EP2120 Internetworking

IK2218 Protocols and Principles of the Internet

Network layer fundamentals
Basic forwarding
IP addressing

Lecture 3

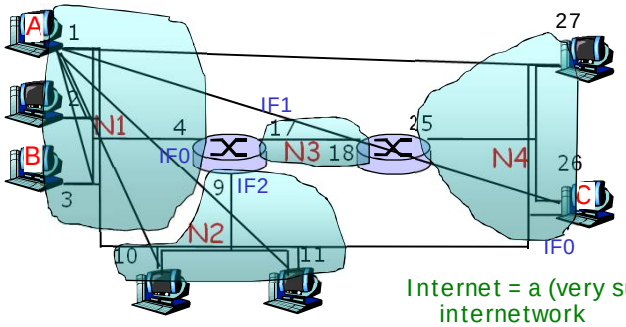
György Dán
KTH/EE/LCN

Literature:

Forouzan, TCP/IP Protocol Suite
(3ed Ch 5,12.2,26)(4ed Ch 5,12.2,26)

Do we need a network layer?

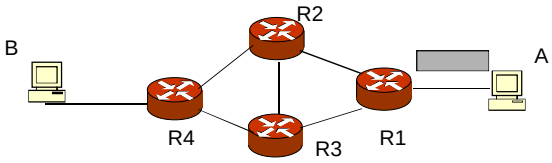
- Network of networks = internetwork
  - Subnets connected by routers
  - Subnets can use different link layer protocols
  - Routers/hosts connected to the subnet via an interface
- Network layer protocols understood by every host and router
  - "Lingua franca" between hosts and routers



Internet = a (very successful) internetwork

Network layer service models

Purpose:
End-to-end delivery of packets independent of the underlying link layer technologies



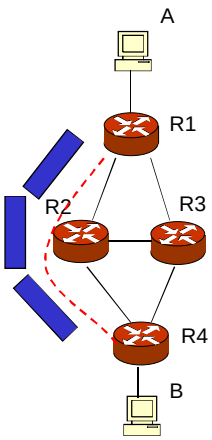
Example service abstractions

- Lossless transmission
- Bounded delay
- Guaranteed bandwidth
- In-order delivery
- Connection-oriented vs. Connectionless

Service model of the Internet?

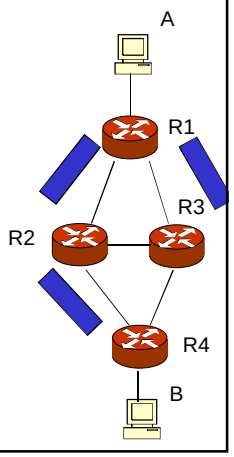
Connection-oriented Service
(Virtual circuit switched)

- Operation
  - Establish connection between source and destination
  - Send packets along the connection
    - Packets carry a VC identifier
  - Tear down connection
- Example: ATM, frame-relay, X.25,MPLS
- Note:
  - The decision about the route is made *once:* at connection establishment
    - Packets follow the same path
  - Routers/switches are *stateful*
  - Resources (link, router buffer, etc) can be allocated to VC



Connectionless Service
(Datagram)

- Operation
  - No connection establishment in network layer
  - Packets routed individually based on some information
    - e.g., destination host address
  - No connection tear down
- Example: Internet (IP)
- Note:
  - Route lookup for each packet
    - Packets may travel on different paths
  - No connection state information in routers
  - Resource allocation challenging



Quiz

- Consider a network like the one below that offers a datagram service. Host A sends 4 packets (numbered 1,2,3,4) to Host B. In what order do the packets arrive at B?

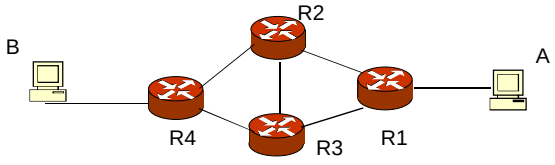
  a) 1,2,3,4

  b) 4,3,2,1

  c) 2,3,4,1

  d) None of the above

  e) Any of the above

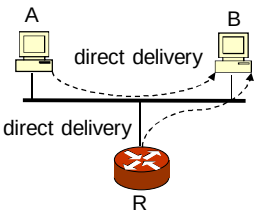


Basic Forwarding

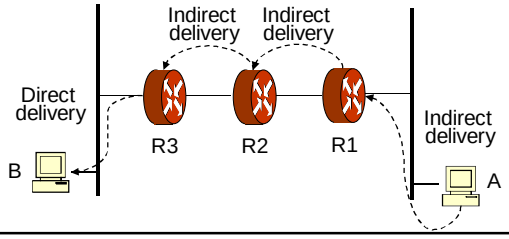
Delivery and Forwarding
at the network layer

