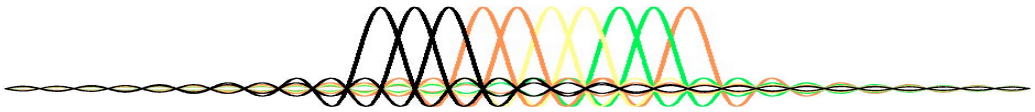


EHB 453, Introduction to Mobile Communications

Lecture 4: ALL IP Networking

Prof. Mustafa Ergen

Outline

- Deep Packet Inspection
- IP Header Compression
- IPSEC
- IP Authentication
- Mobile IP
- SIP
- IMS

Deep Packet Inspection

DPI: Deep Packet Inspection

- A packet is analogous to a physical postal mail message. The address on the outside of the envelope is analogous to the “packet header” and the information inside the envelope is analogous to the “payload.”

- DPI is analogous to taking action on that mail message not only based on the address on the envelope, but also making considerations based on the contents of the envelope.

DPI

DPI provides increased visibility, control, and service creation capabilities to global networks, increasing their operational efficiencies and utility.

- DPI is a significant innovation in networking that forms the foundation of many current and next generation services.

In the enterprise,

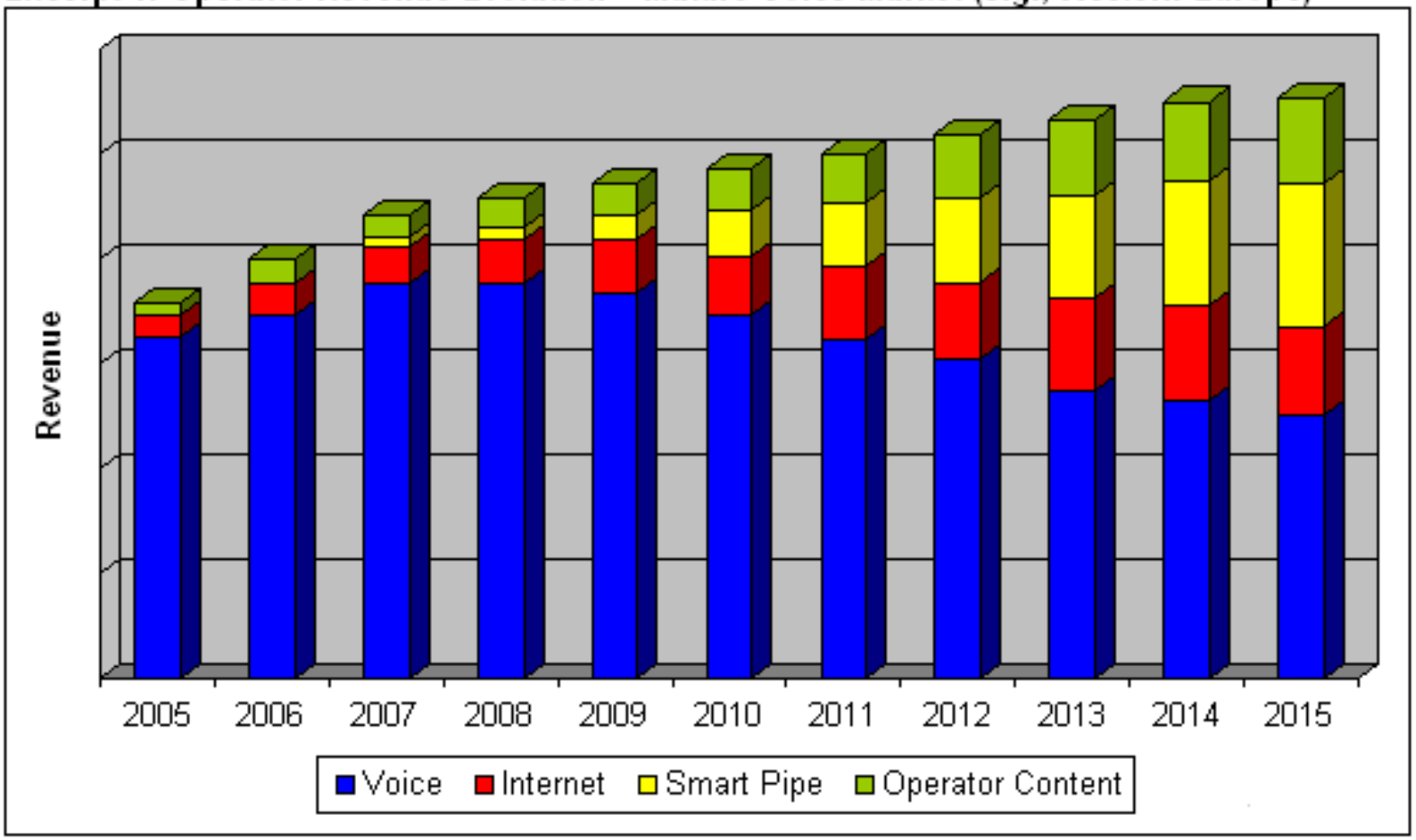
- intelligent switching and routing,
- next generation firewalls, intrusion detection & prevention (IDP),
- regulatory compliance
- leak prevention,
- traffic management,
- application delivery controllers, and more.

In Service Provider networks,

- DPI applications include subscriber-based services,
- content-based billing,
- tiered services,
- advanced P2P traffic management,
- enhanced security, and more.

Operator Revenue Evolution: Smart Pipe

Excerpt 1: Operator Revenue Evolution – Mature Voice Market (e.g., Western Europe)



Source: Unstrung Insider

Shallow Inspection vs Deep Inspection

- Shallow Inspection
  - Considers only the IP Header
  - Insufficient to reach any application specific information
- Deep Inspection
  - Considers also the Payload
  - Brings application awareness

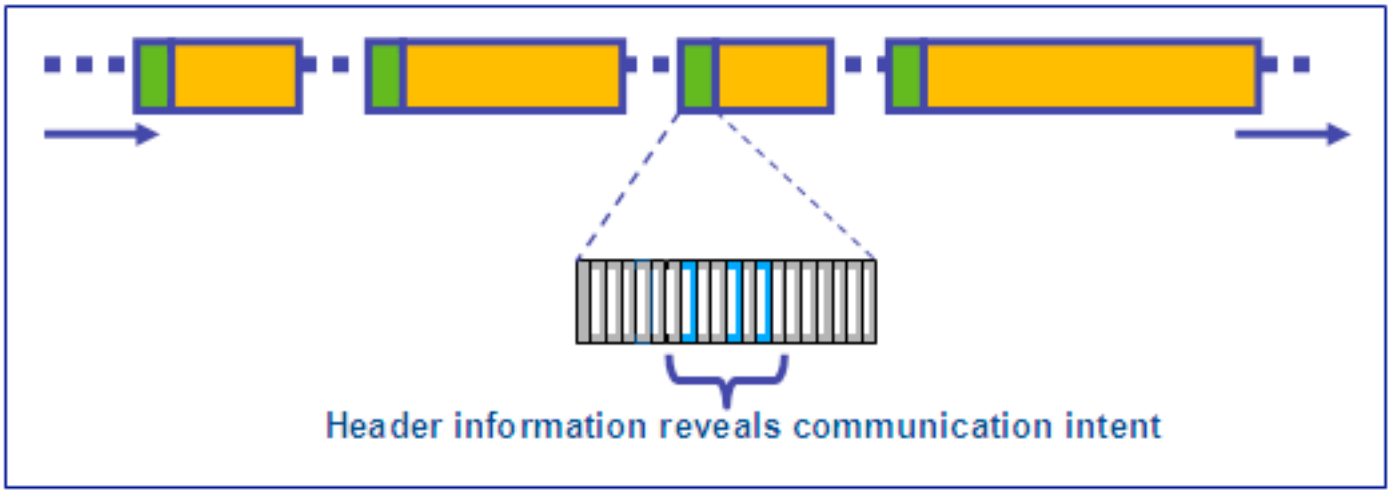


Figure 1: Shallow packet inspection – data from packet headers

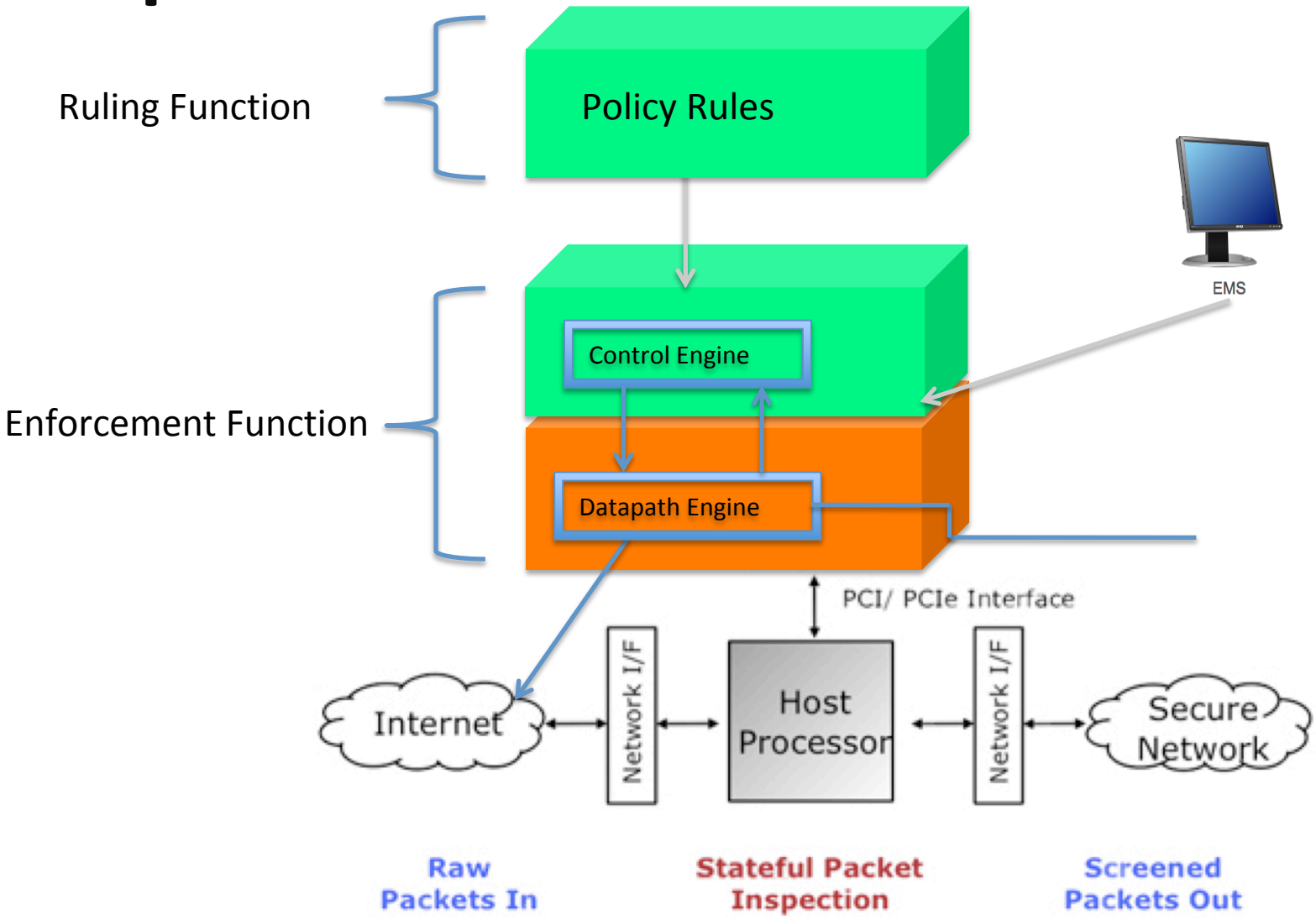
What is DPI?

The DPI box,

- agnostic to any access standard,
- scrutinizes each packet (including the data payload) as traverses,
- rejects, allows, or prioritizes the packet,
- differentiate the charging per session,
- learn about the session to offer related ad, location services, etc.
- The DPI box uses a **ruleset** that is implemented and updated by the operator.

- The **ruleset** is based upon ***signature***-based comparisons

Components


Ruling Function
Policy Rules
EMS
Enforcement Function
Control Engine
Datapath Engine
PCI/ PCIe Interface
Internet
Network I/F
Host Processor
Network I/F
Secure Network
Raw Packets In
Stateful Packet Inspection
Screened Packets Out

Encryption and Obfuscation

- In the DPI world, life is becoming much more difficult with the use of encryption
- RC4, the most popular encryption algorithm used by most P2P protocols. Since the key lengths used are very large, it is almost impossible to reverse engineer and gain some information, like guessing a password.
- The creators of several applications have chosen to conceal their operation by scrambling their related communications.
- This is obfuscation (concealing actions, by making things much more complex than necessary).
  - Ex: eMule (version 0.47c) and BitTorrent.
  - Proprietary technology (e.g., Skype).

Application and Protocol Signatures

- What is a signature? signatures are pattern recipes which are chosen for uniquely identifying an associated application (or protocol).

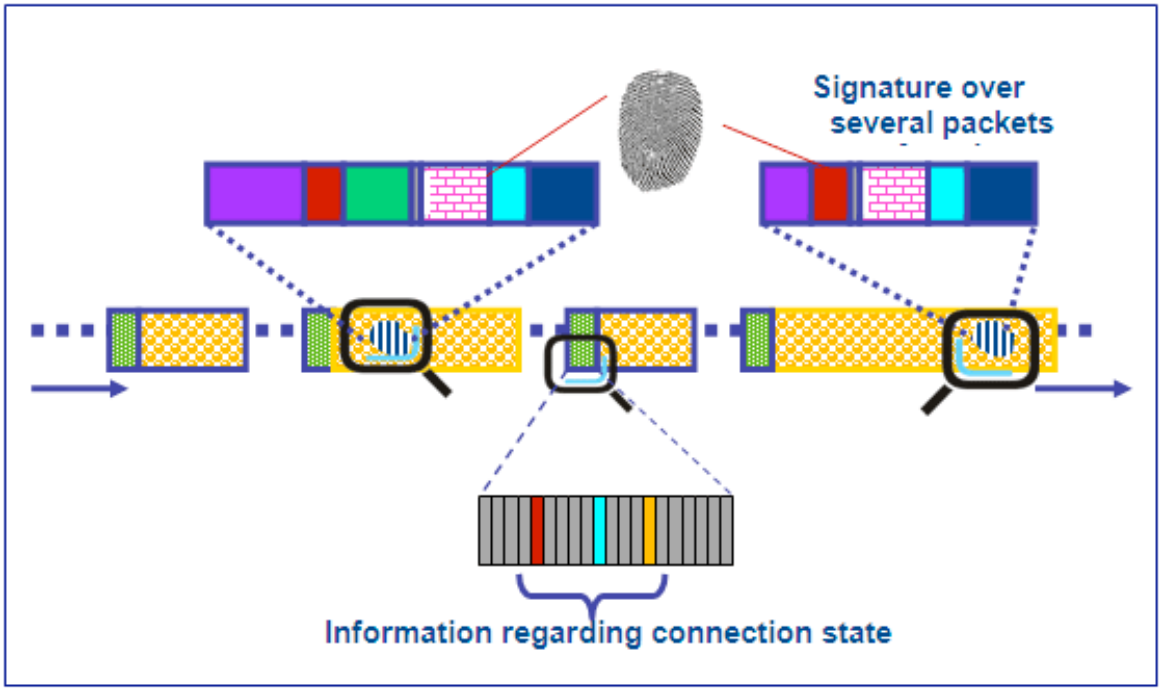


Figure 2: Deep Packet Inspection – analysis of encapsulated content over many packets

What to inspect?

