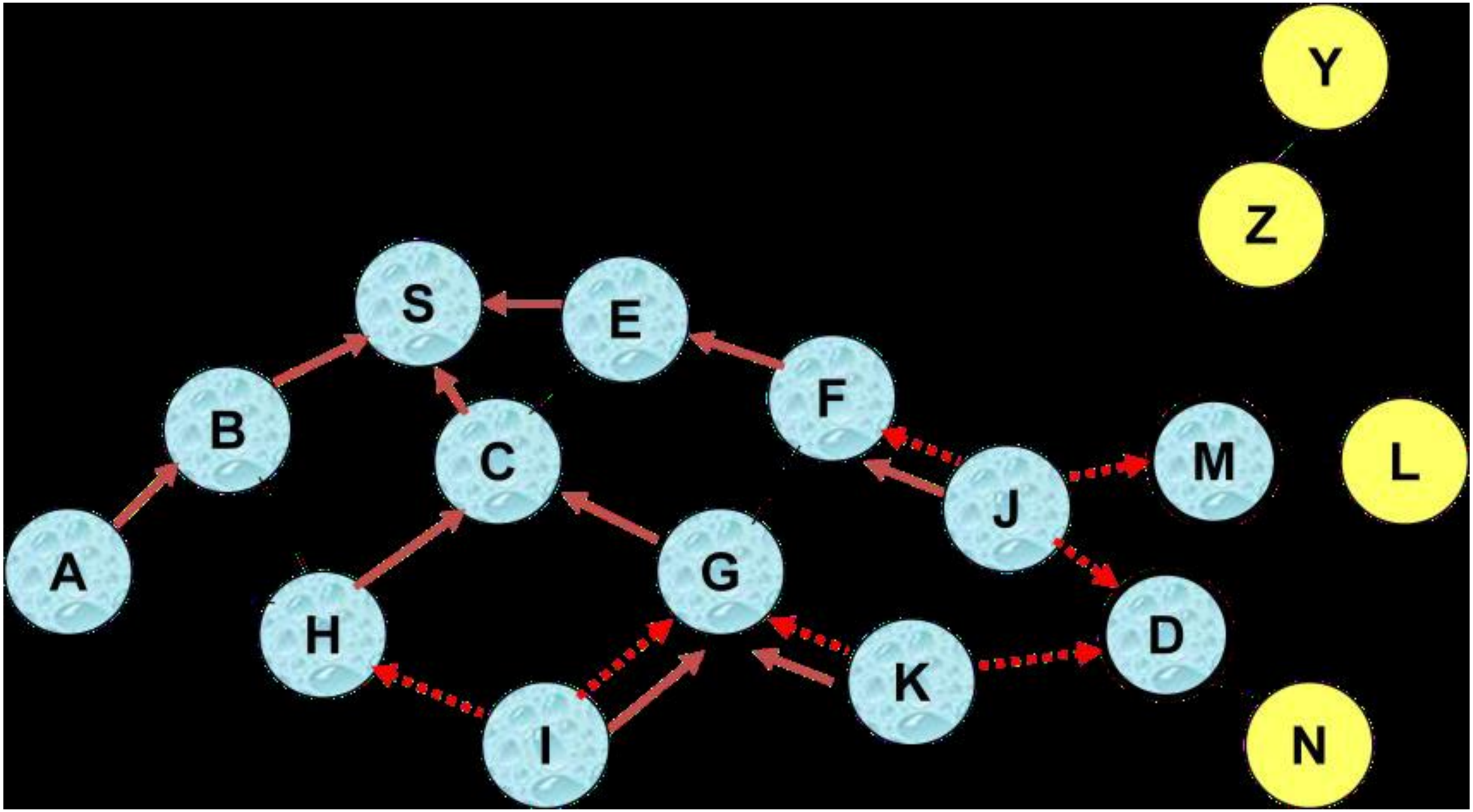


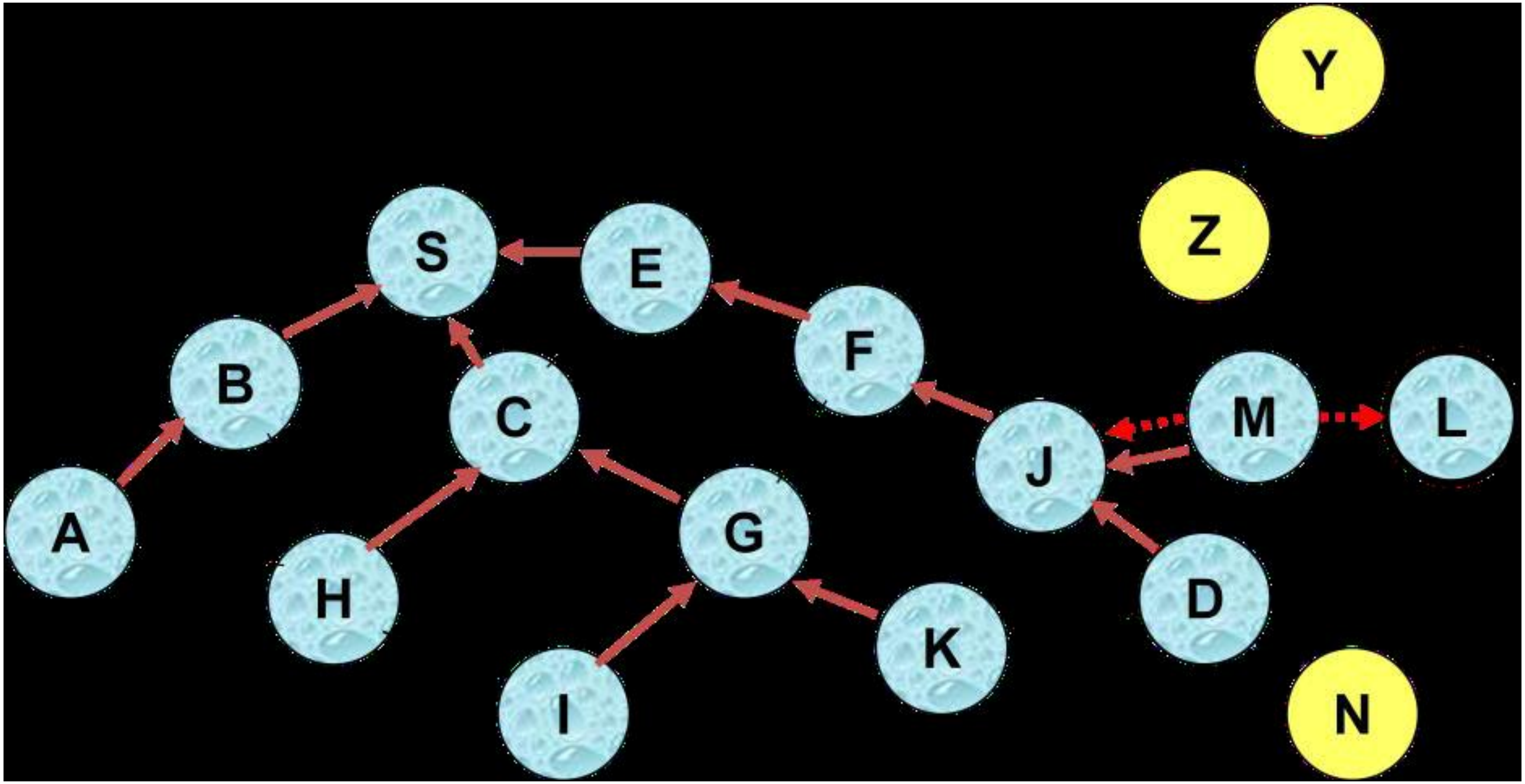
← Represents links on Reverse Path

Reverse Path Setup in AODV



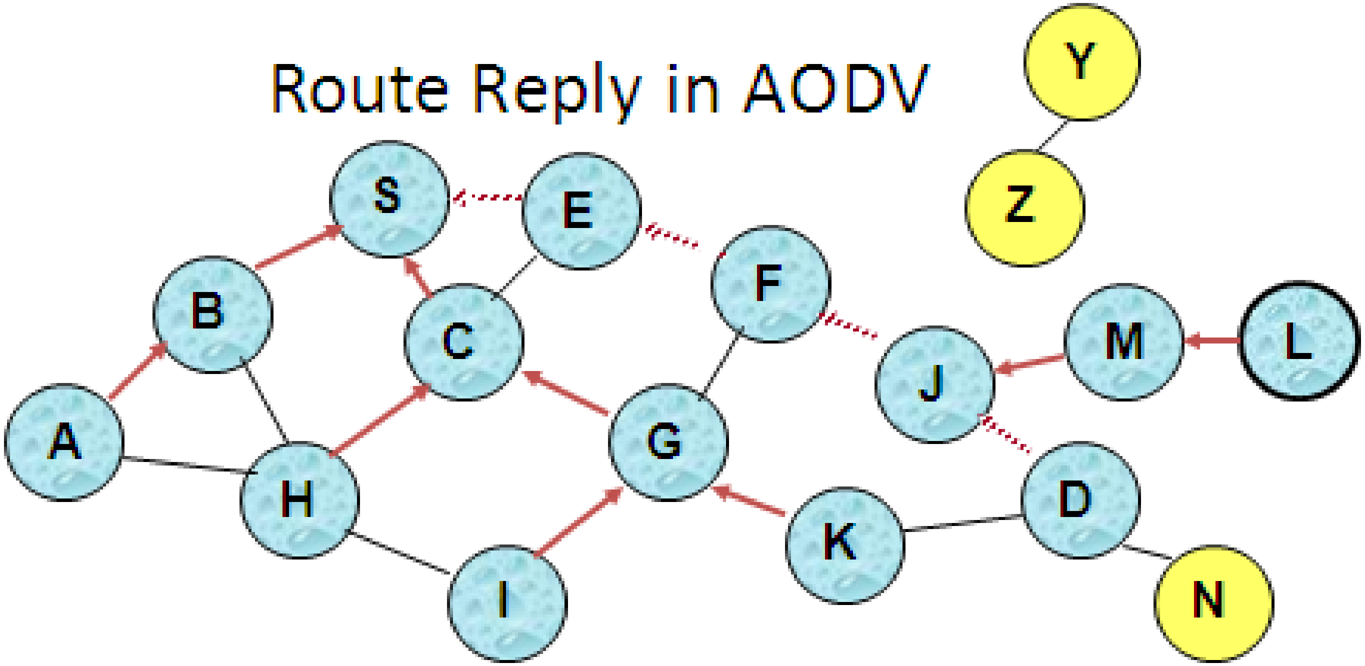
Node C receives RREQ from G and H, but does not forward it again, because node C has already forwarded RREQ once


