

# CSCE 515: Computer Network Programming

----- IP, Ping, Traceroute

Wenyuan Xu

Department of Computer Science and Engineering  
University of South Carolina

## ICMP *Internet Control Message Protocol*

- ICMP is a protocol used for exchanging control messages.
- Two main categories
  - *Query message*
  - *Error message*
- Usage of an ICMP message is determined by *type* and *code* fields
- ICMP uses IP to deliver messages.
- ICMP messages are usually generated and processed by the IP software, not the user process.



20 bytes

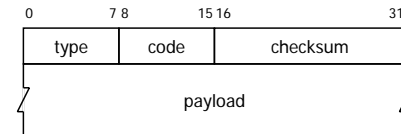
CSCE515 – Computer Network Programming

## IP Datagram

1 byte	1 byte	1 byte	1 byte
VERS	HL	Service	Total Length
Datagram ID		FLAG	Fragment Offset
TTL	Protocol	Header Checksum	
Source Address			
Destination Address			
Options (if any)			
Data			

CSCE515 – Computer Network Programming

## ICMP Message Format



CSCE515 – Computer Network Programming

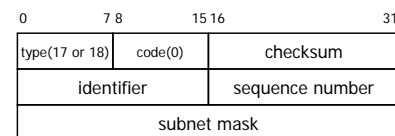
## ICMP Message Types

- Echo Request
- Echo Response
- Destination Unreachable
- Redirect
- Time Exceeded
- there are more ...

CSCE515 – Computer Network Programming

## ICMP Address Mask Request and Reply

- intended for a diskless system to obtain its subnet mask.
- Id and seq can be any values, and these values are returned in the reply.
  - Match replies with request



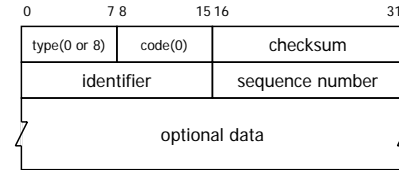
CSCE515 – Computer Network Programming

## ping Program

- Available at `/usr/sbin/ping`
- Test whether another host is reachable
- Send ICMP echo\_request to a network host
- `-n` option to set number of echo request to send
- `-i` option to set TTL
- `-R` option to record route (apollon.cse.sc.edu)
- `-s` option to set timestamp
- `-w` option to set timeout to wait for each reply
- Check manual, different ping versions have different options

CSCE515 – Computer Network Programming

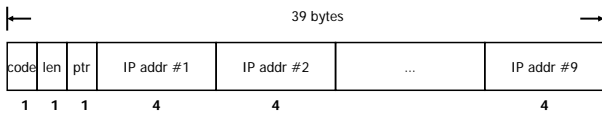
## ICMP Echo Request and Reply



CSCE515 – Computer Network Programming

## IP Record Route Option

- `ping -R`: Record route
  - Every router that handles the datagram adds its IP address to a list in the options field
  - The final destination copies the IP addresses into the outgoing ICMP echo reply
  - All routers on the return path add their IP address to the list
- Problems?



CSCE515 – Computer Network Programming

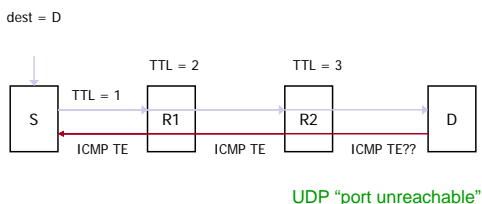
## traceroute Program

- Available at `/usr/sbin/traceroute`
- Display the route that IP datagrams follow from one host to another
- Compare with ping:
  - Doesn't require an special or optional features at any intermediate routers
  - Only requires a working UDP module at the destination
  - uses ICMP and the TTL field in the IP header
- `-g` option to specify intermediate routers to be used with loose source routing (up to 8 times)
- `-G` option to specify intermediate routers to be used with strict source routing (up to 8 times)

CSCE515 – Computer Network Programming

## traceroute Program

- TTL + ICMP
  - Each router decrement the TTL at least by 1
  - A IP datagram whose TTL is either 0 or 1 will not be forwarded.
  - An ICMP "time exceeded" message will be sent back to the originating host.

