T-110.5140 Network Application Frameworks and XML

Routing and mobility

10.2.2009

Tancred Lindholm

Based on slides by Sasu Tarkoma and Pekka Nikander

Contents

- Background
- IP routing and scalability
- Mobility

Background

- What is network architecture?
- Layered architecture
- The original requirements for IP
- Later requirements for IP

Network architecture

- A set of principles and basic mechanisms that guide network engineering
  - Physical links
  - Communication protocols
    - Format of messages
    - The way in messages are exchanged
    - Protocol stack
- Where is the state?

Protocol Stack

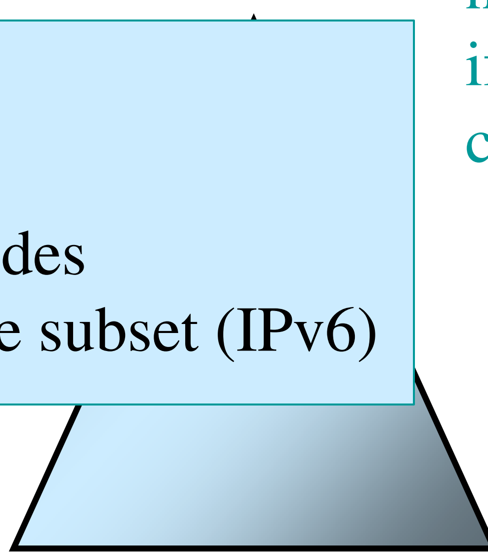
- Layers are part of a network architecture
  - Provide services for layers above
  - Hiding the complexity of the current layer
- Multiple layers are needed in order to reduce complexity
  - Relatively simple interfaces between layers
  - Sometimes lot of complexity inside layer
  - Separation of network functions
  - OSI, TCP/IP
- Protocols are building blocks of a network design

Naming, Addressing, and Routing

NAMING

How to identify and name a node? Even if its address changes.

unicast: to a specific node
broadcast: to all nodes
multicast: to a subset of nodes
anycast: to any one in some subset (IPv6)



ADDRESSING

Where is the node located?

ROUTING

How to route information to the node's address?

TCP/IP Network Stack

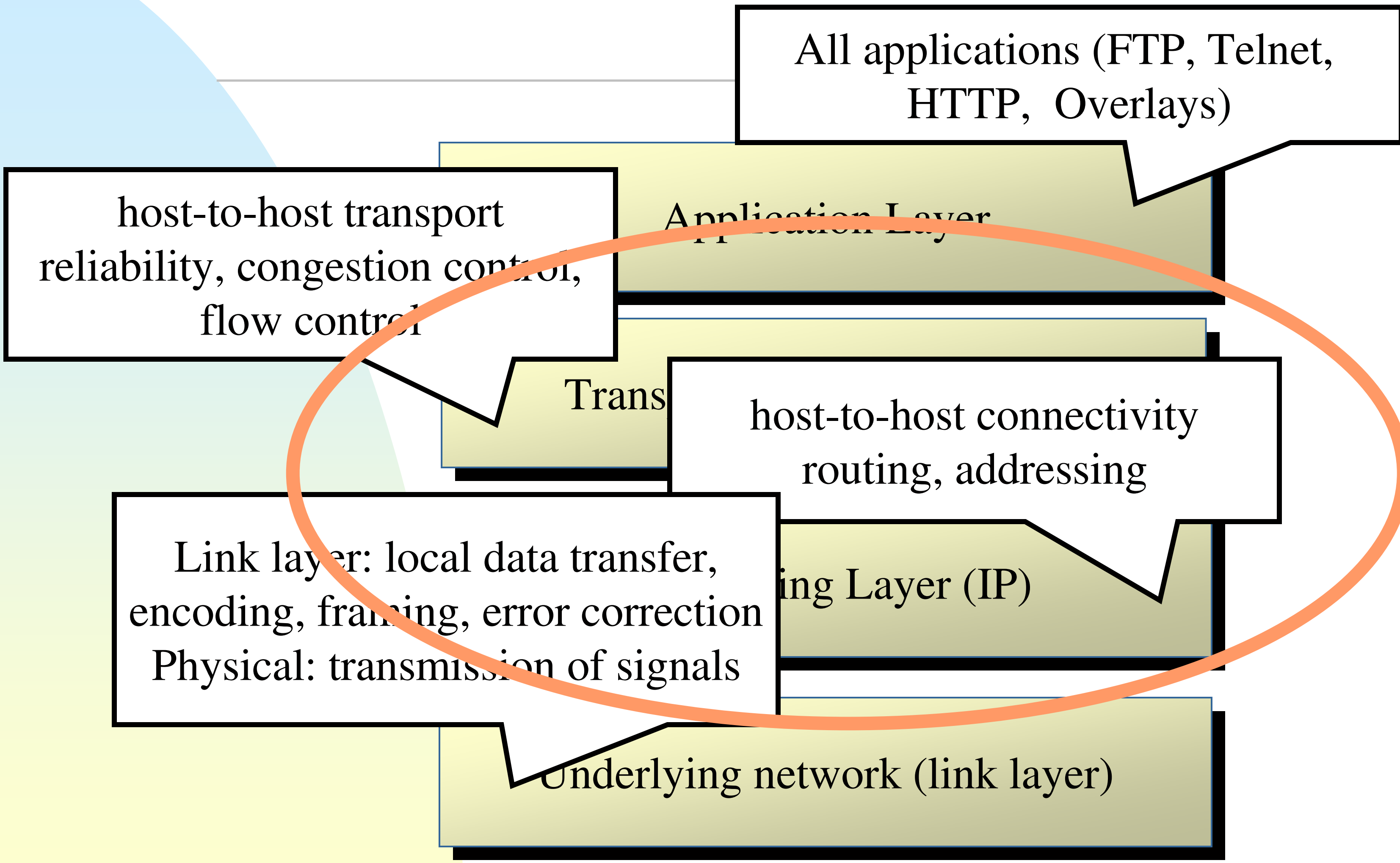
Application Layer

Transport Layer (TCP/UDP)

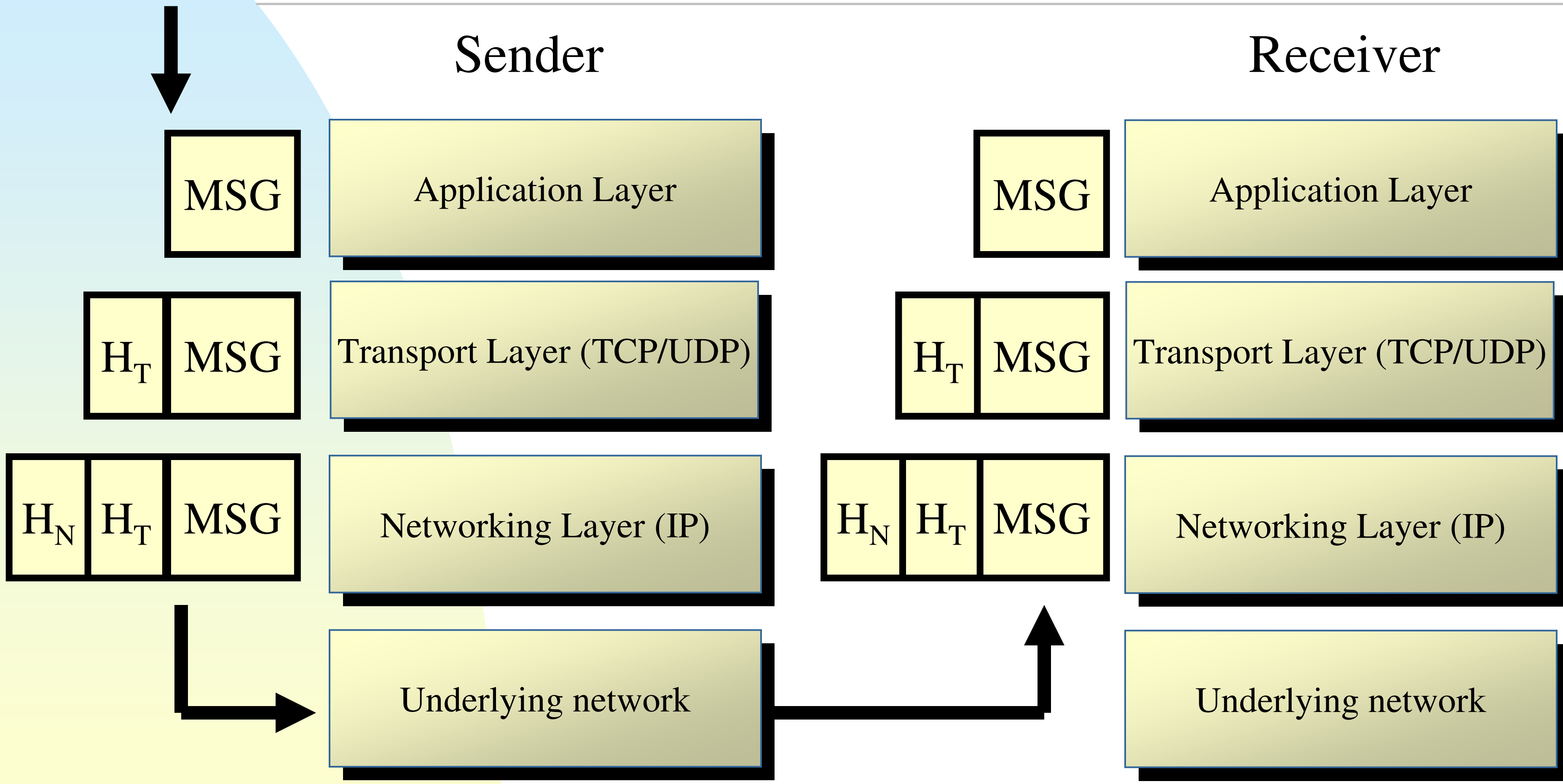
Networking Layer (IP)