Y
Z
S
E
B
C
F
M
L
J
A
G
H
D
K
I
N



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

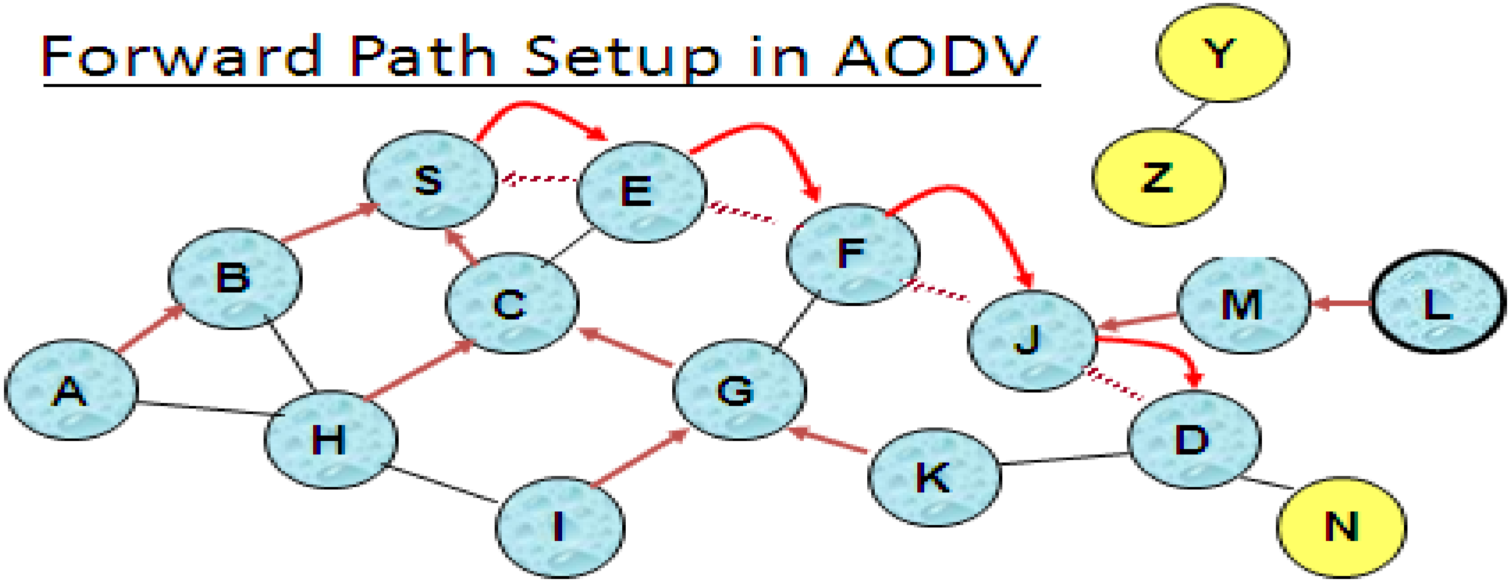
Route Reply in AODV



↶···· Represents links on path taken by RREP

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

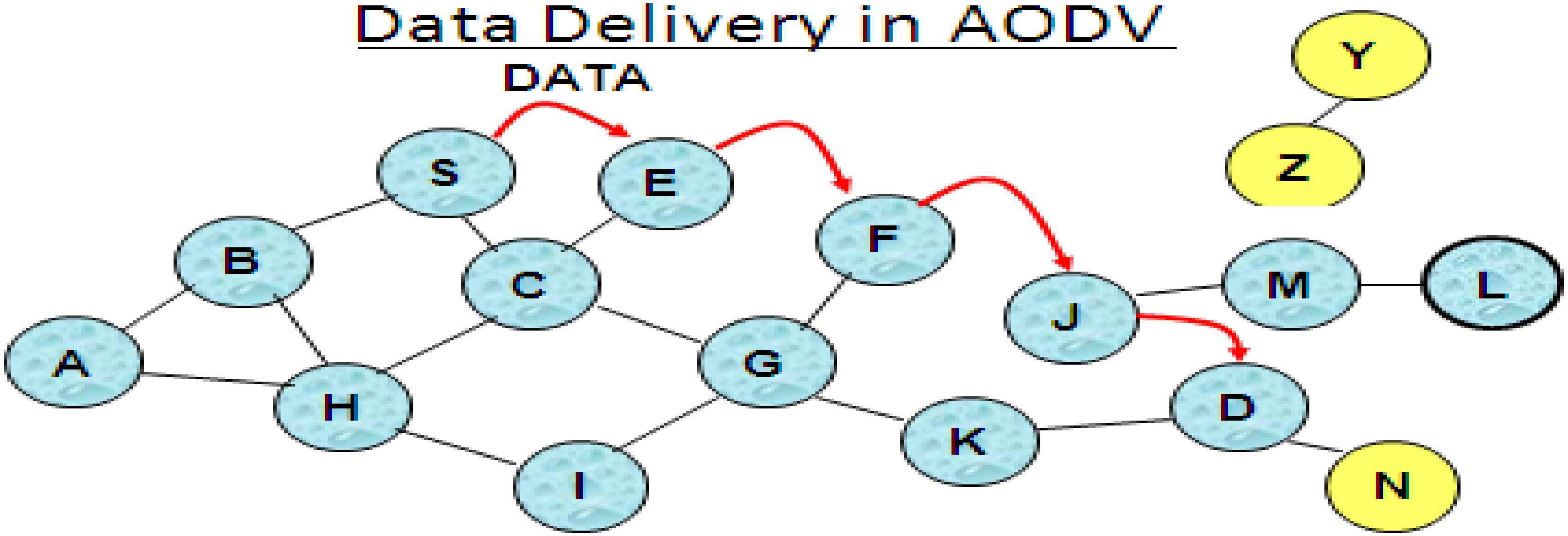
Forward Path Setup in AODV



Forward links are setup when RREP travels along the reverse path

↷ Represents a link on the forward path

Data Delivery in AODV



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

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

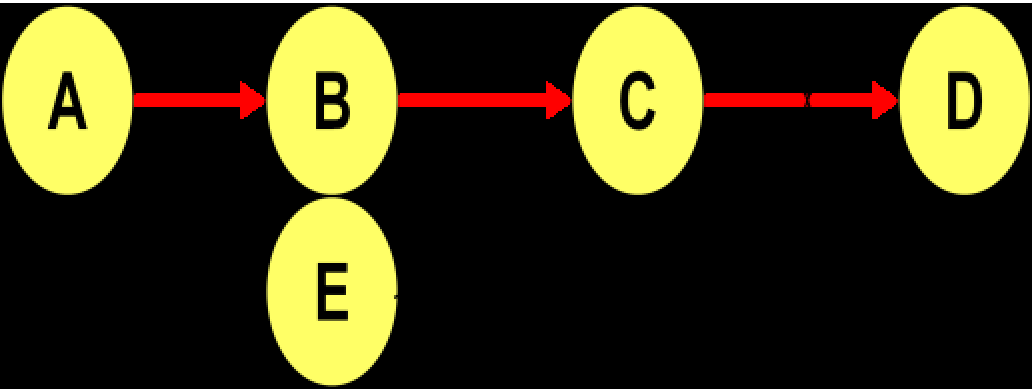
- Sequence numbers are used in AODV to avoid using old/broken routes and to determine which route is newer. Also, it prevents formation of loops.

Assume that A does not know about failur of link C-D because RERR sent by C is lost.

Now C performs a route discovery for D.

Node A receives the RREQ (say, via path C-E-A)

Node A will reply since A knows a route to D via node B resulting



in a loop (for instance, C-E-A-B-C)

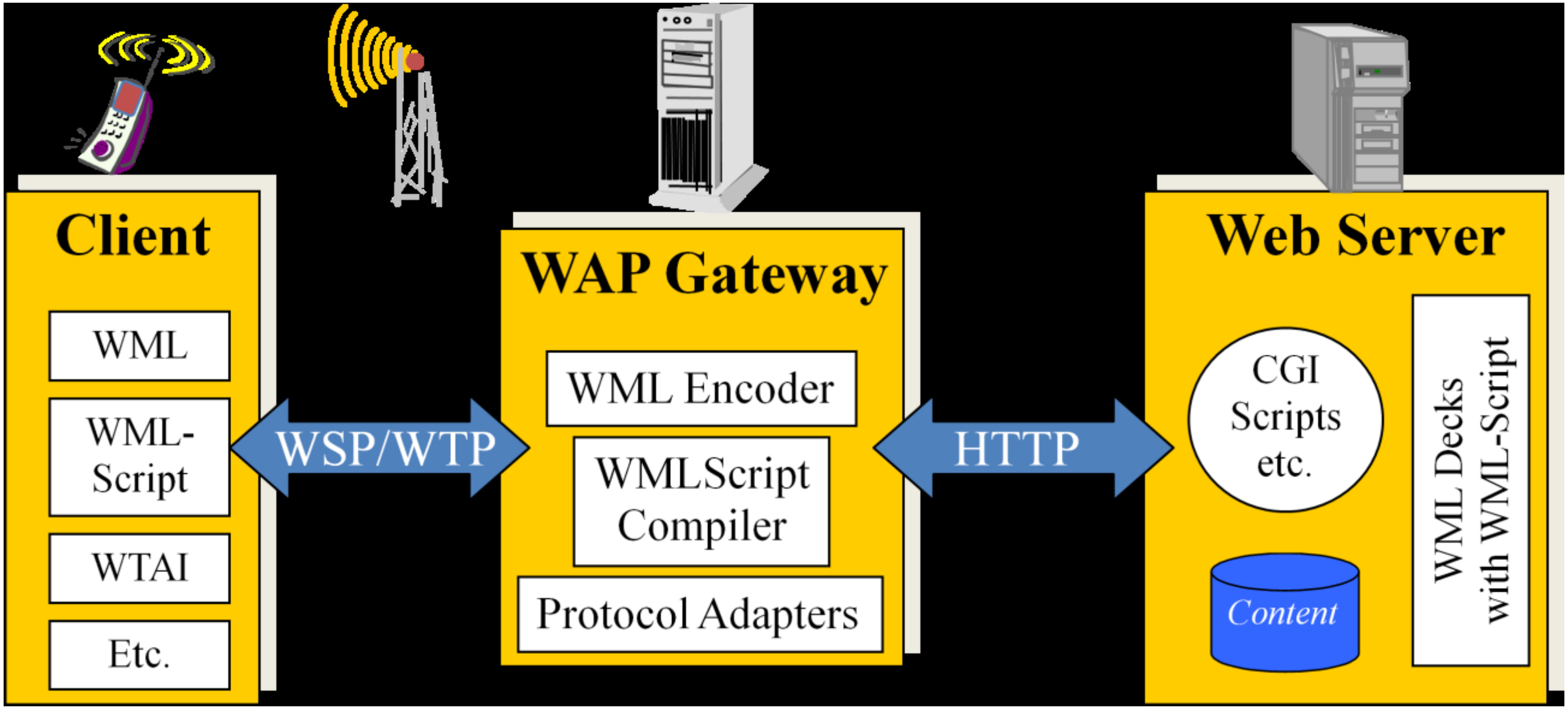
- Neighboring nodes periodically exchange hello message and absence of hello message indicates a link failure. When node X is unable to forward packet P (from node S to node D) on link (X,Y), it generates a **RERR message. Node X increments the destination sequence number for D cached at node X.**