End-to-end Delivery

- Direct delivery
  - Destination and sender connected to the *same* physical network
    - Last delivery is direct
  - Destination address and local interface have same network address (use netmask)

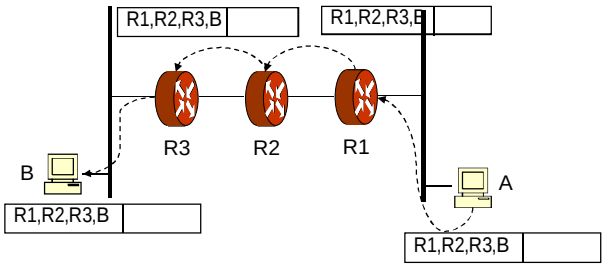


- Indirect delivery
  - From host to router or from router to router
  - Destination address and forwarding table: forwarding



Forwarding: Source routing

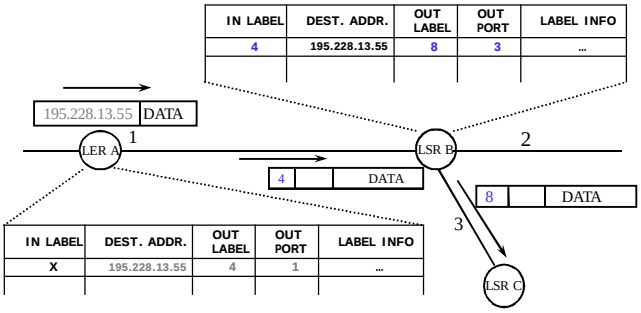
- Source routing
  - Source makes routing decision
  - Packet carries path (e.g., the hops)



  - Drawbacks
    - Topology information needed
    - Space needed in packet

Forwarding: Next-hop Routing I

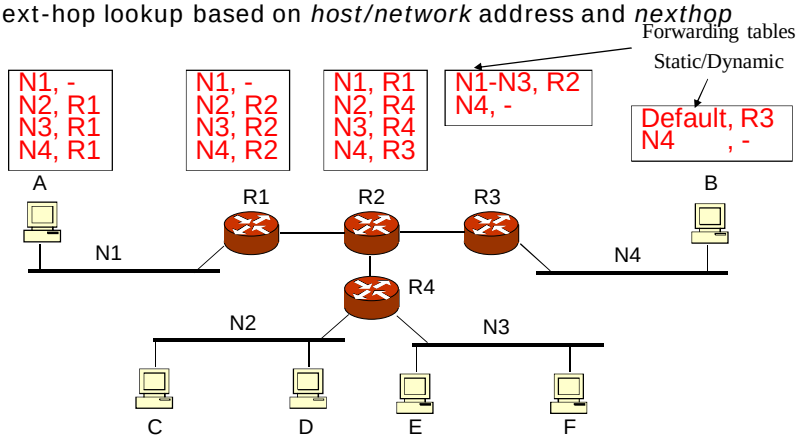
- Virtual circuit ID based (e.g., MPLS, ATM)
  - Packet carries VC identifier (e.g., MPLS label)
    - Can change upon every hop
  - [*VCID, nexthop*] table in every node
  - Next-hop lookup based on VCID



  - Connection establishment needed

Forwarding: Next-hop Routing II

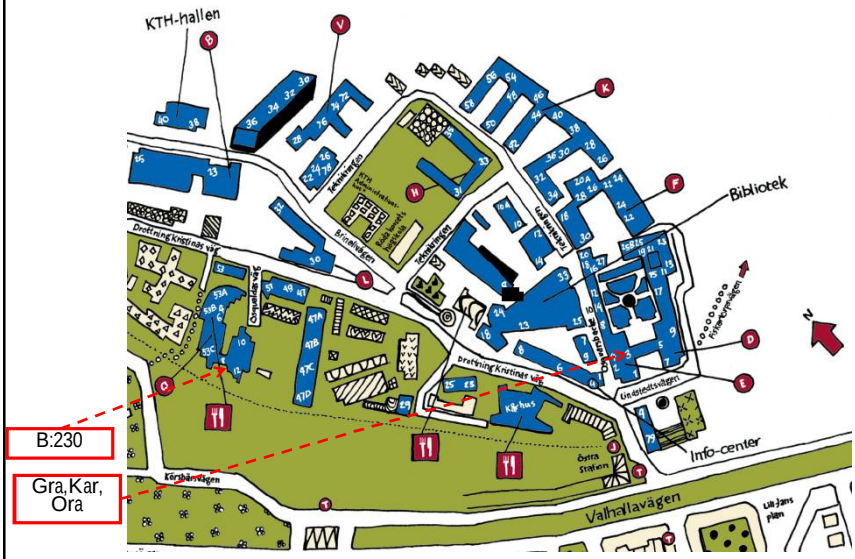
- Destination address based (e.g., IPv4 and IPv6)
  - Packet carries destination address (e.g., IPv4 address)
  - [*host/network address,nexthop*] table in every node
  - Next-hop lookup based on *host/network* address and *nexthop*