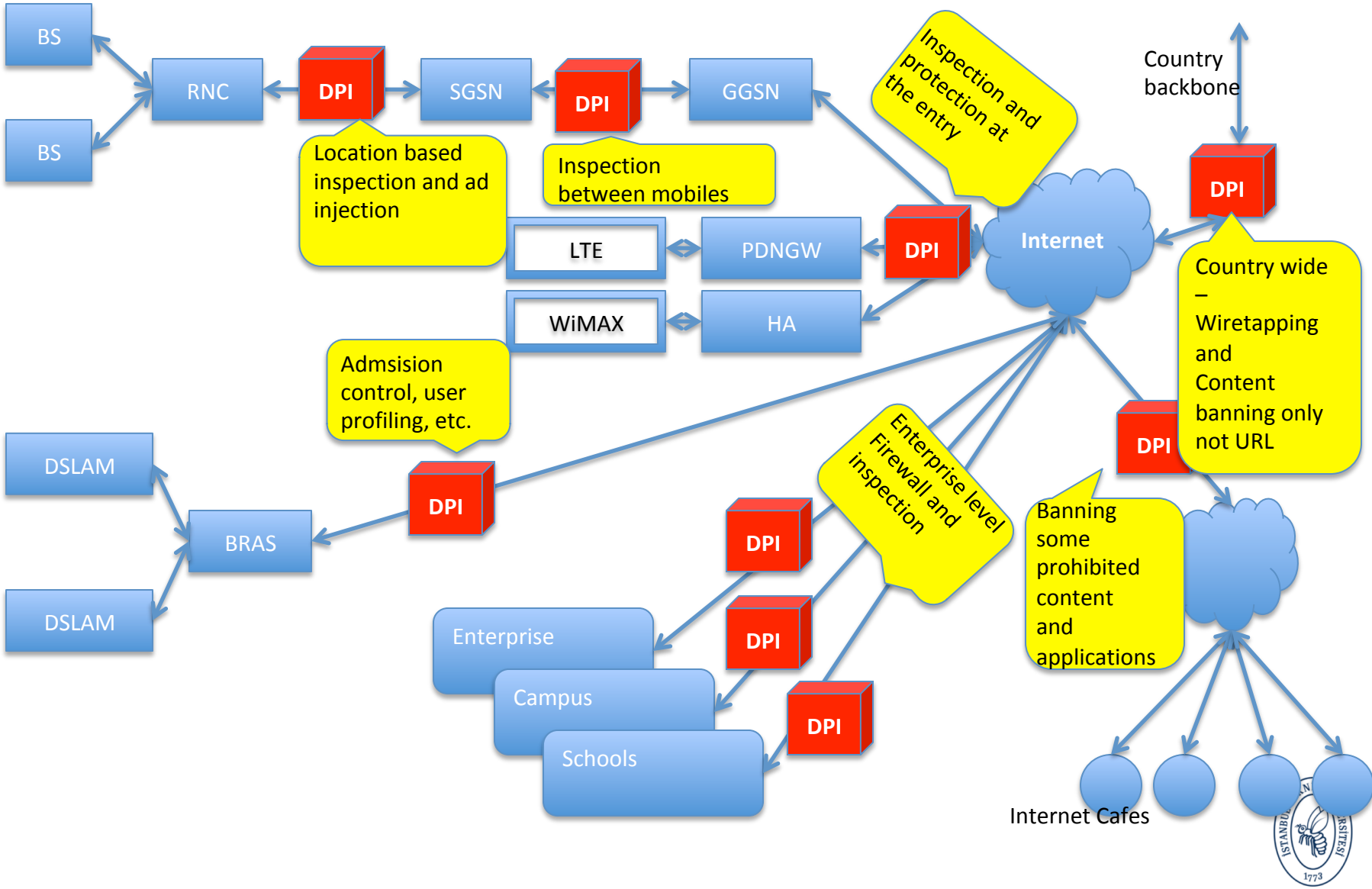
- **P2P:** Gnutella, Imesh, Edonkey, BitTorrent, OFF, etc.
- **Social Applications:** Facebook, Myspace, Twitter, etc.
- **Encrypted P2P:** Skype, Bittorent, Edonkey, Winny, etc.
- **Streaming media:** Real, Flash, RTP/RTSP, Windows Media, Quicktime, MPEG, Joost, iTunes, AVI, PPLive, Sopcast, Youtube, Slingbox, SCTP, etc.
- **Email:** IMAP, POP, Gmail/Yahoo/Webmail, SMTP, Exchange, etc.
- **Collaboration:** Webex, Netmeeting, RDP/VNC, etc.
- **Instant Messaging:** Yahoo, AIM, Google Talk, MSN, Apple talk, QQ, Popo, etc.
- **VoIP:** SIP, H323, MGCP, Skinny, etc.
- **URL Filtering:** Dynamic download link, Popular web sites, etc.
- **Gaming:** PC based games, Xbox, Playstation, Wii, etc.
- **Wireless Apps**: SMS, MMS, WebSMS, etc.
- **Routing**: OSPF, RIP, IGRP, EIGRP, IS-IS, PIM, MPLS, etc.
- **Session:** SSH, Telnet, VNC, Xwindows, Rlogin, RSH, Radming, etc.
- **Security/Tunneling**: L2TP, GRE, IPSEC, IKE/ISKAMP, PPTP, SSL, WAP, RC5DES, SOCKS, etc.
- **Middleware:** Corba, Java RMI, Sun RPC, Java Client, etc.
- **Directory Services:** LDAP, WHOIS, RADIUS, TACACS, WINS, WHOIS, DHCP, Finger, Kerberos, etc.
- **Enterprise Apps**: Oracle, Baan, JDEdwards, SAP, Citrix, SQL, etc.
- **Network Management:** ICMP, SNMP, NTP, Ipcom, RSVP, Timeserver, etc.
- **Productivity:** Zoho, Google Apps, MS Live, etc.
- **Other:** TFTP, HTTP, X400, POP3, SFTP, Netbios, IMAP4, HTTPs, SMTPs, LDAP over SSL, SQL, NETIQ, LDAP over SSL, Real Audio Port, DHCP client/server, Nowell Netware, etc.

1. Network devices need to offer deep packet inspection at gigabit rates.

- 90% fixed signatures
- 10% requires updates

Signature files

Where is DPI?


BS
BS
RNC
DPI
SGSN
DPI
GGSN
Location based inspection and ad injection
Inspection between mobiles
Inspection and protection at the entry
Country backbone
DPI
LTE
PDNGW
DPI
Internet
WiMAX
HA
Admsision control, user profiling, etc.
Country wide – Wiretapping and Content banning only not URL
DPI
DSLAM
DSLAM
BRAS
DPI
DPI
Enterprise level Firewall and inspection
Banning some prohibited content and applications
DPI
Enterprise
Campus
Schools
DPI
Internet Cafes

Methods of Analysis

- **False positives** is the basic terminology referring to misclassification – or in simple terms - the likelihood that an application will be identified as something it is not.

- **False negatives** refers to those cases where it is not possible to consistently identify an application – sometimes the identification is classified, while other times it is missed by the classification tool

Analysis by Port

- Analysis by port is probably the easiest and most well known form of signature analysis. The reasoning is the simple fact that many applications use either default ports or some chosen ports in a specific manner.
  - A good example is POP3 used for an email application. The incoming POP3 typically uses port 110, and if it is secure, it will use port 995. The outgoing SMTP is port 25.

- This is in fact a weakness, particularly because many current applications disguise themselves as other applications.
  - Example: Port 80 is used by many applications to camouflage as pure HTTP traffic.
  - There is often some pattern involved in the port selection process - for example, the first port may be random, but the next will be the subsequent one, and so forth.
  - However in some cases the port selection process may be completely random.

Analysis by String Match

- Search for a sequence of textual characters or numeric values within the contents of the packet.
  - string matches may consist of several strings distributed within a packet or

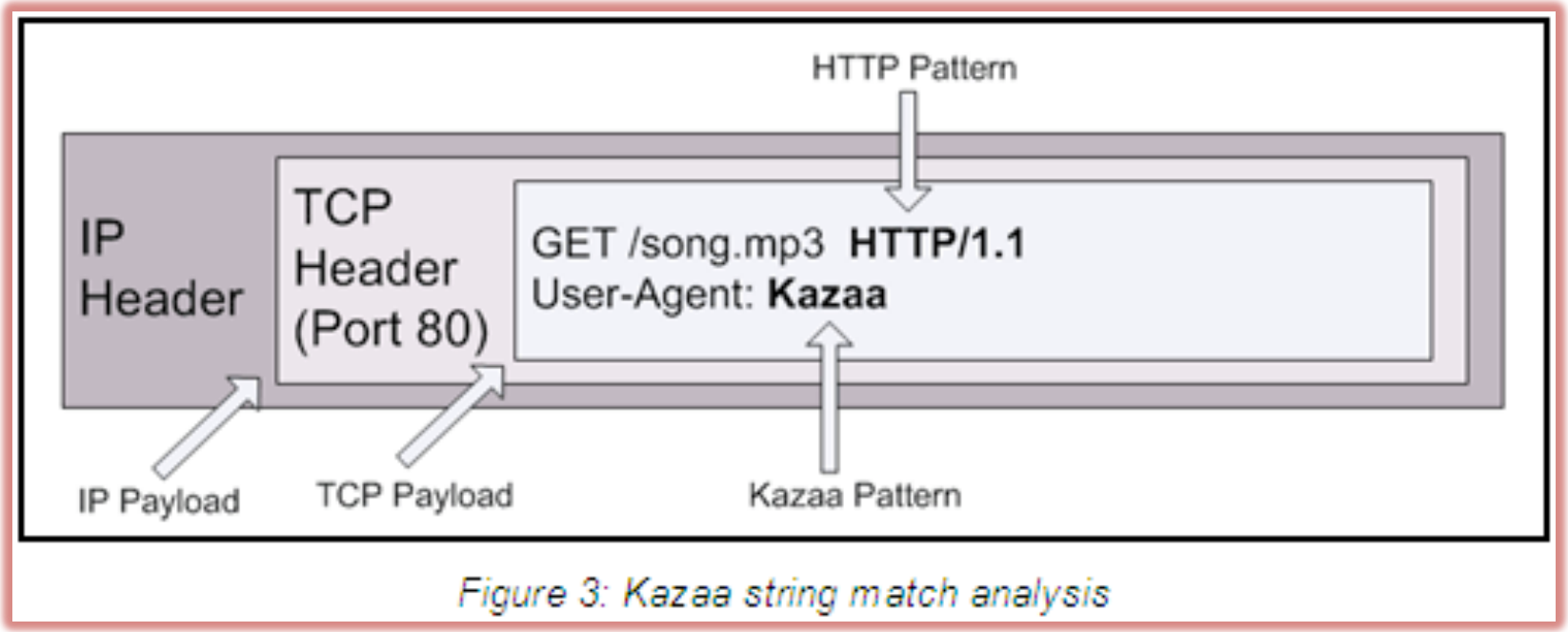


Figure 3: Kazaa string match analysis

Analysis by Numerical Properties

- Investigates the arithmetic and numerical characteristics within a packet, and of a packet or several packets.
  - EX: payload length, the number of packets sent in response to a specific transaction, and the numerical offset of some fixed string (or byte) value within a packet.

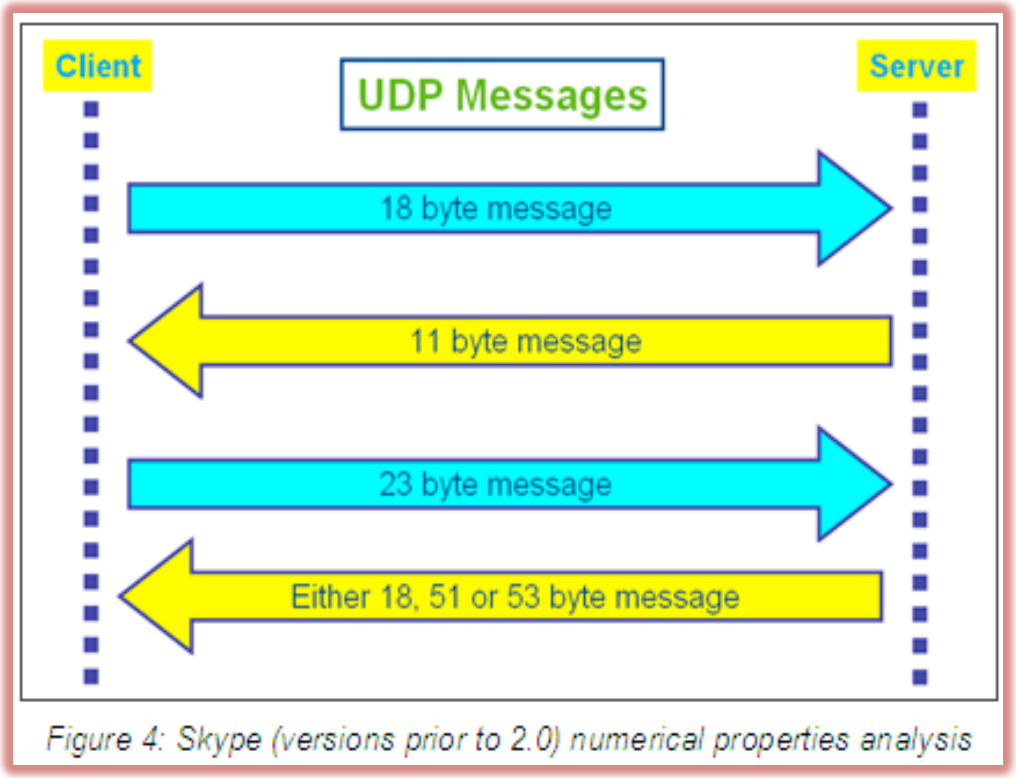


Figure 4: Skype (versions prior to 2.0) numerical properties analysis

Analysis by Behavior and Heuristics

- Behavioral analysis refers to the way a protocol acts and operates.

- Heuristic analysis typically boils down to the extraction of statistical parameters of examined packet transactions.

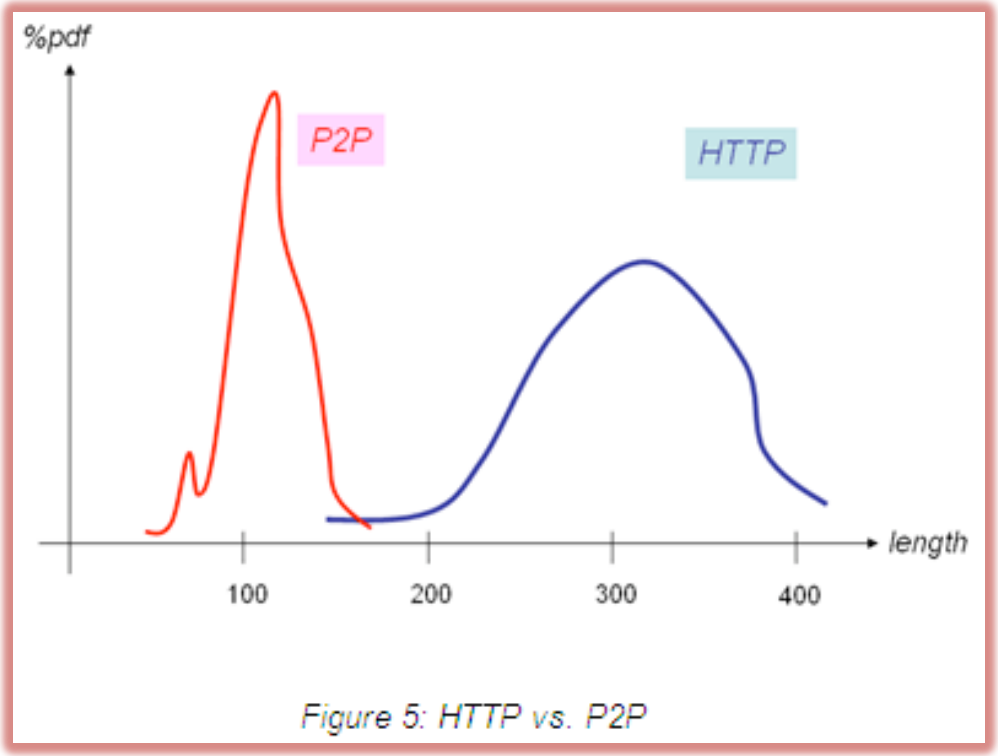
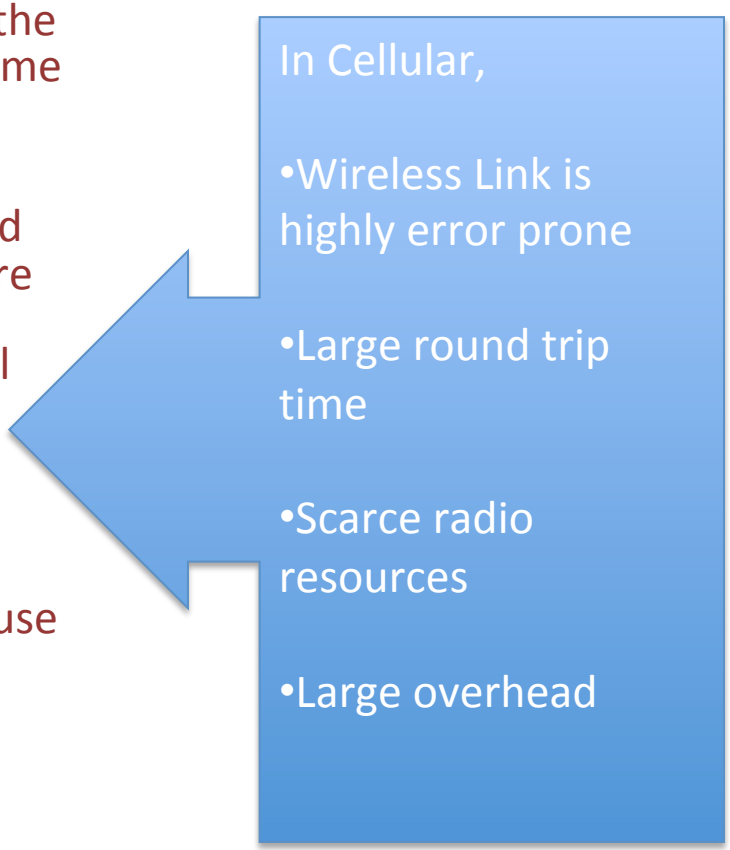