-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

- ***Protocols and Tools:***
  - ***Wireless Application Protocol-WAP***
    - ***(Introduction Protocol architecture, and treatment of protocols of all layers),***
  - ***Bluetooth (User scenarios, physical layer, MAC layer, networking, security, link management)***
  - ***J2ME.***

The Wireless Application Protocol (WAP)

- The Wireless Application Protocol (WAP) is an open, global specification that empowers mobile users with wireless devices to easily access and interact with information and services instantly.

The Wireless Application Protocol (WAP)



The Wireless Application Protocol (WAP)

- WAP is a global standard and is not controlled by any single company.
- Ericsson, Nokia, Motorola, and Unwired Planet founded the **WAP Forum in the summer of 1997 with the initial purpose of defining an industry-wide specification for developing applications over wireless communications networks.**
- **The WAP specifications define a set of protocols in application, session, transaction, security, and transport layers, which enable operators, manufacturers, and applications providers to meet the challenges in advanced wireless service differentiation and fast/flexible service creation.**

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.

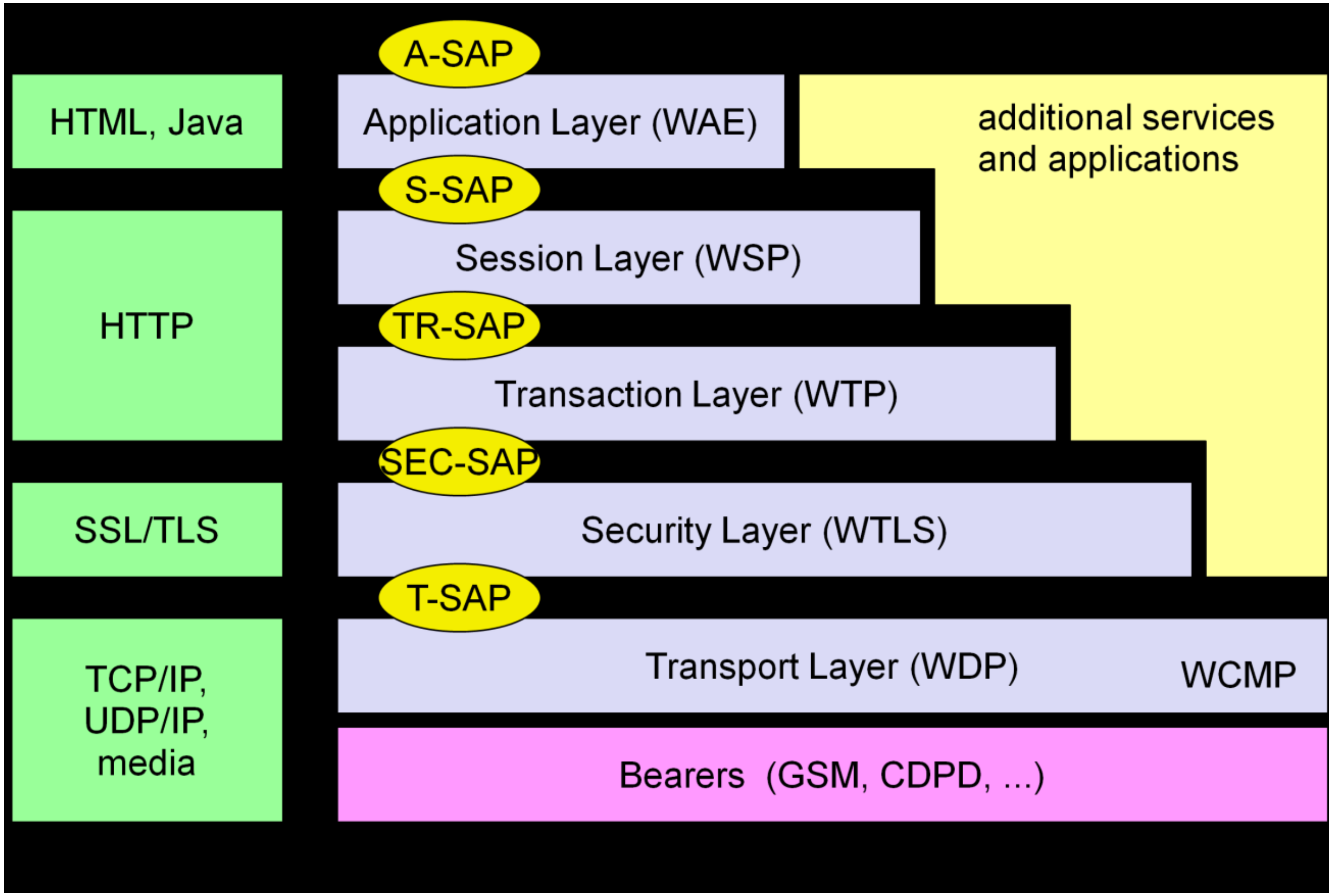
Why Choose WAP?