Underlying network (link layer, physical)

TCP/IP Network Stack



Protocol Layering



Virtual Circuits

- Alternative to datagram routing
- Carries bit streams
- Resources reserved for each session (buffers, bandwidth)
- Guaranteed QoS
- State is stored by intermediate elements (ATM,..)
  - Timing and reliability requirements

Packet Switching

- No connection setup at network layer
- No state about end-to-end connections at routers
- Packets forwarded using destination host address
  - Different paths may exist to a destination
  - Store and forward
- Routing protocol goal
  - Find the best route through the network
  - Link cost: delay, monetary cost, congestion level

Original requirements for IP

- Goal: universal end-to-end connectivity
- Multiplexing
  - Packet switching
- Survivability (robustness)
  - Dynamic adaptation to outages
- Service generality
  - Support widest possible set of applications
- Runs over diverse networking technologies
  - Heterogeneity is unavoidable

Later requirements for IP

- Scalability
  - Exponential growth of # nodes was unplanned
  - Recurrent growth crises
  - Mainly a backbone issue (core routers)
- Distributed management
- Security
- Mobility
- Capacity allocation
  - fairness vs. unfairness

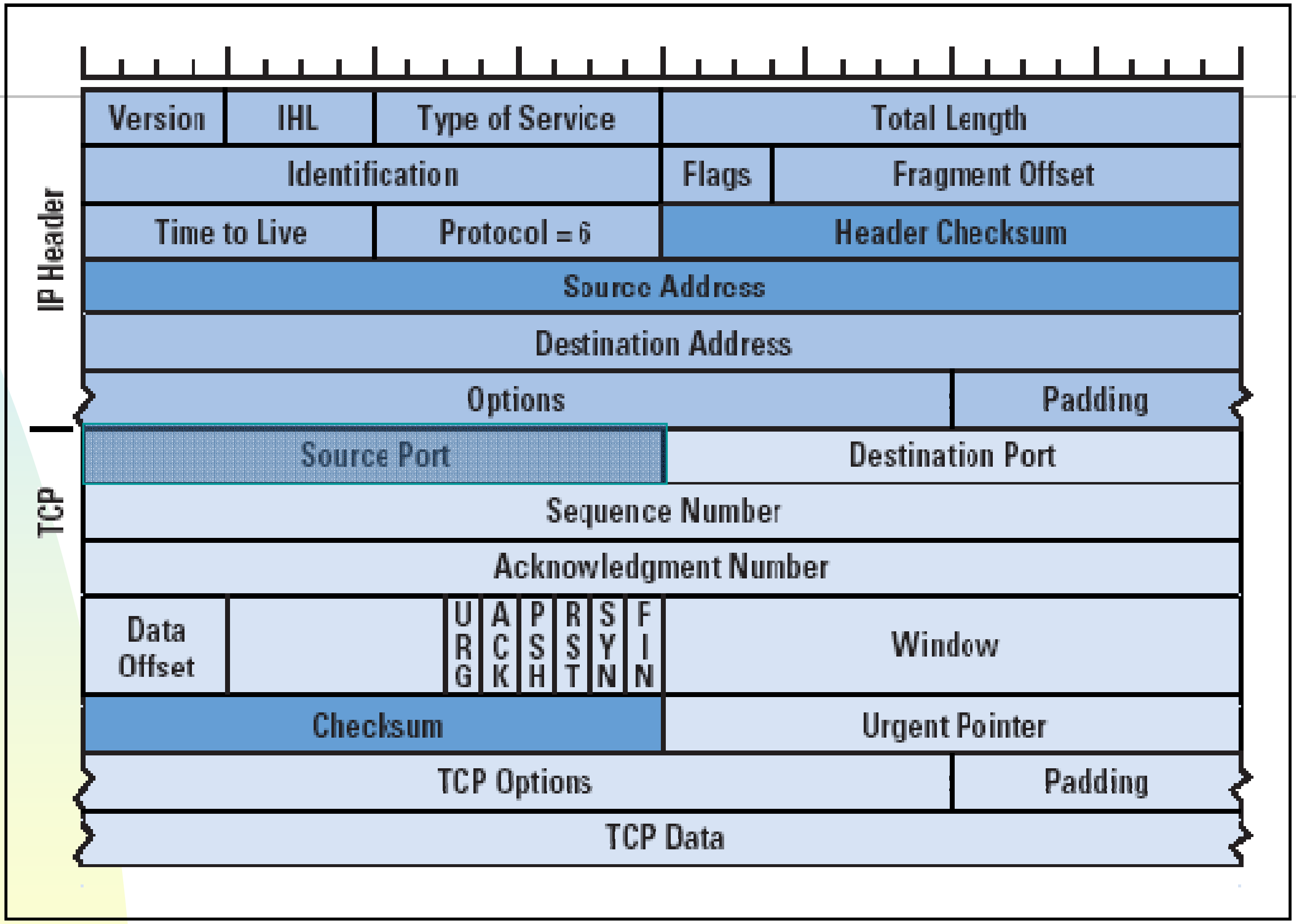
What has changed?

- Permanent IP address
  - Time-varying: DHCP, NAT, mobility
- End-to-end communication
  - Middleboxes, proxies, NATs, ..
- Globally and uniquely routable
  - NAT, firewalls
- Trusted end hosts
  - Hackers, spammers, …
- Four layers
  - Layer splits, cross-layer interactions

Problems with four layers

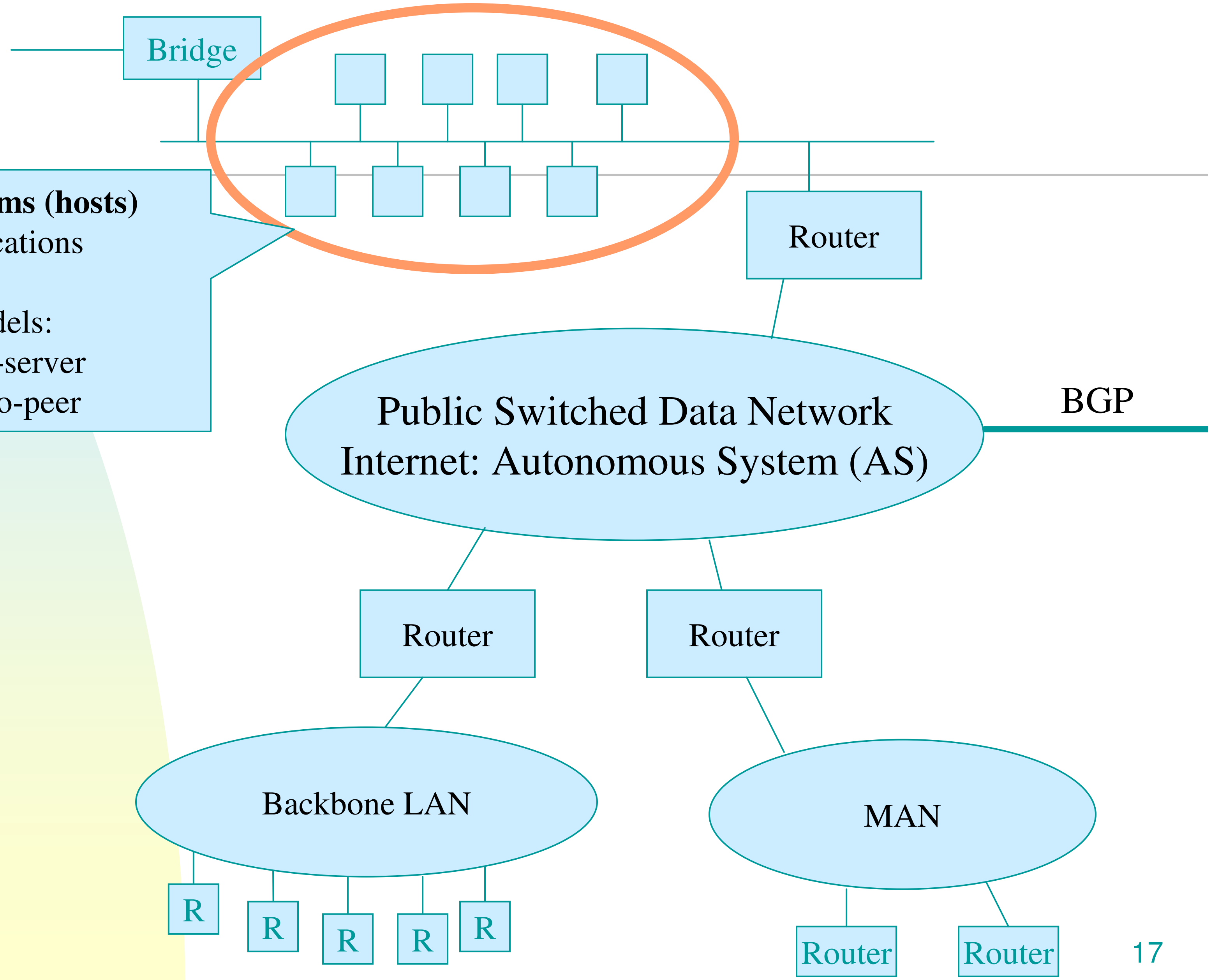
- Layer violations
  - Middleboxes, NATs
- Relation to the theoretical OSI 7 layers
  - What about presentation layer for Internet?
    - XML
  - What about session layer?
    - Separate session management from data delivery
    - For example: SIP

Figure 1: TCP/IP Header Fields Altered by NATs (Outgoing Packet)

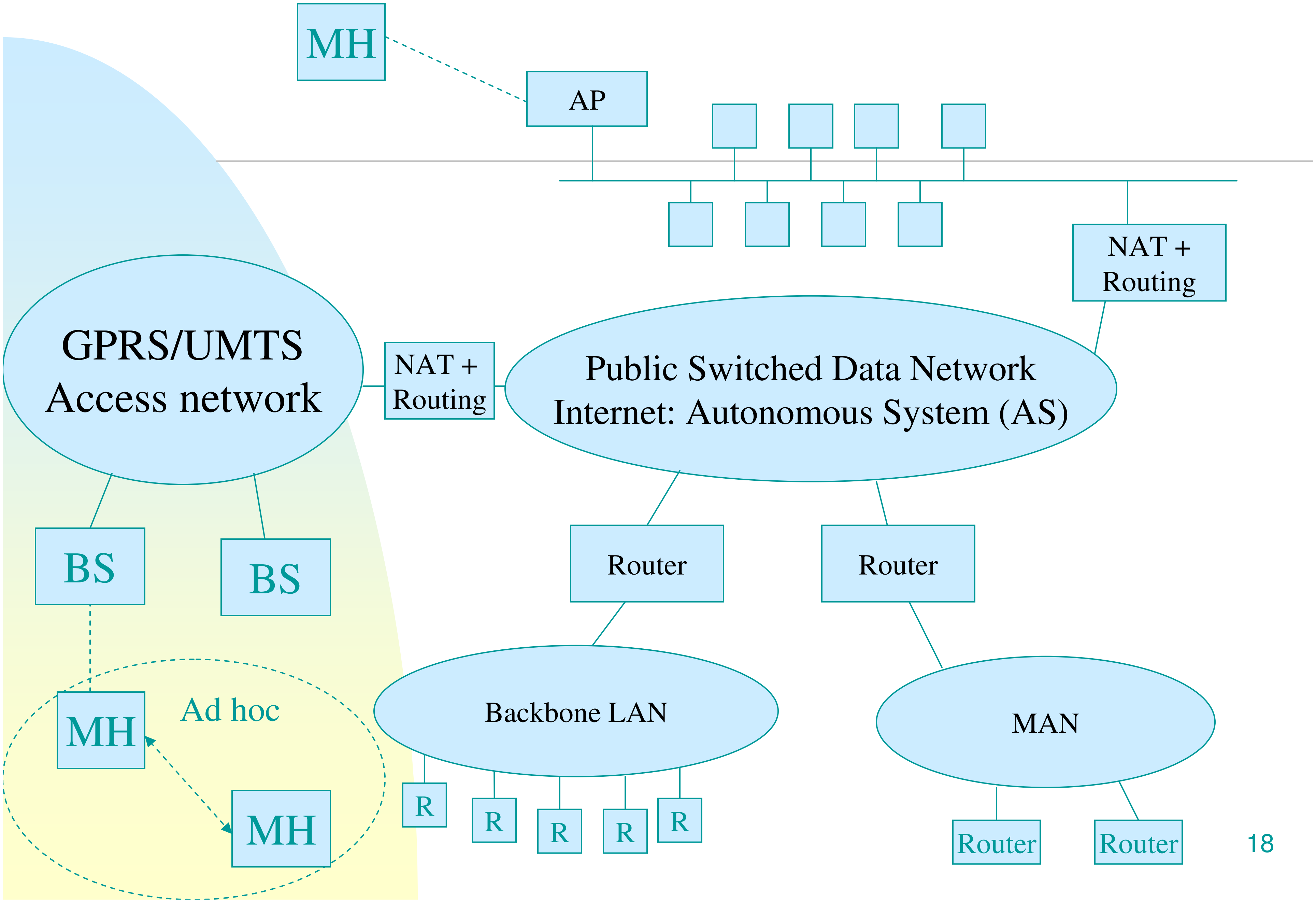


Source: Geoff Huston. Anatomy: A Look Inside Network Address Translators.
The Internet Protocol Journal - Volume 7, Number 3.

Networks: Basics



Networks: Wireless



What is routing?

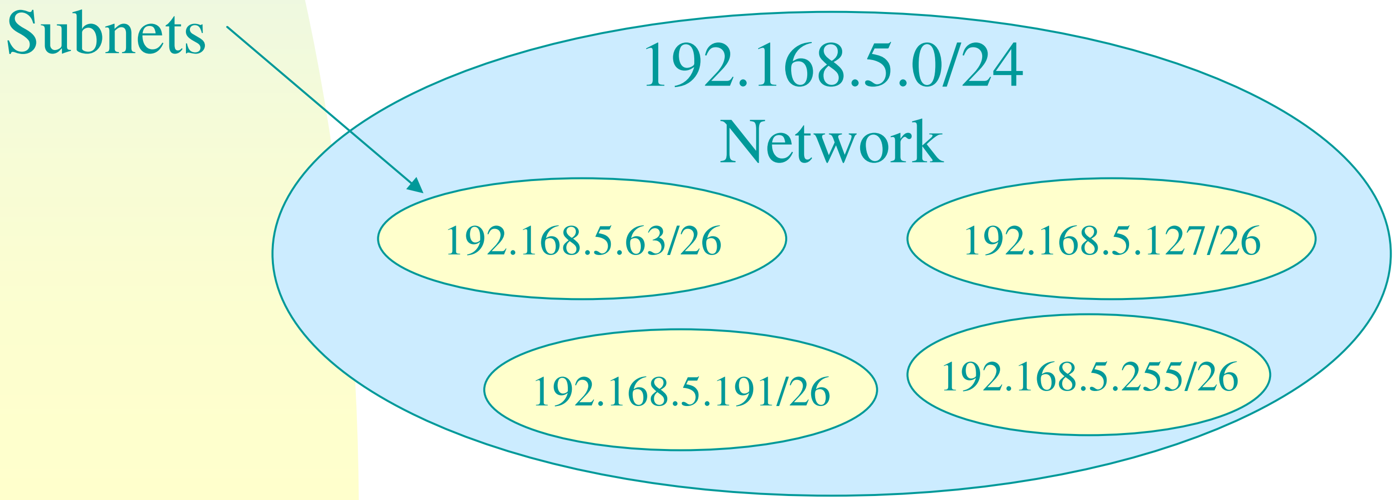
- Selecting the right path towards an address
- Addresses, names of locations or locators
- Routing table used for path selection
- Path selection algorithm
- How to represent topology information?
  - In address vs.in the routing table

IP addresses

- Topological structure is reflected by splitting IP addresses into a host and network part
- Benefits of hierarchical addressing
  - reduced number of routing table entries and localized allocation of addresses.
- Subnetting of networks
  - A subnet takes responsibility for delivering datagrams to a certain range of IP addresses.
  - The network part is now extended to include some bits from the host part.
  - Needed for large networks

Subnetting

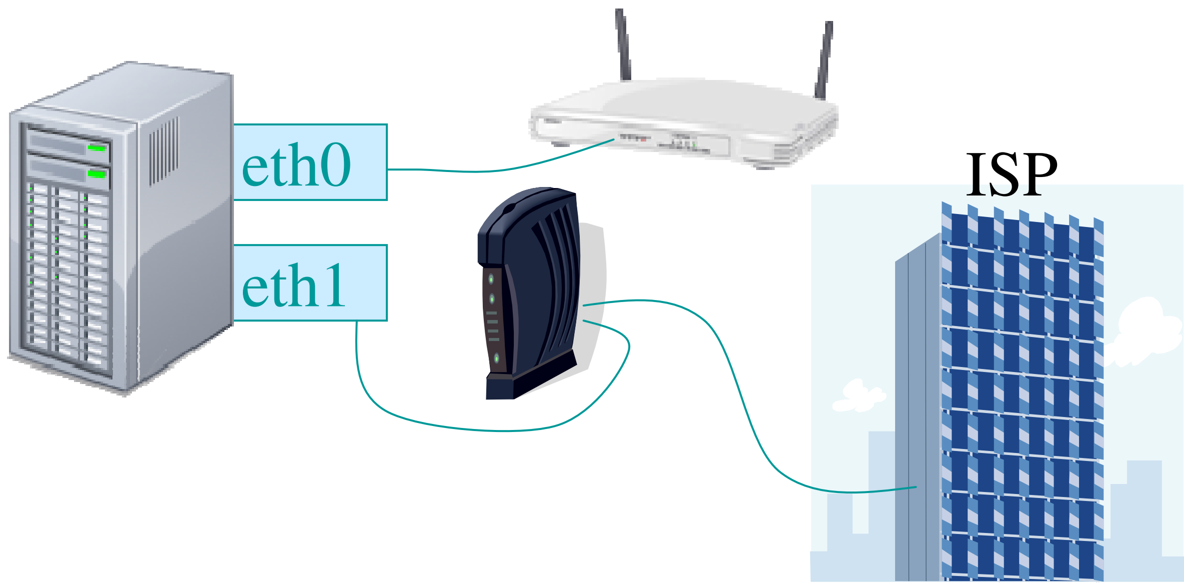
- A subnet mask is a 32-bit value that identifies which bits in an address represent network bits and which represent host bits.
- Note: Subnet-masks affect only internal structure and behaviour of a network!



