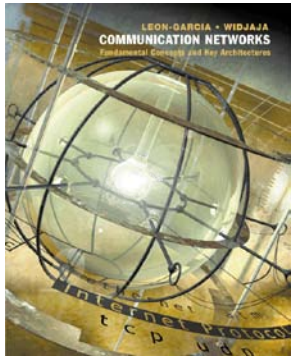
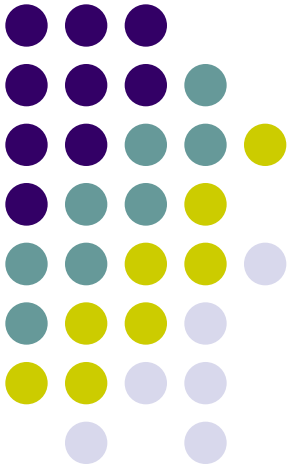
Chapter 8
Communication Networks and Services



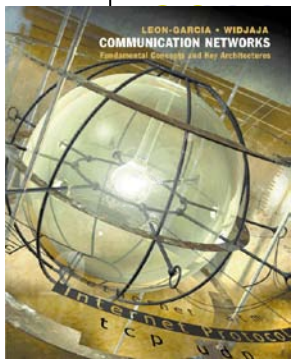


1. IPv6

2. Internet Routing Protocols:
OSPF, RIP, BGP

3. Other protocols:
DHCP, NAT, and Mobile IP

Chapter 8 Communication Networks and Services



IPv6

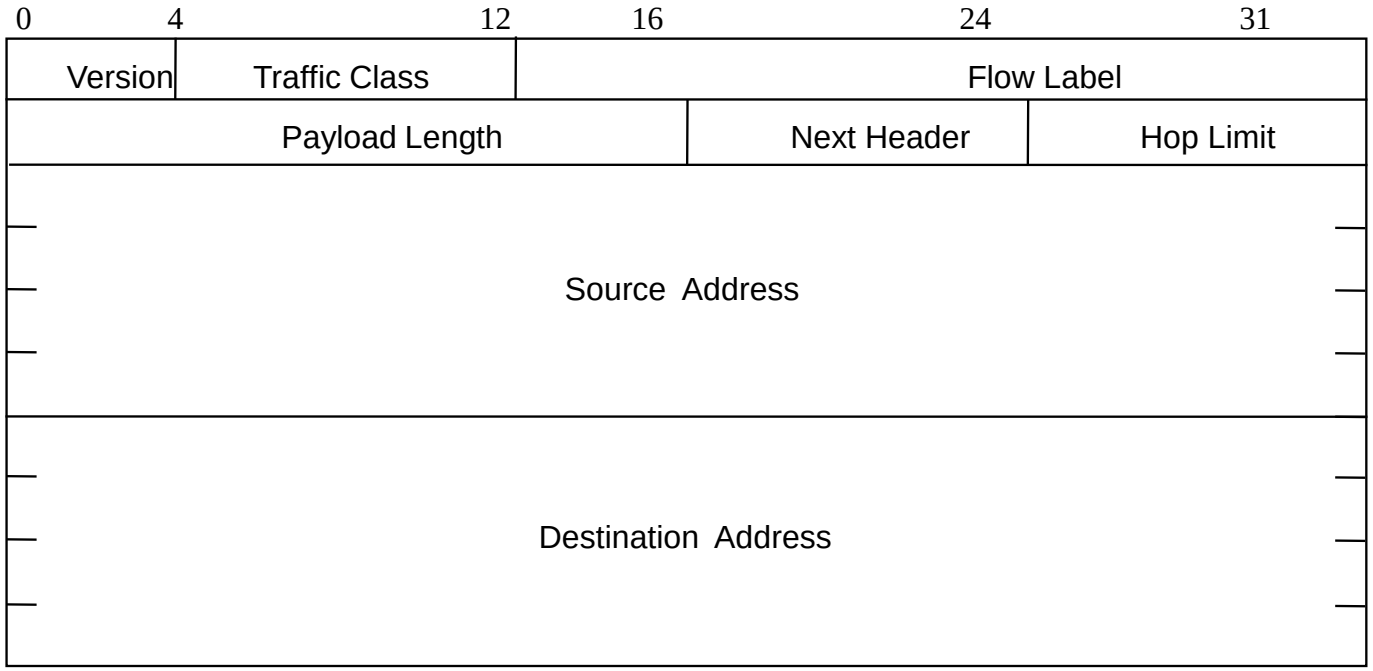
IPv6

- **Longer address field:**
  - **128 bits** can support up to 3.4 x 10^{38} hosts
- **Simplified header format:**
  - Simpler format to speed up processing of each header
    - What processing overhead does IPv4 headers have?
  - All fields are of fixed size
  - IPv4 vs IPv6 fields:
    - Same: Version
    - Dropped: Header length, ID/flags/frag offset, header checksum
    - Replaced:
      - Datagram length by Payload length
      - Protocol type (upper layer) by Next header
      - TTL by Hop limit
      - TOS by traffic class
    - New: Flow label

Other IPv6 Features

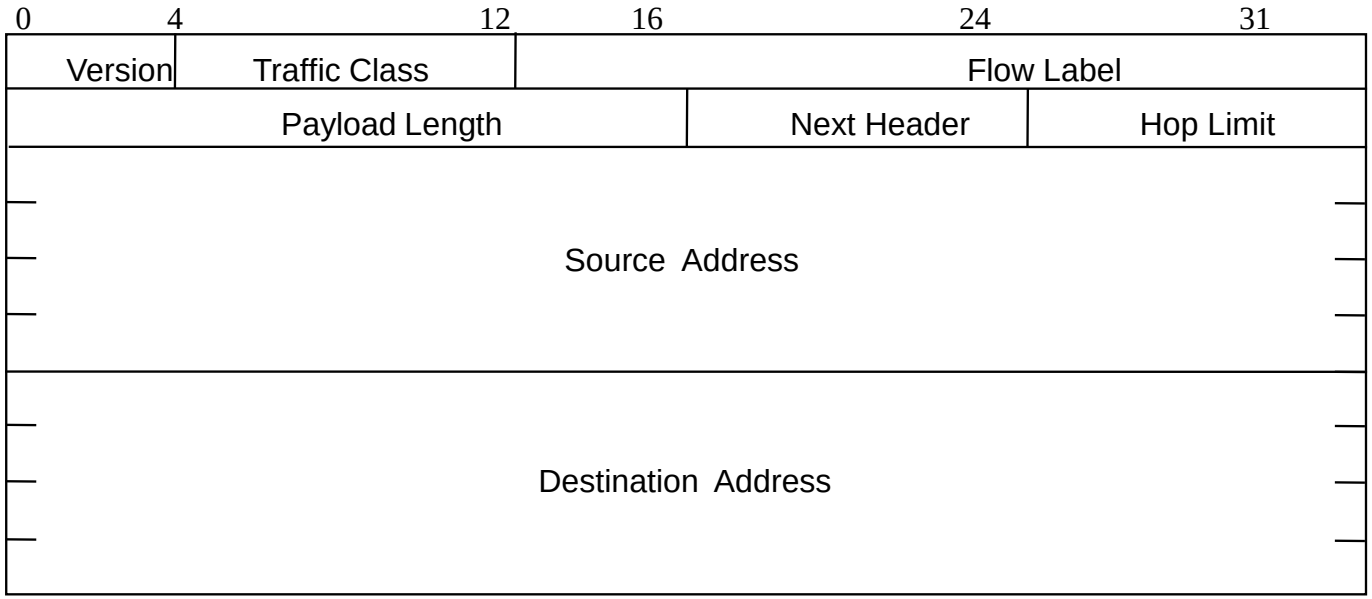
- **Flexible support for options:** more efficient and flexible options encoded in optional *extension headers*
- **Flow label capability:** "flow label" to identify a packet flow that requires a certain QoS
- **Security:** built-in authentication and confidentiality
- **Large packets:** supports payloads that are longer than 64 K bytes, called *jumbo* payloads.
- **Fragmentation at source only:** source should check the minimum MTU along the path
- **No checksum field:** removed to reduce packet processing time in a router

IPv6 Header Format



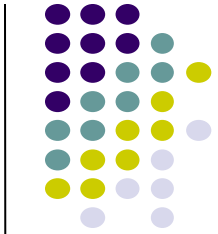
- Version field same size, same location
- Traffic class to support differentiated services
- **Flow**: sequence of packets from a particular source to a particular destination for which source requires special handling
  - Ex: packets belong to the same flow stay on the same path.