- In the past, wireless Internet access has been limited by the capabilities of handheld devices and wireless networks.
- WAP utilizes Internet standards such as XML, user datagram protocol (UDP), and Internet protocol (IP). Many of the protocols are based on Internet standards such as hypertext transfer protocol (HTTP) and TLS but have been optimized for the unique constraints of the wireless environment: low bandwidth, high latency, and less connection stability.
- Internet standards such as hypertext markup language (HTML), HTTP, TLS and transmission control protocol (TCP) are inefficient over mobile networks, requiring large amounts of mainly text-based data to be sent. Standard HTML content cannot be effectively displayed on the small-size screens of pocket-sized mobile phones and pagers.
- WAP utilizes binary transmission for greater compression of data and is optimized for long latency and low bandwidth. WAP sessions cope with intermittent coverage and can operate over a wide variety of wireless transports.

WAP Architecture

- The following 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.**

WAP Architecture



WAP Architecture

- **WDP:**
  - The WAP datagram protocol (WDP) and the additional Wireless control message protocol (WCMP) is the transport layer that sends and receives messages via any available bearer network, including SMS, USSD, CSD, CDPD, IS–136 packet data, and GPRS.
  - The *transport layer service access point (T-SAP) is the common interface to be used by higher layers independent of the underlying network.*

WAP Architecture

- **WTLS:**
  - The next higher layer, the security layer with its wireless transport layer security protocol WTLS offers its service at the *security SAP (SEC-SAP).*
  - *WTLS is based on transport layer security (TLS, formerly SSL, secure sockets layer). WTLS has been optimized for use in wireless networks with narrow-band channels.* It can offer data integrity, privacy, authentication, and (some) denial-of-service protection.

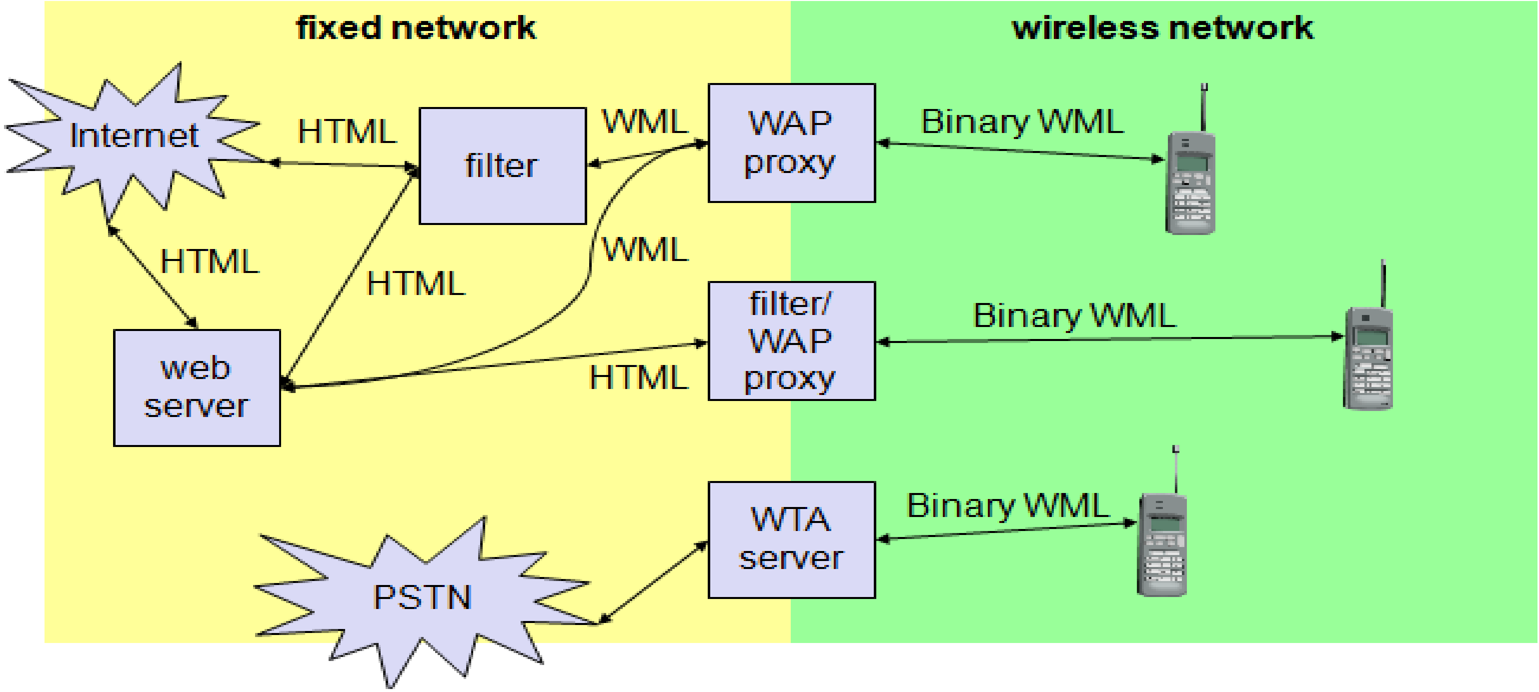
WAP Architecture

- **WTP:**
  - The WAP transaction protocol (WTP) layer provides transaction support, adding reliability to the datagram service provided by WDP at the *transaction SAP (TR-SAP).*
- **WSP:**
  - The session layer with the wireless session protocol (WSP) currently offers two services at the session-SAP (S-SAP), one connection-oriented and one connectionless if used directly on top of WDP. A special service for browsing the web (WSP/B) has been defined that offers HTTP/1.1 functionality, long-lived session state, session suspend and resume, session migration and other features needed for wireless mobile access to the web.
- **WAE:**
  - The application layer with the wireless application environment (WAE) offers a framework for the integration of different www and mobile telephony applications.