Routing Tables

- There are four basic items of information
  - A destination IP address.
  - A gateway IP address.
  - Various flags
    - Usually displayed as U, G, H. U means the route is up. G means the route is via a gateway. H means the destination address is a host address as distinct from a network address.
  - The physical interface identification.
  - Additional info
    - Metrics, protocols

Example Table

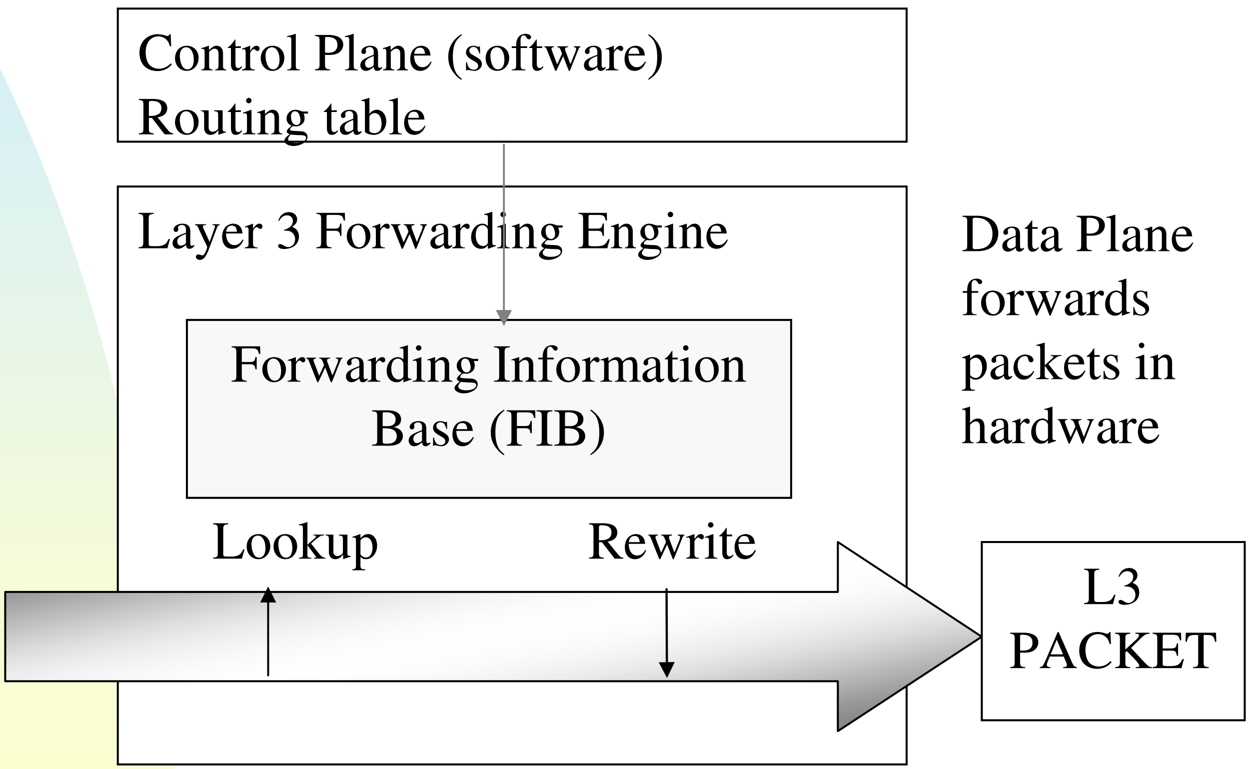


Destination	Gateway	Genmask	Flags	Metric	Ifac
192.168.2.0	0.0.0.0	255.255.255.0	U	0	eth0
82.181.168.0	0.0.0.0	255.255.248.0	U	0	eth1
169.254.0.0	0.0.0.0	255.255.0.0	U	1000	eth0
0.0.0.0	82.181.168.1	0.0.0.0	UG	100	eth1

Host vs. router

- Host: simple static-ish routing
  - First look for the destination address as a host address in the routing table
  - Then look for the destination net address
  - Then use one of the default addresses (there may be several).
- Router
  - Very large routing table, especially in the backbone
  - Routing protocols
    - Interior Gateway Protocols (OSPF)
    - Exterior Gateway Protocols (BGP)...

Fast Routing



Different types of routing

- Source routing
  - Path selection by sender
  - Path encoded in the packet
  - High cost for the sender node
  - Strict source routing vs. loose source routing
- Hop-by-hop routing
  - Router selects the next hop
  - High cost for the backbone routers
- Per-host or per-network routes

Evolution of IP routing

- Class-based routing
  - A ,B and C classes (7,14, 21 bits of network)
  - Routing tables carried entries for all nets
  - No topological aggregation (only network address boundaries)
- Classless routing
  - Using the variable length network mask to aggregate addresses
  - Routers forward mask (longest prefix)
- Too many small networks requiring multiple class C - addresses
  - C class has max 254 hosts
  - Huge routing tables

CIDR

- CIDR (Classless Interdomain Routing)
  - Routing prefixes carry topology information
  - Contiguous blocks of C-class addresses
  - Smaller routing tables
- Addresses two problems
  - Exhaustion of IP address space
  - Size and growth rate of routing tables
- Address format <IP/prefix bits>

CIDR and Route Summarization

- Examples of classless routing protocols
  - RIP version 2 (RIPv2), OSPF, Intermediate System-to-Intermediate System (IS-IS), and Enhanced Interior Gateway Routing Protocol (EIGRP)
- Route summarization: identify common prefix and route for networks

```
172.16.12.0/24 = 172.16.00001100.0
172.16.13.0/24 = 172.16.00001101.0
172.16.14.0/24 = 172.16.00001110.0
172.16.15.0/24 = 172.16.00001111.0
----------------------------------------
 172.16.12.0/22 = Summarized route
```

CIDR and IPv6

- CIDR present in IPv6 (fully classless)
- 128bit IPv6 address has two parts: network and host
  - network address includes the prefix-length
  - a decimal value indicating the number of higher-order bits in the address that belong to the network part
- ISP aggregates all its customers' prefixes into a single prefix and announces that single prefix to the IPv6 Internet

Border Gateway Protocol (BGP)

- BGP (Border Gateway Protocol) first became an Internet standard in 1989
- Know about politics
- BGP selects AS-level paths for inter-domain routing. Each AS may have multiple paths offered by neighbouring ASs.
- BGP-4 supports Classless Inter Domain Routing (CIDR) and is the routing protocol that is used today to route between autonomous systems.
- BGP uses TCP to establish a reliable connection between two BGP speakers on port 179.
- A **path vector** protocol, because it stores routing information as a combination of a destination and attributes of the path to that destination.
- The protocol uses a deterministic route selection process to select the best route from multiple feasible routes

BGP

- Characteristics such as delay, link utilization or router hops **are not** considered in this process.
- BGP runs in two modes: EBGP and IBGP.
  - EBGP (Exterior BGP) is run between different autonomous systems
  - IBGP (Interior BGP) is run between BGP routers in the same autonomous system

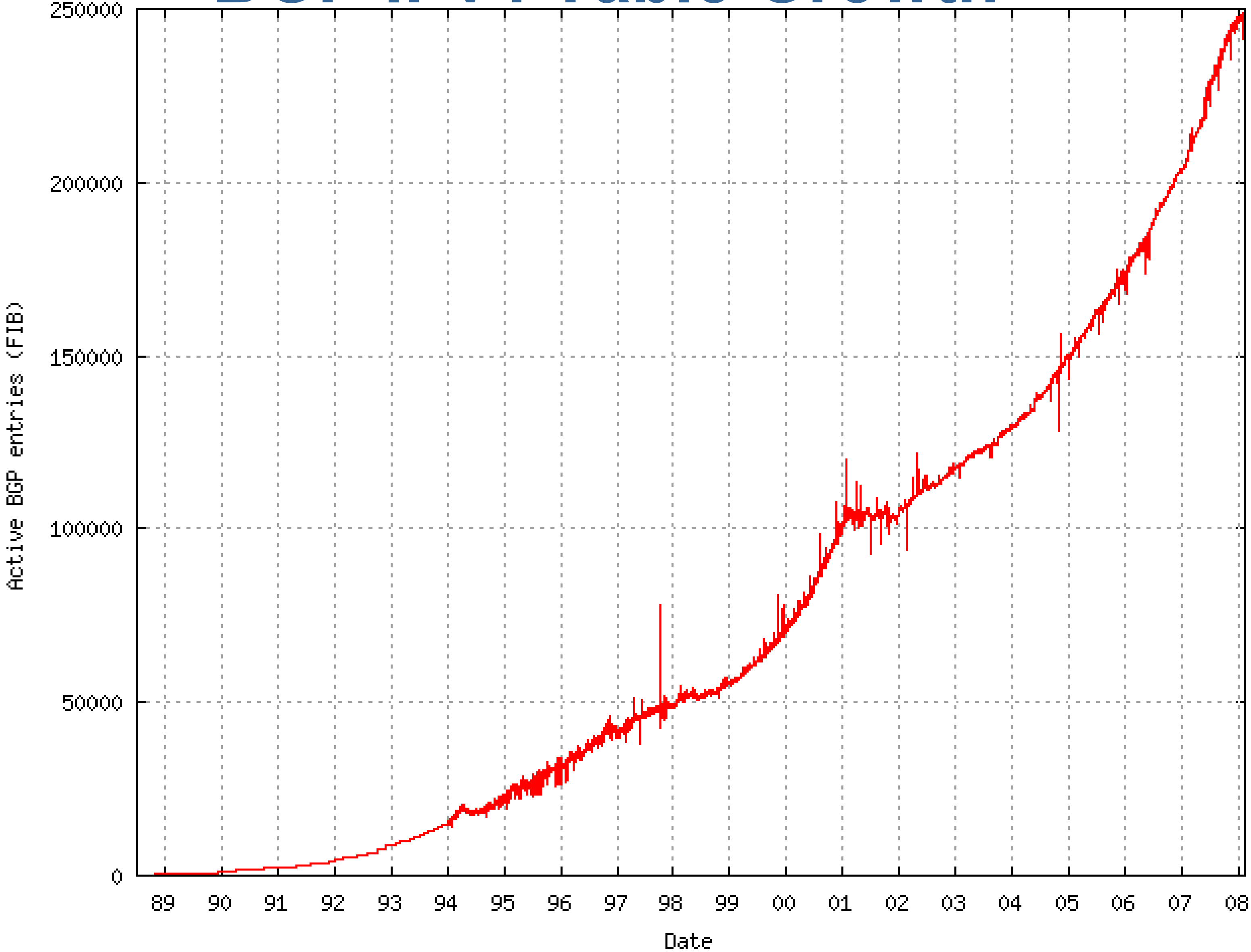
BGP cont.