Addressing in IP

Logical adressing, aggregation, masking

What is an Address?



- The Labs will be in the “Q” and the “E” buildings

ID, Address, Route

- ID/Name – What?
  - Does not change when moving
  - Unique?
- Address/Locator - Where?
  - Changes when moving
  - Unique?
- Route – How?
  - Depends on location

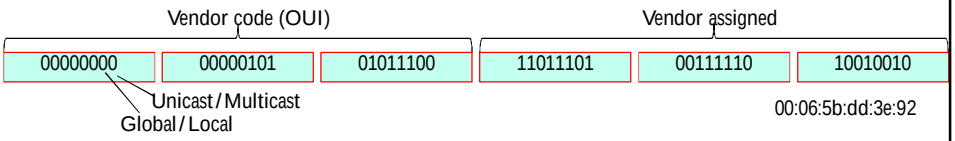
- Challenge
  - Multi-homing: entity with one ID but multiple addresses
  - Mobility: entity with one ID with changing address

Types of Communication & Addresses

- Unicast: one-to-one communication
  - Exactly one destination
- Broadcast: one-to-all communication
  - All destinations (e.g., on a subnet)
- Multicast: one-to-many communication
  - All members of a group
- Anycast: one-to-any communication
  - One member of a group
  - IP: Implemented using unicast addresses shared between several hosts

Physical "Address"

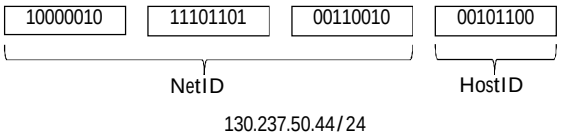
- Link layer
  - Example: IEEE 802 EUI-48 identifier



  - Used in Ethernet, WLAN, Bluetooth
  - Usually globally unique
  - Flat structure – cannot be aggregated
- Interface specific
  - Used on directly connected network
  - Property of the network interface card

Logical Addresses in IPv4

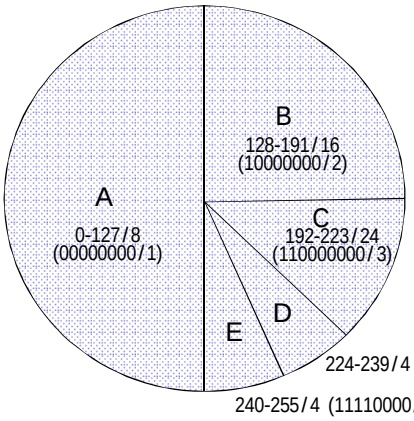
- Assigned to an interface - *not* to a node
- Length: 32 bits ⇒ 2^{32}=4294967296 addresses
- Notation:
  - dotted decimal: 130.237.50.44
  - binary: 10000010 11101101 00110010 00101100
- Hierarchy
  - Network ID – Host ID
  - Classful / Classless (CIDR)



130.237.50.44/24

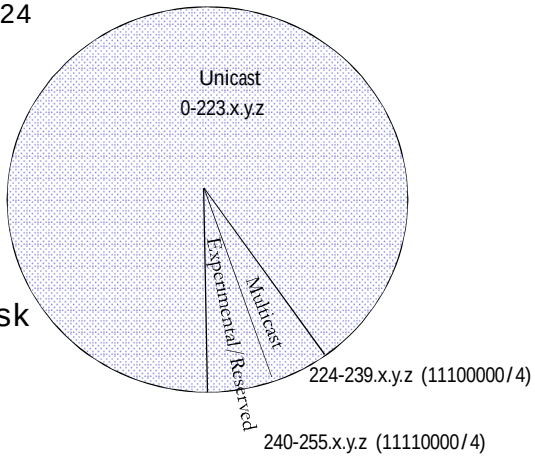
Classful IPv4 Addressing

- Address space partitioned in 5 classes
  - Classes A-C: Unicast
  - Class D: Multicast
  - Class E: Reserved
- Class determines
  - Length of NetID and HostID
- Inefficient
  - Supernetting/ subnetting
  - Obsolete