Figure 5: HTTP vs. P2P

IP Header Compression

IP Header Compression

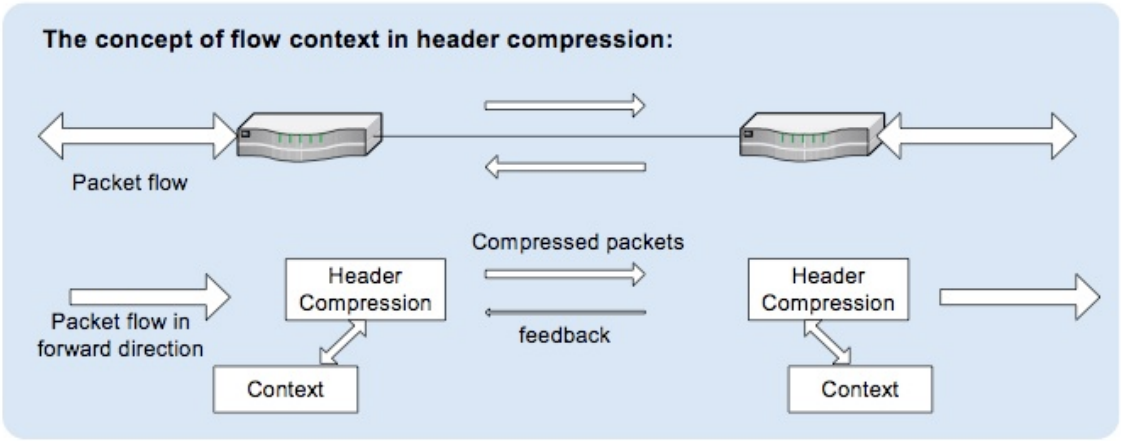
- In many services and applications e.g., Voice over IP, interactive games, messaging etc, the payload of the IP packet is almost of the same size or even smaller than the header.
- Over the end-to-end connection, comprised of multiple hops, these protocol headers are extremely important but over just one link (hop-to-hop) these headers serve no useful purpose.
- It is possible to compress those headers, providing in many cases more than 90% savings, and thus save the bandwidth and use the expensive resources efficiently.
- IP header compression also provides other important benefits, such as reduction in packet loss and improved interactive response time.



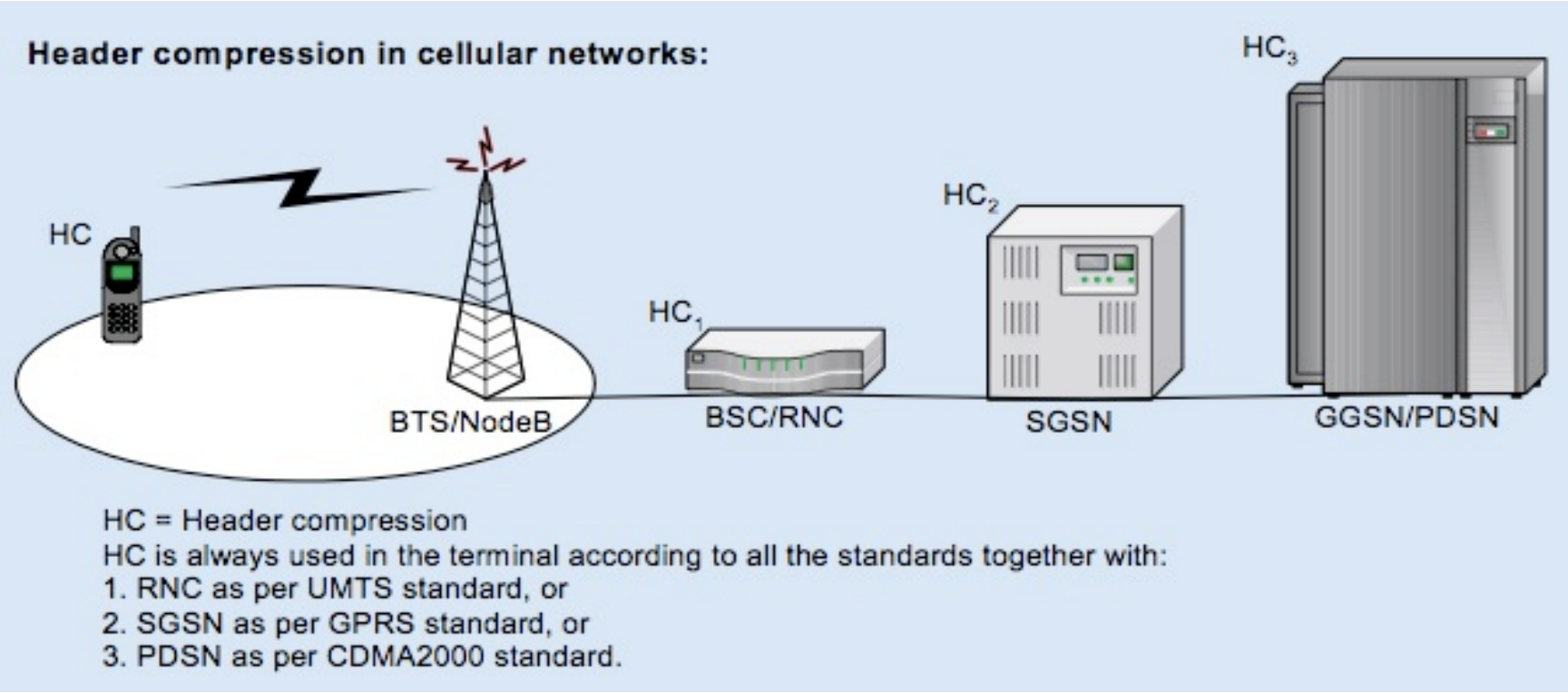
IPHC

- IP header compression is the process of compressing excess protocol headers before transmitting them on a link and uncompressing them to their original state on reception at the other end of the link.

- It is possible to compress the protocol headers due to the redundancy in header fields of the same packet as well as consecutive packets of the same packet stream.



Header Compression in Cellular



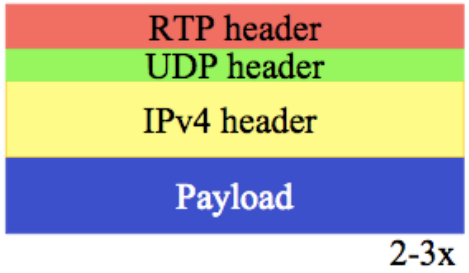
Metrics

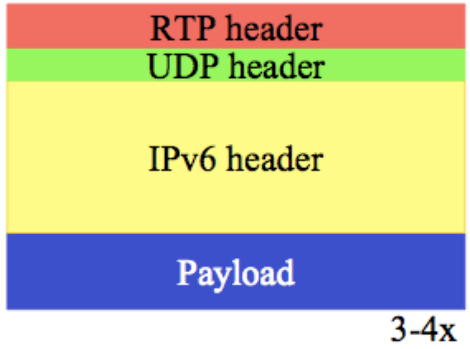
- Compression efficiency
- Compression transparency
- Damage propagation
- Loss propagation
- Residual error
- Robustness

An Example: Voice Communication

- Compression can be done because
  - Significant redundancy betweer header fields in RTP and IP header
  - Significant redundancy betweer header to header fields in one stream, e.g. RTP sequence number
  - Initially send all information then utilize the dependencies to predict the future header information.

- RTP header 12 octets
- UDP header 8 octets
- IPv4 header 20 octets / IPv6 header 40 octets
- Payload 15-20 octets





Header Compression Standards

Comparison of the IETF header compression standards:

IETF standard	RFC 1144 (VJ, CTCP)	RFC 2507 (IPHC)	RFC 2508 (CRTP)	RFC 3095 (ROHC)
Headers	IPv4/TCP	IPv4 (including options and fragments), IPv6 (including extension headers), AH, Minimal Encapsulation header, Tunnelled IP headers, TCP (including options), UDP, ESP	IPv4, IPv6 (including extension headers), AH, Minimal Encapsulation header, Tunnelled IP headers, UDP, RTP	IPv4 (including options and fragments), IPv6 (including extension headers), AH, Minimal Encapsulation headers, GRE, Tunnelled IP headers, UDP, RTP, ESP
Header compressed to minimum	2 bytes	2 bytes	2 bytes	1 byte
Link Type (BER, RTT)	Dial-up (Low, Short)	Dial-up and wireless (Low to medium, Short to medium)	Dial-up and wireless (Low to medium, Short to medium)	Wireless (High, Long)
Encoding	Differential	Differential	Differential	Window-based Least Significant Bit
Error recovery (Feedback)	TCP based (No)	TWICE (Yes)	TWICE (Yes)	Local repair (Yes)
Recommended in (standards)	-	UMTS Release 99 onwards CDMA2000 Release B onwards	-	UMTS Release 4 onwards CDMA2000 Release B onwards

VOIP with IP header compression

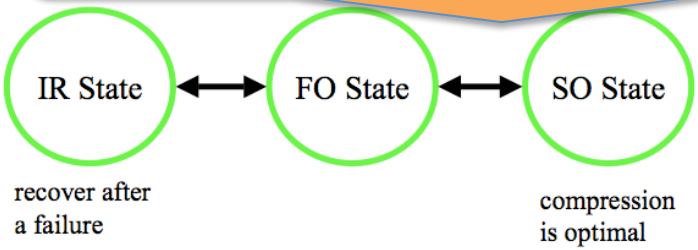
Table 3.4 VoIP packet with header compression: HC and GMH stand for header compression and generic MAC header, respectively

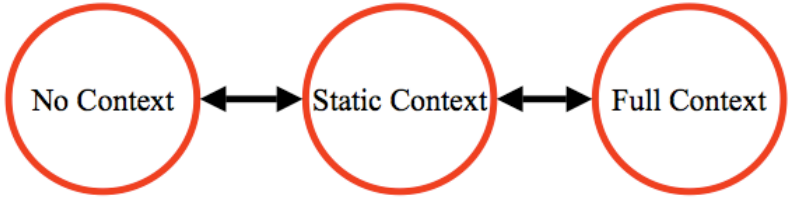
	G.729 with HC	G.729 without HC	AMR with HC	AMR without HC
Voice payload in bytes (inactive/active)	0/20	0/20	7/33	7/33
Headers in bytes (IPv4/IPv6)	2/4	40/60	2/4	40/60
>RIP		12		12
>UDP		8		8
>IPv4/IPv6		20/40		20/40
802.16e GMH	6	6	6	6
CRC	4	4	4	4
Packet size when inactive (IPv4/IPv6)	0/0	0/0	19/21	57/77
Packet size when active (IPv4/IPv6)	32/34	70/90	45/47	83/103

ROHC: Robust Header Compression

- ROCH introduces state information in the compressor and decompressor
- Three states for compressor and decompressor

Three modes of operation: unidirectional (no feedback, periodic timeouts, small compression efficiency), bi-directional (no periodic timeouts, but feedbacks), bi-directional reliable.





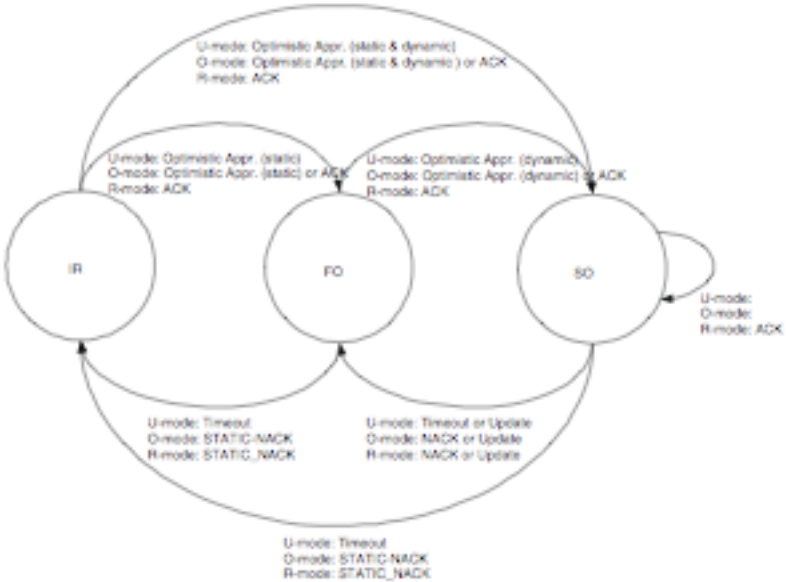
Compressor: Transitions are due to the variation in the packet headers, any feedback from the decompressor, and/or periodically timeouts.

Decompressor: After first successful decompressed packet the decompressor goes into the Full Context State.

ROHC Finite State Machines

Compressor

- IR: Initilization and Refresh
- FO: compressor has detected and stored the static fields- also sending dynamic field differences
- SO: suppressing all dynamic fields and sending only logical sequence number and partial checksum



Decompressor

Encoding

- Least Significant Bit – LSB
- Window Based LSB encoding – WLSB
- Scaled RTP Timestamp encoding
- Timer-based compression of RTP Timestamps
- Offset IP-ID encoding
- Self-describing variable-length values.

Window-based Least Significant Bits Encoding (W-LSB)

- The compressor maintains a sliding window of possible reference values and chooses the minimum number of least significant bits (i) that will produce the original value given those reference values.

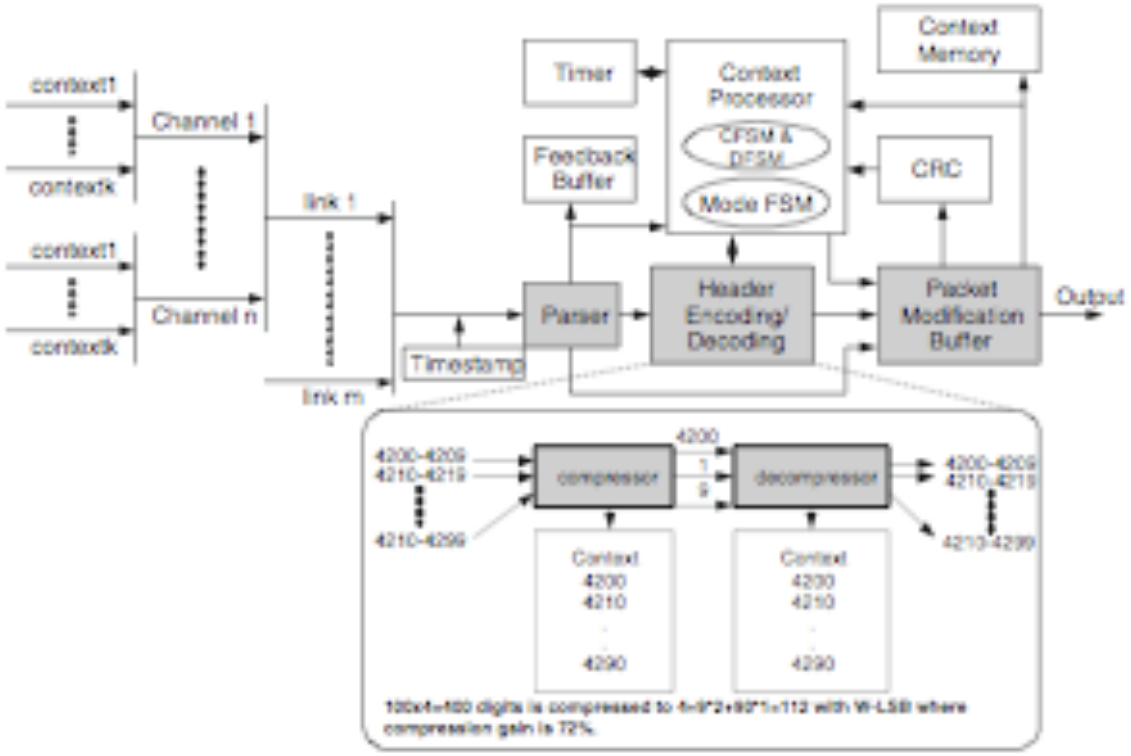


Fig. 10.23 ROHC framework


İSTANBUL TEKNİK ÜNİVERSİTESİ
1773

IPSEC


İSTANBUL TEKNİK ÜNİVERSİTESİ
1773

IPSEC

- IP security was not imminent in the first era of Internet with time; however, it has become essential to secure the network as well as support the necessary protection for mobile applications as well as virtual private networks.