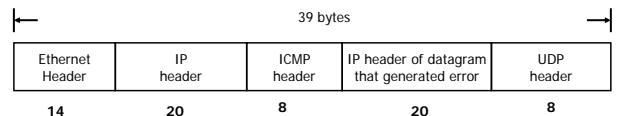


CSCE515 – Computer Network Programming

## UDP port unreachable

- ICMP error message
  - IP header
  - 8 bytes of the IP datagram that caused the error

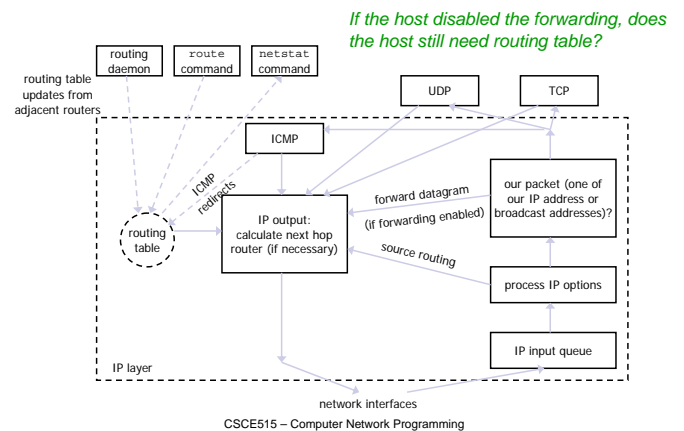
### WHY?



CSCE515 – Computer Network Programming

## Creating Routing Entries

## Kernel Processing at IP Layer



## IP Layer

- Forwarding datagrams generated either on local host or on some other hosts toward their ultimate destination
- Routing:
  - *Static routing*: when network is small, single connection point to other networks, no redundant route existent
    - specified in configuration files
    - not based on measurement or estimates of current traffic and topology
  - *Dynamic routing*: use routing daemon to run routing protocol in order to communicate with other routers

CSCE515 – Computer Network Programming

## Create Routing Table Entries

- Created by default when an interface is configured
  - when the interface's address is set by the `ifconfig`

```
Destination Gateway Flags Ref Use Interface
129.252.130.0 129.252.130.106 U 1 68 eri0
```
- A default router specified in a file, the default is added to the routing table on every reboot.
  - `/etc/defaultrouter`

```
wyxu@altair % cat /etc/defaultrouter
129.252.130.1
```
- Added by `route` command
- Created by an ICMP redirect

CSCE515 – Computer Network Programming

## route Command

- Explicitly add or remove routing table entry from configuration files at bootstrap time
  - `route add default sun 1`
  - `route add slip bsdi 1`

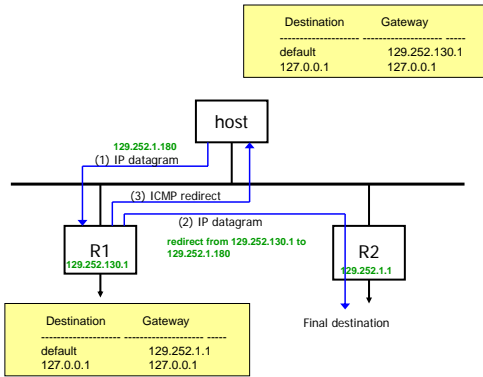
CSCE515 – Computer Network Programming

## ICMP Redirect Error

- Sent by a router to sender of an IP datagram when the datagram should have been sent to a different router
- Used only when the host has a choice of routers to send its datagram to

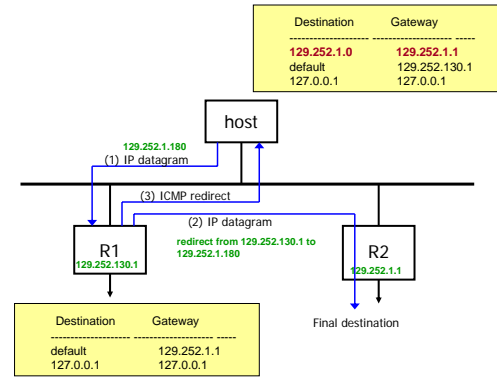
CSCE515 – Computer Network Programming

## Example of ICMP Redirect



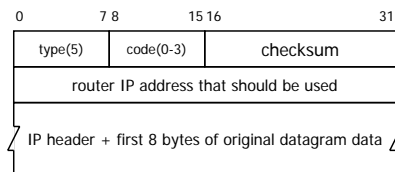
CSCE515 – Computer Network Programming

## Example of ICMP Redirect



CSCE515 – Computer Network Programming

## ICMP Redirect



CSCE515 – Computer Network Programming

## Security concern

- What can you do to take advantage of the ICMP redirect?
  - Redirect to some unknown host
  - Redirect to the host itself
- Sniffing packet
  - Redirect to my own address?
- Greedy router,
  - I don't want to route the packet

CSCE515 – Computer Network Programming

## Security concern- Partial solutions

- The new router must be **on a directly connected network**
- The redirect must be **from the current router** for that destination
- The redirect cannot tell the host to use **itself** as a router
- The route that's being modified must be an **indirect route**

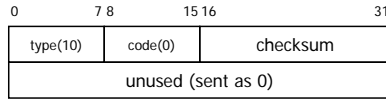
CSCE515 – Computer Network Programming

## ICMP Router Discovery Messages

- After bootstrapping
  - broadcasts / multicasts a **router solicitation** message
  - other routers respond with a **router advertisement** message
- Periodically advertisement
  - broadcasts / multicasts a **router solicitation** message

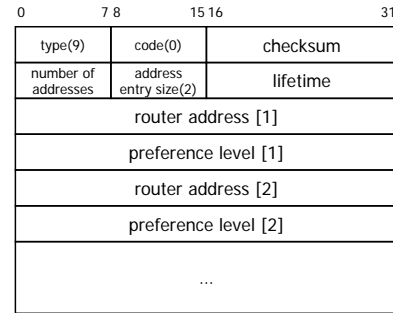
CSCE515 – Computer Network Programming

## ICMP Router Solicitation



CSCE515 – Computer Network Programming

## ICMP Router Advertisement

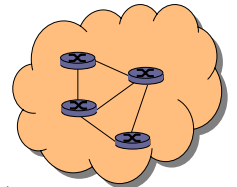


CSCE515 – Computer Network Programming

## Routing protocols

## Autonomous Systems

- Collection of networks with same policy
- Single routing protocol
- Usually under single administrative control



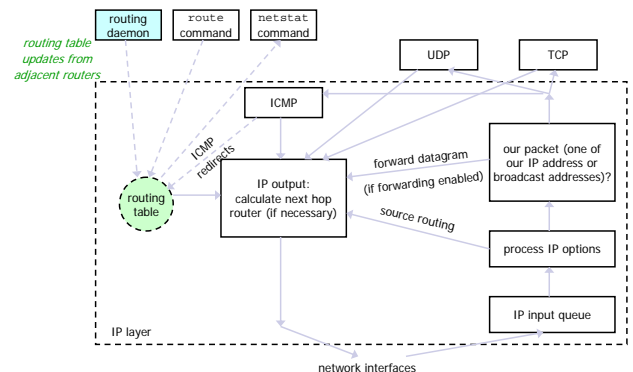
CSCE515 – Computer Network Programming

## Routing classification

- Interior gateway protocols (IGPs)
  - RIP (Routing Information Protocol)
  - OSPF (Open Shortest Path First)
- Exterior gateway protocols (EGPs)
  - BGP: border gateway protocol
    - Used between NSFNET backbone and some of the regional networks