- When the BGP router receives its neighbors' full BGP routing table (>100k routes),
  - Requires approx. 70 MB.
  - With the AS-PATH filters applied
    - 32k routes in 28 MB. 60% space decrease while preserving optimal routing.
- Problems
  - multihomed customers forget to stop reannouncing redirect routes from upstream A to upstream B
  - peer networks leak full tables to their peers
  - A misconfigured router leaks out all internal more specific routes (/48, /64, /128 prefixes)

BGP Problems

- Convergence time
- Limited policies
- Security problems
  - In 2008, Pakistan wanted to block YouTube
  - BGP update leaked, all Webwide youtube traffic directed to Pakistani /dev/null
  - YouTube became temporarily unusable

BGP IPv4 Table Growth



Source: http://www.cidr-report.org/#General_Status

Autonomous System Numbers

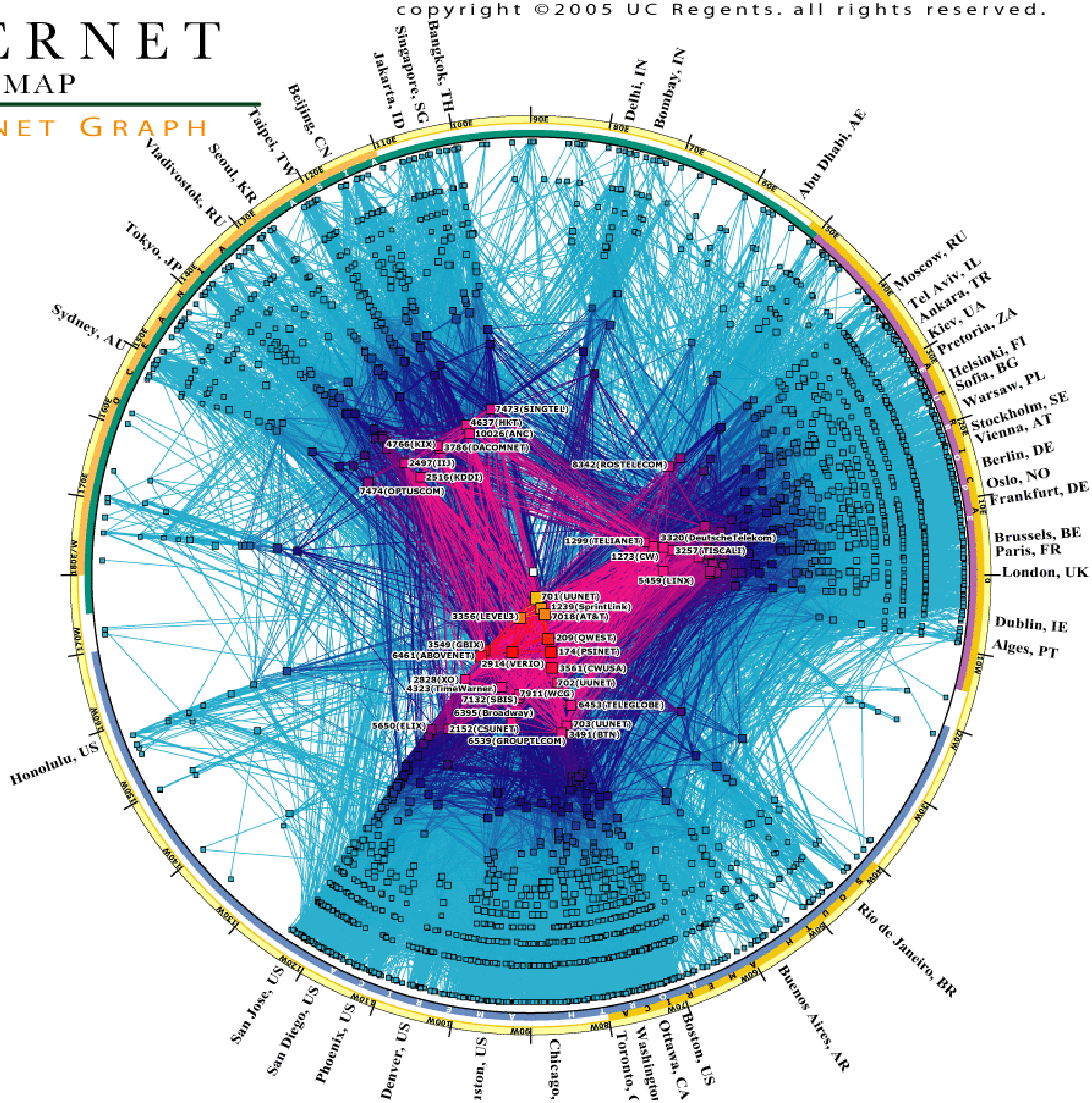
- Autonomous systems identified by 16-bit AS numbers
- Current estimate is that limit will be reached on February 2011
- IETF standards action in November 2006
  - IANA extended the AS number field to 32 bits
    - 65536 to 4,294,967,296 values
    - From Jan, 2007 32bit values have been available from the Regional Internet Number Registries (RIR)