Working of WAP

- WAP does not always force all applications to use the whole protocol architecture.
- Applications can use only a part of the architecture.
- For example, 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.**

Working of WAP



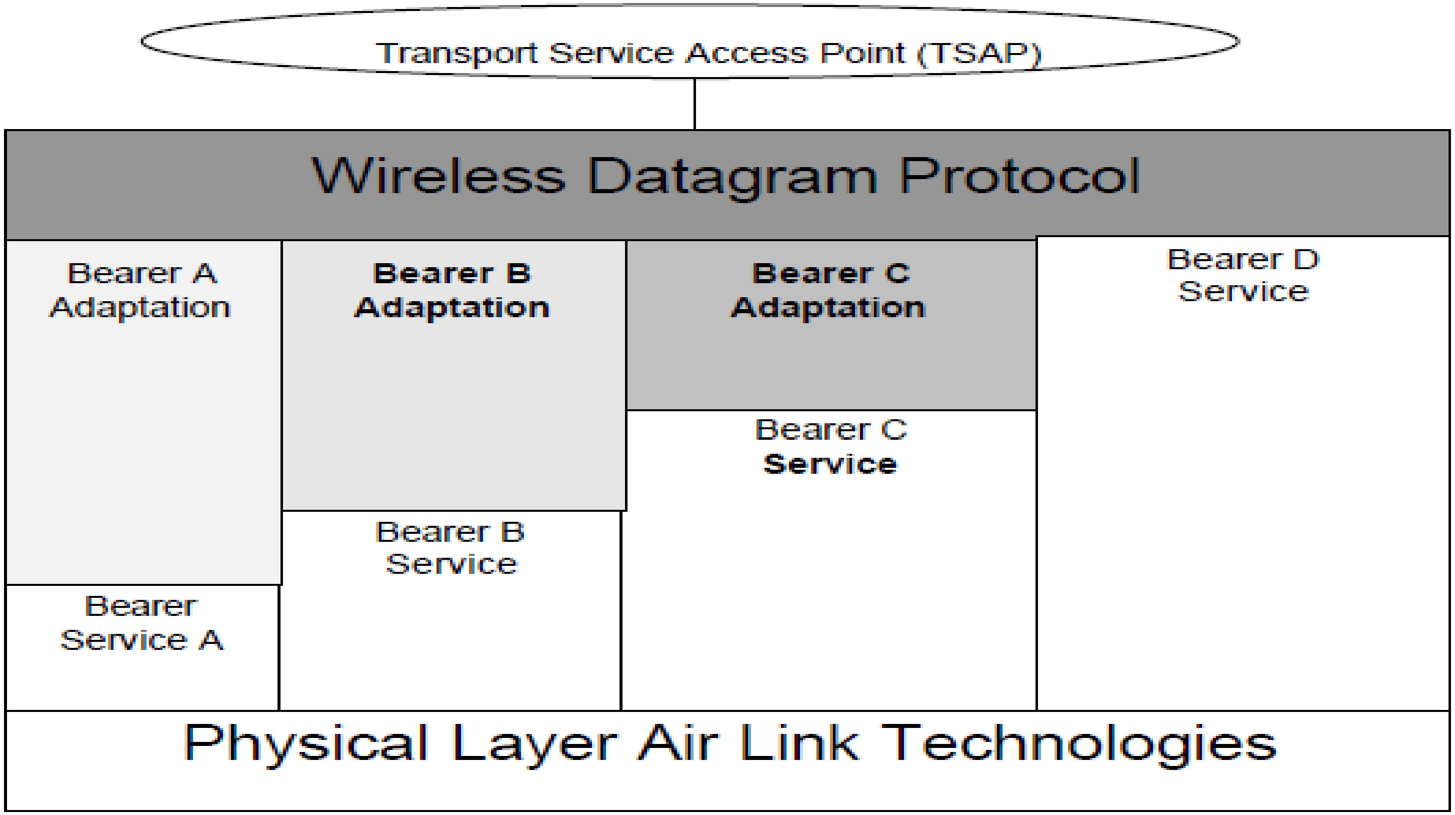
Working of WAP

- Different scenarios are possible for the integration of WAP components into existing wireless and fixed networks.
- On the left side, different fixed networks, such as the traditional internet and the public switched telephone network (PSTN), are shown.
- One cannot change protocols and services of these existing networks so several new elements will be implemented between these networks and the WAP-enabled wireless, mobile devices in a wireless network on the right-hand side.

Wireless Datagram Protocol (WDP)

- **Wireless Datagram Protocol defines the movement of information from receiver to the sender and resembles the User Datagram Protocol in the Internet protocol suite.**

WDP



Wireless Datagram Protocol Architecture

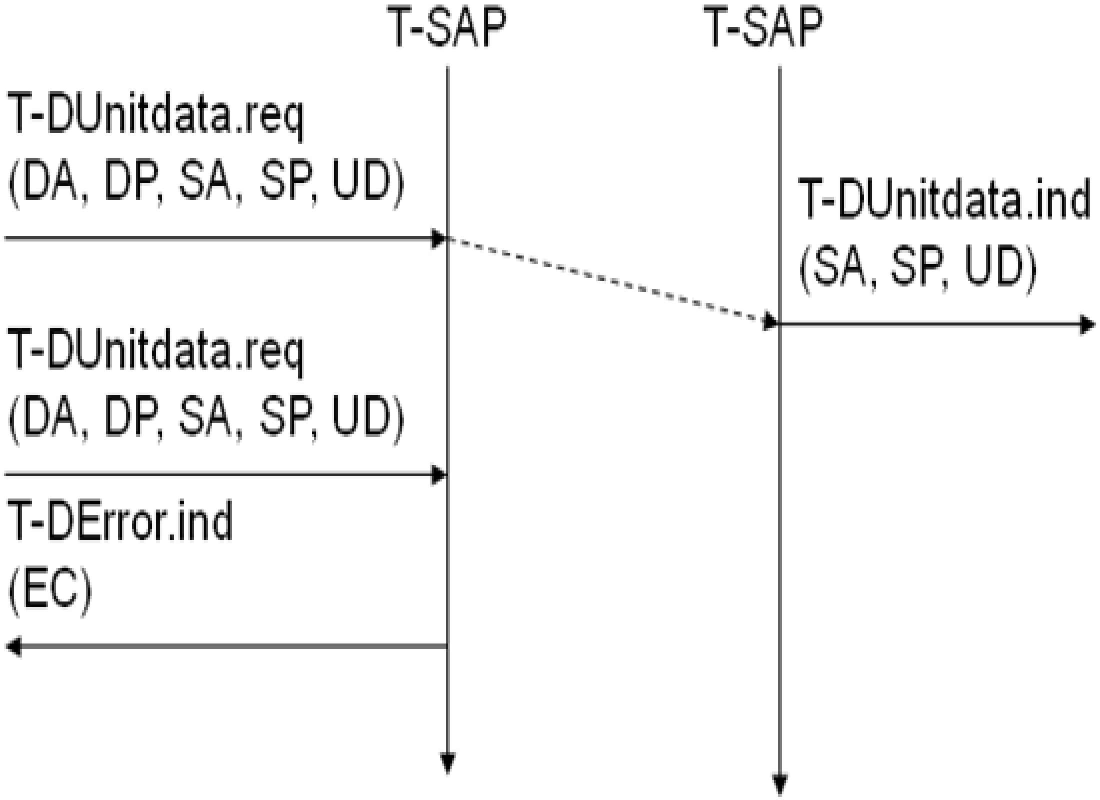
WDP

- WDP offers a consistent service at the Transport Service Access Point to the upper layer protocol of WAP.
- This consistency of service allows for applications to operate transparently over different available bearer services.
- WDP can be mapped onto different bearers, with different characteristics. In order to optimize the protocol with respect to memory usage and radio transmission efficiency, the protocol performance over each bearer may vary.

WDP

- 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.
- Destination and source address are unique addresses for the receiver and sender of the user data.
- These could be MSISDNs (i.e., a telephone number), IP addresses, or any other unique identifiers.
- The T-DUnitdata.ind service primitive indicates the reception of data. Here destination address and port are only optional parameters.

WDP



SAP: Service Access Point

DA: Destination Address

DP: Destination Port

SA: Source Address

SP: Source Port

UD: User Data

EC: Error Code

WDP service primitives

WDP

- If a higher layer requests a service the WDP cannot fulfil, this error is indicated with the T-DError.ind service primitive.
- An error code (EC) is returned indicating the reason for the error to the higher layer. WDP is not allowed to use this primitive to indicate problems with the bearer service.
- It is only allowed to use the primitive to indicate local problems, such as a user data size that is too large.
- If any errors happen when WDP datagrams are sent from one WDP entity to another, the wireless control message protocol (WCMP) provides error handling mechanisms for WDP and should therefore be implemented.
- WCMP contains control messages that resemble the internet control message protocol messages and can also be used for diagnostic and informational purposes.
- WCMP can be used by WDP nodes and gateways to report errors.