Classless IPv4 Addressing (CIDR)

- CIDR notation:
  - e.g., 130.237.15.44/24
- Prefix length/netmask provides
  - NetID/HostID

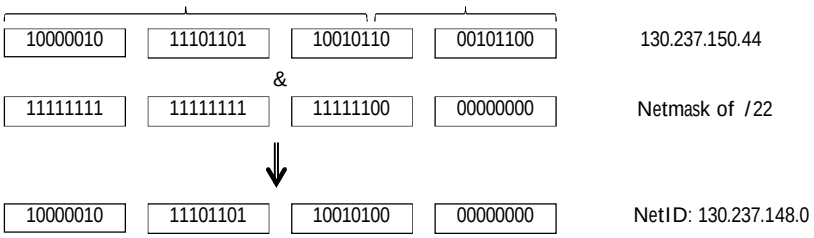


Network Mask

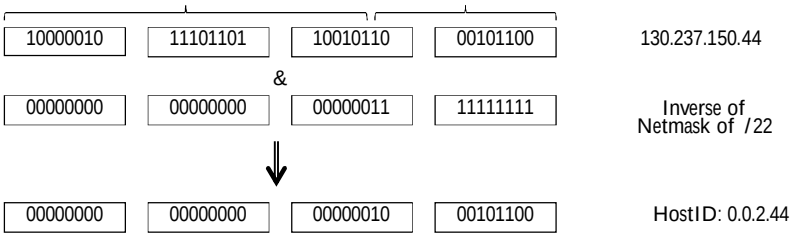
- Used to compute NetID, HostID and broadcast address
  - Address & Mask = NetID (network address)
  - Address & !Mask = HostID (host address)
  - Address & Mask | !Mask = Directed broadcast address
- IPv4: 32-bit binary number
  - Prefix notation /24
  - Binary: 11111111 11111111 11111111 00000000
  - Hex: FF FF FF 00
  - Dotted decimal: 255.255.255.0
- IPv6: 128-bit, same use

10000010	11101101	00110010	00101100	130.237.50.44
11111111	11111111	11111111	00000000	Netmask: /24

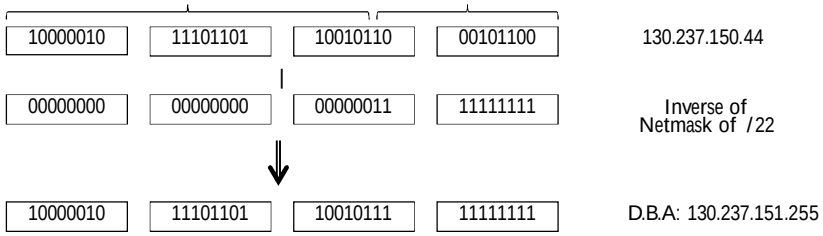
Computing the NetID



Computing the HostID

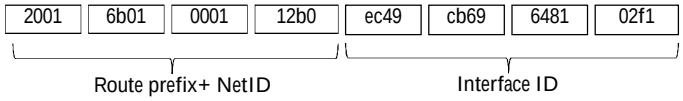


Computing the Directed Broadcast Address



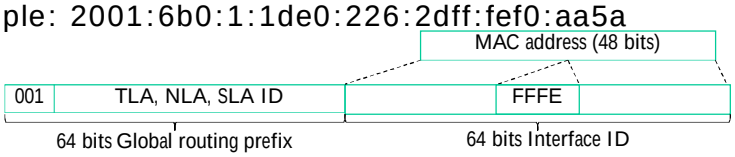
Logical Addresses in IPv6

- Assigned to an interface – can be *more* than 1 per interface
- Length: 128 bits ⇒ 2^{128} addresses (~5×10^{28}/person on Earth)
- Notation
  - Colon hexadecimal: 2001:06b0:0001:12b0:ec49:cb69:6481:02f1
  - Shortening
    - Leading zeros can be omitted
      - 2001:6b0:1:12b0:ec49:cb69:6481:2f1
    - Zero compression: one of the series of zeros replaced by ::



IPv6 Global Unicast Addresses

- Address format: 001_b prefix (2::/3)
  - 64 bit network ID
    - 48/32 bits public prefix, 16/32 bits site prefix
  - 64 bit interface ID
    - Derived from 48 bit IEEE 802 MAC address (EUI 48) – privacy!
    - Assigned at random or through IP configuration (see later)
- Example: 2001:6b0:1:1de0:226:2dff:fef0:aa5a



- Network address: obtained using netmask (as in IPv4)

Private/Unique Local Addresses

- Not globally unique unicast address
- Two uses
  - Isolated network
  - Behind NAT (e.g., most WiFi routers)
- IPv4 (RFC1918)

Class	NetID	Range
A	10.0.0.0/8	10.0.0.0 - 10.2555.255.255
B	172.16/12	172.16.0.0 – 172.31.255.255
C	192.168/16	192.168.0.0 – 192.168.255.255

- IPv6 (RFC4193)
  - FC00::/7 – (FD00::/8 for /48 bit prefix, 41 bits randomly generated)

Multicast Addresses

- IPv4 (rfc5771)
  - 1110/4 and 28 bit multicast group ID
  - Class D: 224.0.0.0 – 239.255.255.255
- IPv6 (rfc4291)
  - ff00::/8 (bits 8-15: flags and scope)
- Reserved addresses registered by IANA

Description	IPv4 Address	IPv6 Address
Local network control block (not forwarded)	224.0.0.0 – 224.0.0.255	ffx2::/16
All hosts on subnet	224.0.0.1	ff02::1
All routers on subnet	224.0.0.2	ff02::2
OSPF All Routers	224.0.0.5	ff02::5
Source-specific multicast block	232.0.0.0-232.255.255.255	FF3x::/32
	...	...

Link Local Addresses

- Used for “Link local unicast”
  - When you do not have an address yet (instead of 0.0.0.0)
    - Automatic address configuration, neighbor discovery
  - Isolated network
- Routers do not forward packets with such a destination
- Reserved prefix
  - IPv4: 169.254.0.0/16
  - IPv6: FE80:0:0:0:/64 - coexists with routable unicast address
- How to choose
  - Random with duplication detection
  - MAC derived (in IPv6 only)

Special IPv4 and IPv6 Addresses

	Source or Destination	IPv4			IPv6
		NetID	HostID	Example	
Network Address	None	X	All 0's	130.237.148.0	-||-
Directed Broadcast	Destination	X	All 1's	130.237.151.255	
Limited Broadcast	Destination	All 1's	All 1's	255.255.255.255	
Particular host on this network	Source	All 0's	Y	0.0.2.44	
This host on this network	Source	All 0's	All 0's	0.0.0.0	::
Loopback address	Destination	127	Any	127.0.0.1	::1

- Martian address
  - Address reserved by IANA that should not be used (240/8)
  - bogon – address not yet allocated by IANA*

Example from 'whois' database

IPv4

```
% Information related to '130.237.0.0  - 130.237.63.255'

inetnum:        130.237.0.0 - 130.237.63.255
netname:        KTHLAN-0-63
descr:          Royal Institute of Technology
country:        SE
org:            ORG-SA58-RIPE
admin-c:        riot1-ripe
tech-c:         riot1-ripe
status:         ASSIGNED PI
mnt-by:         RIPE-NCC-HM-PI-MNT
mnt-lower:      RIPE-NCC-HM-PI-MNT
mnt-by:         SUNET-MNT
mnt-routes:     SUNET-MNT
mnt-domains:    SUNET-MNT
source:         RIPE # Filtered
.....
% Information related to '130.237.0.0/18AS2839'

route:          130.237.0.0/18
descr:          KTH Royal Institute of Technology
origin:         AS2839
mnt-by:         SUNET-MNT
source:         RIPE # Filtered

% This query was served by the RIPE Database Query
Service version 1.19.9 (WHOIS4)
```

IPv6

```
% Information related to '2001:06B0:0001::/48'

inet6num:     2001:06B0:0001::/48
netname:      SE-KTH-1
descr:        Royal Institute of Technology
country:      SE
admin-c:      AH94
tech-c:       RASU1-RIPE
tech-c:       AH94
status:       NLA
mnt-by:       SUNET-MNT
source:       RIPE # Filtered
.....

% This query was served by the RIPE Database
Query Service version 1.19.9 (WHOIS2)
```

Quiz

- Consider the following parameters
  IP Address: 130.237.15.44
  Netmask: 255.255.192.0

- The netID, hostID and the DBA are
  - a) 130.237.15.0, 0.0.1.44, 130.238.255.255
  - b) 130.237.0.0, 0.0.15.44, 130.237.63.255
  - c) 130.224.255.255, 0.13.15.44, 130.224.255.255
  - d) None of the above.

DBA: Directed Broadcast Address

Global IP Address Allocation

- Internet Assigned Numbers Authority (IANA@ICANN) manages
  - Public IP addresses
  - Autonomous system (AS) numbers
- IANA allocates blocks to Regional Internet Registries (RIR)
  - Réseaux IP Européens Network Coordination Centre (RIPE NCC)
  - American Registry for Internet Numbers (ARIN)
  - Latin American and Caribbean Internet Addresses Registry (LACNIC)
  - Asia Pacific Network Information Centre (APNIC)
  - African Network Information Centre (AfriNIC)
- RIRs allocate blocks to National/Local Internet Registry (LIR)
  - Internet Service Providers (ISP), Institutions
- LIRs *assign* addresses to end users

Source: Wikipedia

IPv4 Address Exhaustion

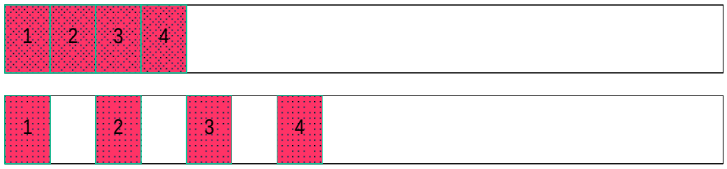
- Last IPv4 address block allocated by IANA
  - 3 February 2011
- Already exhausted (/8)
  - APNIC (15 Apr 2011) - allocating 1 /22 block per member
  - RIPE (14 Sep 2012)
  - Caribbean (10 Jun 2014)
- Others to be exhausted soon
  - 80% assignment rule
- Future
  - Transition to IPv6
  - More efficient use of IP addresses – policies
  - Address block trading $$$
  - Network Address Translation (NAT) – private addresses

How to Allocate Addresses?

- Number of addresses in a block
  - 2^{32-n} (n is prefix length) – always power of 2
  - Not all addresses are usable (by hosts or routers)
    - Network address – first address of the block
    - Directed broadcast address – last address of the block
- Example: 130.237.48.0/22
  - Address range: 130.237.48.0-130.237.51.255
  - Special addresses – not usable
    - Network address: 130.237.48.0
    - Directed broadcast address: 130.237.51.255
  - Number of usable addresses: 2^{10}-2

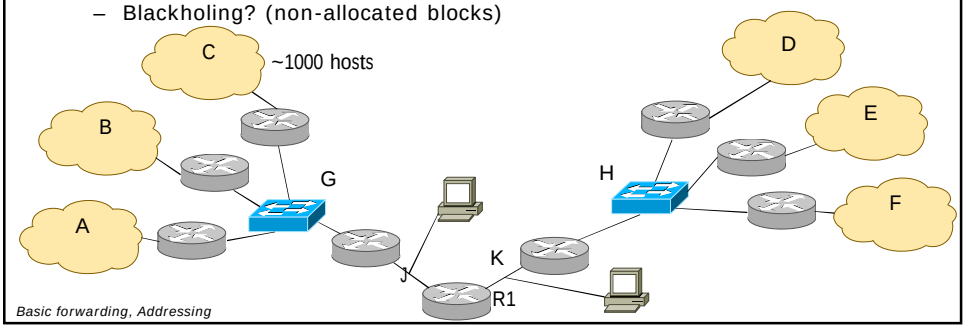
How to Allocate Address Blocks

- Consider an institution with address block 130.237.0.0/18
- Allocate addresses to the labs/departments/schools
  - 128 labs require 64 addresses each
  - 32 departments require 256 addresses each
  - 8 schools require 1024 addresses each
- What is the winning strategy?
  - Allocate blocks sequentially – expansion?
  - Spread out the blocks – inefficient use – new customer?
  - Remember NetID and Directed broadcast address



Exercise: Address Allocation

- Use the following block of addresses to allocate addresses to the network shown below
  - Address block: 121.100.128.0/19
- Answer the following questions
  - What are the network and broadcast addresses?
  - What are the router and host addresses?
  - What network is announced by R1?
  - Blackholing? (non-allocated blocks)