IPv4 INTERNET
TOPOLOGY MAP
AS-level INTERNET GRAPH

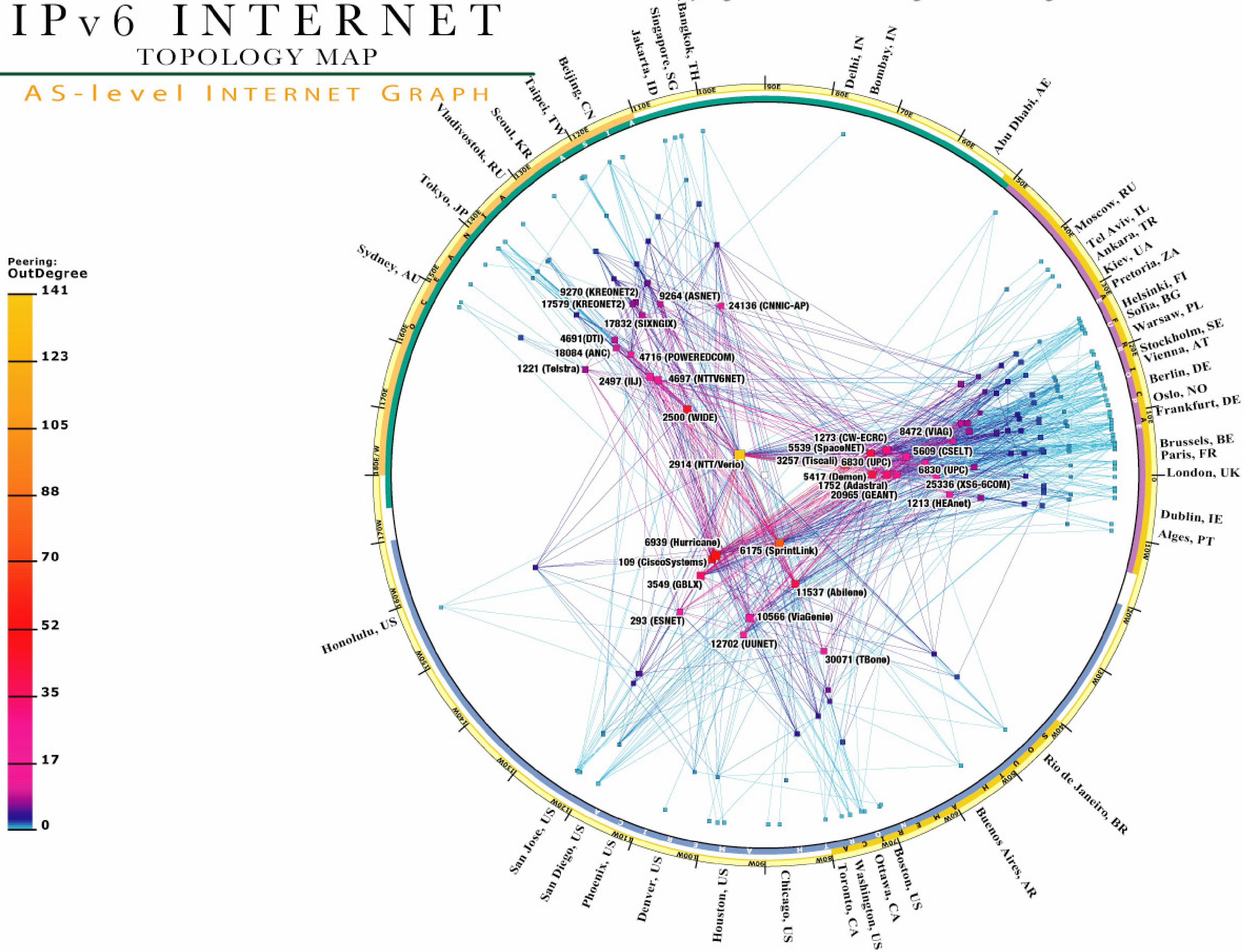
PDF Version

Peering: OutDegree

1659
1451
1244
1037
830
622
415
207
0



IPv6 INTERNET
TOPOLOGY MAP
AS-level INTERNET GRAPH



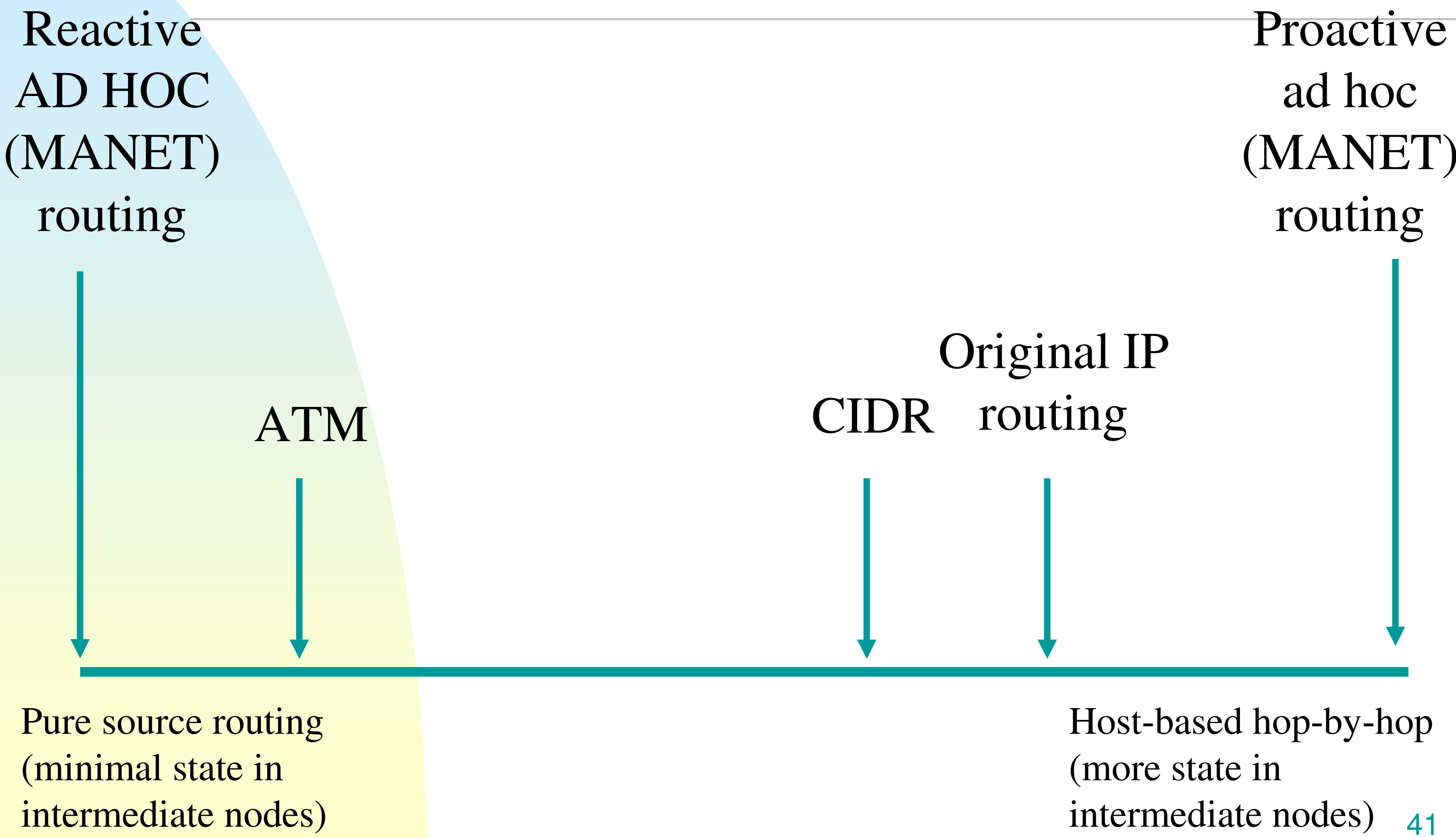
Mobile Ad Hoc Networks (MANET)

- Mobile Ad Hoc Networks
  - Routing for dynamic environments
  - Proactive protocols (maintain routing table)
    - continuously evaluate routes
    - no latency in discovery
    - possibly a lot of entries not used
    - large capacity to keep current info
  - Reactive protocols (on demand)
    - route discovery using global search
    - high latency
    - possibly not suited for real-time

MANET cont.

- IETF MANET Working Group
  - The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)
    - Source routing (route discovery & maintenance)
    - Route cache
      - Only communicating nodes cache a route
  - Ad Hoc On Demand Distance Vector (AODV) Routing (RFC 3561)
    - Route table
      - Also intermediary nodes keep a distance vector
    - Multicast
- Other protocols
  - Hierarchical, geographical, multicast, power-aware
- What is the expected size of the network?
- Feasibility of wireless multi-hop?
  - Capacity showed to be low.

Topology in address vs. routing table



Difficult Issues

- Convergence time of routing information
- State in the network
  - Per-connection state is bad? (e.g. NAT)
- Security of routing information
  - Whom to trust? How to represent authorization?
- QoS routing

Mobility

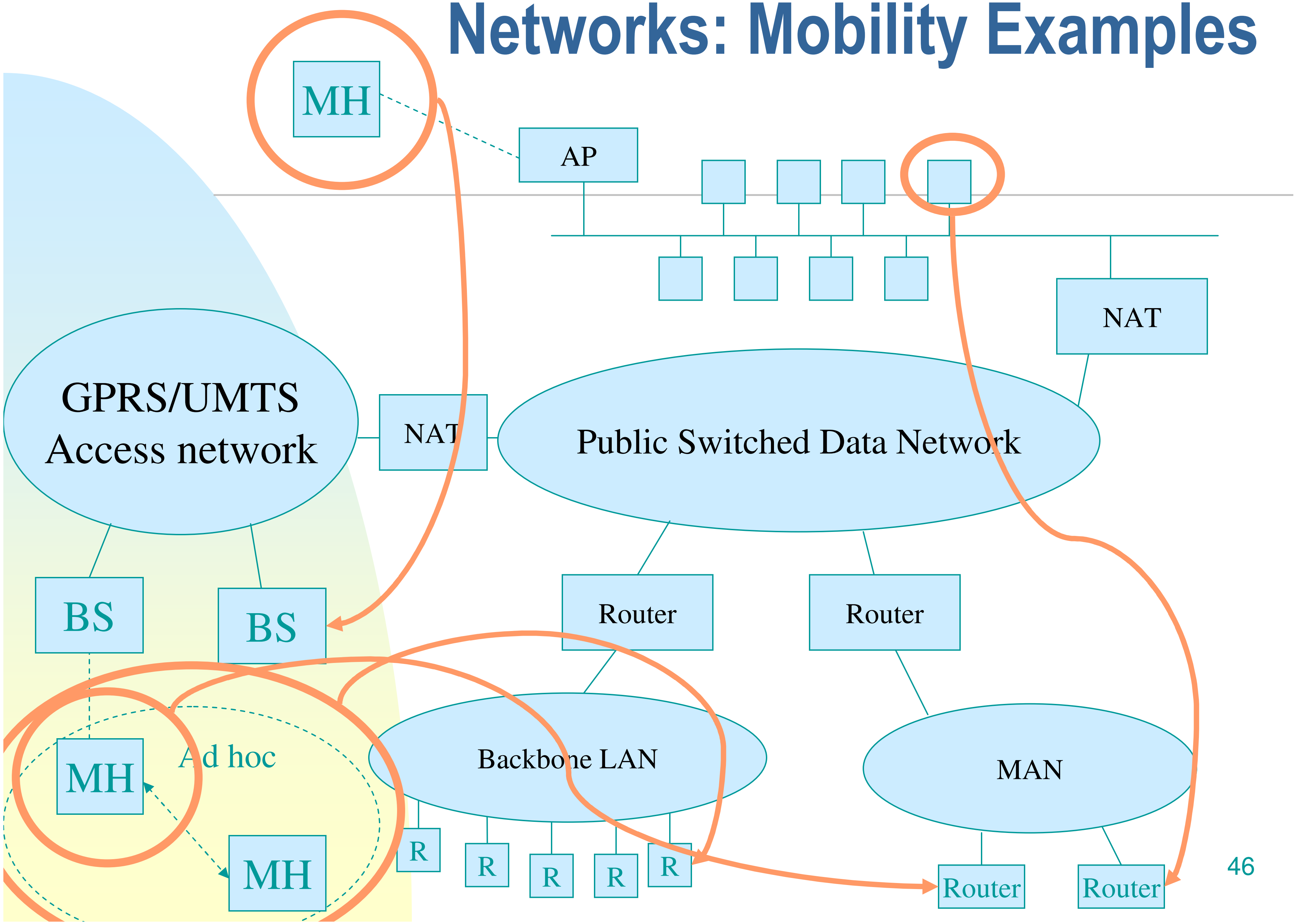
Mobility in Internet

- Routing from the mobility perspective
- Mobility on various layers
  - Mobile IP approach
  - Transport and application - level mobility
- Separating identifiers and locators
- Mobility management and rendezvous
- Security issues
- Lessons to learn

Routing vs. mobility

- Topology data aggregation is necessary
  - Cannot track all hosts in the world
  - IP addresses determined by topology
    - Network gives the routing prefix
- Mobile hosts must change their IP addresses
  - Causes sockets / connections to break
- How to communicate address changes?
- Goal of a mobility protocol
  - Transport and applications do not see address changes
  - Mobility transparency

Networks: Mobility Examples



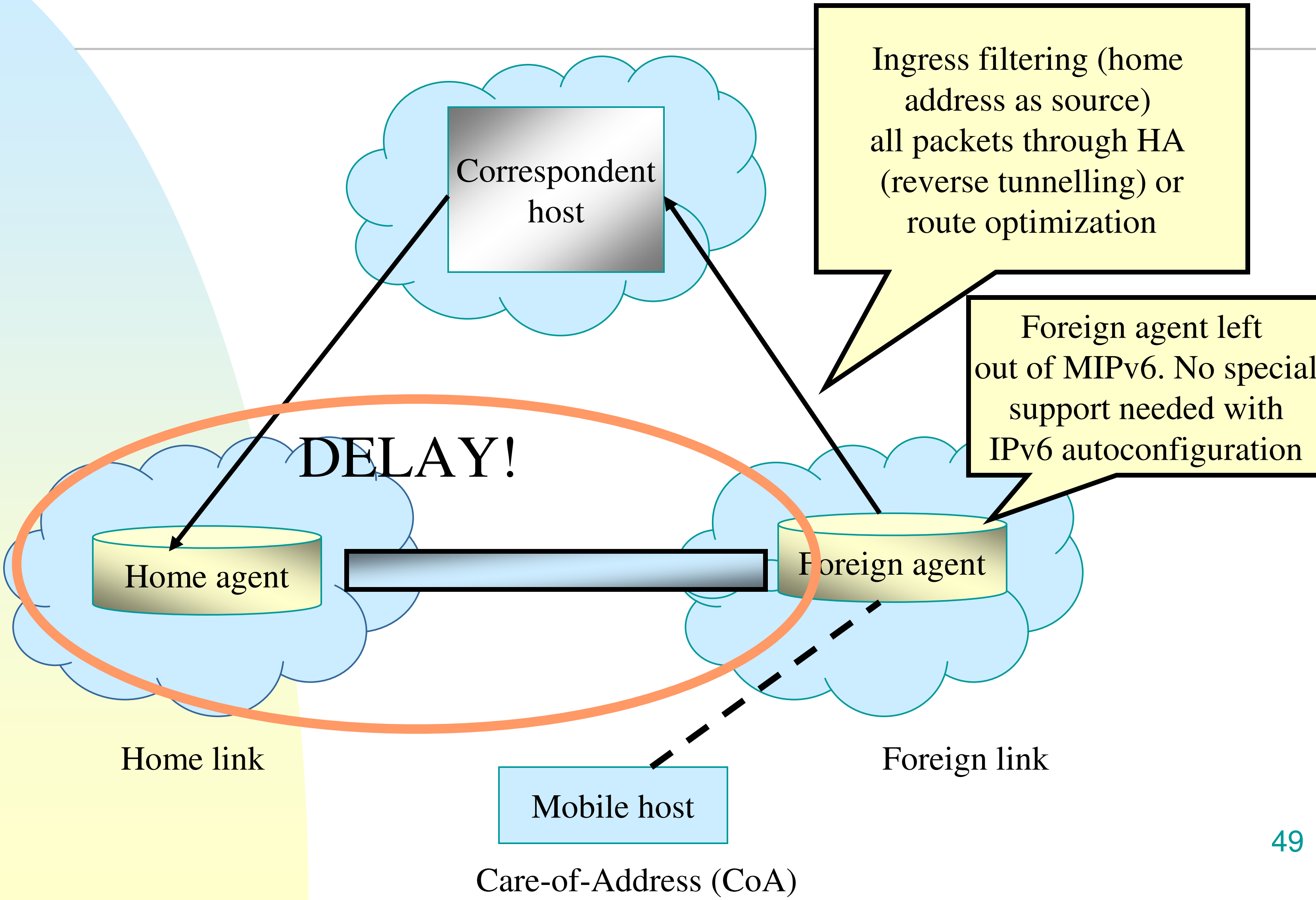
Mobile IP

- Two versions
  - IPv4 (optional)
  - integrated into IPv6 (with IPSec security)
- Home Agent (HA)
  - Home address
  - Initial reachability
  - Triangular routing / reverse tunneling
- Route optimization
  - Tunnels to bypass HA