WDP

- Typical WCMP messages are destination unreachable (route, port, address unreachable), parameter problem (errors in the packet header), message too big, reassembly failure, or echo request/reply.
- An additional WDP management entity supports WDP and provides information about changes in the environment, which may influence the correct operation of WDP.

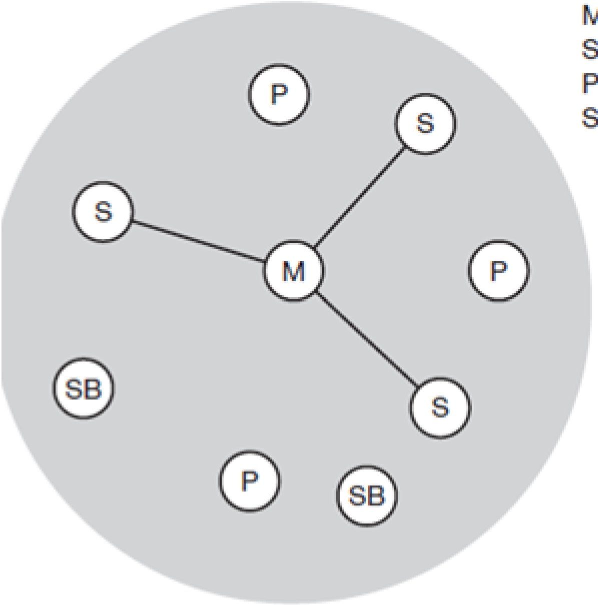
BLUETOOTH

- "Bluetooth" was the nickname of Harald Blåtland II, king of Denmark from 940 to 981, who united all of Denmark and part of Norway under his rule. **Bluetooth is a proprietary open wireless technology standard for exchanging data over short distances (using short wavelength radio transmissions in the ISM band from 2400-2480 MHz) from fixed and mobile devices, creating personal area networks (PANs) with high levels of security. The Bluetooth technology aims at so-called ad-hoc piconets, which are local area networks with a very limited coverage and without the need for an infrastructure.**

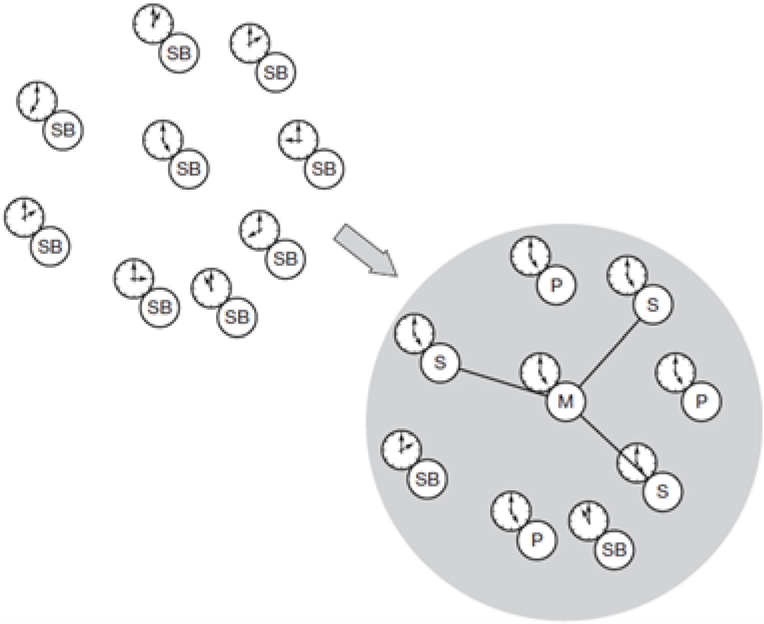
▪

- Device synchronization: Seamless connectivity among PDAs, computers, and mobile phones allows applications to update information on multiple devices automatically when data on any one device changes.
- □ Peripheral connectivity.
- □ Car kits: Hands-free packages enable users to access phones and other devices without taking their hands off the steering wheel
- □ Mobile payments: Your Bluetooth-enabled phone can communicate with a Bluetooth-enabled vending machine to buy a can of Diet Pepsi, and put the charge on your phone bill.
- The 802.11b protocol is designed to connect relatively large devices with lots of power and speed, such as desktops and laptops, where devices communicate at up to 11 Mbit/sec, at greater distances (up to 300 feet, or 100 meters). By contrast, Bluetooth is designed to connect small devices like PDAs, mobile phones, and peripherals at slower speeds (1 Mbit/sec), within a shorter range (30 feet, or 10 meters), which reduces power requirements. Another major difference is that 802.11b wasn't designed for voice communications, while any Bluetooth connection can support both data and voice communications.
- User scenarios

- Networking in Bluetooth
- Bluetooth operates on 79 channels in the 2.4 GHz band with 1 MHz carrier spacing. Each device performs frequency hopping with 1,600 hops/s in a pseudo random fashion. A piconet is a collection of Bluetooth devices which are synchronized to the same hopping sequence. One device in the piconet can act as **master (M), all other devices connected to the master must act as slaves (S). The master determines the hopping pattern in the piconet and the slaves have to synchronize to this pattern. Each piconet has a unique hopping pattern. If a device wants to participate it has to synchronize to this. A typical piconet is shown below:**



M = Master
S = Slave
P = Parked
SB = Standby

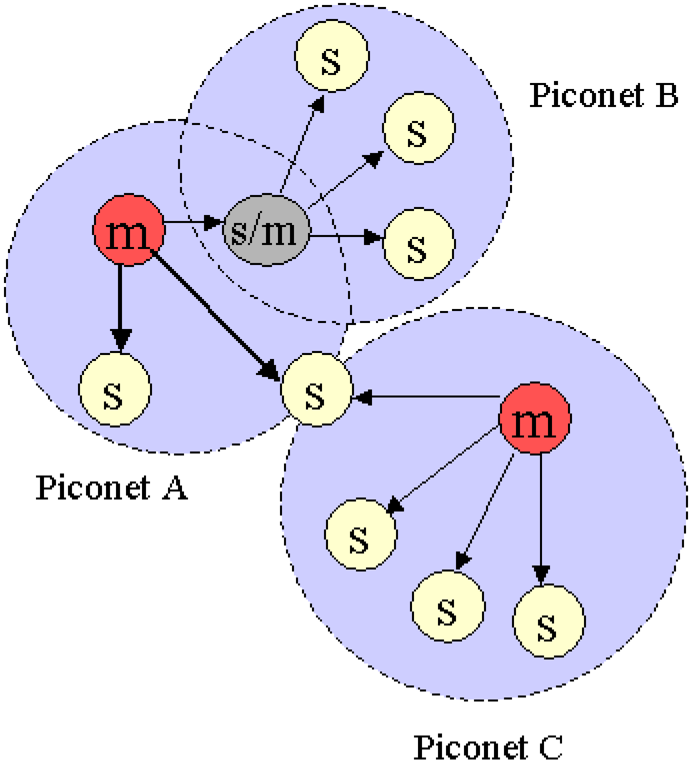


Simple Bluetooth piconet

Forming a Bluetooth piconet

- Parked devices (P) can not actively participate in the piconet (i.e., they do not have a connection), but are known and can be reactivated within some milliseconds. Devices in stand-by (SB) do not participate in the piconet. Each piconet has exactly one master and up to seven simultaneous slaves. More than 200 devices can be parked. The first step in forming a piconet involves a master sending its clock and device ID. All the Bluetooth devices have the same capability to become a master or a slave and two or three devices are sufficient to form a piconet. The unit establishing the piconet automatically becomes the master, all other devices will be slaves. The hopping pattern is determined by the device ID, a 48-bit worldwide unique identifier.
- The phase in the hopping pattern is determined by the master’s clock. After adjusting the internal clock according to the master a device may participate in the piconet. All active devices are assigned a 3-bit **active member address (AMA). All parked devices use an 8-bit parked member address (PMA). Devices in stand-by do not need an address.**