CSCE515 – Computer Network Programming

## Kernel Processing at IP Layer



CSCE515 – Computer Network Programming

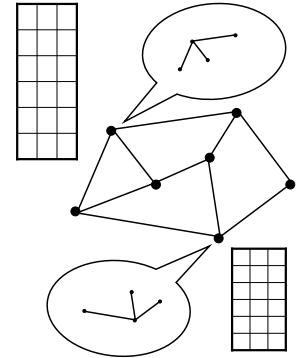
## Routing Protocols

- Executed by routing daemon to communicate routing information with other routers
- Two types of routing algorithms (IGPs)
  - Distance-vector routing
  - Link-state routing

CSCE515 – Computer Network Programming

## Distance-vector Protocols

- Maintain a vector of distances
- Each router updates its routing table based on vector of distances received from neighbors
- Example: RIP
  - most widely used routing protocol
  - the metrics used: hop count



CSCE515 – Computer Network Programming

## Problem: Count-to-infinity

- With distance vector routing, good news travels fast, but bad news travels slowly
- When a router goes down, it takes can take a really long time before all the other routers become aware of it

CSCE515 – Computer Network Programming

## Count-to-infinity

A	B	C	D	E	
X					Initially
	1	2	3	4	After 1 exchange
	3	2	3	4	After 2 exchanges
	3	4	3	4	After 3 exchanges
	5	4	5	4	After 4 exchanges
	5	6	5	6	After 5 exchanges
	7	6	7	6	After 5 exchanges
					etc... to infinity

CSCE515 – Computer Network Programming

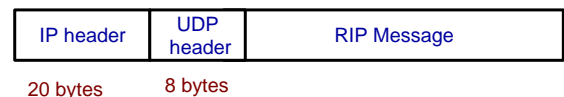
## Improvements

- Split Horizon
  - Don't tell neighbor about routes obtained from it
- Triggered updates as opposed to periodic updates
- Path vectors, Store vectors or complete path as opposed to just next hop

CSCE515 – Computer Network Programming

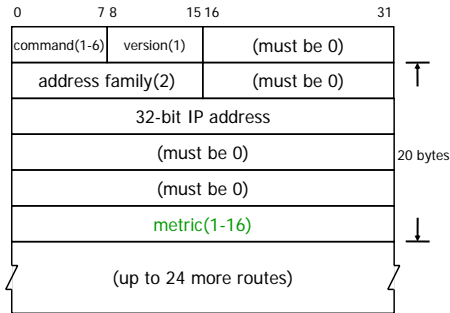
## Routing Information Protocol (RIP)

- Most widely used routing protocol
- Carried in **UDP** datagrams
- Routing daemon:
  - routed
  - gated



CSCE515 – Computer Network Programming

## RIP Message Format



CSCE515 – Computer Network Programming

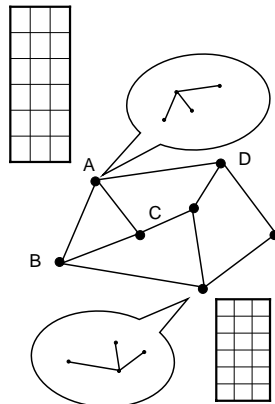
## RIP Metrics

- RIP uses hop count as its metric
- If there are multiple paths, router chooses the one with smallest hop count, and ignores other paths

CSCE515 – Computer Network Programming

## RIP Operation

- Initialization
- Request received
- Response received
- Regular routing updates
- Triggered updates



CSCE515 – Computer Network Programming

## Problems with RIP

- Has no knowledge about subnet addressing
- Take long time to stabilize after a router or link failure
- Maximum of metric limits network size
- No security protection

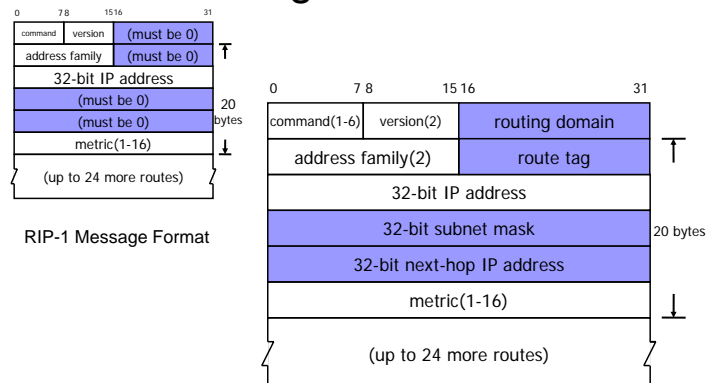
CSCE515 – Computer Network Programming

## RIP Version 2

- Fix some deficiencies of RIP
- Support multiple domain
- Include subnet mask
- Some simple authentication scheme added

CSCE515 – Computer Network Programming

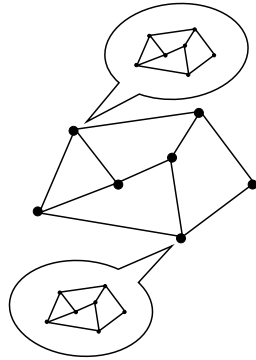
## RIP-2 Message Format



CSCE515 – Computer Network Programming

## Link-state Protocols

- Each router maintains a complete routing table of the network
- Example: Open Shortest Path First (OSPF)



CSCE515 – Computer Network Programming

## A link-state routing protocol

- Discover neighbors
- Measure the delay or cost to each of its neighbors
- Flood routing information and link costs
  - To control flooding, the sequence numbers are used by routers to discard flood packets they have already seen from a given router
  - The age field in the packet is an expiration date. It specifies how long the information in the packet is good for.
- Once a router receives all the link state packets from the network, it can reconstruct the complete topology and compute a shortest path between itself and any other node using Dijkstra's algorithm (shortest path).

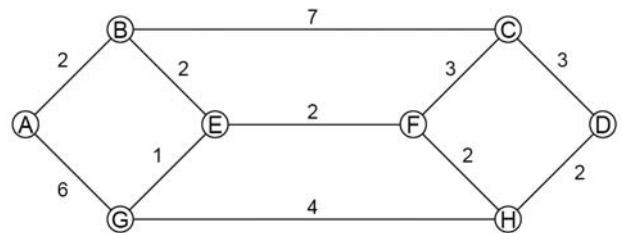
CSCE515 – Computer Network Programming

## Computing the Shortest Path

- Dijkstra's Shortest Path Algorithm:
  - Step 1: Draw nodes as circles. Fill in a circle to mark it as a "temporary node."
  - Step 2: Set the current node equal to the source node
  - Step 3: For the current node:
    - – Mark the cumulative distance from the current node to each temporary adjacent node. Also mark the name of the current node. Erase this marking if the adjacent node already has a shorter cumulative distance marked
    - – Mark the temporary node with the shortest listed cumulative distance as permanent by marking the <cost of the best known path from Source, Previous Hop> and set the current node equal to it. Repeat step 3 until all nodes are marked permanent.

CSCE515 – Computer Network Programming

## Dijkstra's Shortest Path Algorithm



CSCE515 – Computer Network Programming

## Open Shortest Path First (OSPF)

- Routing algorithm now used in the Internet
- OSPF uses the *Link State Routing algorithm* with modifications to support:
  - Multiple distance metrics (geographical distance, delay, throughput)
  - Support for real-time traffic
  - Support for subnets
  - Hierarchical routing
  - Security – a simple authentication scheme
- Use IP to carry its message
- Provide features superior to RIP

CSCE515 – Computer Network Programming

## OSPF: Modified Link State Routing

- Recall:
  - In link state routing, routers flood their routing information to all other routers in the network
- In OSPF, routers only send their information to "adjacent routers", not to all routers.
- Adjacent does NOT mean nearest-neighbor in OSPF
- One router in each area is marked as the "designated router"
- Designated routers are considered adjacent to all other routers in the area
- OSPF combines link state routing with centralized adaptive routing

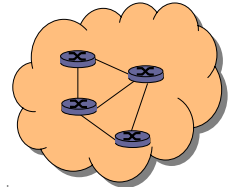
Someone know the topology of network

CSCE515 – Computer Network Programming

# BGP

## Autonomous Systems

- Collection of networks with same policy
- Single routing protocol
- Usually under single administrative control

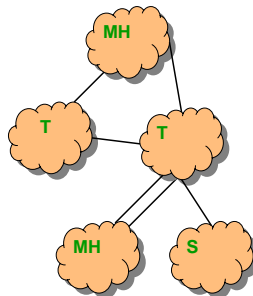


CSCE515 – Computer Network Programming

## Autonomous Systems

### ■ Three categories of AS

- **Stub AS**
  - Carry only local traffic
- **Multihomed AS**
  - Connected to more than one AS
  - Still local traffic
- **Transit AS**
  - Carries local and non-local traffic



CSCE515 – Computer Network Programming

## IGP and EGP

- Each AS selects its **interior gateway protocol** (IGP) for communications between routers in this AS
  - E.g. RIP, OSPF
  - Within AS, local routing protocols used (optimize path metric)
- Multiple AS's use **exterior gateway protocol** (EGP) for communications between routers in different AS's
  - E.g. EGP, BGP
  - Inter-AS concerned with reachability and policy implementation
  - Usually \$\$ involved with relationships

CSCE515 – Computer Network Programming

## Interior vs. Exterior Routing Protocols

### ■ Interior

- Automatic discovery
- Generally trust your IGP routers
- Routes go to all IGP routers

### ■ Exterior

- Specifically configured peers
- Connecting with outside networks
- Set administrative boundaries

CSCE515 – Computer Network Programming

## Why do we need an EGP?

- Scaling to large network
  - Hierarchy
  - Limit scope of failure
- Policy
  - Control reachability to prefixes
- Allow policy-based routing
  - No Transit traffic through certain ASes
  - Never put Iraq on a route starting at the Pentagon
  - Traffic starting or ending at IBM should not transit Microsoft

CSCE515 – Computer Network Programming

## Border Gateway Protocol (BGP)

- An exterior gateway protocol
- It's neither a distance-vector nor a link-state protocol
  - Distance-vector protocol but enumerates route to each destination
- Typically static metrics (DELAY or BANDWIDTH)
- Use **TCP** to transport its messages

CSCE515 – Computer Network Programming

## BGP protocol

- BGP uses **TCP** as its transport protocol, on port 179. On connection start, BGP peers exchange complete copies of their routing tables, which can be quite large. However, only changes (deltas) are then exchanged, which makes long running BGP sessions more efficient than shorter ones.
- Four Basic messages:
  - **Open:**
    - Establishes BGP session (uses TCP port #179)
  - **Notification:**
    - Report unusual conditions
  - **Update:**
    - Inform neighbor of new routes that become active
    - Inform neighbor of old routes that become inactive
  - **Keepalive:**
    - Inform neighbor that connection is still viable

CSCE515 – Computer Network Programming

## OPEN Message

- Each AS has:
  - one or more border routers
    - Handles inter-AS traffic
  - one BGP *speaker* for an AS that participates in routing
- During session establishment, two BGP speakers exchange their
  - AS numbers
  - BGP identifiers (usually one of the router's IP addresses)
- A BGP speaker has option to refuse a session
- Select the value of the hold timer:
  - maximum time to wait to hear something from other end before assuming session is down.
- authentication information (optional)

CSCE515 – Computer Network Programming

## NOTIFICATION and KEEPALIVE Messages

- **NOTIFICATION**
  - Indicates an error
  - terminates the TCP session
  - gives receiver an indication of why BGP session terminated
  - Examples: header errors, hold timer expiry, bad peer AS, bad BGP identifier, malformed attribute list, missing required attribute, AS routing loop, etc.
- **KEEPALIVE**
  - protocol requires some data to be sent periodically. If no UPDATE to send within the specified time period, then send KEEPALIVE message to assure partner that connection still alive

CSCE515 – Computer Network Programming

## UPDATE Message

- withdrawn routes
- attributes
- advertised routes

CSCE515 – Computer Network Programming

## Update Messages..

- Network reachability information
  - network prefix/length
  - Example :
    - 131.108/16
    - 131.108.0.0 255.255.0.0
    - 198/8
    - 198.0.0.0 255.0.0.0

CSCE515 – Computer Network Programming

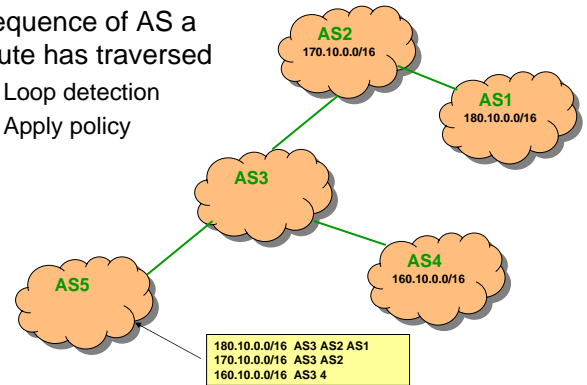
## BGP Attributes

- What is an attribute?
  - AS path
  - Next hop
  - Local preference
  - Multi-Exit Discriminator (MED)

CSCE515 – Computer Network Programming

## AS-Path

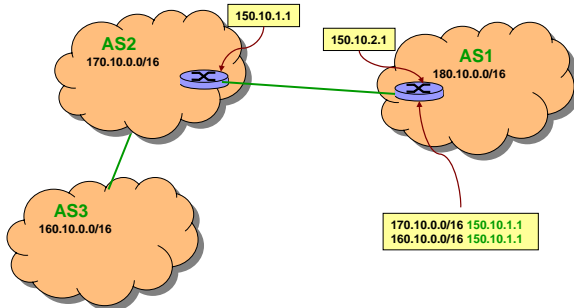
- Sequence of AS a route has traversed
  - Loop detection
  - Apply policy



CSCE515 – Computer Network Programming

## Next hop

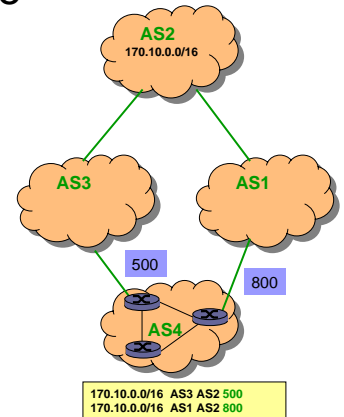
- Next hop to reach a network



CSCE515 – Computer Network Programming

## Local Preference

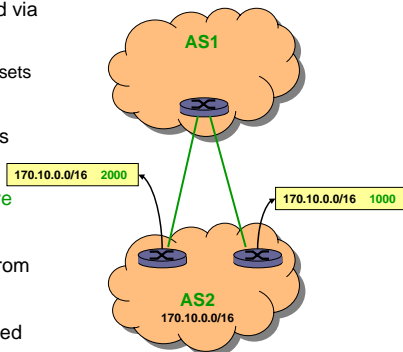
- Used to indicate preference among multiple paths for the same prefix *anywhere* in the internet.
- The higher the value the more it is preferred
- Default value is 100
- Local to the AS
- Often used to select a specific exit point for a particular destination
- Used when AS path lengths are same
- Valid within a AS only



CSCE515 – Computer Network Programming

## Multi-Exit Discriminator

- When AS's interconnected via 2 or more links
  - AS path length are same
  - AS announcing a prefix, sets MED value
- Enables AS2 to indicate its preference (*lower MED is better*)
- Used to convey the relative preference of entry points
- Comparable if paths are from same AS
- IGP metric can be conveyed as MED



CSCE515 – Computer Network Programming

## BGP Decision Process

1. Choose route with highest LOCAL-PREF
2. If have more than 1 route, select route with shortest AS-PATH
3. If have more than 1 route, select according to lowest ORIGIN type where  $IGP < BGP < default$
4. If have more than 1 route, select route with lowest MED value
5. Select min cost path to NEXT HOP using IGP metrics
6. If have multiple internal paths, use BGP Router ID to break tie.

See: <http://www.cisco.com/warp/public/459/37.html>

CSCE515 – Computer Network Programming



## Assignment & Next time

- Reading:

- TI Ch 7, 8, 9 \*\*;

- Next Lecture:

- DNS