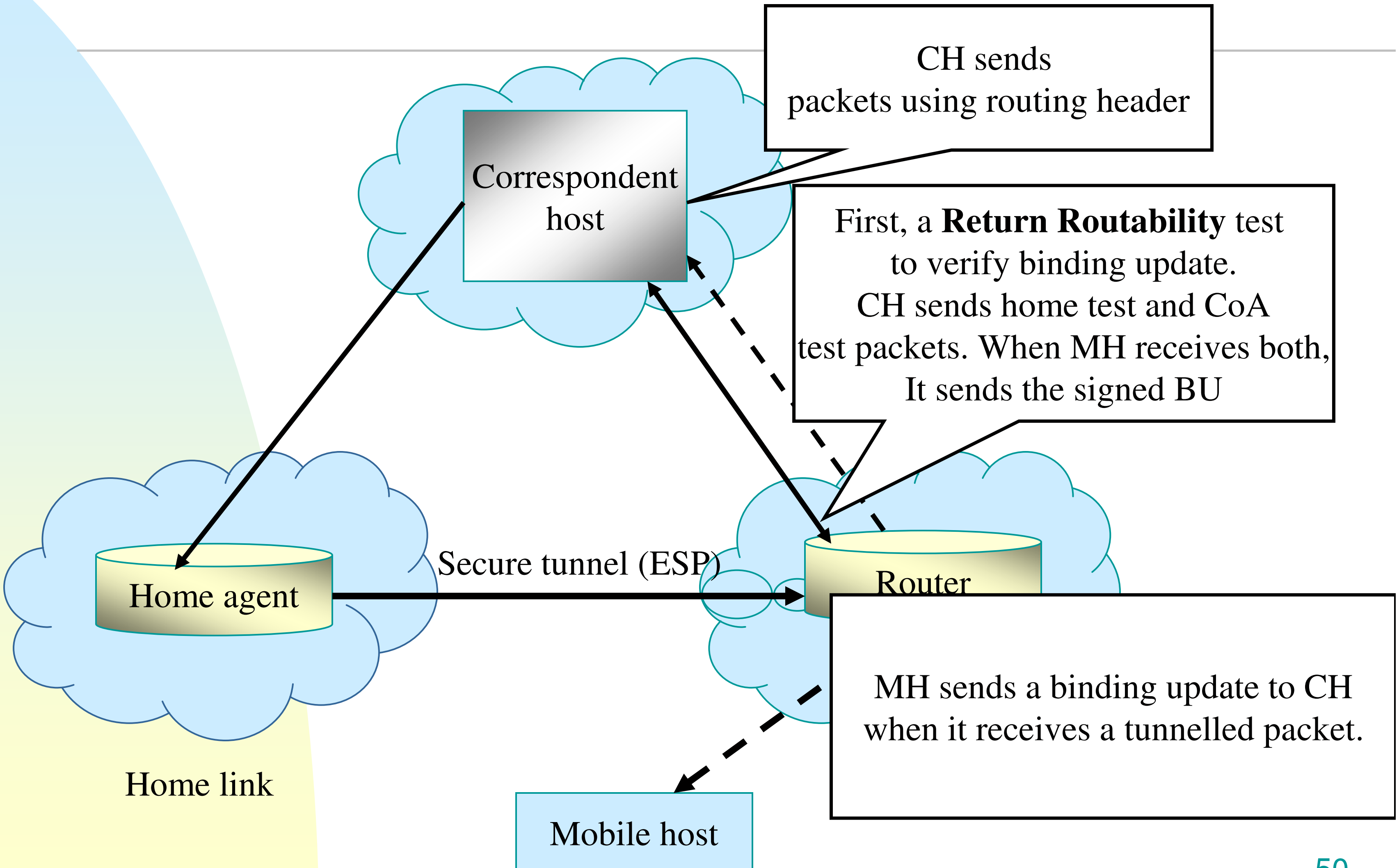
Security issues

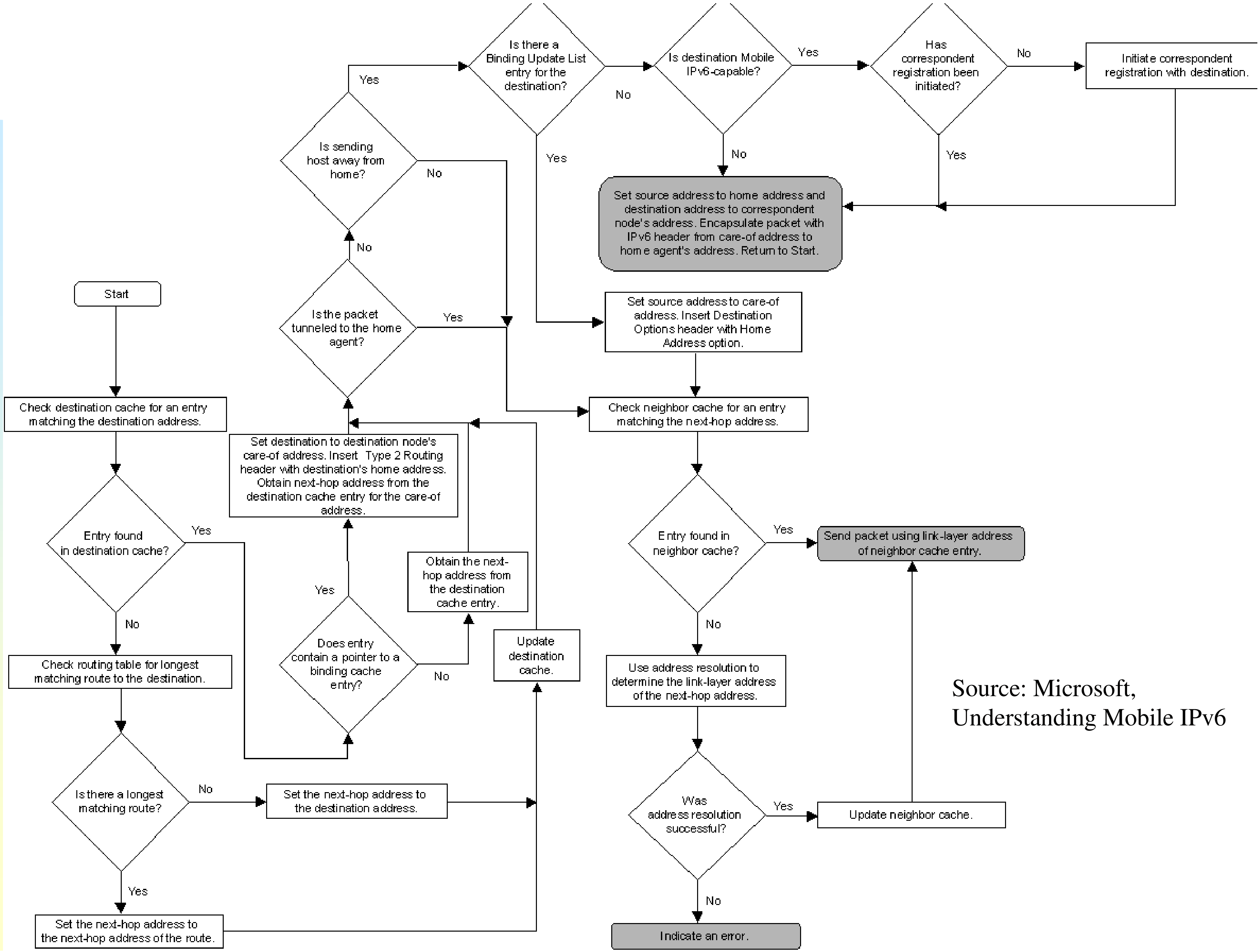
- Address stealing
  - Alice and Bob communicate
  - Mallory, sitting at C, tells Alice
    - Bob is now at C
- Address flooding
  - Mallory downloads at a high rate from A
  - Mallory tells A "I have moved to C"
    - C gets (temporarily) flooded

Mobility Example:Mobile IP Triangular Routing



Mobility Example:Mobile IPv6 Route Optimization





Source: Microsoft, Understanding Mobile IPv6

Security in Mobile IP

- MIPv6 RFC 3775/3776
  - Protection of Binding Updates
  - IPsec extension headers or the binding authorization data option
  - Binding management key, Kbm, which is established through return routability procedure
  - Protection of the mechanisms that MIPv6 uses for transporting data
- Protecting binding updates
  - Must be secured through IPsec
  - ESP is used for updates and acks
- Shoulds: init messages, some others

Rendezvous ("Meeting Point")

- How to find the moving end-point?
  - Tackling double jump
    - What if both hosts move at the same time?
    - With binding updates, this requires a rendezvous point
- Rendezvous point = well known "meeting point" (routable address)
    - In MIP, home agent can be the rzd point
- Mobility management is needed
  - Initial rendezvous to find mobile node
  - Can be based on directories
  - Requires fast updates to directories
    - Does not work well for DNS

Multi-layer Operation

- Mobility and multi-homing can be realized on different layers
  - Network
    - Mobile IP
  - Between network and transport
    - Host Identity Protocol (HIP)
  - Transport (SCTP)
    - TCP extensions
  - Application
    - SIP, Wireless CORBA, overlays
    - Re-establish TCP-sessions after movement

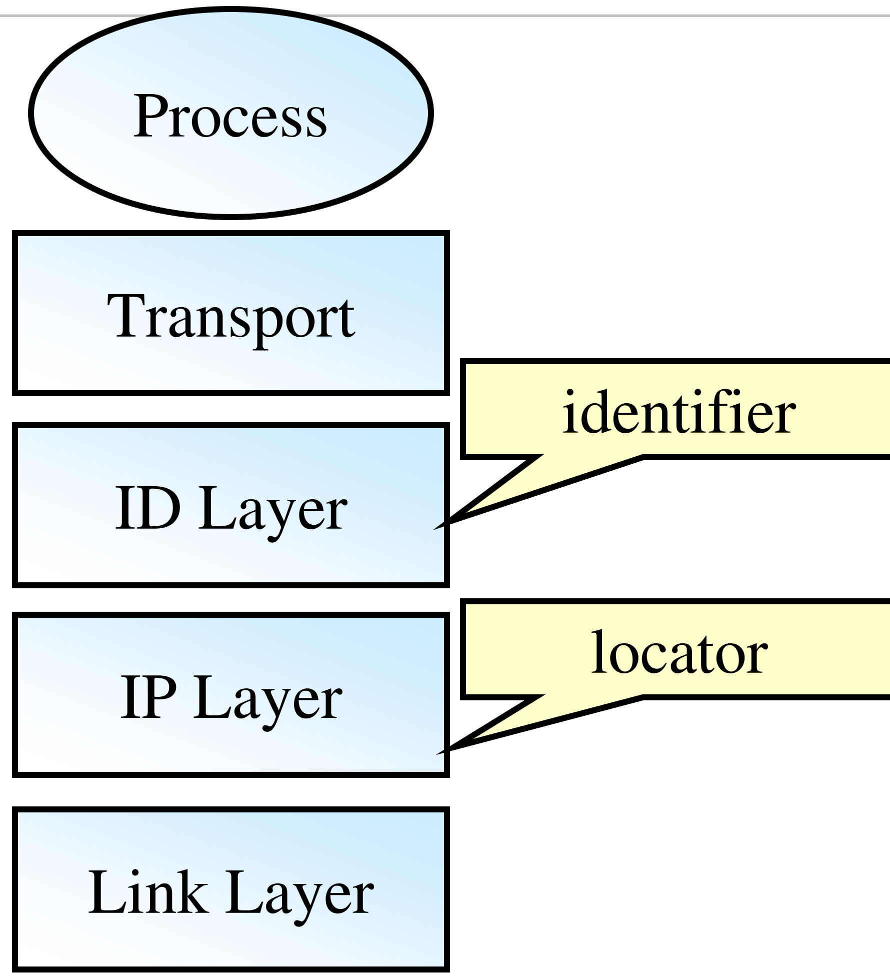
Separating Identifiers and Locators

- Problem: machine.domain.com is both name **and** address (b/c DNS limitations, early resolution to IP address)
- New name space for entity IDs
  - Maybe based on DNS?
  - Maybe a separate namespace?
  - Maybe IP addresses are used for location?
- Communication end-points (sockets) bound to identifiers, not addresses

Host Identity Protocol

- New cryptographic namespace
- Connection endpoints mapped to 128 bit host identity tags (hashes of public keys)
- Mapping at HIP layer
- 4-phase Base Exchange with cryptographic puzzle for DoS prevention
- IPSec for network-level security

Identity/Locator split



Application-layer mobility

- Many application-layer protocols are, in principle, similar to Mobile IP
- Moving entity may differ
  - Instead of host we have object, session, entity, or interests
- For example:
  - Object mobility
    - Wireless CORBA
  - Session mobility
    - SIP
  - Interest mobility
    - Content-based routing
  - Generic mobility
    - i3 overlay, service composition

Indirection Points

- Mobility may be characterized by indirection points
  - Mobile IP
    - Single fixed indirection point (Home Agent)
  - Location / Identity split
    - Single indirection point (ID->IP resolution)
  - Content-based routing
    - Many indirection points
    - Stepwise fingerprint -> IP resolution

Lessons to learn

- Hierarchical routing likely to stay
  - Addresses carry topological information
  - Efficient and well established
- Applications face changing connectivity
  - QoS varies
  - periods of non-connectivity
- Identifiers and locators likely to split
- Mobility management is needed
- Probably changes in directory services
  - Overlays have been proposed

Summary

- Topology based routing is necessary
- Mobility causes address changes
- Address changes preferrably signalled end-to-end
  - Alternative: use triangular routing as in Mobile IP
- Mobility management
  - Initial rendezvous: maybe a directory service
  - Double jump problem: rendezvous needed
- Various engineering trade-offs