IPv6 Header Format



- Payload length: length of data excluding header, up to 65535 B
  - 16-bit length limitation in UDP and the MSS (Maximum Segment Size) limitation of TCP
- Next header: type of extension header that follows basic header to support more features
- Hop limit: # hops packet can travel before being dropped by a router

Special Purpose Addresses

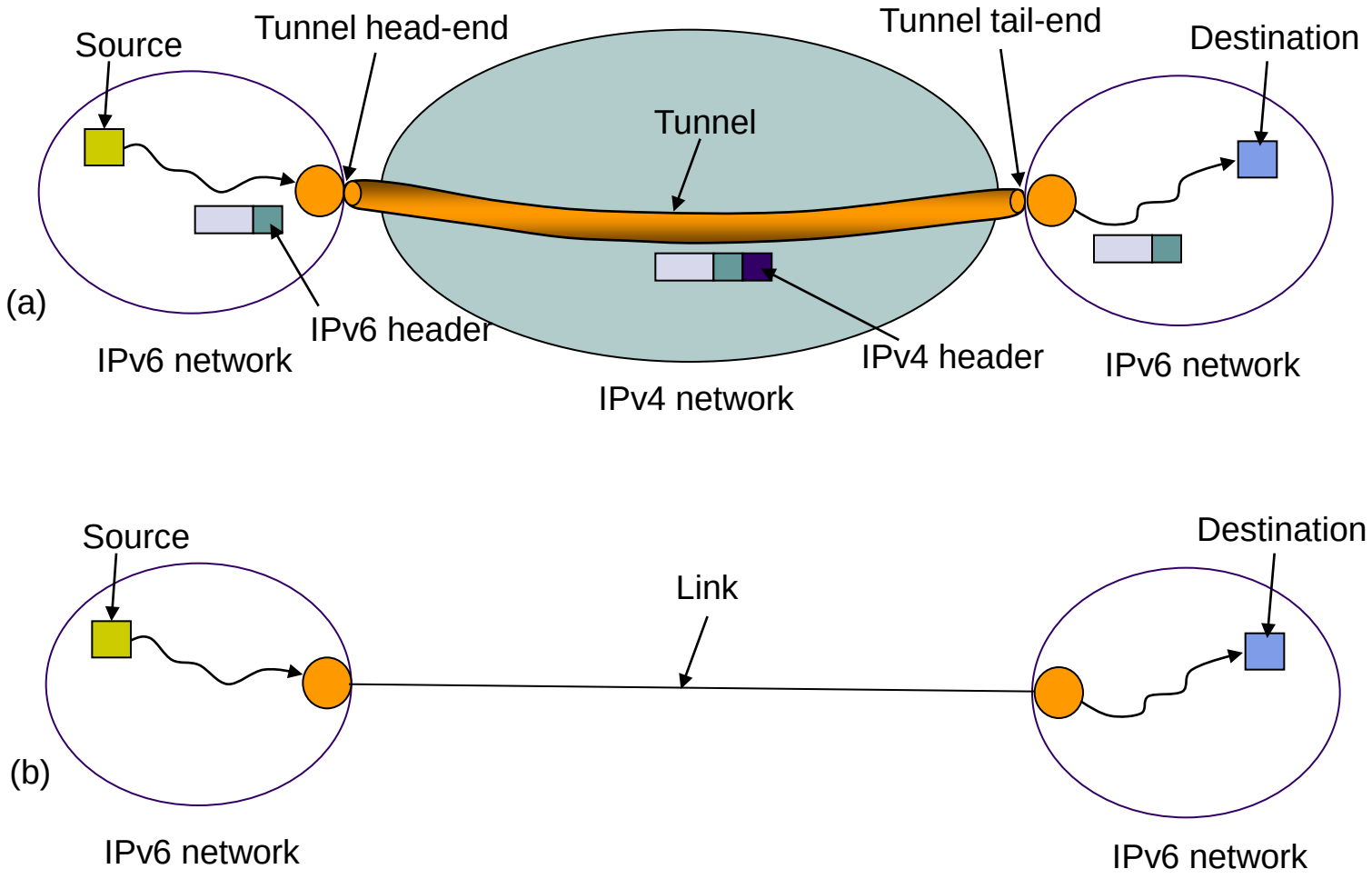


- *Unspecified Address*: 0::0
  - Used by source station to learn own address
- *Loopback Address*: ::1
- *IPv4-compatible addresses*: 96 0's + IPv4
  - For tunneling by IPv6 routers connected to IPv4 networks
  - ::135.150.10.247
- *IP-mapped addresses*: 80 0's + 16 1's + IPv4
  - Denote IPv4 hosts & routers that do not support IPv6

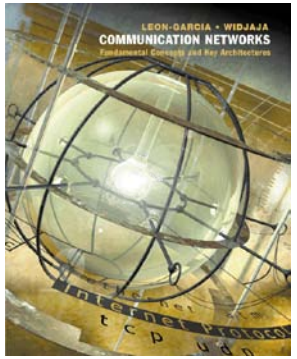
Migration from IPv4 to IPv6

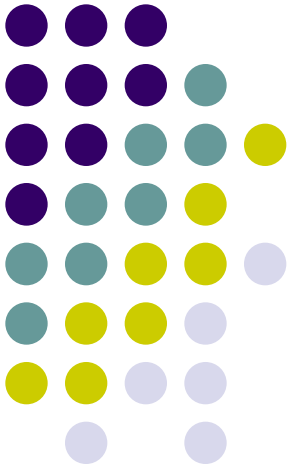
- Gradual transition from IPv4 to IPv6
- Dual IP stacks: routers run IPv4 & IPv6
  - Type field used to direct packet to IP version
- IPv6 islands can tunnel across IPv4 networks
  - **Encapsulate** user packet insider IPv4 packet
  - Tunnel endpoint at source host, intermediate router, or destination host
  - Tunneling can be recursive

Migration from IPv4 to IPv6



Chapter 8
Communication Networks and Services





Internet Routing Protocols

Outline

- Basic Routing
- Routing Information Protocol (RIP)
- Open Shortest Path First (OSPF)
- Border Gateway Protocol (BGP)

Routing vs. Forwarding

- Routing → control plane
  - How to determine the routing table entries?
    - Carried out by routing daemon
    - Routers exchange information using routing protocols to develop the routing tables
- Forwarding → data plane
  - Moving an arriving packet
  - IP datagram: Look up routing table & forward packet from input to output port
    - Longest-prefix matching
    - Carried out by IP layer
  - VC: Look up VCI and VC table
  - MPLS: Look up labels

Host Behavior

- Every host must do IP forwarding
- For datagram generated by own higher layers
  - if destination connected through point-to-point link or on shared network, send datagram directly to destination
  - Else, send datagram to a default router
- For datagrams received on network interface
  - if destination address, own address, pass to higher layer
  - if destination address not own, discard “silently”

Router Behavior

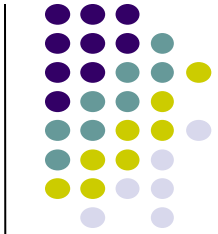
Router's IP layer

- can receive datagrams from own higher layers
- can receive datagram from a network interface
  - if destination IP address own or broadcast address, pass to layer above
  - else, forward the datagram to next hop
- routing table determines handling of datagram

Routing Table Entries

- Destination IP Address:
  - complete host address or network address
- IP address of
  - next-hop router or directly connected network
- Flags
  - Is destination IP address a network address or a host address?
  - Is next hop, a router or directly connected?
- Network interface on which to send packet

Forwarding Procedure



- Does routing table have entry that matches complete destination IP address? If so, use this entry to forward
- Else, does routing table have entry that matches the longest prefix of the destination IP address? If so, use this entry to forward
- Else, does the routing table have a default entry? If so, use this entry.
- Else, packet is undeliverable

Autonomous Systems

- Link-state and distance vector algorithms conceputually consider a flat network topology.
- In practice, global Internet viewed as collection of autonomous systems.
- **Autonomous system (AS)** is a set of routers or networks administered by a single organization, e.g., ISP
- Intra-AS routing vs. inter-AS routing:
  - An AS should present a *consistent picture of what ASs are reachable* through it
- **Stub AS:** has only a single connection to the outside world.
- **Multihomed AS:** has multiple connections to the outside world, but refuses to carry transit traffic
- **Transit AS:** . If one AS is an ISP for another, then the former is a transit AS. Ex: net A can use net B, the transit AS, to connect to net C.

Inter and Intra Domain Routing

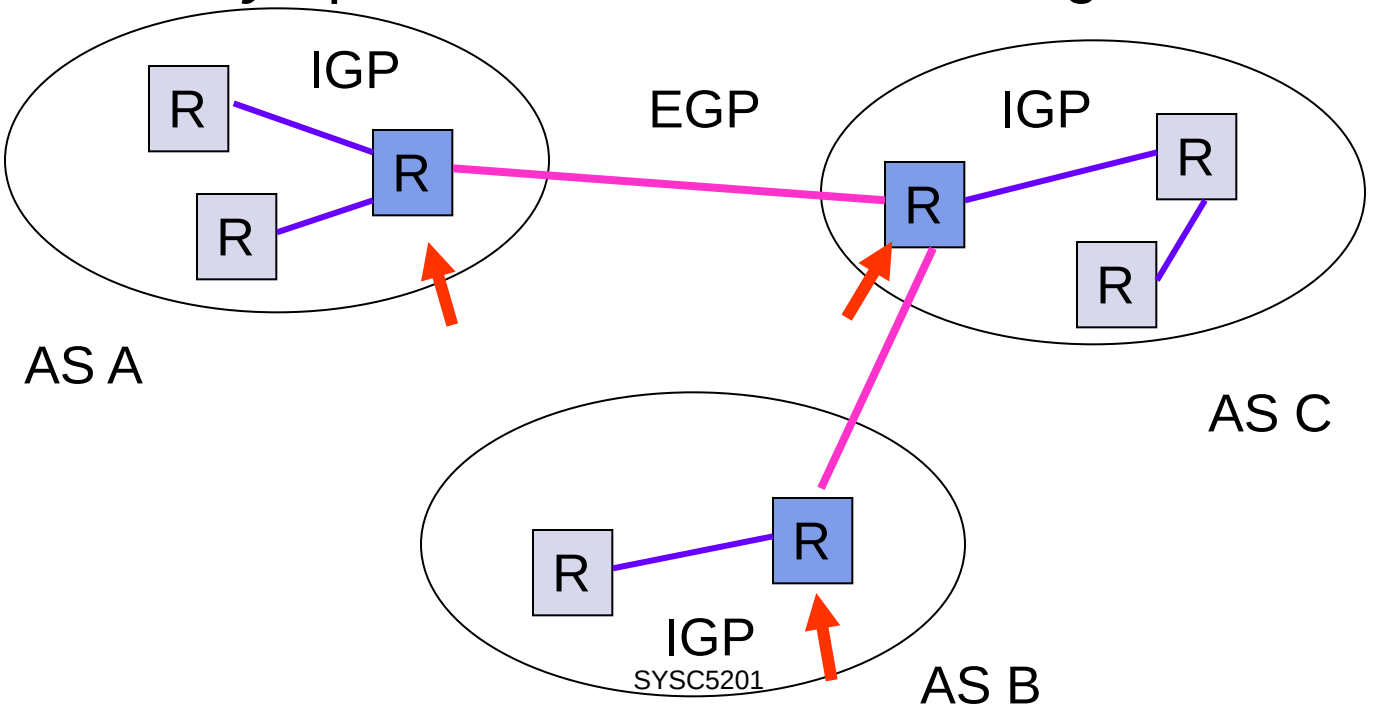
Interior Gateway Protocol (IGP): routing within AS

- RIP, OSPF, IS-IS
- Intra-domain size: roughly 70 routers (Cisco, may chage)

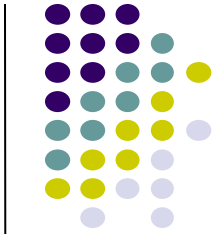
Exterior Gateway Protocol (EGP): routing between AS's

- BGPv4

Border Gateways perform IGP & EGP routing



Outline



- Basic Routing
- Routing Information Protocol (RIP)
- Open Shortest Path First (OSPF)
- Border Gateway Protocol (BGP)

Routing Information Protocol (RIP)

- RFC 1058
- Uses the **distance-vector algorithm**
- Runs on **top of UDP**, port number 520
- Metric: number of hops

  Max no of hops is limited to 15

  - suitable for small networks (local area environments)
  - value of 16 is reserved to represent infinity
  - small number limits the *count-to-infinity* problem

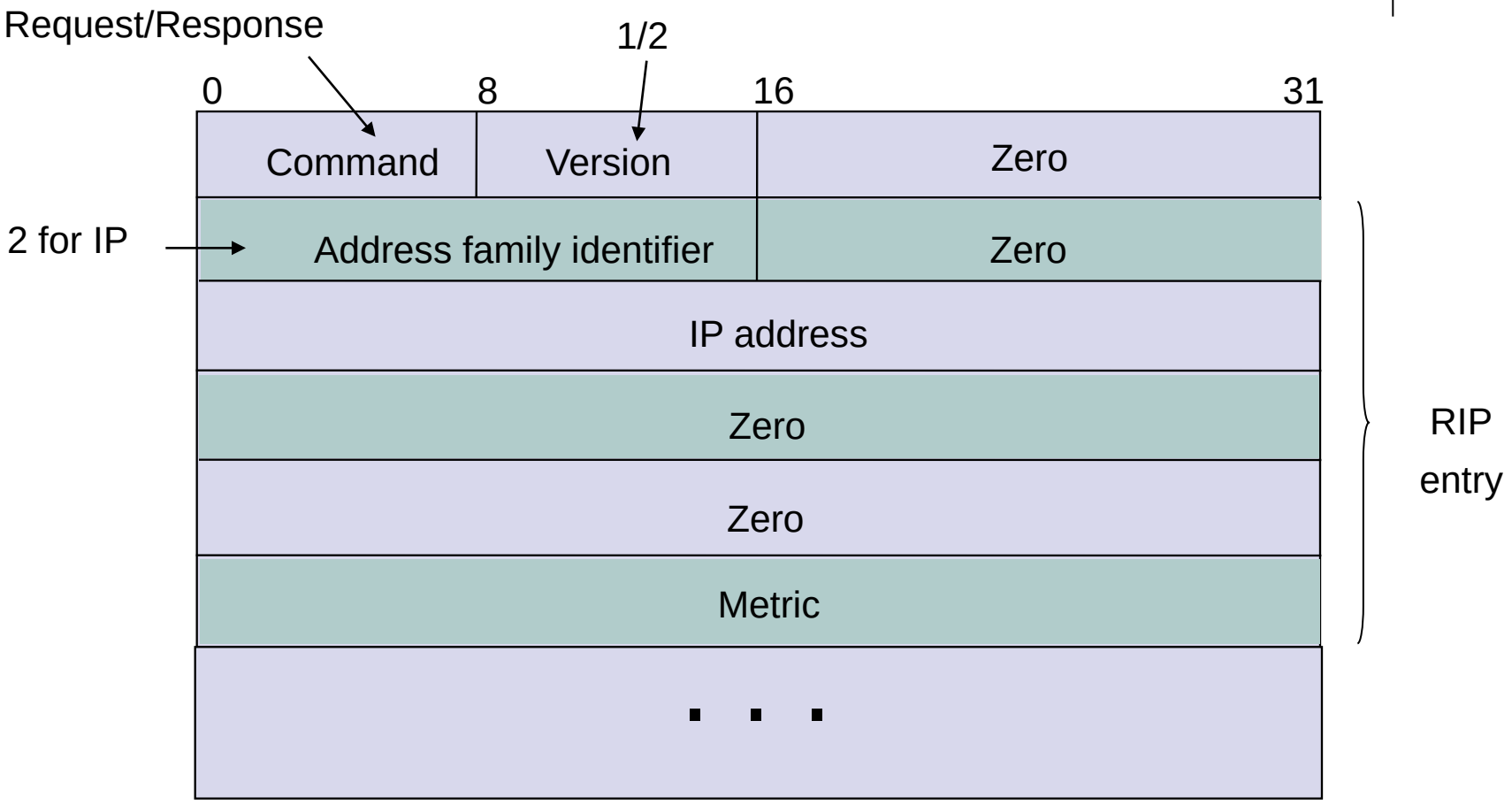
RIP Operation

- Router sends update message to neighbors every 30 sec (usually configurable)
- A router expects to receive an update message from each of its neighbors within 180 seconds in the worst case
- If router does not receive update message from neighbor X within this limit, it assumes the link to X has failed and sets the corresponding minimum cost to 16 (infinity)
- Uses ***split horizon with poisoned reverse***
- Convergence speeded up by triggered updates
  - neighbors notified immediately of changes in distance vector table

RIP Protocol

- Routers run RIP in active mode (advertise distance vector tables)
- Hosts can run RIP in passive mode (update distance vector tables, but do not advertise)
- Two RIP packet types:
  - ***request*** to ask neighbor for distance vector table
  - ***response*** to advertise distance vector table

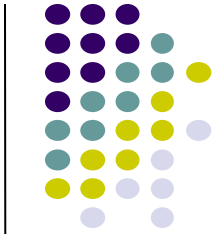
RIP Message Format



Up to 25 RIP entries per message

Slide is for self study

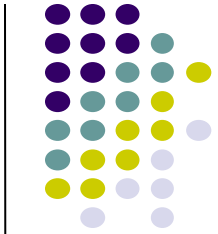
RIP Message Format



- Command: request or response
- Version: v1 or v2
- One or more of:
  - Address Family: 2 for IP
  - IP Address: network or host destination
  - Metric: number of hops to destination
- Version 1 does not send subnet mask
- Version 2 sends subnet mask and support CIDR (variable subnet masks)
- still uses max cost of 16

Slide is for self study

Outline



- Basic Routing
- Routing Information Protocol (RIP)
- Open Shortest Path First (OSPF)
- Border Gateway Protocol (BGP)

Open Shortest Path First

- RFC 2328 (v2)
- Fixes some of the deficiencies in RIP
- Enables each router to learn **complete network topology**
- Each router monitors the ***link state*** to each neighbor and floods the link-state information to other routers
- Each router builds an identical *link-state database*
- Allows router to build shortest path tree with router as root
- OSPF typically **converges faster** than RIP when there is a failure in the network

OSPF Features

- *Multiple routes* to a given destination, one per type of service
- Support for ***variable-length subnetting*** by including the subnet mask in the routing message
- More *flexible link cost* which can range from 1 to 65,535
- Distribution of traffic over ***multiple paths* of equal cost**
- *Authentication* to ensure routers exchange information with trusted neighbors
- Uses *notion of* ***area*** to partition sites into subsets
- Support *host-specific routes* as well as net-specific routes
- *Designated router* to minimize table maintenance overhead

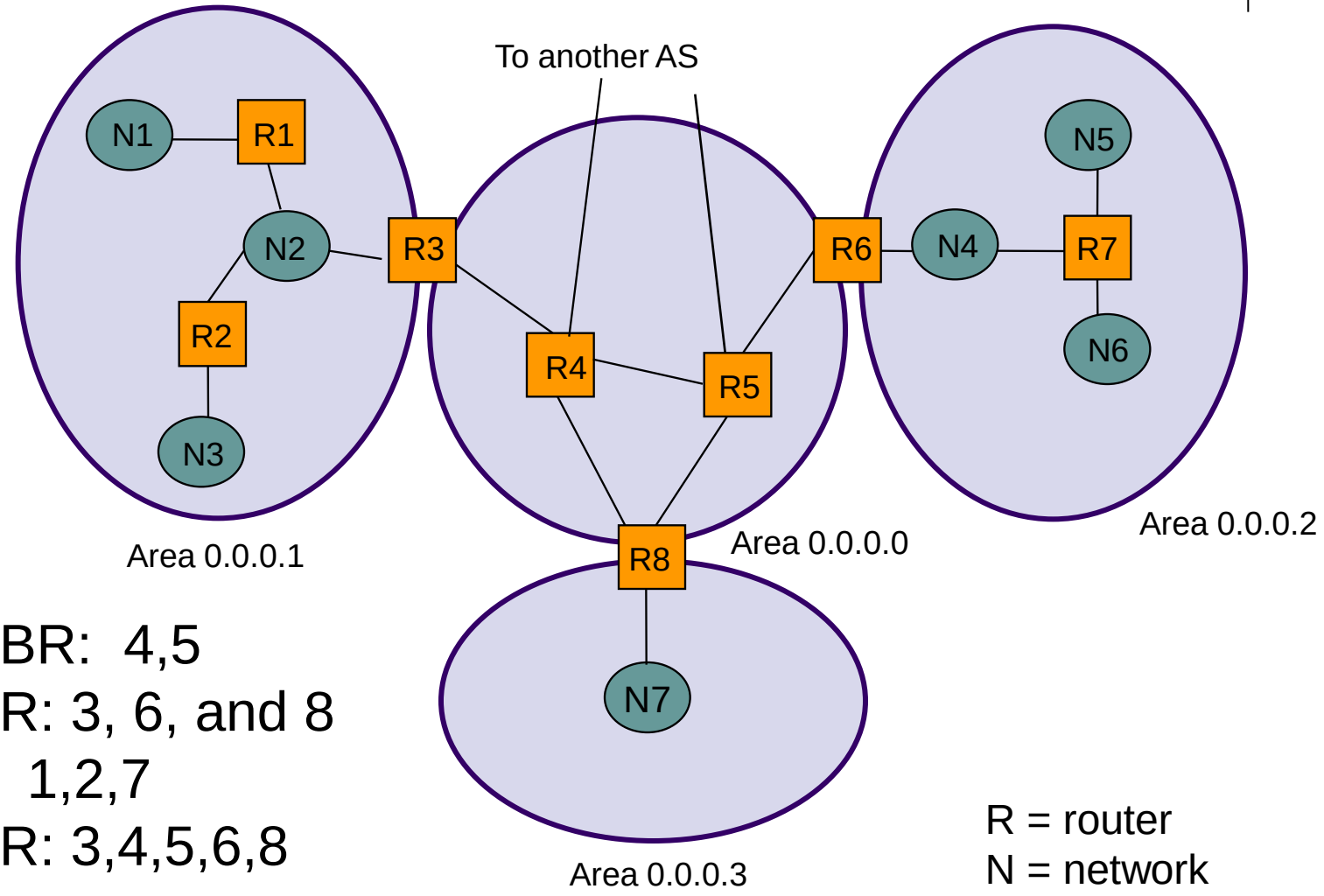
Flooding

- Used in OSPF to distribute link state (LS) information
- Forward incoming packet to all ports except where packet came in
- Packet eventually reaches destination as long as there is a path between the source and destination
- Generates exponential number of packet transmissions
- Approaches to limit # of transmissions:
  - Use a TTL at each packet; won't flood if TTL is reached
  - Each router adds its identifier to header of packet before it floods the packet; won't flood if its identifier is detected
  - Each packet from a given source is identified with a unique sequence number; won't flood if sequence number is same

OSPF Network

- To improve **scalability**, AS may be partitioned into ***areas***
  - Area is identified by 32-bit Area ID
  - Router in area only knows complete topology inside area & limits the flooding of link-state information to an area
  - ***Area border routers*** summarize info from other areas
- Each area must be connected to *backbone area* (***BBR***) (0.0.0.0)
  - Distributes routing info between areas
- *Internal router* (***IR***) has all links to nets within the same area
- *Area border router (**ABR**)* has links to more than one area
- *backbone router* has links connected to the backbone
- *Autonomous system boundary router (**ASBR**)* has links to another autonomous system.

OSPF Areas



ASBR: 4,5
ABR: 3, 6, and 8
IR: 1,2,7
BBR: 3,4,5,6,8

R = router
N = network

Neighbor, Adjacent & Designated Routers

- *Neighbor routers*: two routers that have interfaces to a common network
  - Neighbors are discovered dynamically by *Hello protocol*
- Each neighbor of a router described by a state
  - down, attempt, init, 2-way, Ex-Start, Exchange, Loading, Full
- *Adjacent router*: neighbor routers become adjacent when they synchronize topology databases by exchange of link state information
  - Neighbors on point-to-point links become adjacent
  - Routers on multiaccess nets become adjacent only to *designated & backup designated routers*
    - Reduces size of topological database & routing traffic

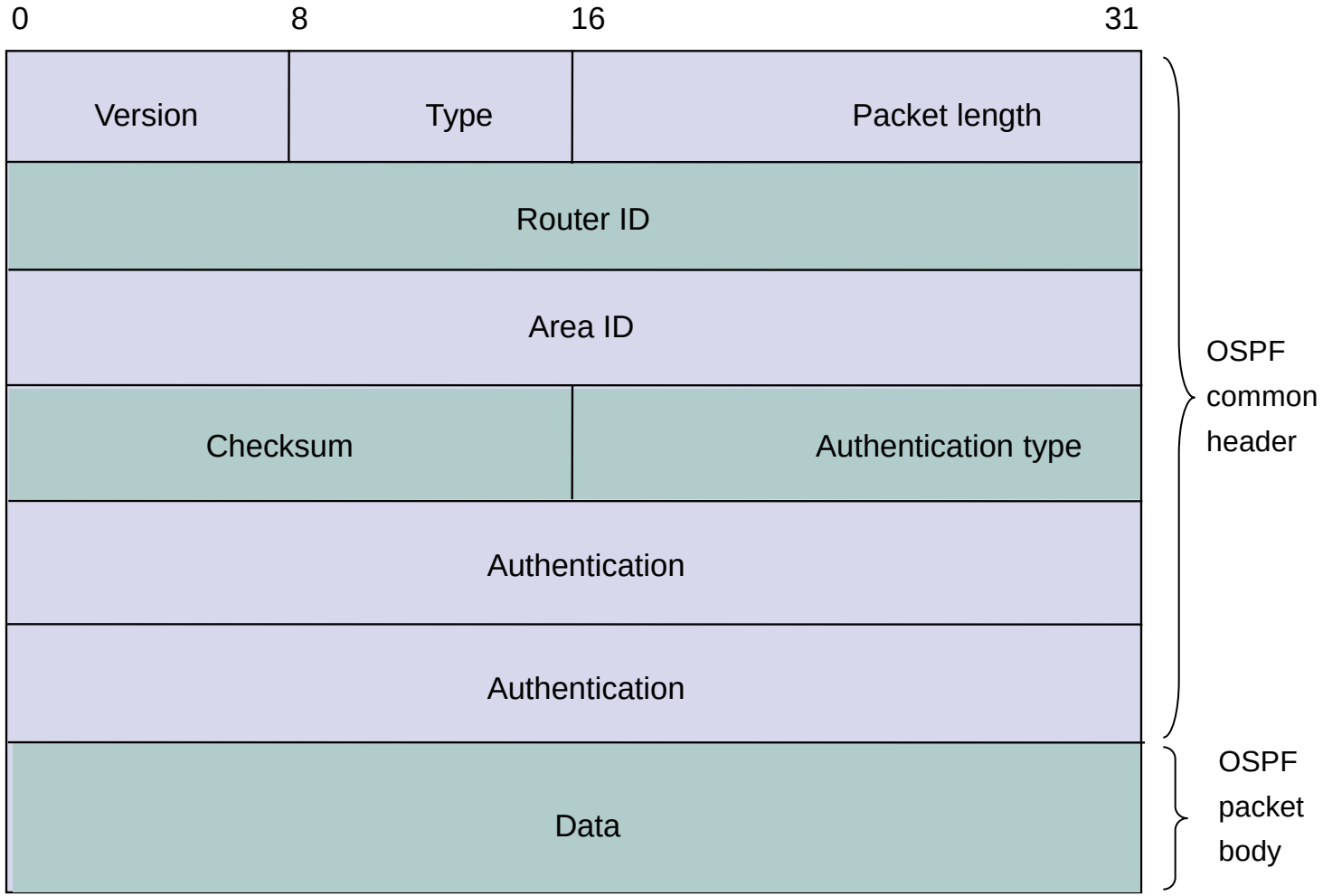
Link State Advertisements

- Link state info exchanged by adjacent routers to allow
  - area topology databases to be maintained
  - inter-area & inter-AS routes to be advertised
- *Router link ad*: generated by all OSPF routers
  - state of router links within area; flooded within area only
- *Net link ad*: generated by the designated router
  - lists routers connected to net: flooded within area only
- *Summary link ad*: generated by area border routers
  - 1. routes to dest in other areas; 2. routes to ASB routers
- *AS external link ad*: generated by ASB routers
  - describes routes to destinations outside the OSPF net
  - flooded in all areas in the OSPF net

OSPF Protocol

- OSPF packets transmitted **directly on IP datagrams**; Protocol ID 89
- TOS 0, IP precedence field set to internetwork control to get precedence over normal traffic
- OSPF packets sent to multicast address 224.0.0.5 (allSPFRouters on pt-2-pt and broadcast nets)
- OSPF packets sent on specific IP addresses on non-broadcast nets
- Five OSPF packet types:
  - *Hello*
  - *Database description*
  - *Link state request; Link state update; Link state ack*

OSPF Header



- Type: Hello, Database description, Link state request, Link state update, Link state acknowledgements

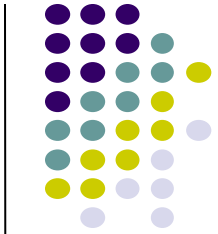
OSPF Stages

1. Discover neighbors by sending Hello packets (every 10 sec) and designated router elected in multi-access networks

2. Adjacencies are established & wait for their LSDBs to be synchronized
   - OSPF technique:
     - Source sends only LSA headers, then
     - Neighbor requests LSAs that it does not have
     - Those LSAs are sent over
     - After sync, the neighbors are said to be “fully adjacent”

3. Link state information is propagated & routing tables are calculated

Outline



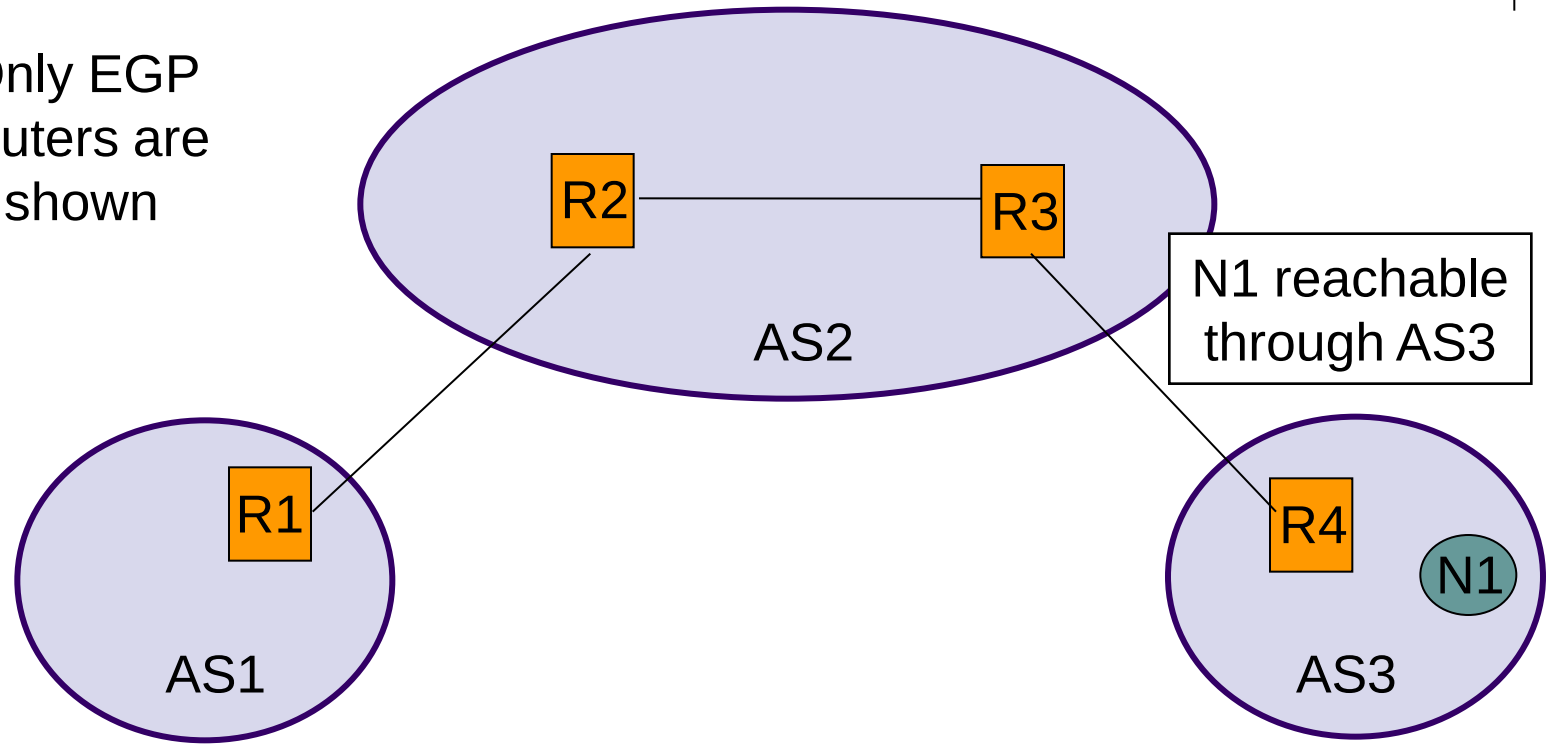
- Basic Routing
- Routing Information Protocol (RIP)
- Open Shortest Path First (OSPF)
- Border Gateway Protocol (BGP)

Exterior Gateway Protocols

- Within each AS, there is a consistent set of routes connecting the constituent networks
- The Internet is woven into a coherent whole by *Exterior Gateway Protocols (EGPs)* that operate between AS's
- EGP enables two AS's to exchange routing information about:
  - The networks that are contained within each AS
  - The AS's that can be reached through each AS
- EGP path selection guided by **policy** rather than path optimality
  - Trust, peering arrangements, etc

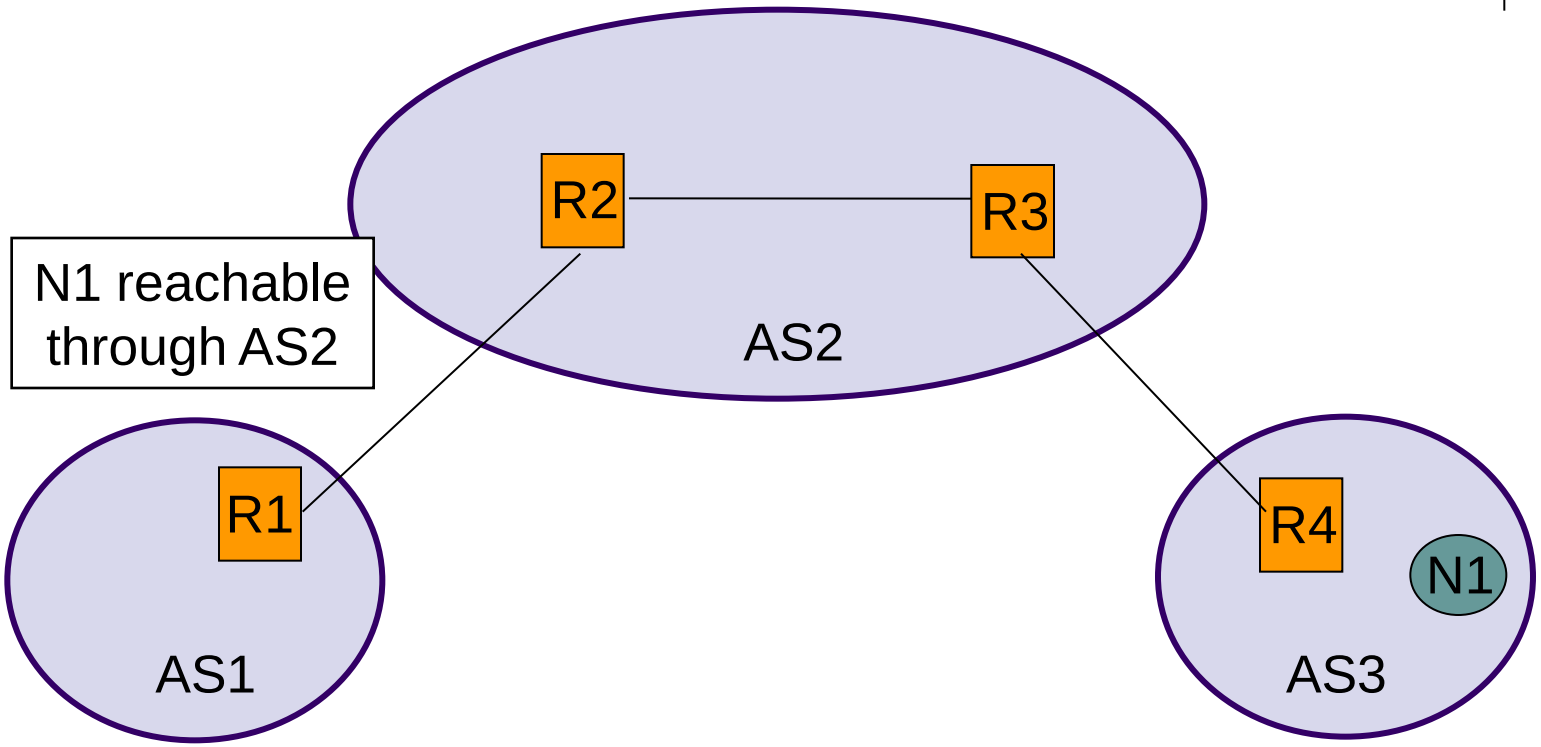
EGP Example

Only EGP routers are shown



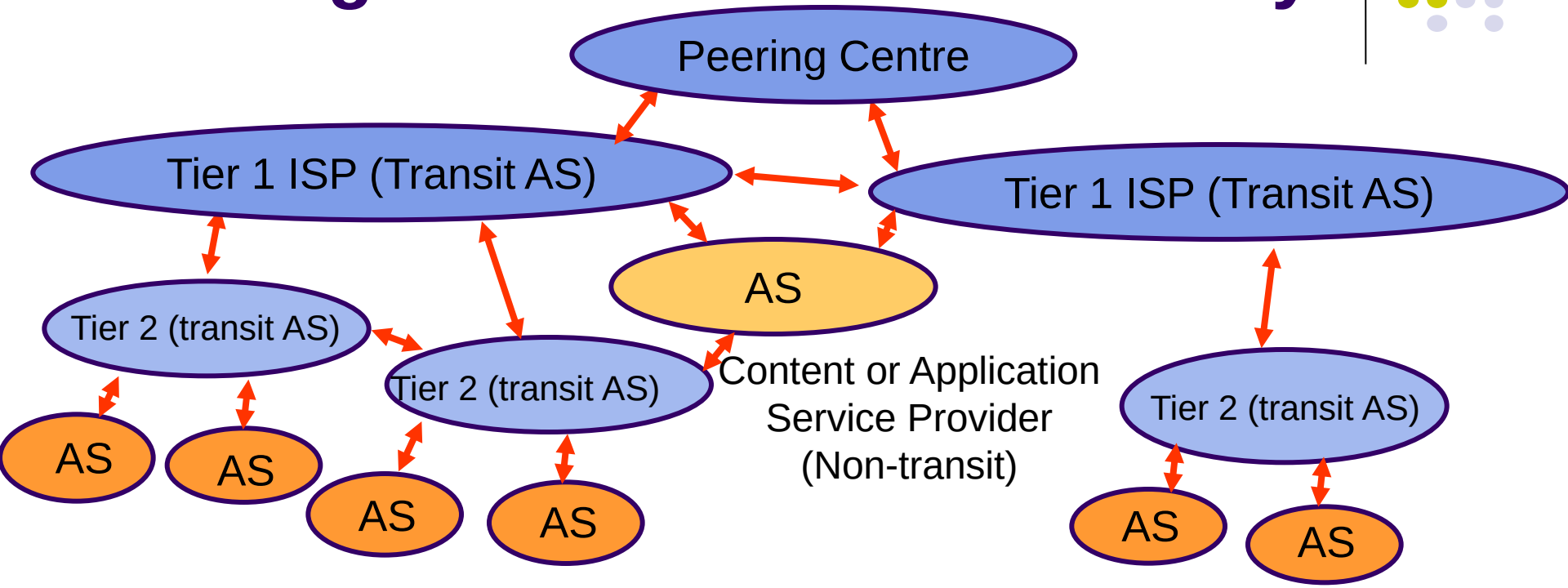
- R4 advertises that network N1 can be reached through AS3
- R3 examines announcement & applies *policy* to decide whether it will forward packets to N1 through R4
- If yes, routing table updated in R3 to indicate R4 as next hop to N1
- IGP propagates N1 reachability information through AS2

EGP Example



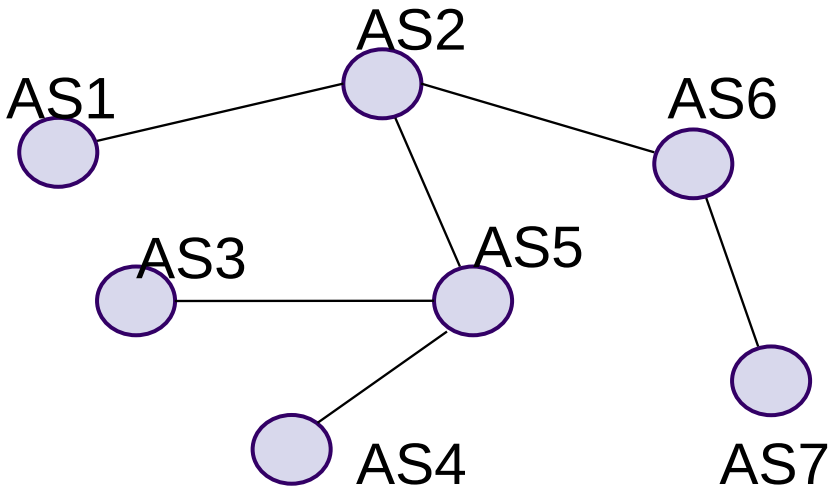
- EGP routers within an AS, e.g. R3 and R2, are kept consistent
- Suppose AS2 willing to handle *transit* packets from AS1 to N1
- R2 advertises to AS1 the reachability of N1 through AS2
- R1 applies its policy to decide whether to send to N1 via AS2

Peering and Inter-AS connectivity



- Non-transit AS's (stub & multihomed) do not carry transit traffic
- Tier 1 ISPs peer with each other, privately or through peering centers
- Tier 2 ISPs peer with each other & obtain transit services from Tier 1s; Tier 1's carry transit traffic between their Tier 2 customers
- Client AS's obtain service from Tier 2 ISPs

Border Gateway Protocol v4



- BGP (RFC 1771) an EGP routing protocol to exchange network reachability information among BGP routers (also called ***BGP speakers***)
- Network reachability info contains a sequence of ASs that packets traverse to reach a destination network
- Info exchanged between BGP speakers allows a router to construct a graph of AS connectivity

BGP Features

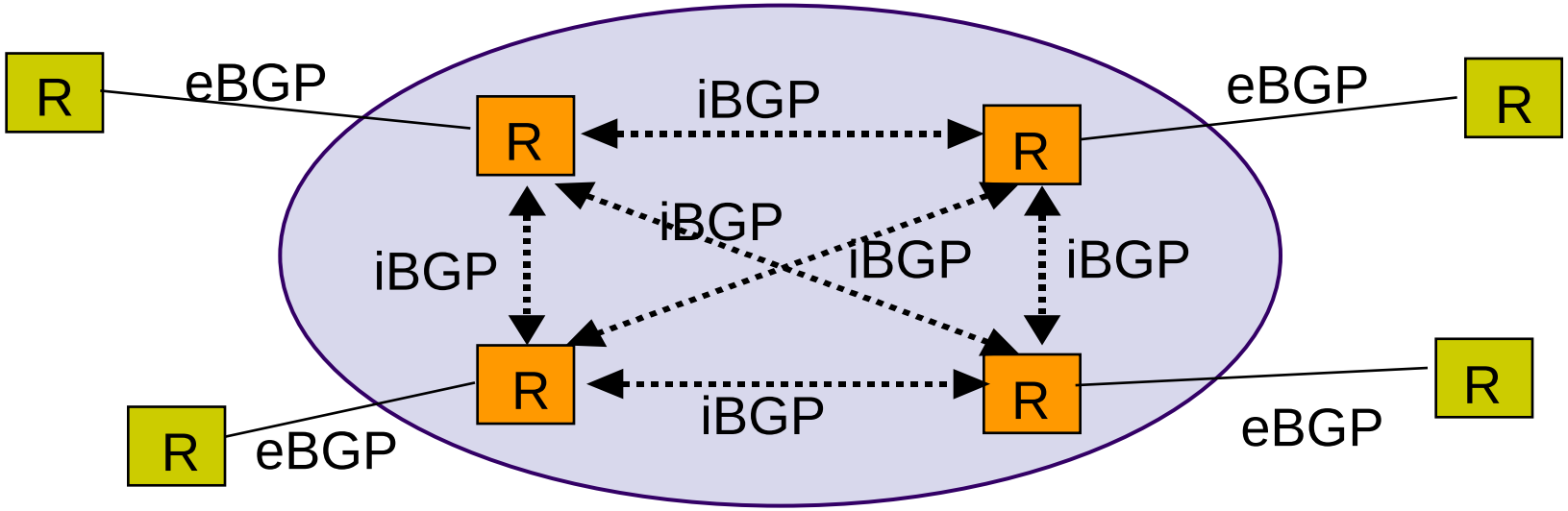
- BGP is ***path vector protocol***: advertises sequence of AS numbers (AS1, AS6, and AS7) to the destination network (10.10.1.0/24)
- Path vector info used to prevent routing loops
- BGP enforces policy through selection of different paths to a destination and by control of redistribution of routing information
- Uses CIDR to support aggregation & reduction of routing information

BGP Speaker & AS Relationship

- *BGP speaker*: a router running BGP
- *Peers or neighbors*: two speakers exchanging information on a connection
- **BGP peers use TCP** (port 179) to exchange messages
- Initially, BGP peers exchange entire BGP routing table
  - Incremental updates sent subsequently
  - Reduces bandwidth usage and processing overhead
  - Keepalive messages sent periodically (30 seconds)
- *Internal BGP* (iBPG) between BGP routers in same AS
- *External BGP* (eBGP) connections across AS borders

Slide is for self study

iBGP & eBGP



- eBGP to exchange reachability information in different AS's
  - eBGP peers directly connected
- iBGP to ensure net reachability info is consistent among the BGP speakers in the same AS
  - usually not directly connected
  - iBGP speakers exchange info learned from other iBGP speakers, and thus fully meshed

Path Selection

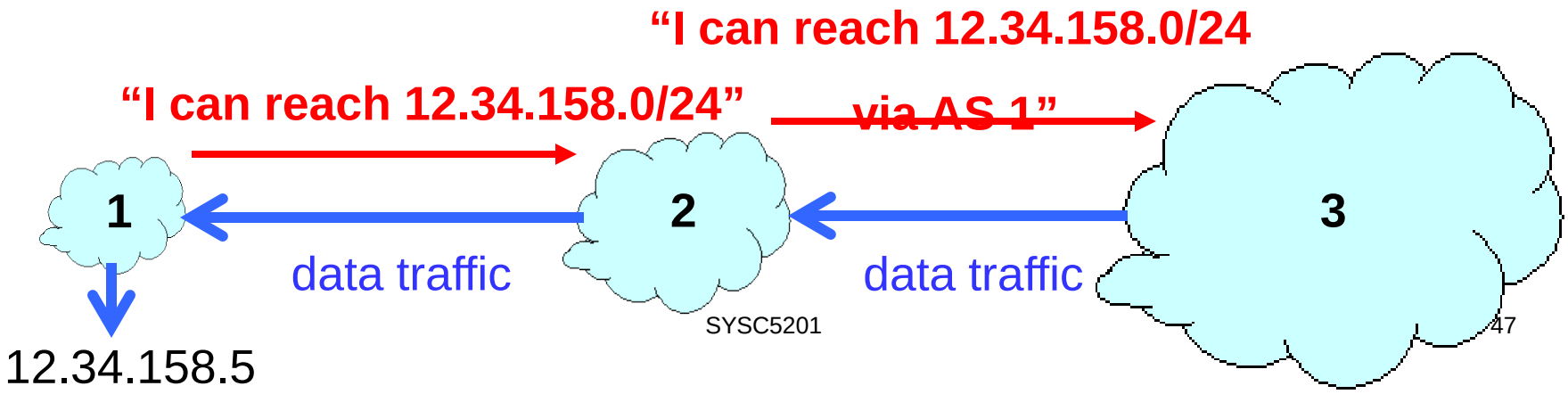
- Each BGP speaker
  - Evaluates paths to a destination from an AS border router
  - Selects the best that **complies with policies**
  - Advertises that route to all BGP neighbors
- BGP assigns a preference order to each path & selects path with highest value; BGP does not keep a cost metric to any path
- When multiple paths to a destination exist, BGP maintains all of the paths, but only advertises the one with highest preference value

BGP Policy

- Examples of policy:
  - Never use AS X
  - Never use AS X to get to a destination in AS Y
  - Never use AS X and AS Y in the same path
- ***Import policies*** to accept, deny, or set preferences on route advertisements from neighbors
- ***Export policies*** to determine which routes should be advertised to which neighbors
  - A route is advertised only if AS is willing to carry traffic on that route

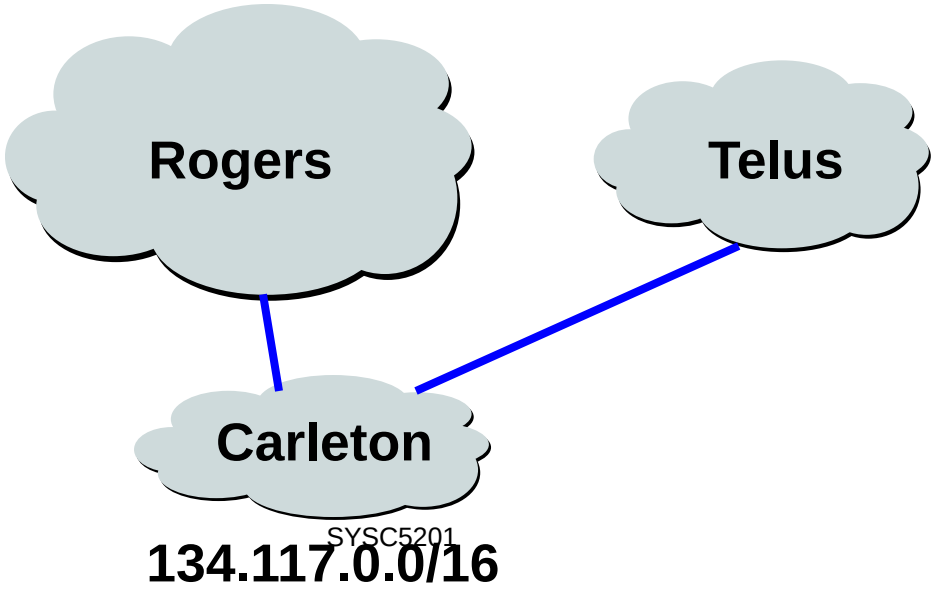
Border Gateway Protocol

- ASes exchange reachability information
  - IP prefix: block of destination addresses
  - AS path: sequence of ASes along the path
- Policies configured by the network operator
  - Path selection: which of the paths to use?
  - Path export: which neighbors to tell?



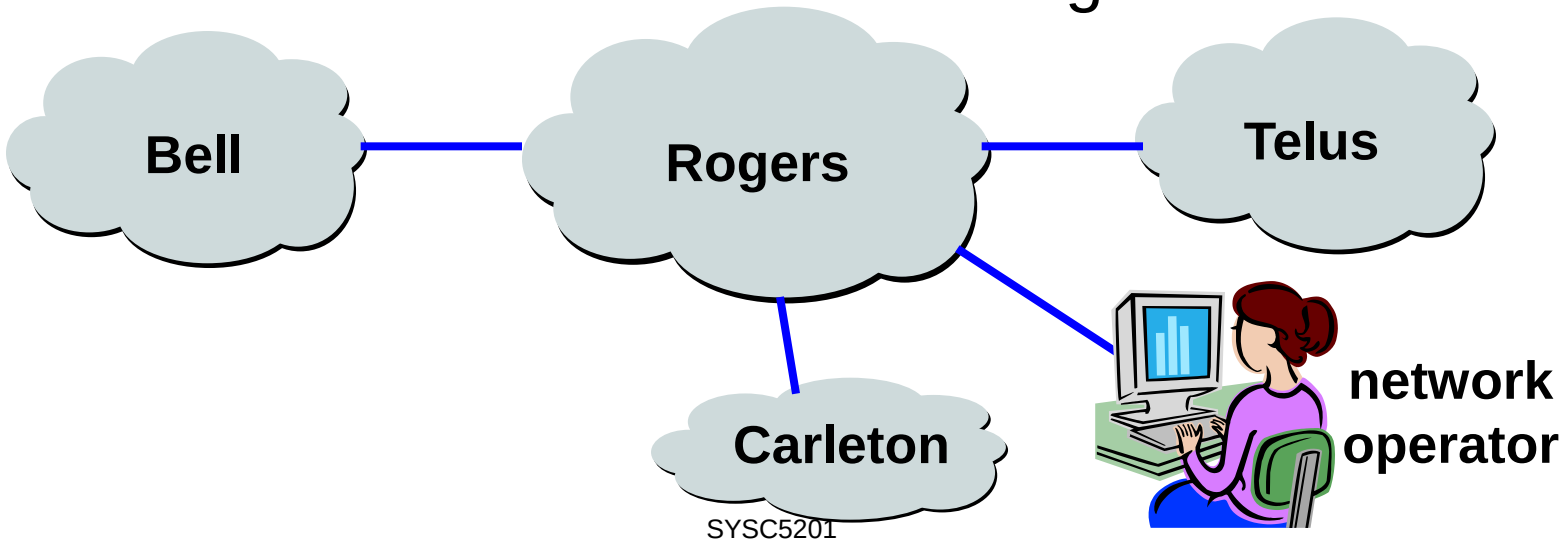
Import Policy: Filtering

- Discard some route announcements
  - Detect configuration mistakes and attacks
- Examples on session to a customer
  - Discard route if prefix not owned by the customer
  - Discard route with other large ISP in the AS path

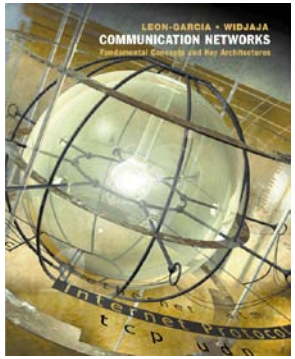


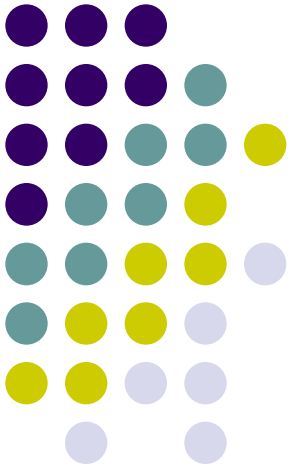
Export Policy: Filtering

- Discard some route announcements
  - Limit propagation of routing information
- Examples
  - Don’t announce routes from one peer to another
  - Don’t announce routes for management hosts



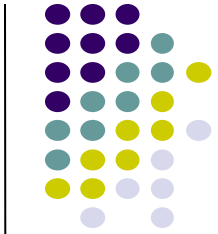
Chapter 8
Communication Networks and Services





DHCP, NAT, and Mobile IP

DHCP



- Dynamic Host Configuration Protocol (RFC 2131)
- BOOTP (RFC 951, 1542) allows a diskless workstation to be remotely booted up in a network
  - UDP port 67 (server) & port 68 (client)
- DHCP builds on BOOTP to **allow servers to deliver configuration information to a host**
  - Used extensively to assign temporary IP addresses to hosts
  - Allows ISP to maximize usage of their limited IP addresses

DHCP Operation

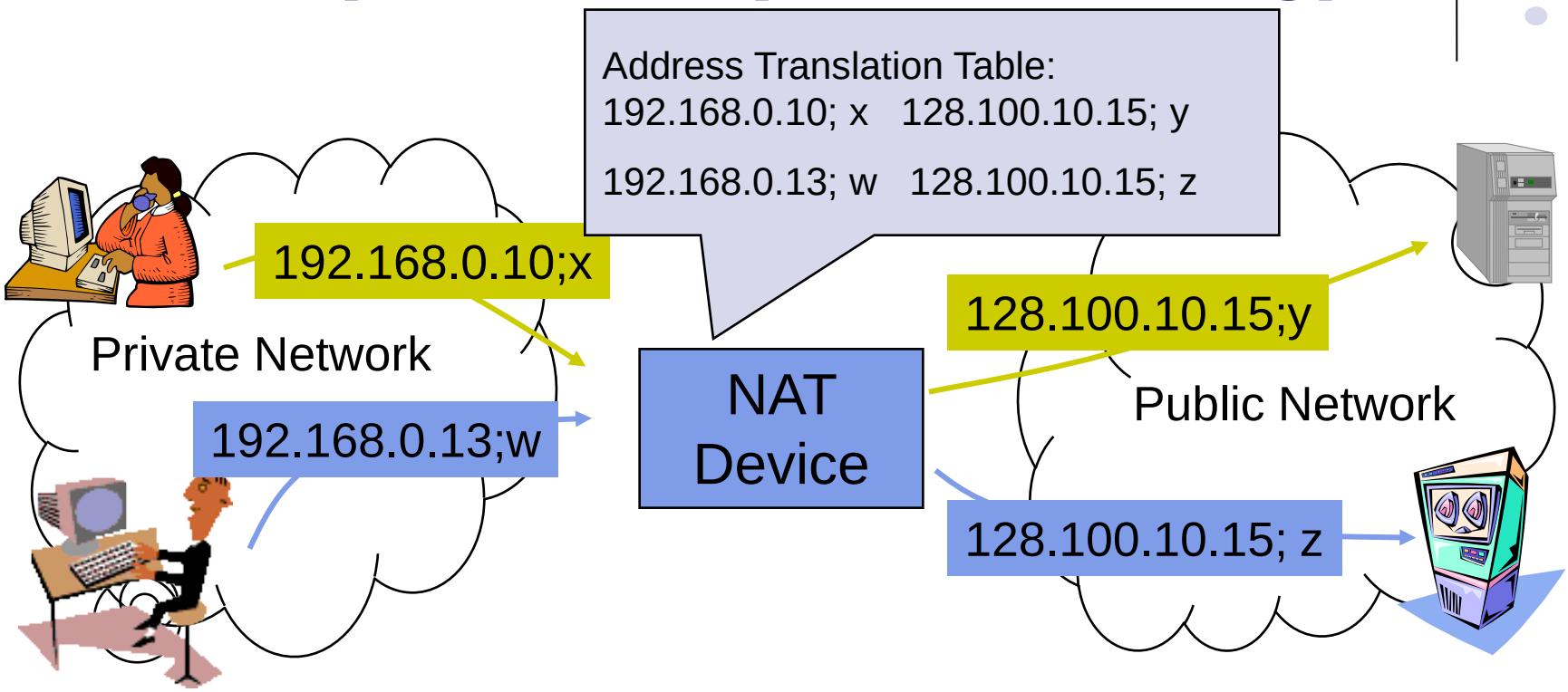
- Host broadcasts DHCP *Discover* message on its physical network
- Server replies with *Offer* message (IP address + configuration information)
- Host selects one offer and broadcasts *DHCP Request* message
- Server allocates IP address for lease time T
  - Sends DHCP ACK message with T, and threshold times T1 (=1/2 T) and T2 (=.875T)
- At T1, host attempts to renew lease by sending DHCP Request message to original server
- If no reply by T2, host broadcasts DHCP Request to *any* server
- If no reply by T, host must relinquish IP address and start from the beginning

Network Address Translation (NAT)

- Class A, B, and C addresses have been set aside for use within private internets
  - Packets with private ("unregistered") addresses are discarded by routers in the global Internet

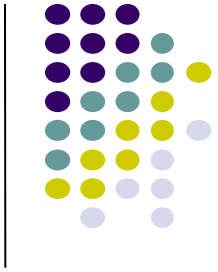
- NAT (RFC 1631): method for mapping packets from hosts in private internets into packets that can traverse the Internet
  - A device (computer, router, firewall) acts as an agent between a private network and a public network
  - A number of hosts can share a limited number of registered IP addresses
    - Static/Dynamic NAT: map unregistered addresses to registered addresses
    - Overloading: maps multiple unregistered addresses into a single registered address (e.g. Home LAN)

NAT Operation (Overloading)



- Hosts inside private networks generate packets with private IP address & TCP/UDP port #s
- NAT maps each private IP address & port # into shared global IP address & available port #
- Translation table allows packets to be routed unambiguously

Mobile IP



- Proliferation of mobile devices: smart phones, laptops
- As user moves, point-of-attachment to network necessarily changes
- Problem: IP address specifies point-of-attachment to Internet
  - Changing IP address involves terminating all connections & sessions
- *Mobile IP (RFC 2002)*: device can change point-of-attachment while retaining IP address and maintaining communications