Exercise: Address Allocation

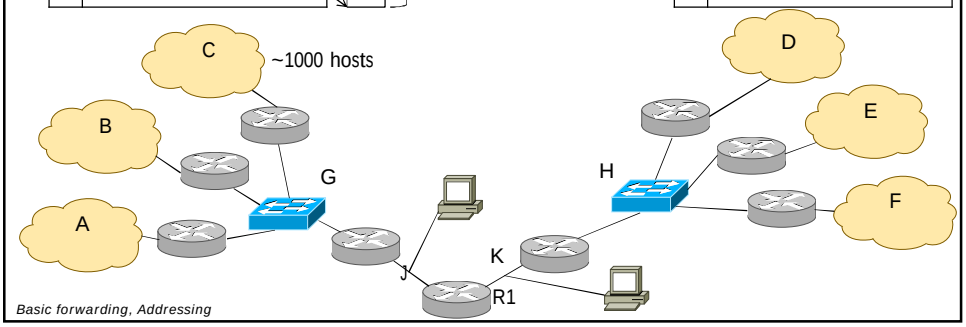
- Block of addresses: 121.100.128.0/19
- Split into two /20 blocks: 121.100.128.0/20 and 121.100.144.0/20

J	128.100.140.8/29
G	128.100.140.0/29
C	128.100.136.0/22
B	128.100.132.0/22
A	128.100.128.0/22

non-allocated

121.100.128.0/20

D	128.100.144.0/22
E	128.100.148.0/22
F	128.100.152.0/22
H	128.100.156.0/29
K	128.100.156.8/29

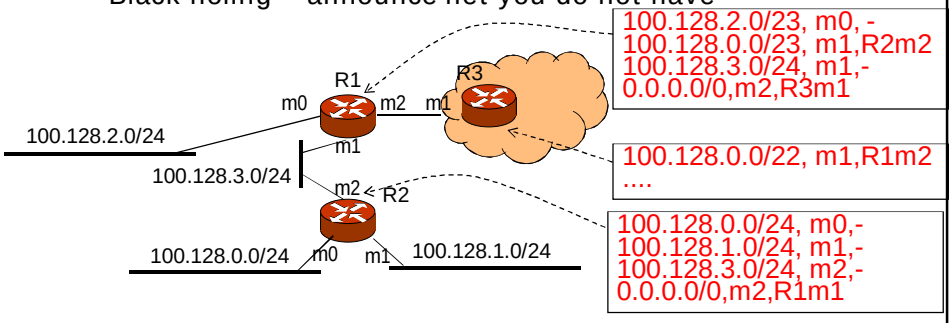


Basic Forwarding Continued

Delivery and Forwarding in IP

Address aggregation and forwarding

- Aggregate NetIDs
  - Shorter prefix – bigger address block
  - Less RIB and FIB entries
  - Black holing – announce net you do not have



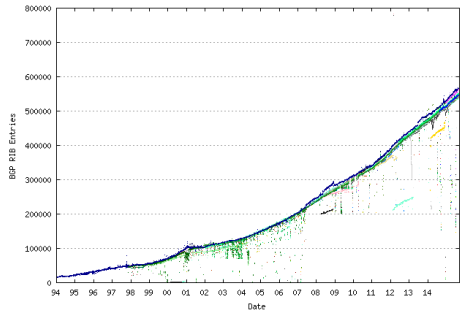
- Longest prefix matching!

Address aggregation in practice

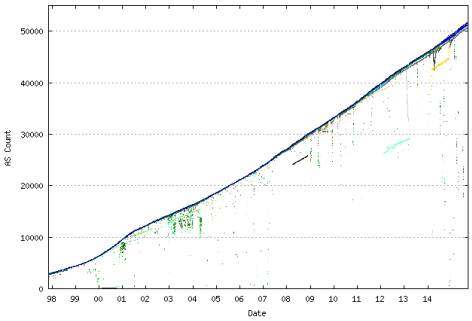
- Effective address allocation policy needed
  - e.g., based on geographic location
    - IANA → RIR → (NIR) → LIR → end users
      - E.g., 071/8 - ARIN (~N.America), 061/8 – APNIC (Asia-Pacific)
      - 61.213.162.230 - Tokyo, Japan (NTT)
      - 61.120.145.198 – Tokyo, Japan (NTT)
      - 61.1.3.1 – New Delhi, India (BSNLNET)
      - 61/8 is in Asia – yet different routes
- Caveats
  - Multi-homing
  - Lack of IP addresses⇒ Allocation of long (/24) prefixes
  - Enforcement needed (e.g., RIPE 80% rule)
- Current forwarding tables
  - # of entries ~500000 (~60% are /24 prefixes)

Forwarding Table Sizes

- IPv4 networks announced in BGP (FIB)



- Autonomous systems (AS)



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

Forwarding Table – Common Fields

Mask	Network Address	Next-hop Address	Interface	Flags	Reference count	Use
....	....	....	....	....	....	.... ..

- Mask – netmask applied for the entry [255.255.0.0]
- Network address – destination network [145.168.0.0]
- Next-hop address – next router [130.237.43.1]
- Interface – outgoing interface [eth0]
- Flags – status/info [U(p), G(ateway), H(ost-specific)...]
- Reference count – # of users using this route
- Use – # of packets transmitted for this destination

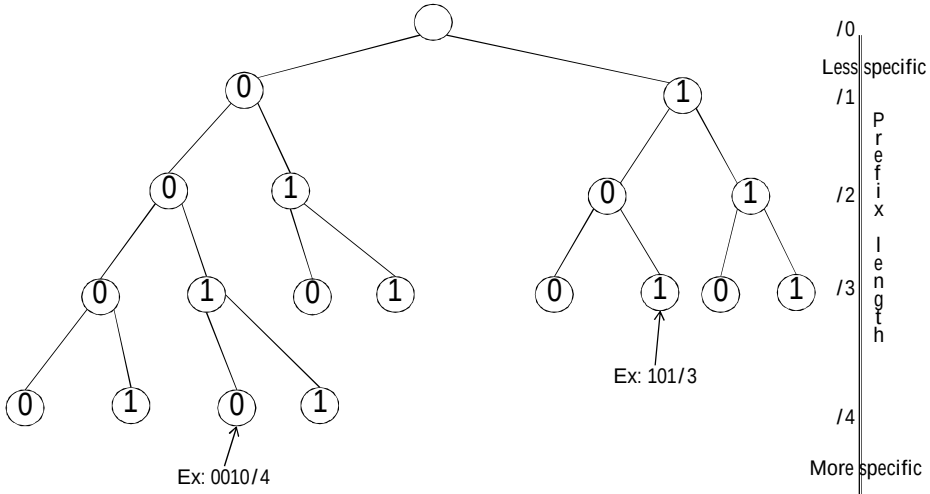
Exercise: Forwarding table

- A router has the following forwarding table

Destination	Next Hop	Flags	Interface
111.0.0.0/8	-	U	m0
193.14.5.160/27	-	U	m2
193.14.5.192/27	-	U	m1
194.17.21.16/32	111.20.18.14	UGH	m0
192.16.7.0/24	111.15.17.32	UG	m0
194.17.21.0/24	111.20.18.14	UG	m0
0/0	111.30.31.18	UG	m0

- Determine the next-hop address and the outgoing interface for each packet that arrives to the router if the packet's destination address is
  - a) 111.45.32.16
  - b) 192.16.7.31
  - c) 194.17.21.45
  - d) 220.7.14.7
  - e) 193.14.5.16
  - f) 193.14.5.196
  - g) 192.168.130.25
  - Try to sketch the network as seen from the router, based on the routing table.

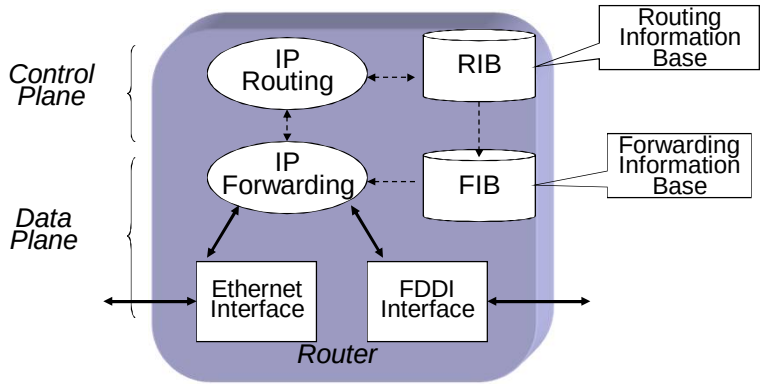
Fast Address Lookup: Trie



- Longest prefix matching: find most specific match

S. Nilsson, G. Karlsson, "IP-Address Lookup Using LC-Tries," IEEE JSAC, vol. 17, no. 6, 1999

IP Router Model



- Two planes of router operation
  - *data plane*: fast and special purpose
    - handles packet *forwarding* in real-time
  - *control plane*: general purpose
    - handles *routing* in the background

IP Forwarding Summary

- Router forwards packets between network interfaces
  - Extract header information from the incoming datagram
    - Destination IP address
  - Lookup in the forwarding information base (match networks)
    - Next-Hop IP address,
    - Outgoing interface,...
  - Modify datagram header (why?)
  - Send out on outgoing interface
- Router can perform much more than address lookup
  - Access lists, filtering
  - Traffic management
  - Other protocols: Bridging, MPLS, IPv6, ...