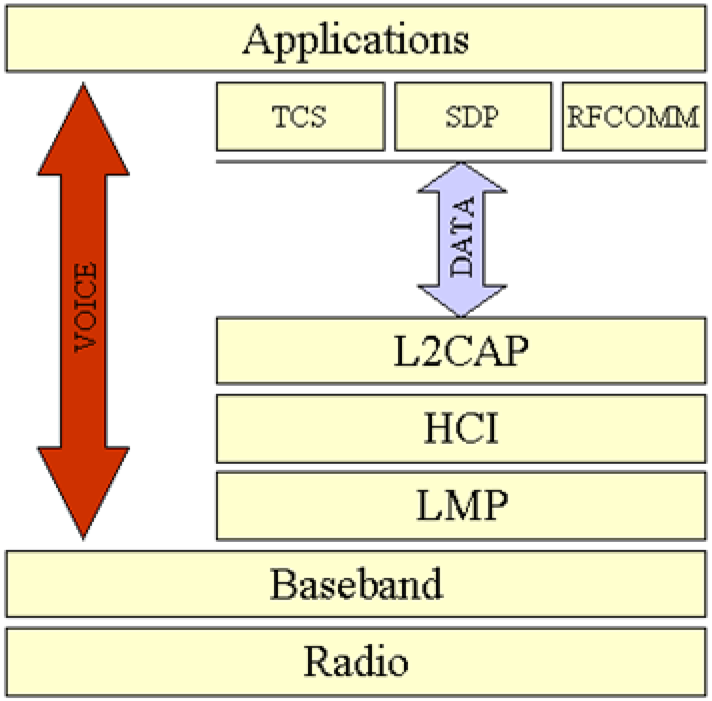
- A device in one piconet can communicate to another device in another piconet, forming a **scatternet. A master in one piconet may be a slave in another piconet. Both piconets use a different hopping sequence, always determined by the master of the piconet. Bluetooth applies FH-CDMA for separation of piconets.**

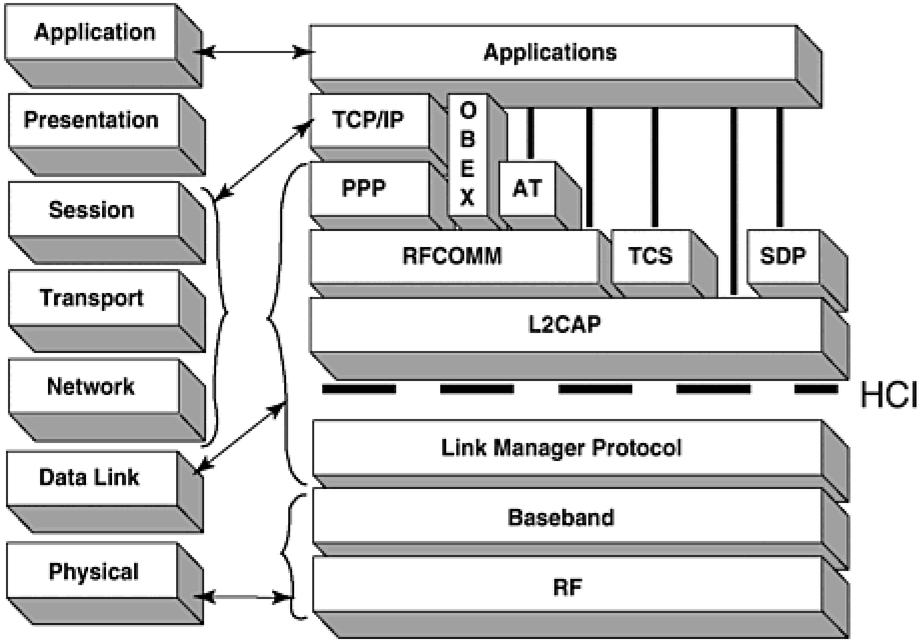

s
s
Piconet B
m
s/m
s
s
s
m
Piconet A
s
s
s
Piconet C

- A collision occurs if two or more piconets use the same carrier frequency at the same time. This will probably happen as the hopping sequences are not coordinated. If a device wants to participate in more than one piconet, it has to synchronize to the hopping sequence of the piconet it wants to take part in. If a device acts as slave in one piconet, it simply starts to synchronize with the hopping sequence of the piconet it wants to join. After synchronization, it acts as a slave in this piconet and no longer participates in its former piconet. To enable synchronization, a slave has to know the identity of the master that determines the hopping sequence of a piconet. Before leaving one piconet, a slave informs the current master that it will be unavailable for a certain amount of time. The remaining devices in the piconet continue to communicate as usual.

Bluetooth Protocol Stack

- The Bluetooth protocol stack can be divided into a **core specification, which describes the protocols from physical layer to the data link control together with management functions, and profile specifications describing many protocols and functions needed to adapt the wireless Bluetooth technology to legacy and new applications.**
- A high-level view of the architecture is shown. The responsibilities of the layers in this stack are as follows:
- □ *The radio layer is the physical wireless connection. To avoid interference with other devices that communicate in the ISM band, the modulation is based on fast frequency hopping. Bluetooth divides the 2.4 GHz frequency band into 79 channels 1 MHz apart (from 2.402 to 2.480 GHz), and uses this spread spectrum to hop from one channel to another, up to 1600*
- times a second. The standard wavelength range is 10 cm to 10 m, and can be extended to 100 m by increasing transmission power.


Applications
TCS
SDP
RFCOMM
VOICE
DATA
L2CAP
HCI
LMP
Baseband
Radio

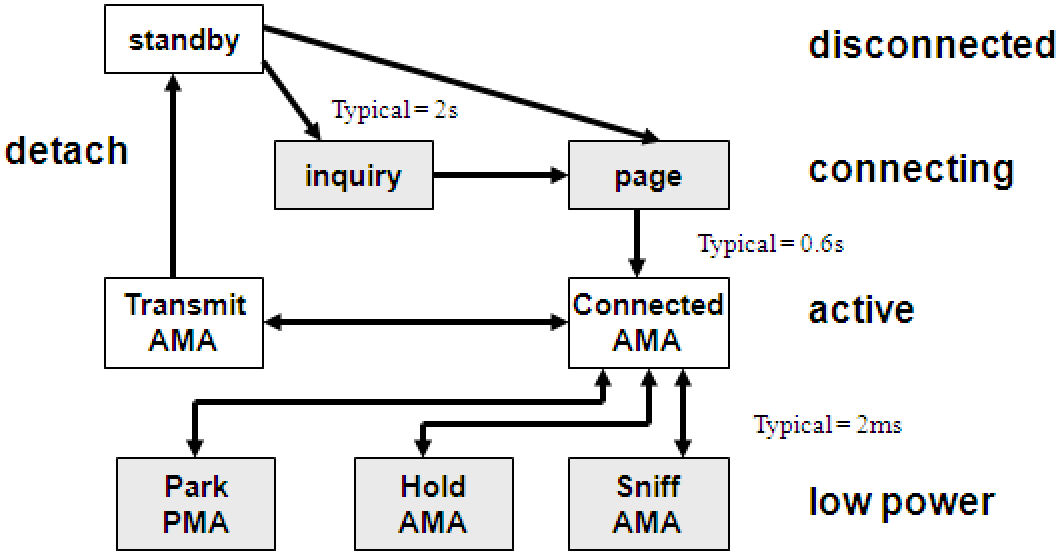

Application
Presentation
Session
Transport
Network
Data Link
Physical
Applications
TCP/IP
OBEX
PPP
AT
RFCOMM
TCS
SDP
L2CAP
HCI
Link Manager Protocol
Baseband
RF

- The baseband layer is responsible for controlling and sending data packets over the radio link. It provides transmission channels for both data and voice. The baseband layer maintains Synchronous Connection-Oriented (SCO) links for voice and Asynchronous Connectionless (ACL) links for data. SCO packets are never retransmitted but ACL packets are, to ensure data integrity. SCO links are point-to-point symmetric connections, where time slots are reserved to guarantee timely transmission. A slave device is allowed to respond during the time slot immediately following an SCO transmission from the master. A master can support up to three SCO links to a single slave or to multiple slaves, and a single slave can support up to two SCO links to different slaves. Data transmissions on ACL links, on the other hand, are established on a per-slot basis (using slots not reserved for SCO links). ACL links support point-to-multipoint transmissions. After an ACL transmission from the master, only a slave addressed specifically may respond during the next time slot; if no device is addressed, the message is treated as a broadcast.