- The framework for IP security (IPSec) has begun in the 1990s by IETF. The IPSec standard is optional for IPv4, however mandatory feature of IPv6 network.

- An unsecure IP packet is vulnerable to the following; it can be spoofed by eavesdroppers (confidentiality); sender and destination address may be altered along the network (authentication); data can be modified (integrity).

- These are addressed in IPSec framework by the following services:
  - Authentication
  - Integrity
  - Privacy
  - Protection against replays
  - Compression

[11] IPSec is defined by several RFCs:

- RFC2401: Security architecture for the Inter
- RFC2402: Authentication header
- RFC2411: IP security document road map
- RFC2406: Encapsulation security payload

IPSEC Methods

- **AH:** AH provides data integrity, data source verification, and protection against replay.
  - The host on a secure LAN digitally signs the packet to authenticate it.
  - The receiving host will check the signature and either accept or reject the packet.
  - If the packet has been altered during its journey, the digital signature will not concur with the packet contents.
  - The contents are not encrypted as they travel across the Internet.

- **ESP:** In addition to these, ESP also provides data confidentiality.
  - The host on a secure LAN may encrypt the packet for its journey across the Internet.
  - Anybody that happens listening to packets using a network analyzer will be able to receive the packet but will not be able to decipher its contents since the entire payload is encrypted or authenticated or both.

Key management is not part of the protocol; however, it can be provided manu- ally or through Internet Key Exchange (IKE) protocol [RFC2409], which is based on public-key-based approach for automatic key management. Other automated key distribution techniques such as Kerberos and SKIP may be used as well.

IPSec Modes of Operation:

- Transport mode is basically with two peers without any intermediate nodes in between.
- Tunnel mode is to protect entire IP datagram when packets traverse through security gateways.
- AH and ESP modes of operation can support transport and tunnel modes.
- They can be applied individually or in combination with each other to provide a set of security services in IPv4 and IPv6.

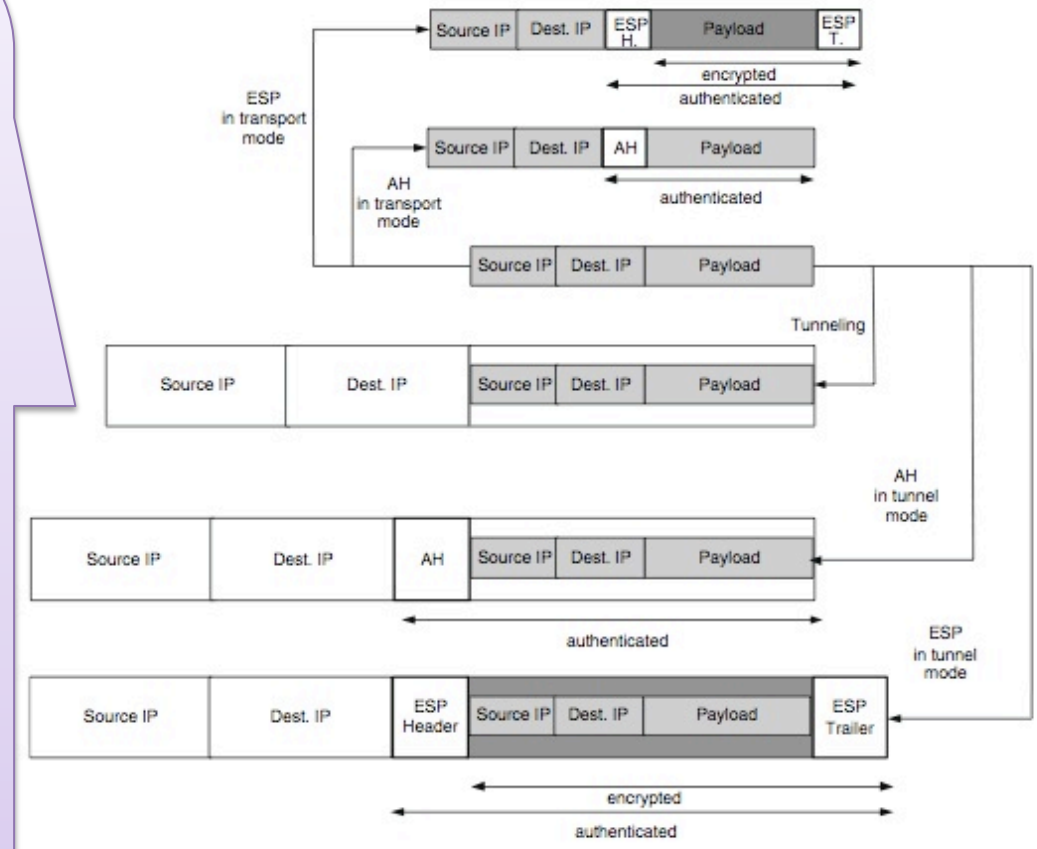
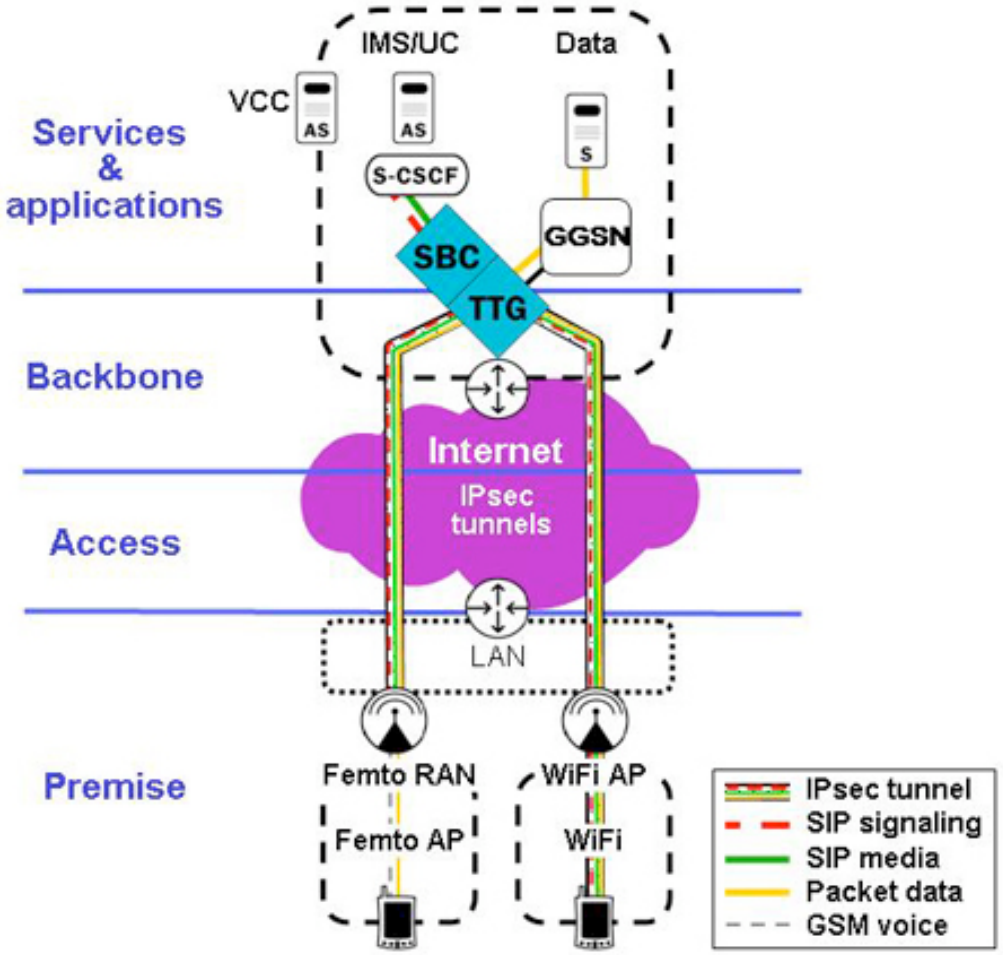


Fig. 3.7 IPSec procedures

IPSEC and Devices behind NAT: e.g., Femtocells



Source: AcmePackets

- Security as well as NAT Traversal
  - If device is behind a NAT, for access from outside, there has to be a mechanism

IP Tunneling

One hop communication behavior, tunneling is introduced.

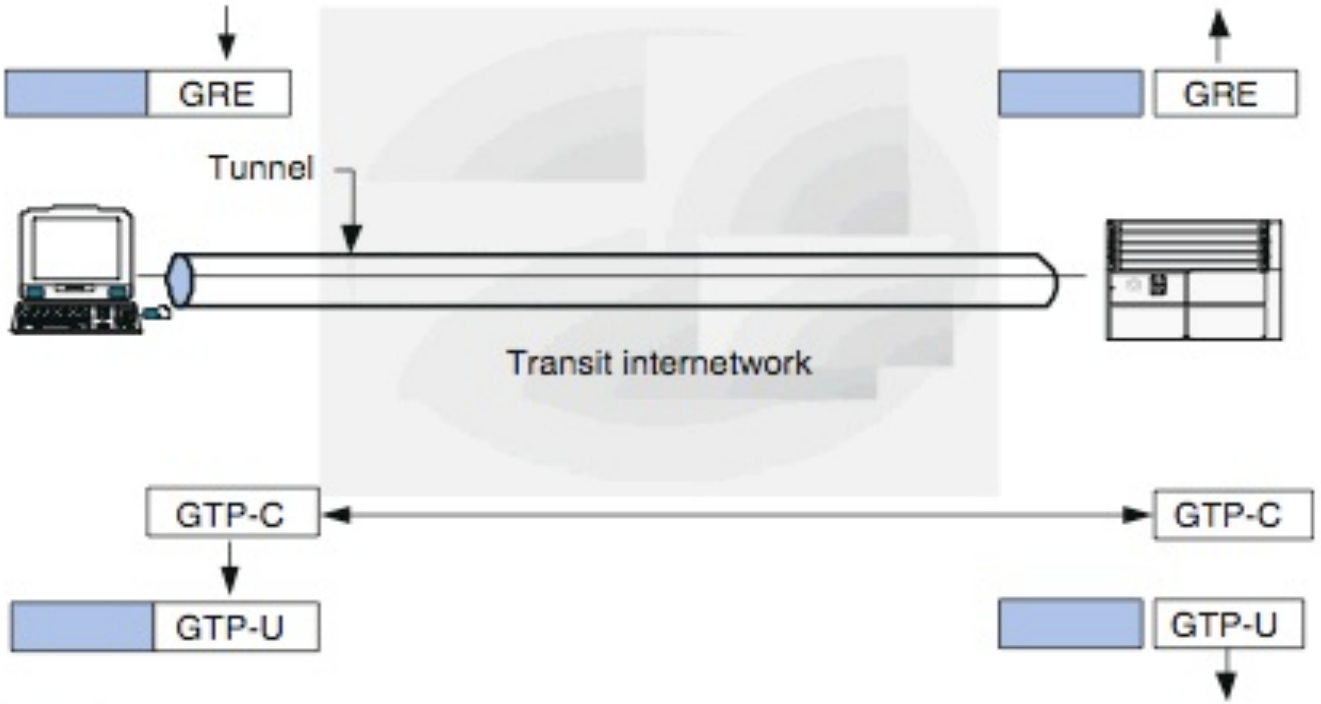
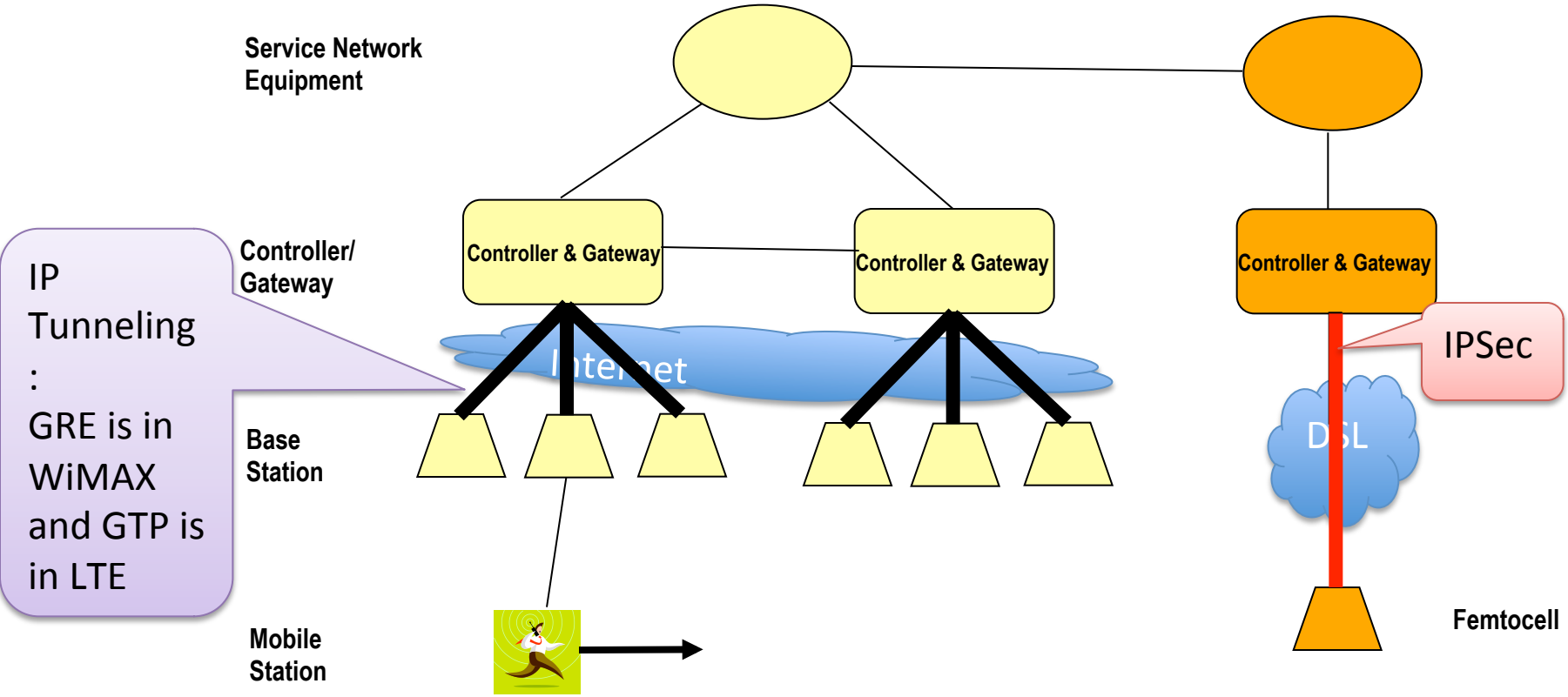


Fig. 3.8 Tunneling

IP Tunneling in Mobile Broadband

- The encapsulated packets are then routed between tunnel endpoints over the tunnel. In the destination, the frame is decapsulated and forwarded to the final destination. Hereby, tunneling includes this entire process: encapsulation, transmission, and decapsulation.



Authentication, Authorization & Accounting

AAA/RADIUS

- Authentication, Authorization and Accounting
  - Used to verify identity of the remote user
  - Introduced RADIUS: Remote Authentication Dial-In User service
  - RADIUS is introduced in dial-up era, new protocol is DIAMETER backward compatible with RADIUS
  - AAA client and AAA server communicates to perform the functionalities.
  - Uses NAI to identify the AAA server: Network Address Identifier
    - name@home.network

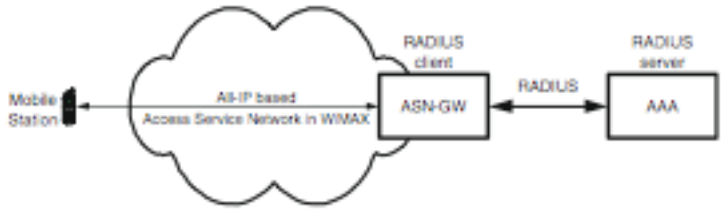


Fig. 3.9 RADIUS configuration

DIAMETER

- Addresses the flaws of RADIUS
  - Composed of base protocol and set of protocol extensions
  - Peer to peer protocol unlike RADIUS which client/server
  - A node can be a client, server, or an agent
  - Introduces peer discovery to avoid manual configuration of the AAA server.

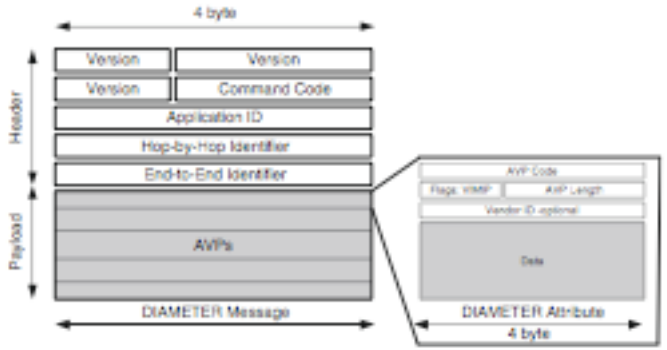


Fig. 3.10 DIAMETER packet format

EAP

- Extensible Authentication Protocol
  - IEEE 802.1x introduced supplicant, authenticator, and authentication server (RADIUS server).
  - The goal of 802.1x/EAP is to distribute the shared key between the supplicant and the authentication server.
  - EAP method varies according to the protocol selected.

EAP-TLS
PEAP,
EAP-TTLS
EAP-FAST
EAP-SIM
EAP-AKA
EAP-MD5
EAP-PSK
EAP-IKEv2, etc.

EAP Protocols

- EAP-TLS [RFC5216] is SSL/TSL procedure.
  - Prevent eavesdropping, tampering, or message forge
  - Handshake to agree on secu parameters
  - Uses public key cryptograph calculate the shared keys
  - Requires client-side certifica (in smartcards)
  - Lacks identity protection and unprotected EAP success/failure messages.

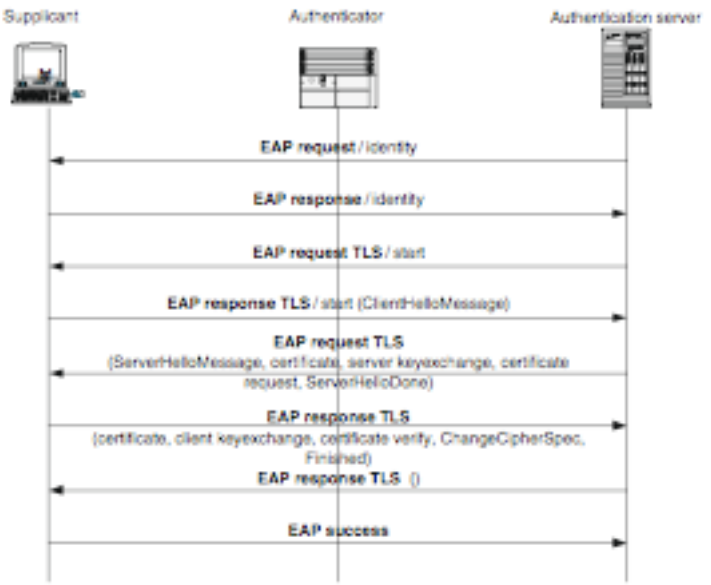


Fig. 3.11 EAP-TLS

EAP Protocols

- EAP-TTLS is tunneled TSL procedure.
  - Removes the need for client authentication
  - TTLS server establishes a secure tunnel to authenticate the client.
  - Protection from eavesdropping, and man-in-the-middle attacks.
  - Provides identity protection and data ciphering suite negotiation.

EAP Protocols

- EAP-AKA is for 3G and successor of EAP-SIM of GSM.
  - EAP-SIM only secures mobile station and base station and assumes the connection secure beyond base station
  - EAP-AKA extends this beyond base station and provides no area with clear data transmission.
  - International Mobile Subscriber Identity (IMSI) is used for 3G
  - NAI is used for IMS
    - 15 digit number

EAP-AKA

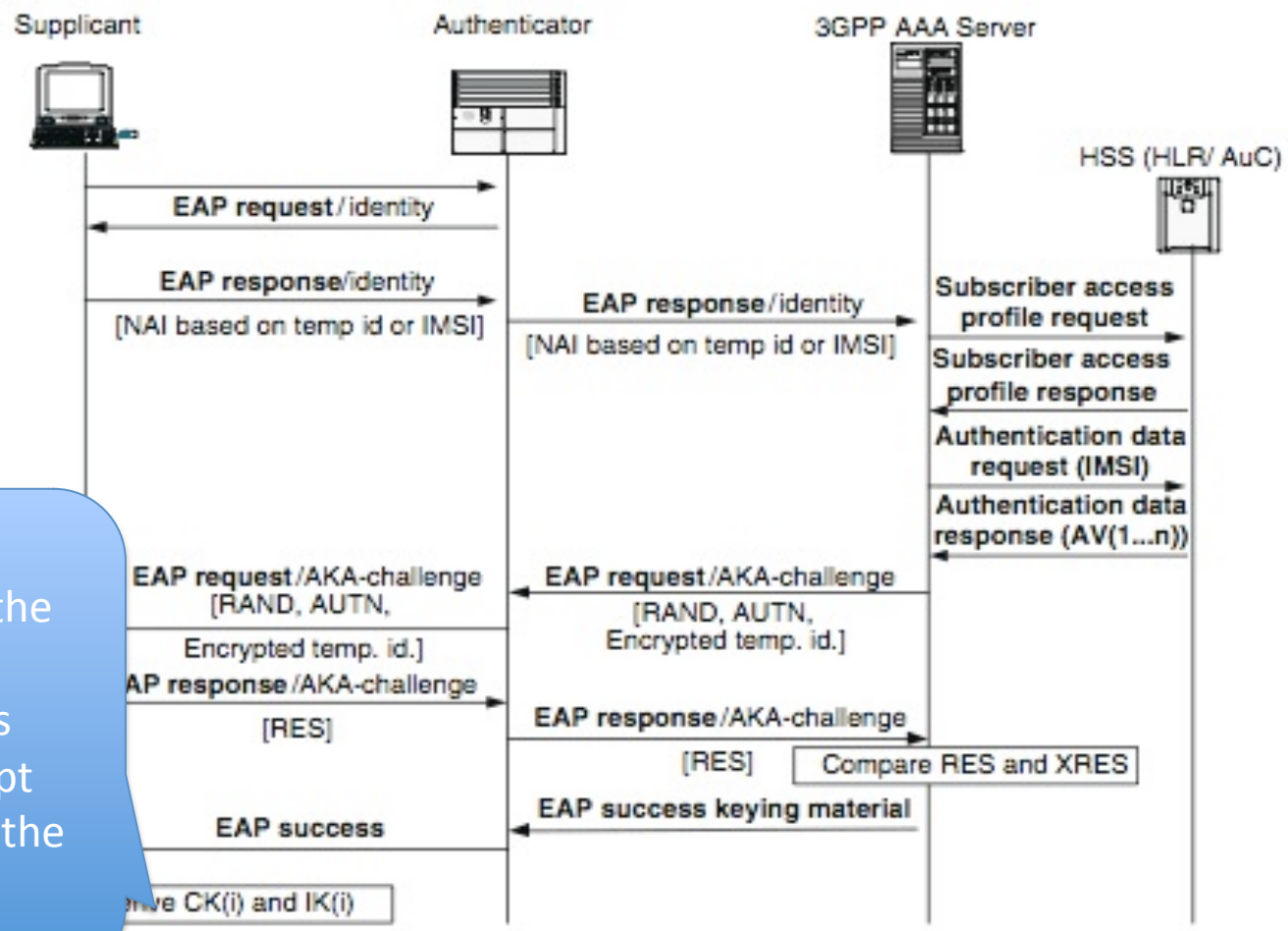


Fig. 3.12 EAP-AKA

IK is used to authenticate the signaling message, CK is used to encrypt the data over the air


İSTANBUL TEKNİK ÜNİVERSİTESİ
1773

Mobile IP


İSTANBUL TEKNİK ÜNİVERSİTESİ
1773

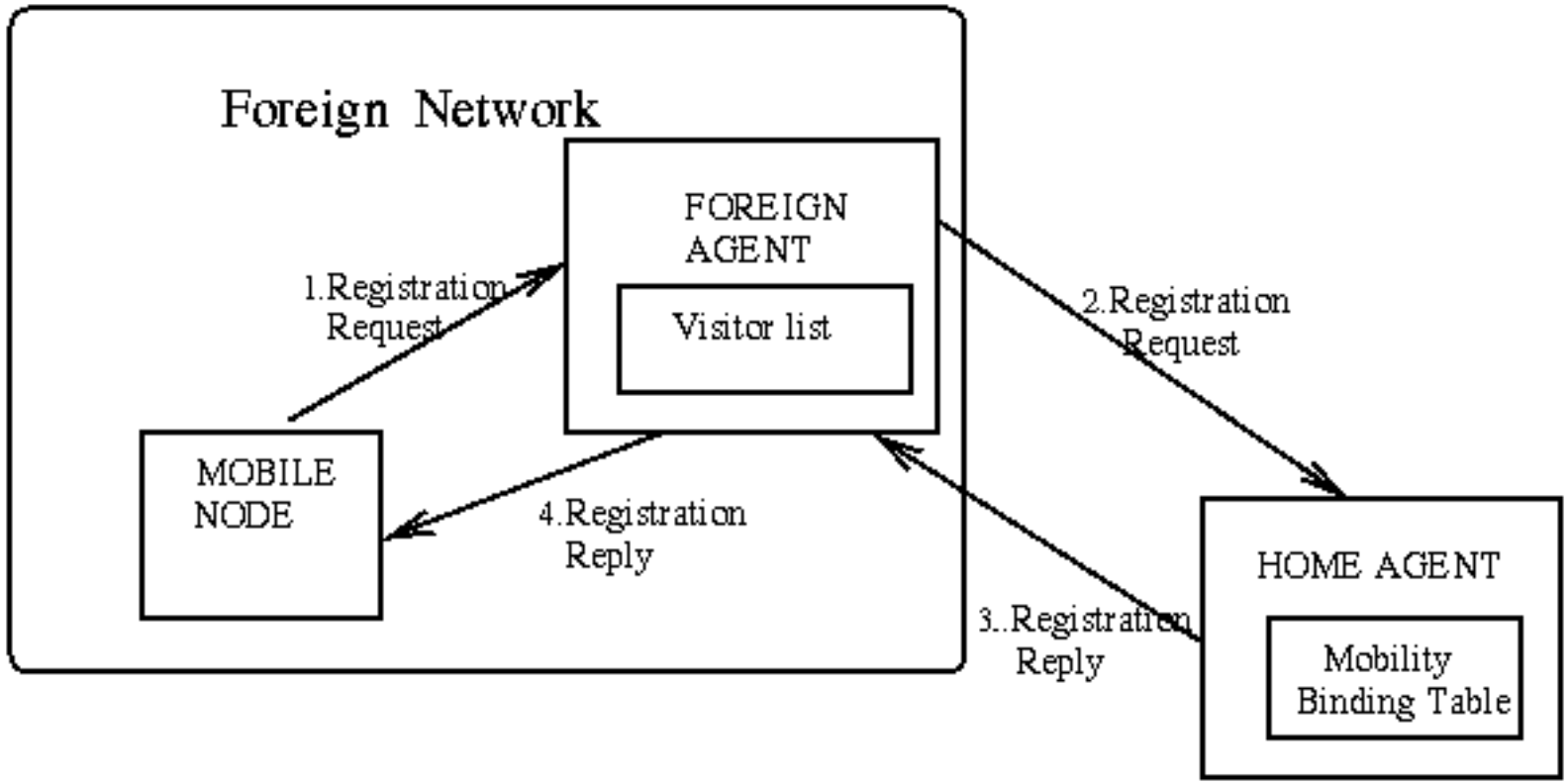
Mobility and Standard IP Routing

- IP assumes end hosts are in fixed physical locations
  - What happens if we move a host between networks?
- IP addresses enable IP routing algorithms to get packets to the correct network
  - Each IP address has network part and host part
    - This keeps host specific information out of routers
  - DHCP is used to get packets to end hosts in networks
    - This still assumes a fixed end host
- What if a user wants to roam between networks?
  - Mobile users don't want to know that they are moving between networks
  - Why can't mobile users change IP when running an application?

Mobile IP

- Mobile IP was developed as a means for transparently dealing with problems of mobile users
  - Enables hosts to stay connected to the Internet regardless of their location
  - Enables hosts to be tracked without needing to change their IP address
  - Requires no changes to software of non-mobile hosts/routers
  - Requires addition of some infrastructure
  - Has no geographical limitations
  - Requires no modifications to IP addresses or IP address format
  - Supports security
    - Could be even more important than physically connected routing

Components


Foreign Network
FOREIGN AGENT
Visitor list
1.Registration Request
2.Registration Request
MOBILE NODE
4.Registration Reply
3..Registration Reply
HOME AGENT
Mobility Binding Table

Mobile IP Entities

- Mobile Node (MN)
  - The entity that may change its point of attachment from network to network in the Internet
    - Detects it has moved and registers with “best” FA
  - Assigned a permanent IP called its *home address* to which other hosts send packets regardless of MN’s location
    - Since this IP doesn’t change it can be used by long-lived applications as MN’s location changes
- Home Agent (HA)
  - This is router with additional functionality
  - Located on home network of MN
  - Does mobility binding of MN’s IP with its COA
  - Forwards packets to appropriate network when MN is away
    - Does this through encapsulation

Mobile IP Entities contd.

- Foreign Agent (FA)
  - Another router with enhanced functionality
  - If MN is away from HA the it uses an FA to send/receive data to/from HA
  - Advertises itself periodically
  - Forward's MN's registration request
  - Decapsulates messages for delivery to MN
- Care-of-address (COA)
  - Address which identifies MN's current location
  - Sent by FA to HA when MN attaches
  - Usually the IP address of the FA
- Correspondent Node (CN)
  - End host to which MN is corresponding (eg. a web server)

Mobile IP Support Services

- Agent Discovery
  - HA's and FA's broadcast their presence on each network to which they are attached
    - Beacon messages via ICMP Router Discovery Protocol (IRDP)
  - MN's listen for advertisement and then initiate registration
- Registration
  - When MN is away, it registers its COA with its HA
    - Typically through the FA with strongest signal
  - Registration control messages are sent via UDP to well known port
- Encapsulation – just like standard IP only with COA
- Decapsulation – again, just like standard IP

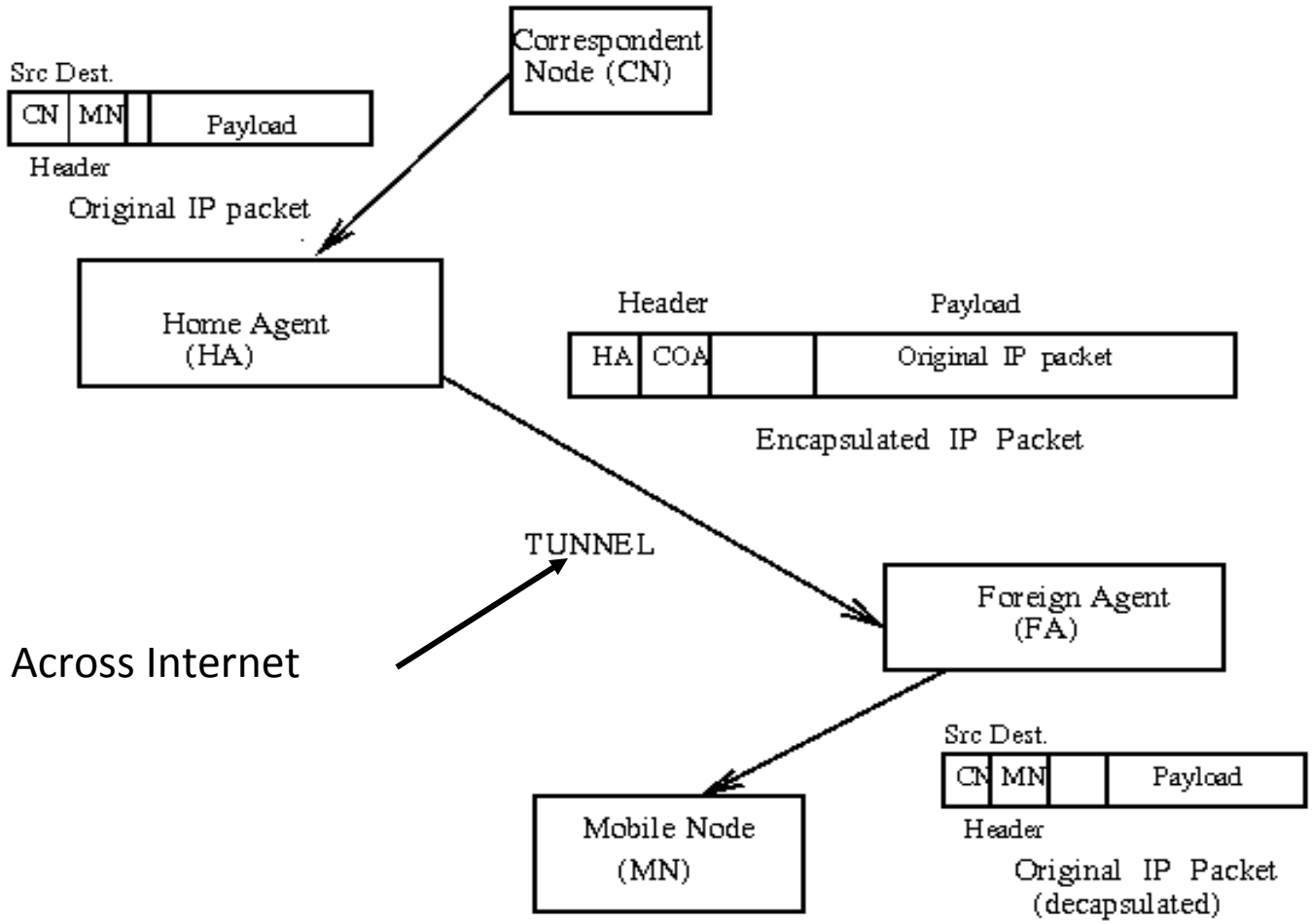
Mobile IP Operation

- A MN listens for agent advertisement and then initiates registration
  - If responding agent is the HA, then mobile IP is not necessary
- After receiving the registration request from a MN, the HA acknowledges and registration is complete
  - Registration happens as often as MN changes networks
- HA intercepts all packets destined for MN
  - This is simple unless sending application is on or near the same network as the MN
  - HA masquerades as MN
  - There is a specific lifetime for service before a MN must re-register
  - There is also a de-registration process with HA if an MN returns home

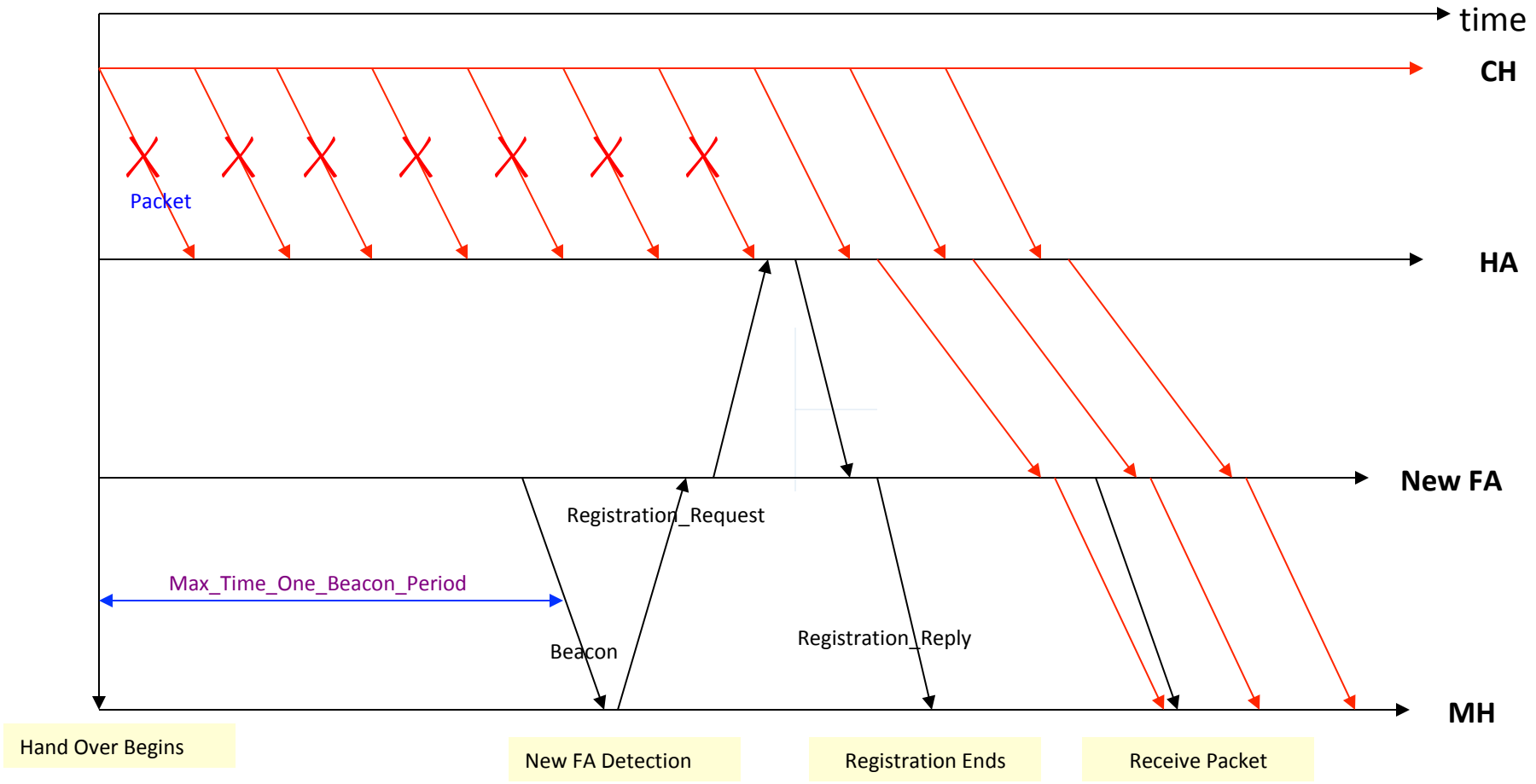
Mobile IP Operation contd.

- HA then encapsulates all packets addressed to MN and forwards them to FA
  - IP tunneling
- FA decapsulates all packets addressed to MN and forwards them via hardware address (learned as part of registration process)
- NOTE that the MN can perform FA functions if it acquires an IP address eg. via DHCP
- Bidirectional communications require tunneling in each direction

Mobile IP Tunneling



Time Diagram of Handover



A Client Mobile IP

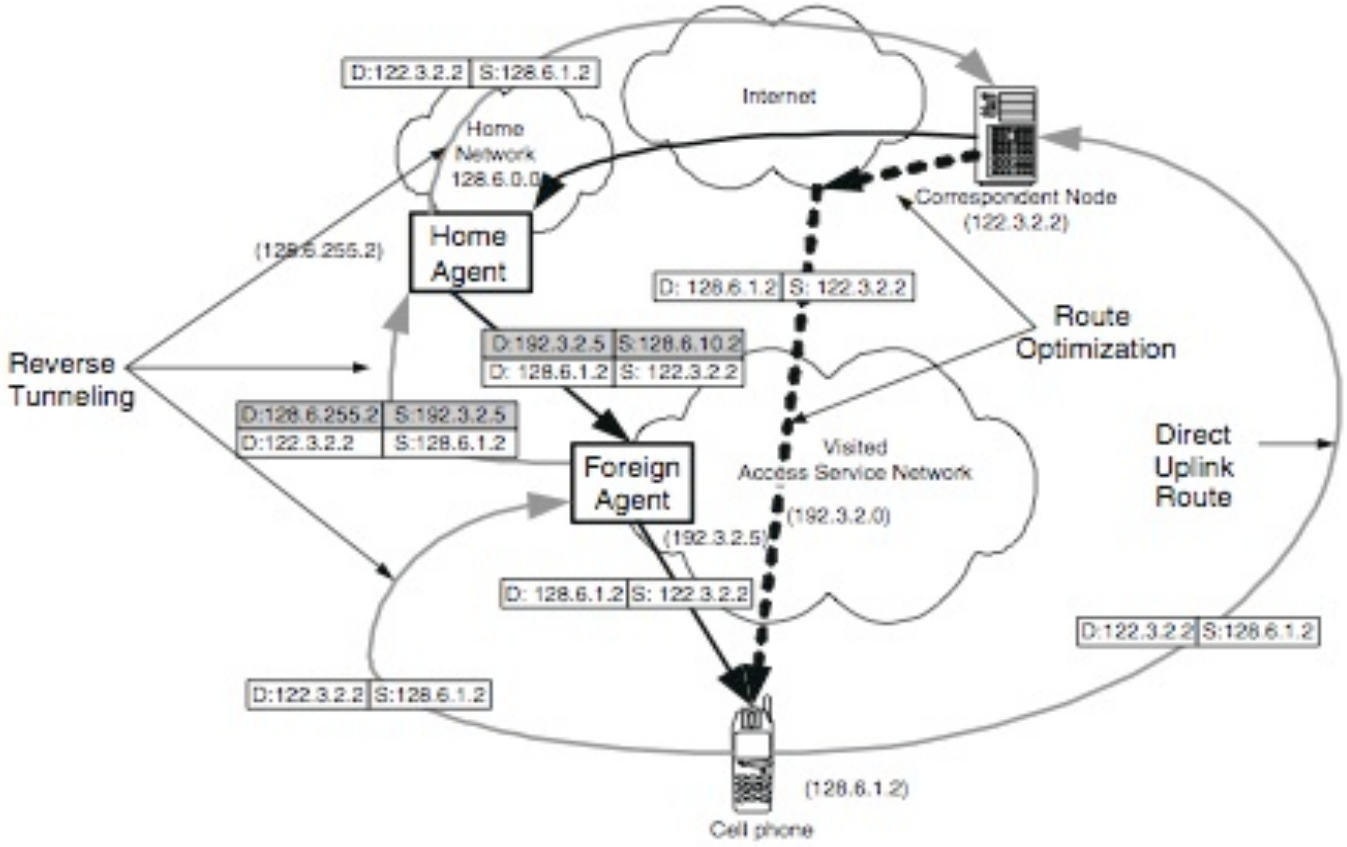


Fig. 3.13 An mobile IP

ProxyMIP

- Proxy Mobile IP
  - Mobile IP Client is in the Access Gateway
  - No requirement on the device

Mobile IP in WiMAX

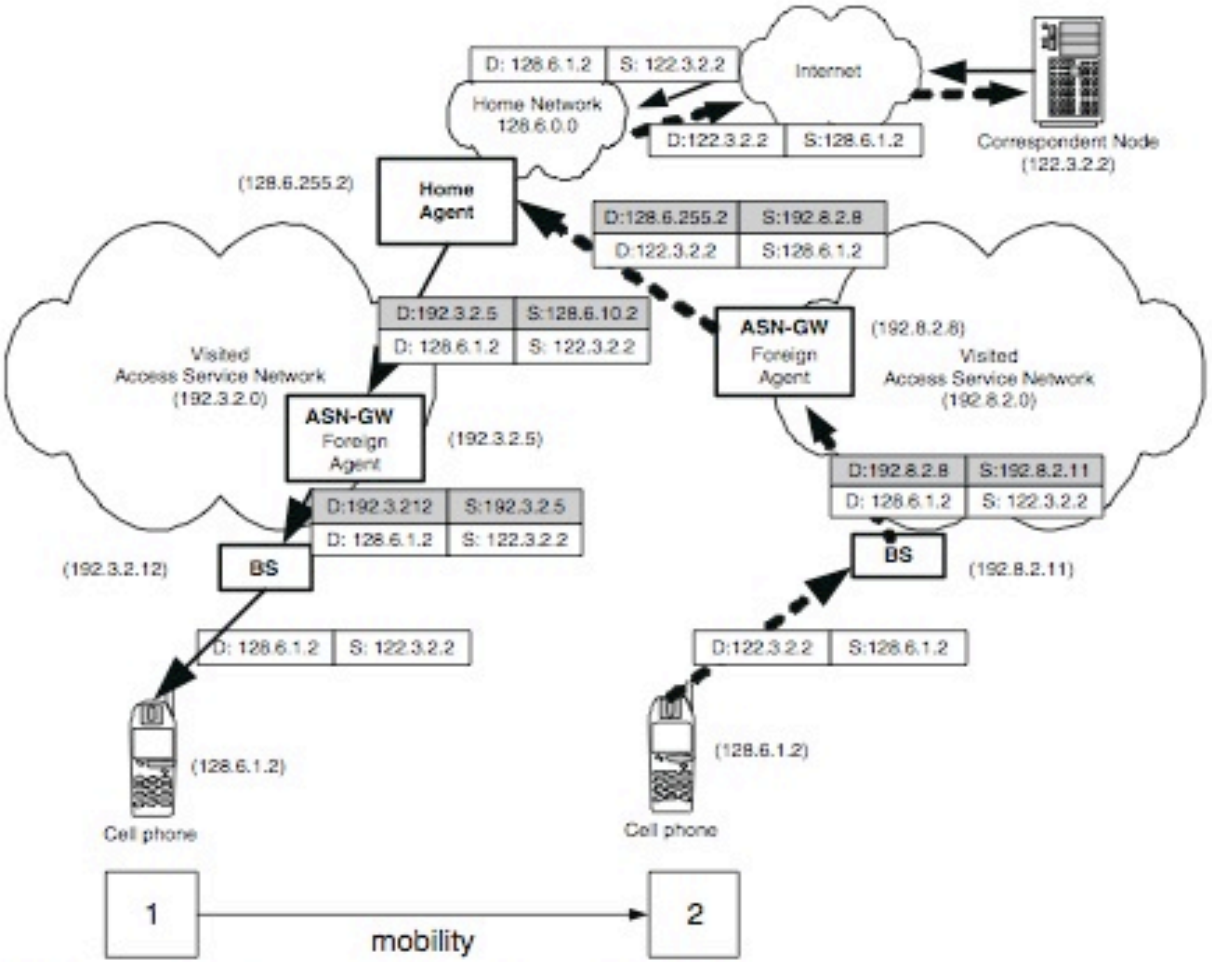


Fig. 3.14 An example for mobile IP procedure in WiMAX

Security in Mobile IP

- Authentication can be performed by all parties
  - Only authentication between MN and HA is required
  - Keyed MD5 is the default
- Replay protection
  - Timestamps are mandatory
  - Random numbers on request reply packets are optional
- HA and FA do not have to share any security information.

Problems with Mobile IP

- Suboptimal “triangle” routing
  - What if MN is in same subnetwork as the node to which it is communicating and HA is on the other side of the world?
    - It would be nice if we could directly route packets
  - Solution: Let the CN know the COA of MN
    - Then the CN can create its own tunnel to MN
    - CN must be equipped with software to enable it to learn the COA
    - Initiated by HA who notifies CN via “binding update”
    - Binding table can become stale

Other Mobile IP Problems

- Single HA model is fragile
  - Possible solution – have multiple HA
- Frequent reports to HA if MN is moving
  - Possible solution – support of FA clustering
- Security
  - Connection hijacking, snooping…
- Many open research questions

Mobility in IPv6

- Route Optimization is a fundamental part of Mobile IPv6
  - Mobile IPv4 it is an optional set of extensions that may not be supported by all nodes
- Foreign Agents are not needed in Mobile IPv6
  - MNs can function in any location without the services of any special router in that location
- Security
  - Nodes are expected to employ strong authentication and encryption
- Other details…

Session Initiation Protocol

Transition to IP

- Changing Business Models
  - For carriers, wireline voice revenue is in decline
  - Wireless carriers have had explosive growth, but also seek new revenue sources
- Enterprises have moved toward a converged voice and data network
- Traditional circuit switched technology is in decline, being replaced by Voice over IP
  - After years of argument, SIP (Session Initiation Protocol) is the choice for VoIP.

SIP

- An Application-layer control (signaling) protocol for creating, modifying and terminating sessions with one or more participants.
- Sessions include Internet multimedia conferences, Internet telephone calls and multimedia distribution.
- Members in a session can communicate via multicast or via a mesh of unicast relations, or a combination of these.
- Text based , Model similar to HTTP : uses client-server model

SIP Basic Functionality

Supports 5 facets of communication:

- **User location**: determination of the end system to be used for communication;
- **User capabilities**: determination of the media and media parameters to be used;
- **User availability**: determination of the willingness of the called party to engage in communications;
- **Call setup**: "ringing", establishment of call parameters at both called and calling party;
- **Call handling**: including transfer and termination of calls.

SIP Functionality (cont.)

- SIP can also **initiate multi-party calls** using a multipoint control unit (MCU) or fully-meshed interconnection instead of multicast.
- Internet telephony gateways that connect Public Switched Telephone Network (PSTN) parties can also use SIP to **set up calls** between them.

Development of SIP

- SIP developed by Handley, Schulzrinne, Schooler, and Rosenberg
  - Submitted as Internet-Draft 7/97
- Assigned RFC 2543 in 3/99
- Goals: Re-use of & Maximum Interoperability with existing protocols

- Alternative to ITU's H.323
  - H.323 used for IP Telephony since 1994
  - Problems: No new services, addressing, features
  - Concerns: scalability, extensibility

SIP Philosophy

- Internet Standard
  - IETF - http://www.ietf.org
- Reuse Internet addressing (URLs, DNS, proxies)
  - Utilizes rich Internet feature set
- Reuse HTTP coding
  - Text based
- Makes no assumptions about underlying protocol:
  - TCP, UDP, X.25, frame, ATM, etc.
  - Support of multicast

SIP Architecture

- SIP uses client/server architecture
- Elements:
  - SIP User Agents (SIP Phones)
  - SIP Servers (Proxy or Redirect - used to locate SIP users or to forward messages.)
    - Can be stateless or stateful
  - SIP Gateways:
    - To PSTN for telephony interworking
    - To H.323 for IP Telephony interworking
- Client - originates message
- Server - responds to or forwards message

SIP Entities

- **User Agents**
  - User Agent Client (UAC): Initiates SIP requests
  - User Agent Server (UAS): Returns SIP responses
- **Network Servers** (diff. types may be co-located)
  - *Proxy*: Decides next hop and forwards request, relays call signaling , operates in a transactional manner, saves no session state
  - *Redirect*: Sends address of next hop back to client, redirects callers to other servers

- *Registrar*: Accepts REGISTER requests from clients, maintains users' whereabouts at a location server

SIP Operation

1. SIP Addressing
2. Locating a SIP Server
3. Sending SIP Requests : SIP Transactions
4. SIP Methods
5. SIP Responses
6. Subsequent Requests and Responses

Step #1:SIP Addressing

Uses Internet URLs

- Uniform Resource Locators
- Supports both Internet and PSTN addresses
- General form is **name@domain**
- To complete a call, needs to be resolved down to User@Host
- Examples:

  sip:alan@wcom.com

  sip:J.T. Kirk <kirk@starfleet.gov>

  sip:+1-613-555-1212@wcom.com;user=phone

  sip:guest@10.64.1.1

  sip:790-7360@wcom.com;phone-context=VNET

Step#2: Locating a SIP server

- A caller first locates the appropriate server
- When client wants to send a request URI

  client will either send it to

  - Locally configured Proxy server or to
  - IP address & port corresponding to the

    request URI [similar to the one in step#1]
- Client must determine IP address, port of server and the protocol to be used.

Locating (cont.)

Client

1. Should try to contact a server at the port listed in request URI. If no port specified then try port 5060

   - Use specified protocol if applicable

   - o.w. use UDP if supported,

   - if UDP fails or o.w. use TCP

2. Send the request to the server's IP address if the host part of request URI is an IP address o.w.
3. Find one or more address of server by querying DNS,
4. Results MAY be cached.
5. Capability to interpret ICMP messages must exist

Step# 3:Send a SIP request

- Once the host part has been resolved to a SIP server, - client sends 1 / more SIP requests to that server &
  - receives 1 / more responses from the server.
- SIP Request-line (Messages) defined as:

  <Method> <SP> Request-URI <SP>SIP-Version <CRLF>

  (SP=Space, CRLF=Carriage Return and Line Feed)

  (Method = “INVITE” | “ACK” | “OPTIONS” |

  “BYE” | “CANCEL” | “REGISTER”)
- Example:

  INVITE sip:picard@wcom.com SIP/2.0

Order of Operation

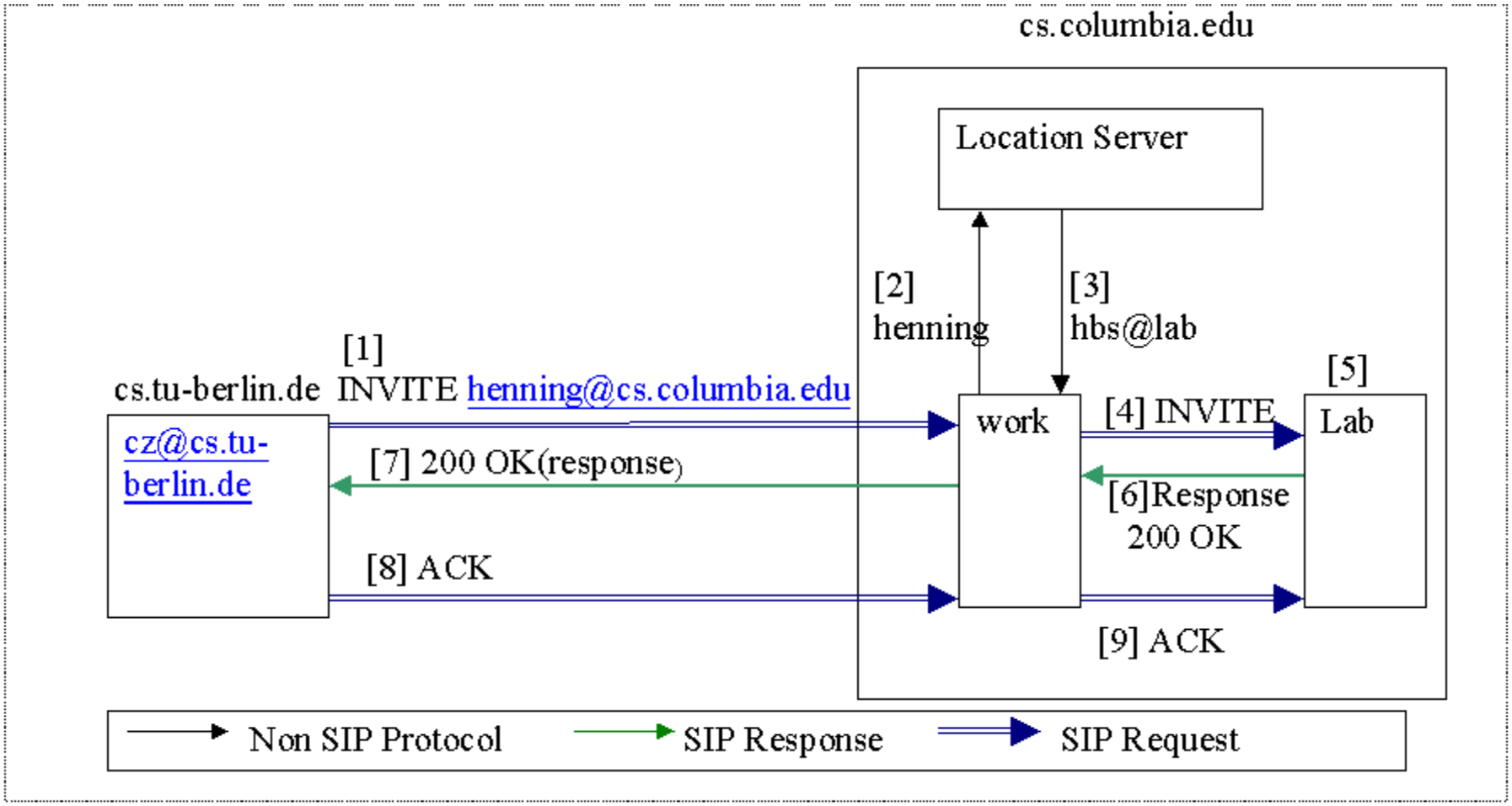
- Step1: Caller Issues Initial INVITE Request
- Step 2: Callee Issues Response
- Step 3: Caller Receives Response to Initial request
- Step 4: Caller or Callee Generate Subsequent requests
- Step 5: Receive Subsequent Requests
- Step 6: BYE to end session
- Step x: CANCEL may be issued

Methods (cont.)

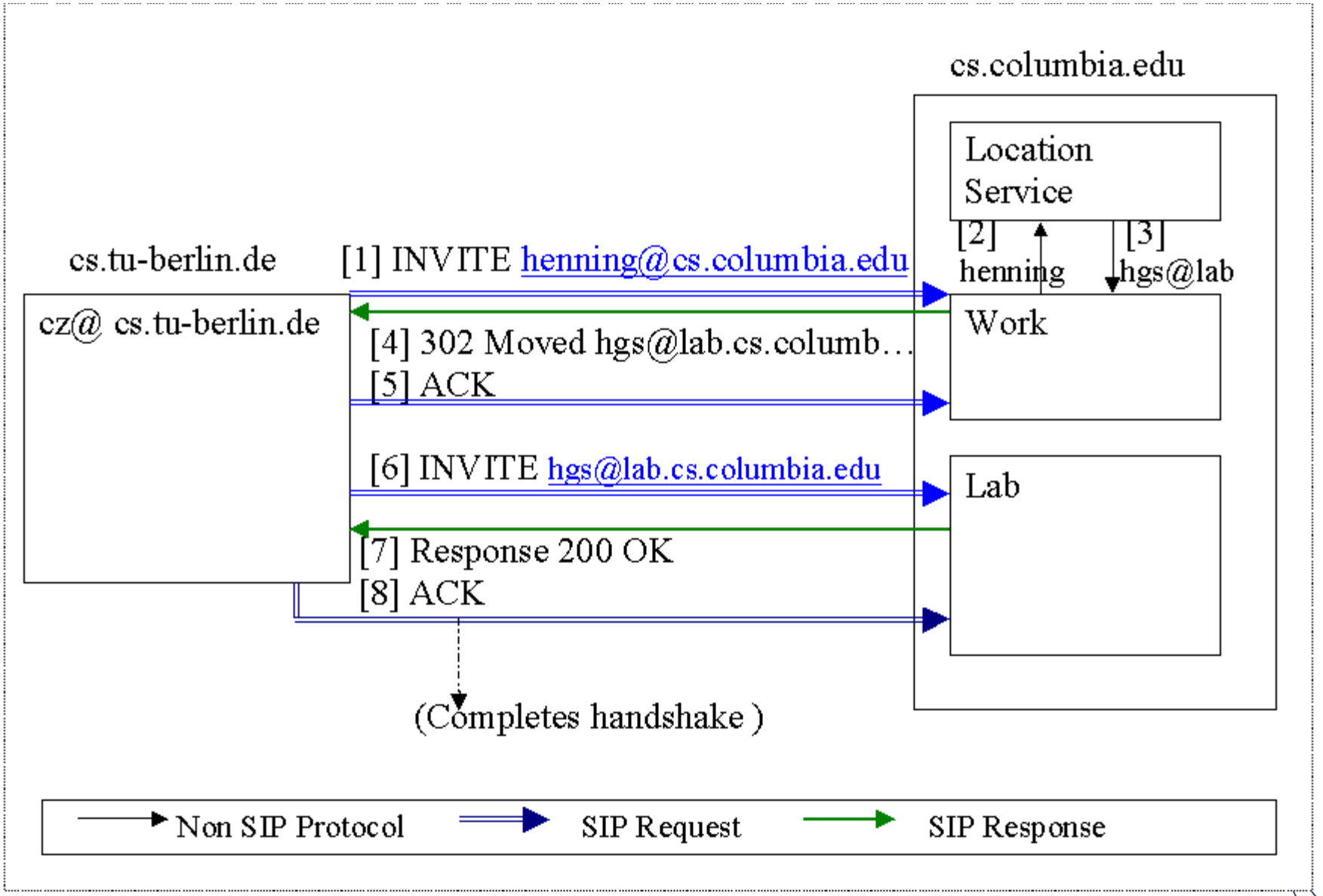
- INVITE: Initiates sessions

  - Session description included in message body
- Re-INVITEs used to change session state
- ACK confirms session establishment, can only be used with INVITE
- BYE terminates a session (hanging up)
- CANCEL cancels a pending invite
- REGISTER: binds a permanent address to current location, may convey user data
- OPTIONS: capability inquiry

Proxy Server Example



Redirect Server Example



SIP Responses

- SIP Responses defined as (HTTP-style):

  SIP-Version SP Status-Code SP Reason-Phrase CRLF

  (SP=Space, CRLF=Carriage Return and Line Feed)
- Example:

  SIP/2.0 404 Not Found
- First digit gives Class of response

SIP Responses (cont.)

- 1xy – Informational

  request received , continuing to process request
- 2xy – Success

  action successfully recvd., understood & accepted
- 3xy – Redirection

  Further action to be taken to complete the request
- 4xy – Client error

  request contains syntax error or cant be completed at this server
- 5xy – Server error

  server fails to fulfill an apparently valid request
- 6xy – global failure,

  request is invalid at any server

Step #4 Generate Subsequent Requests

- Once the call has been established, either the caller or callee may generate INVITE or BYE requests to change or terminate the call.
- For the desired call leg the headers are set as follows (both including any tags):

  - the **To** header field is set to the **remote** address, and

  - the **From** header field is set to the **local** address.
- The Contact header field ***may*** be different than the Contact header field sent in a previous response or request. The Request-URI ***may*** be set to the value of the Contact header field received in a previous request or response from the remote party, or to the value of the remote address. [supports mobility]

SIP Requests Example

Required Headers (fields):

```
INVITE sip:picard@wcom.com SIP/2.0
Via: SIP/2.0/UDP host.wcom.com:5060
From: Alan Johnston <sip:alan.johnston@wcom.com>
To: Jean Luc Picard <sip:picard@wcom.com>
Call-ID: 314159@host.wcom.com
CSeq: 1 INVITE
```

} Uniquely identify this session request

- **Via**: Shows route taken by request.
- **Call-ID**: unique identifier generated by client.
- **CSeq**: Command Sequence number
  - generated by client
  - Incremented for each successive request

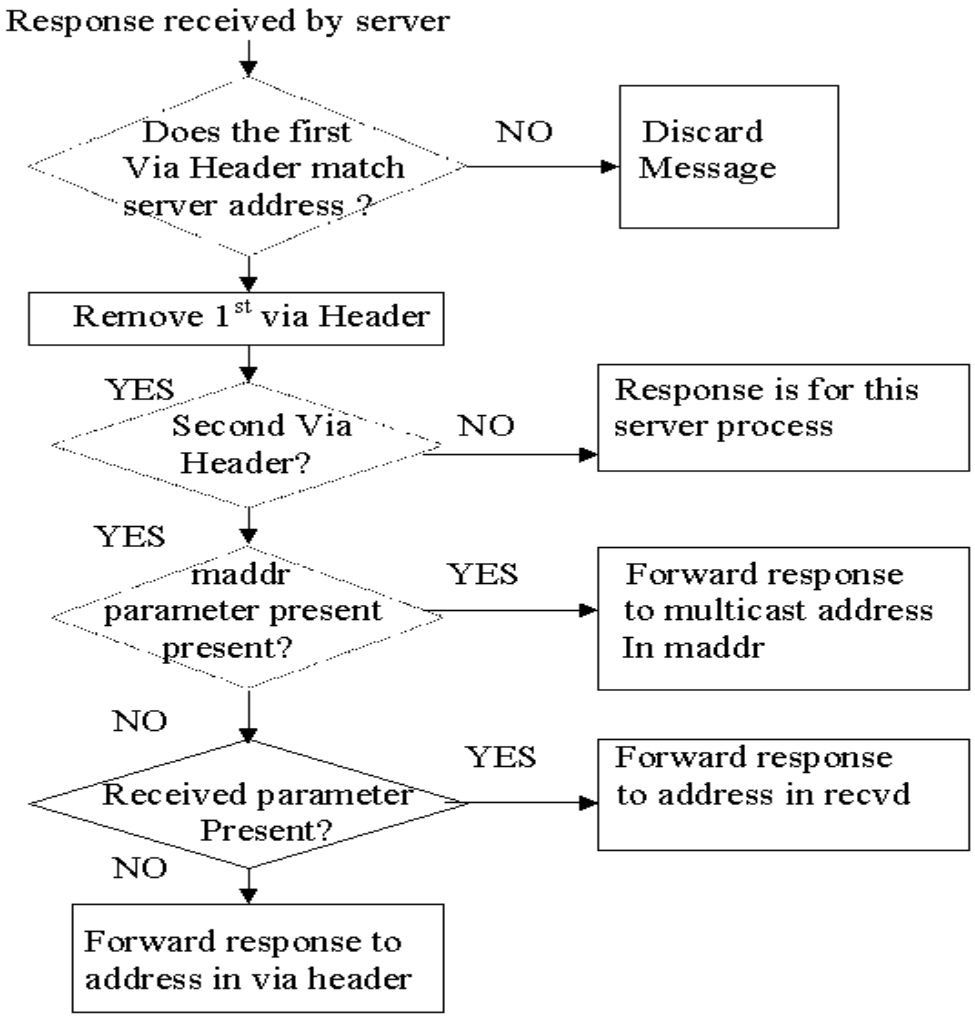
Via Field in Header

- The Request headers include a Via field
- The Via field indicates the path taken by the request so far.
- Every proxy adds a Via Header with its address to make sure that responses within a transaction take the same path (to avoid loops, or to make sure that same firewall will be hit on the way back)
- This prevents request looping and ensures replies take the same path as the requests, which assists in firewall traversal and other unusual routing situations.

Via Headers and Routing

- **Via** headers are used for routing SIP messages
- Requests
  - Request initiator puts address in **Via** header
  - Servers check **Via** with sender's address, then add own address, then forward. (if different, add "**received**" parameter)
- Responses
  - Response initiator copies request **Via** headers.
  - Servers check **Via** with own address, then forward to next **Via** address
- All Via headers are copied from request to response in order
- Response is sent to address in top Via header

Via Header (cont.)



Step #5 Receiving Subsequent Requests

- Subsequent to receipt, the following checks are made:

1. If the Call-ID is new,
   - the request is for a new call, regardless of the values of the To and From header fields.
2. If the Call-ID exists,
   - the request is for an existing call.
   - If the To, From, Call-ID, and CSeq values exactly match (including tags) those of any requests received previously,
     - the request is a retransmission.
3. If there was no match to the previous step,
   - To & From fields compared against existing call leg local and remote addresses.
   - If there is a match, & the CSeq in the request > last CSeq received on that leg,
     - the request is a new transaction for an existing call leg.

Reliability

- If UDP is used:

  -SIP client should retransmit a BYE, CANCEL, OPTIONS, or REGISTER request, exponential backoff, starting at a T1 second interval, doubling the interval for each packet, and capping off at a T2 second interval.

  -Retransmit a INVITE request with an interval that starts at T1 seconds, exponential back off, cease retransmissions if a provisional or definitive response recvd., or once it has sent a total of 7 request packets

- Clients using TCP do not need to retransmit requests

Authentication & Encryption

- SIP supports a variety of approaches:
  - end to end encryption
  - hop by hop encryption
- Proxies can require authentication:
  - Responds to **INVITE**s with **407 Proxy-Authentication Required**
  - Client re-**INVITE**s with **Proxy-Authorization** header.
- SIP Users can require authentication:
  - Responds to **INVITE**s with **401 Unathorized**
  - Client re-**INVITE**s with **Authorization** header

IP Multimedia Subsystem

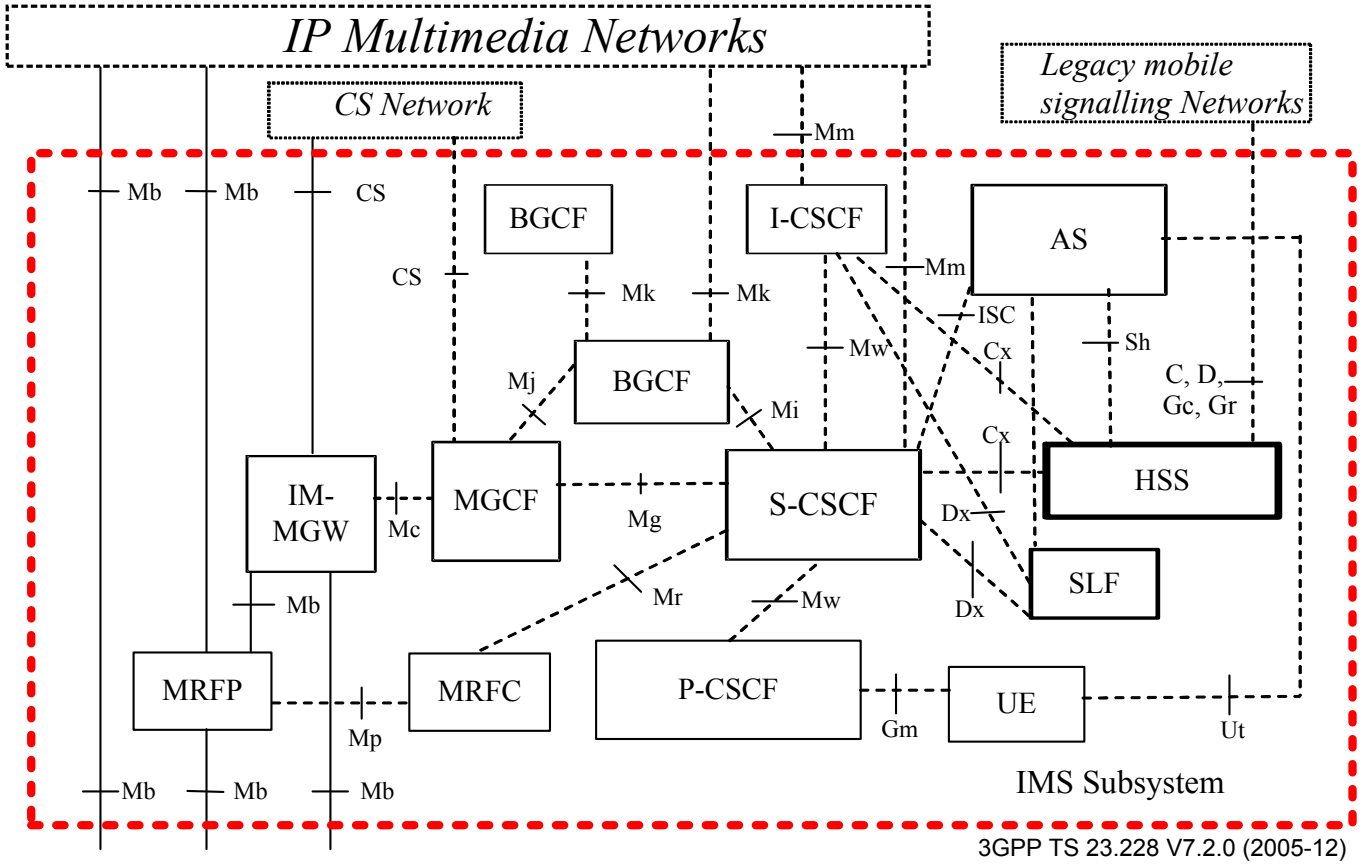
Introducing IMS

- Standards experts sought to solve these issues and move to VoIP for services
  - Resulting architecture is called IP Multimedia Subsystem or IMS
  - IMS began in the wireless community (3GPP/3GPP2), but is now being accepted by a variety of carriers and industry organizations
    - The IETF, ETSI/TISPAN, CableLabs, ITU-T support it as a framework for IP multimedia applications and services

What is IMS…

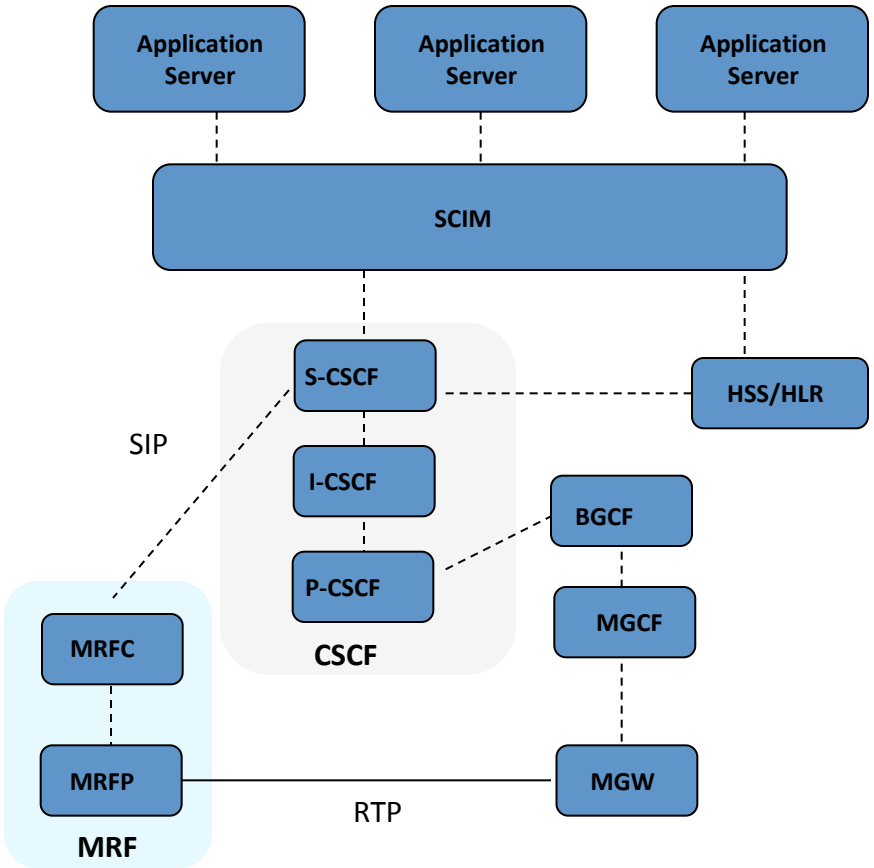
- IP Multimedia Subsystem (IMS) is an architecture that enables wireline, wireless and cable operators to offer a new generation of rich multimedia services
  - Across both circuit switched and packet switched networking infrastructures
- IMS defines a architecture of logical elements using SIP for call signaling between network elements
  - Provides a layered approach with defined service, control, and transport planes

The IMS Architecture



3GPP TS 23.228 V7.2.0 (2005-12)

IMS – Simplified View



Key Elements:

- AS – Application Server
- SCIM - Service Capability Interaction Manager
- MRFC - Multimedia Resource Function Controller
- MRFP - Multimedia Resource Function Processor
- MRF – Media Resource Function
- CSCF- Call Session Control Function
- BGCF - Breakout Gateway Control Function
- MGCF - Media Gateway Control Function
- MGW - Media Gateway
- HSS - Home Subscription Server
- HLR - Home Location Register

An IMS Call Flow

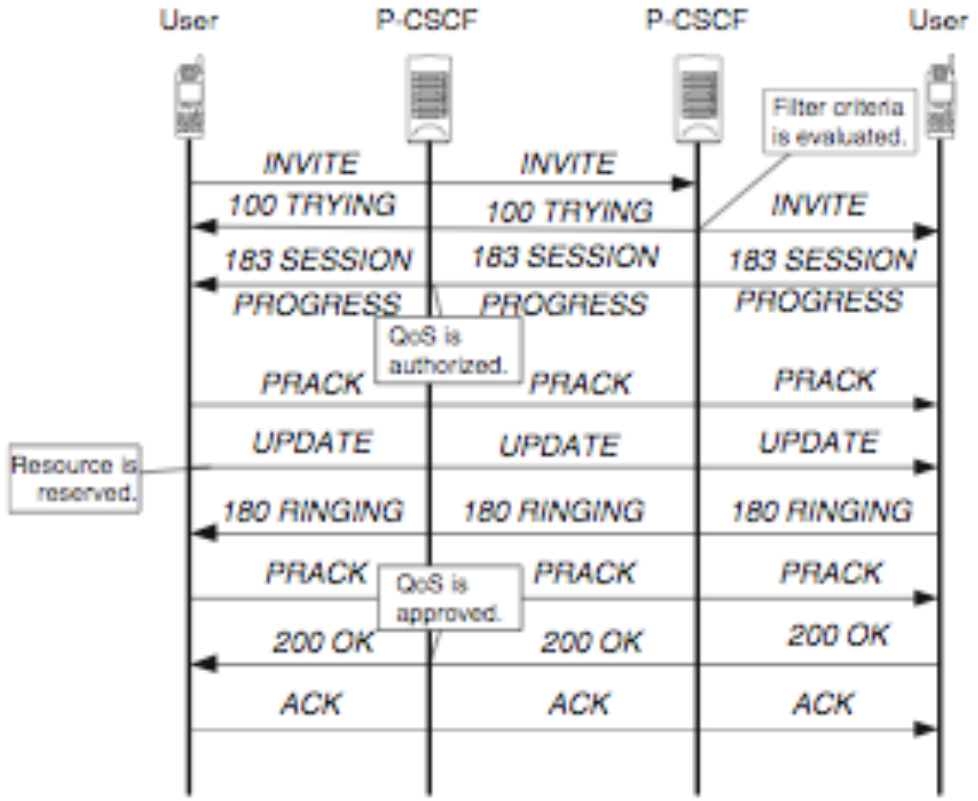


Fig. 3.18 IMS call flow

Other Key IMS Concepts

- Multiple Plane Architecture
  - Makes use of separate planes:
    - Application, Transport and Session Control
- Common Security and Login functions
  - Makes use of Diameter protocol and HSS (Home Subscriber Server) to validate users
- Applications and Services are independent of Access Method
  - Enables support for 3G mobile, WiFi, DSL, etc.

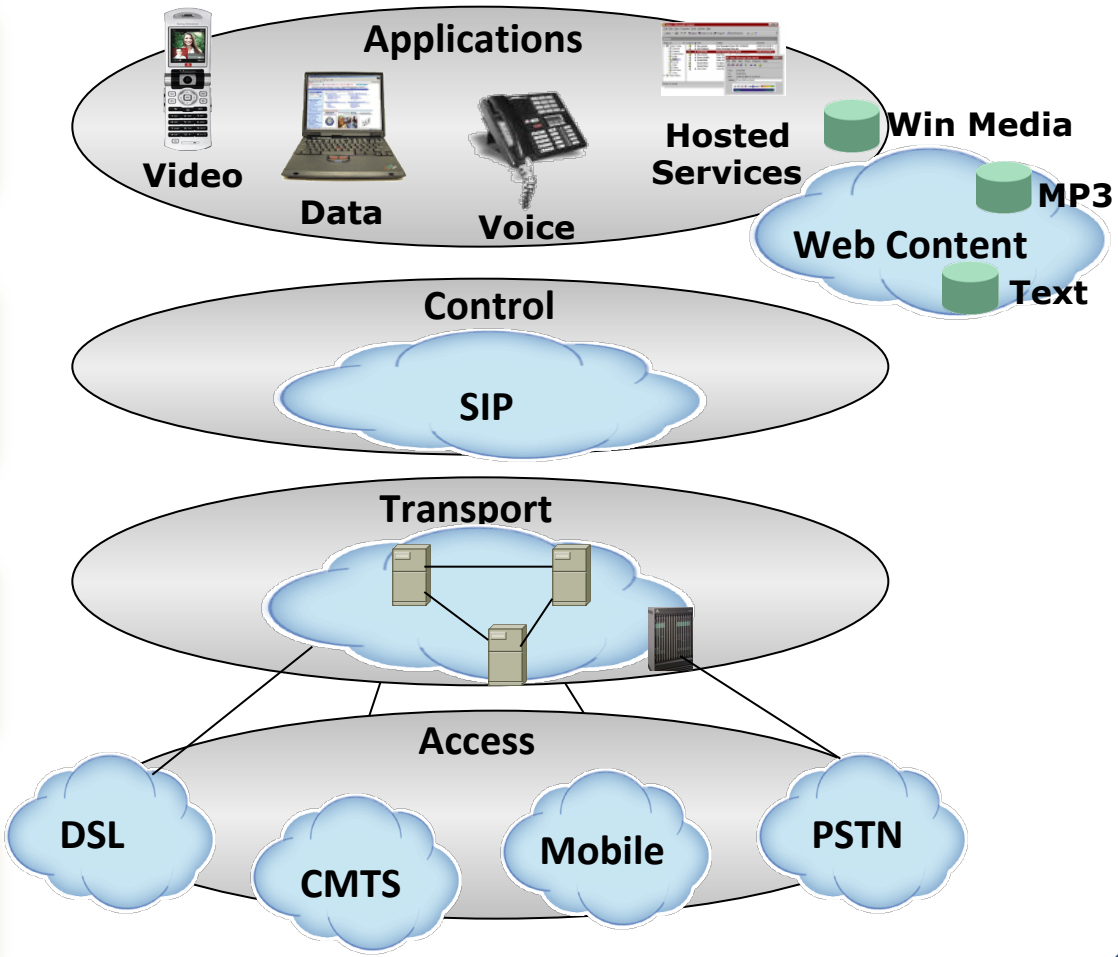
IMS Benefits

Converged Applications
- Across Networks
- Reduced development costs and time
- Voice, Video and data services
- Write once / use many

Shared Resources
- Media server resources
- Common user data
- Single user profile across applications
- Integrated applications

Session Control
- Common Session Control (SIP)
- Provides common service policies
- Leverages investments across multiple applications

Access Network Agnostic
- Eliminates multiple service solutions
- Network transparency
- Consistent services across networks



References

- Mobile Broadband, Ergen
- IMS, J. Rafferty
- Internet Telephony based on SIP, H. Sinnreich, A. Johnston
- A Multi-gigabit Rate Deep Packet Inspection Algorithm using TCAM, J-S Sung, et. al.
- CS40 Lecture 6: Security, R. Johari
- SIP, N. V. Pandrye
- Security Evolution on the Edge, W. Wilkening
- Qos in Data Networks, O. Ruso
- DPI, dpacket.org
- ROHC for multimedia services, F. Fitzek, ASU