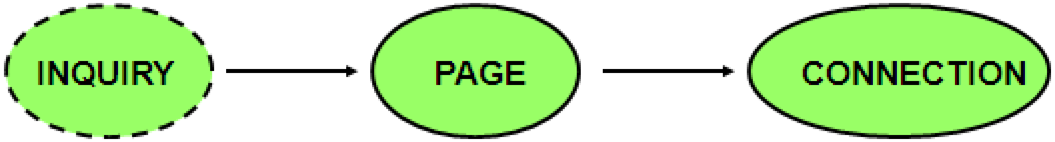
- by the application, or through certain support protocols provided to ease the burden on application programmers.
- □ *The Logical Link Control and Adaptation Protocol (L2CAP) receives application data and adapts it to the Bluetooth format. Quality of Service (QoS) parameters are exchanged at this layer.*

- Link Manager Protocol
- The link manager protocol (LMP) manages various aspects of the radio link between a master and a slave and the current parameter setting of the devices. LMP enhances baseband functionality, but higher layers can still directly access the baseband. The following groups of functions are covered by the LMP:
- □ **Authentication, pairing, and encryption: Although basic authentication is handled in the baseband, LMP has to control the exchange of random numbers and signed responses. LMP is not directly involved in the encryption process, but sets the encryption mode (no encryption, point-to-point, or broadcast), key size, and random speed.**
- □ **Synchronization: Precise synchronization is of major importance within a Bluetooth network. The clock offset is updated each time a packet is received from the master.**
- □ **Capability negotiation: Not only the version of the LMP can be exchanged but also information about the supported features. Not all Bluetooth devices will support all features that are described in the standard, so devices have to agree the usage of, e.g., multi-slot packets, encryption, SCO links, voice encoding, park/sniff/hold mode, HV2/HV3 packets etc.**
- □ **Quality of service negotiation: Different parameters control the QoS of a Bluetooth device at these lower layers. The poll interval, i.e., the maximum time between transmissions from a master to a particular slave, controls the latency and transfer capacity. A master can also limit the number of slots available for slaves' answers to increase its own bandwidth.**
- □ **Power control: A Bluetooth device can measure the received signal strength. Depending on this signal level the device can direct the sender of the measured signal to increase or decrease its transmit power.**
- □ **Link supervision: LMP has to control the activity of a link, it may set up new SCO links, or it may declare the failure of a link.**
- □ **State and transmission mode change: Devices might switch the master/slave role, detach themselves from a connection, or change the operating mode**

- Bluetooth defines several low-power states for a device. The following figure shows the major states of a Bluetooth device and typical transitions. Every device, which is currently not participating in a piconet (and not switched off), is in standby mode. This is a low-power mode where only the native clock is running. The next step towards the inquiry mode can happen in two different ways. Either a device wants to establish a piconet or a device just wants to listen to see if something is going on.


standby
disconnected
Typical=2s
detach
inquiry
page
connecting
Typical=0.6s
Transmit AMA
Connected AMA
active
Typical=2ms
Park PMA
Hold AMA
Sniff AMA
low power

- A device wants to establish a piconet: A user of the device wants to scan for other devices in the radio range. The device starts the inquiry procedure by sending an inquiry access code (IAC) that is common to all Bluetooth devices. The IAC is broadcast over 32 so-called wake-up carriers in turn.
- □ Devices in standby that listen periodically: Devices in standby may enter the inquiry mode periodically to search for IAC messages on the wake-up carriers. As soon as a device detects an inquiry it returns a packet containing its device address and timing information required by the master to initiate a connection. From that moment on, the device acts as slave.


INQUIRY
PAGE
CONNECTION

- During the page state two different roles are defined. After finding all required devices the master is able to set up connections to each device, i.e., setting up a piconet. As soon as a device synchronizes to the hopping pattern of the piconet it also enters the connection state. The connection state comprises the active state and the low power states: park, sniff, and hold. In the active state the slave participates in the piconet by listening, transmitting, and receiving. ACL and SCO links can be used. A master periodically synchronizes with all slaves. All devices being active must have the 3-bit active member address (AMA). To save battery power, a Bluetooth device can go into one of three low power states:

- L2CAP
- The logical link control and adaptation protocol (L2CAP) is a data link control protocol on top of the baseband layer offering logical channels between Bluetooth devices with QoS properties. L2CAP is available for ACLs only.
- L2CAP provides three different types of logical channels that are transported via the ACL between master and slave:
- □ *Connectionless: These unidirectional channels are typically used for broadcasts from a master to its slave(s).*
- □ *Connection-oriented: Each channel of this type is bi-directional and supports QoS flow specifications for each direction. These flow specs follow RFC 1363 and define average/peak data rate, maximum burst size, latency, and jitter.*

- Signaling: This third type of logical channel is used to exchanging signaling messages between L2CAP entities.
- Each channel can be identified by its **channel identifier (CID). Signaling channels always use a CID value of 1, a CID value of 2 is reserved for connectionless channels. For connection-oriented channels a unique CID (>= 64) is dynamically assigned at each end of the channel to identify the connection.**
- The following figure shows the three packet types belonging to the three logical channel types.

- The length field indicates the length of the payload (plus PSM for connectionless PDUs). The CID has the multiplexing/demultiplexing function. For connectionless PDUs a protocol/service multiplexor (PSM) field is needed to identify the higher layer recipient for the payload. For connection-oriented PDUs the CID already fulfills this function. Several PSM values have been defined, e.g., 1 (SDP), 3 (RFCOMM), 5 (TCS-BIN). Values above 4096 can be assigned dynamically. The payload of the signaling PDU contains one or more commands. Each command has its own code (e.g., for command reject, connection request, disconnection response etc.) and an ID that matches a request with its reply. The length field indicates the length of the data field for this command.

- The first step, called pairing, is necessary if two Bluetooth devices have never met before. To set up trust between the two devices a user can enter a secret PIN into both devices. This PIN can have a length of up to 16 byte. Based on the PIN, the device address, and random numbers, several keys can be computed which can be used as link key for authentication. The authentication is a challenge-response process based on the link key, a random number generated by a verifier (the device that requests authentication), and the device address of the claimat (the device that is authenticated).

- Based on the link key, and again a random number an encryption key is generated during the encryption stage of the security architecture. This key has a maximum size of 128 bits and can be individually generated for each transmission. Based on the encryption key, the device address and the current clock a payload key is generated for ciphering user data. The payload key is a stream of pseudo-random bits. The ciphering process is a simple XOR of the user data and the payload key**.**
- All Bluetooth-enabled devices must implement the Generic Access Profile, which contains all the Bluetooth protocols and possible devices. This profile defines a security model that includes three security modes.

Bluetooth Profiles

- Bluetooth profiles are intended to ensure interoperability among Bluetooth-enabled devices and applications from different manufacturers and vendors. A profile defines the roles and capabilities for specific types of applications. Profiles represent default solutions for a certain usage model. They use a selection of protocols and parameter set to form a basis for interoperability. Protocols can be seen as horizontal layers while profiles are vertical slices as shown below:

- The following basic profiles have been specified: generic access, service discovery, cordless telephony, intercom, serial port, headset, dialup networking, fax, LAN access, generic object exchange, object push, file transfer, and synchronization. Additional profiles are: advanced audio distribution, PAN, audio video remote control, basic printing, basic imaging, extended service discovery, generic audio video distribution, hands-free, and hardcopy cable replacement. Some of the profiles are given below**:**
- The Generic Access Profile defines connection procedures, device discovery, and link management. It also defines procedures related to use of different security models and common format requirements for parameters accessible on the user interface level. At a minimum all Bluetooth devices must support this profile.

-

Java 2 Micro Edition (J2ME)

- Sun Microsystems defines J2ME as "a highly optimized Java run-time environment targeting a wide range of consumer products, including pagers, cellular phones, screen-phones, digital set-top boxes and car navigation systems." J2ME brings the cross-platform functionality of the Java language to smaller devices, allowing mobile wireless devices to share applications. Java 2 Micro Edition maintains the qualities that Java technology has become known for:
- □ built-in consistency across products in terms of running anywhere, anytime, on any device
- □ the power of a high-level object-oriented programming language with a large developer base;
- □ portability of code;
- □ safe network delivery; and
- □ upward scalability with J2SE and J2EE

- class libraries that are far more domain-specific than the class libraries provided in a configuration.

Configurations

- A configuration is a subset of profile. A configuration defines a Java platform for a “horizontal” category or grouping of devices with similar requirements on total memory budget and other hardware capabilities. More specifically, a configuration:
- **specifies the Java programming language features supported,**
- **specifies the Java virtual machine features supported,**
- **specifies the basic Java libraries and APIs supported.**

- To avoid fragmentation, there will be a very limited number of J2ME configurations. Currently, the goal is to define two standard J2ME configurations: