Digital Communication in the Modern World

Network Layer: Routing in the Internet

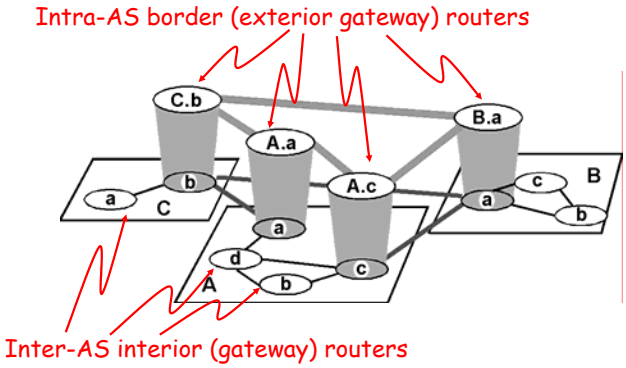
http://www.cs.huji.ac.il/~com1
com1@cs.huji.ac.il

Some of the slides have been borrowed from:
Computer Networking: A Top Down Approach Featuring the Internet,
2nd edition.
Jim Kurose, Keith Ross
Addison-Wesley, July 2002.

Routing in the Internet

- The Global Internet consists of Autonomous Systems (AS) interconnected with each other:
  - **Stub AS**: small corporation: one connection to other AS's
  - **Multihomed AS**: large corporation (no transit): multiple connections to other AS's
  - **Transit AS**: provider, hooking many AS's together
- Two-level routing:
  - **Intra-AS**: administrator responsible for choice of routing algorithm within network
  - **Inter-AS**: unique standard for inter-AS routing: BGP

Internet AS Hierarchy



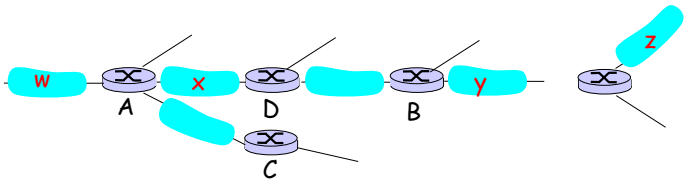
Intra-AS Routing

- Also known as Interior Gateway Protocols (IGP)
- Most common Intra-AS routing protocols:
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First
  - (IGRP: Interior Gateway Routing Protocol - Cisco proprietary)

RIP (**R**outing **I**nformation **P**rotocol)

- Distance Vector algorithm
- Included in BSD-UNIX Distribution in 1982
- Distance metric: # of hops (max = 15 hops)
- Distance vectors: exchanged among neighbors every 30 sec via Response Message (also called **advertisement**)
- Each advertisement: list of up to 25 destination **nets** within AS

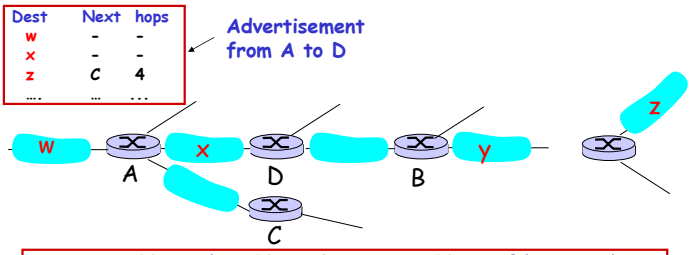
RIP: Example



Destination Network	Next Router	Num. of hops to dest.
w	A	2
y	B	2
z	B	7
x	--	1
….	….	….

Routing table in **D**

RIP: Example



Destination Network	Next Router	Num. of hops to dest.
w	A	2
y	B	2
z	~~B~~ A	~~7~~ 5
x	--	1
….	….	….

Routing table in D

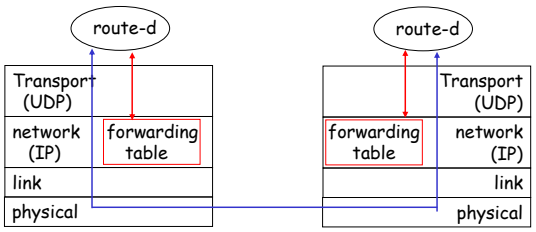
RIP: Link Failure and Recovery

If no advertisement heard after 180 sec --> neighbor/link declared dead

- routes via that neighbor invalidated
- new advertisements sent to neighbors
- neighbors in turn send out new advertisements (if tables changed)
- link failure info quickly propagates to entire net
- **poisoned reserve** used to prevent ping-pong loops (infinite distance = 16 hops)

RIP Table processing

- RIP routing tables managed by **application-level** process called route-d (Unix daemon)
- advertisements sent in UDP packets



RIP Table example (continued)

netstat -r at router *giroflee.eurocom.fr*

Destination	Gateway	Flags	Ref	Use	Interface
127.0.0.1	127.0.0.1	UH	0	26492	lo0
192.168.2.	192.168.2.5	U	2	13	fa0
193.55.114.	193.55.114.6	U	3	58503	le0
192.168.3.	192.168.3.5	U	2	25	qaa0
224.0.0.0	193.55.114.6	U	3	0	le0
default	193.55.114.129	UG	0	143454	

- Three attached class C networks (LANs)
- Router only knows routes to attached LANs
- 'Default router' used to go to unlisted destinations
- Router multicast address: 224.0.0.0
- Loopback interface (for debugging)

OSPF (Open Shortest Path First)

- RIP not sufficient for large nets, inherited from ARPANET
- In 1979 IETF started replacing RIP with a link state gateway routing protocol
- In 1988 IETF started the design of a successor called OSPF which became a standard in 1990
- OSPF was designed to deal with a variety of issues:
  - Routing based on type of service
  - Enable variable distance metrics
  - Load balancing
  - Security
  - Scalability

OSPF

- "open": publicly available
- Uses Link State algorithm
  - LS packet dissemination (diffusion)
  - Topology map at each node
  - Route computation using Dijkstra's algorithm
- OSPF advertisement carries one entry per neighbor router
- Advertisements disseminated to entire AS (via flooding)
  - Carried in OSPF messages directly over IP (rather than TCP or UDP

OSPF

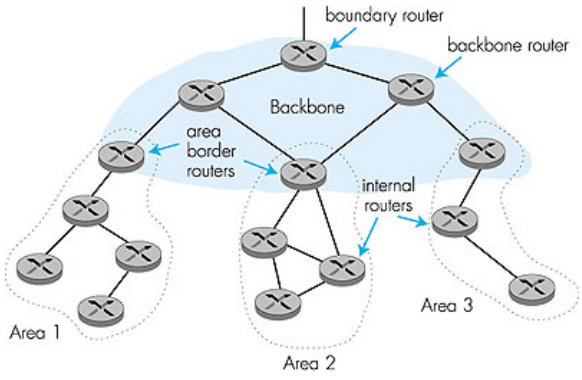
The five types of OSPF messages:

Message type	Description
Hello	Used to discover who the neighbors are
Link state update	Provides the sender's costs to its neighbors
Link state ack	Acknowledges link state update
Database description	Announces which updates the sender has
Link state request	Requests information from the partner

OSPF "advanced" features (not in RIP)

- Security: all OSPF messages authenticated (to prevent malicious intrusion)
- Multiple same-cost paths allowed (only one path in RIP); can use next-shortest path first for load balancing
- For each link, multiple cost metrics (e.g., satellite link cost set to "low" for best effort; high for real time)
- Integrated uni- and multicast support:
  - Multicast OSPF (MOSPF) uses same topology data base as OSPF
- Hierarchical OSPF in large domains.

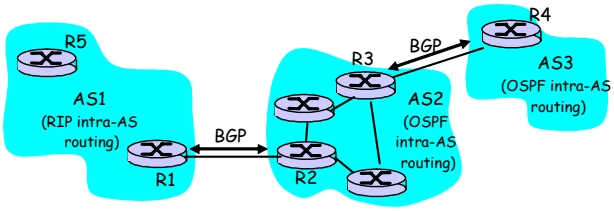
Hierarchical OSPF



Hierarchical OSPF

- Two-level hierarchy: local area, backbone.
  - Link-state advertisements only in area
  - each node has detailed area topology; only know direction to nets in other areas.
- **Area border routers:** "summarize" distances to nets in own area, advertise to other Area Border routers.
- **Backbone routers:** run OSPF routing limited to backbone.
- **Boundary routers:** connect to other AS's.

Inter-AS routing in the Internet: BGP



Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol): *the* de facto standard
- **Path Vector** protocol:
  - similar to Distance Vector protocol
  - each Border Gateway broadcast to neighbors (peers) *entire path* (i.e., sequence of AS's) to destination
  - BGP routes to networks (ASs), not individual hosts
  - E.g., Gateway X may send its path to dest. Z:

Path (X,Z) = X,Y1,Y2,Y3,…,Z

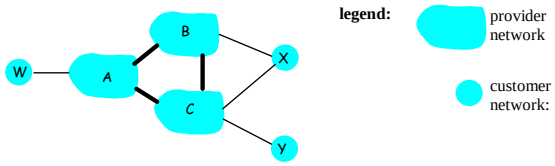
Internet inter-AS routing: BGP

Suppose: gateway X sends its path to peer gateway W

- W may or may not select path offered by X due to:
  - cost, policy (don't route via competitors AS), loop prevention reasons.
- If W selects path advertised by X, then:

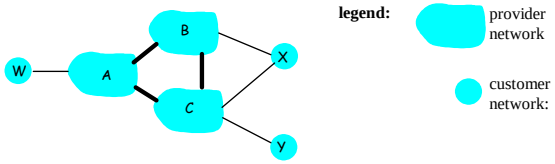
  Path (W,Z) = Path (W,X) + Path (X,Z)
- Note: X can control incoming traffic by controlling its route advertisements to peers:
  - e.g., don't want to route traffic to Z? => don't advertise any routes to Z!

BGP: controlling who routes to you



- A,B,C are provider networks
- x,w,y are customers (of the provider networks)
- x is dual-homed: attached to two networks
  - x does not want to route from B (via x) to C
  - .. so x will not advertise to B a route to C

BGP: controlling who routes to you



- A advertises to B the path Aw
- B advertises to x the path BAw
- Should B advertise to C the path BAw?
  - No way! B gets no "revenue" for routing CBAw since neither w nor C are B's customers
  - B wants to force C to route to w via A
  - B wants to route *only* to/from its customers!

BGP operation

Q: What does a BGP router do?

- Receiving and filtering route advertisements from directly attached neighbors
- Route selection
  - To route to destination X, which path will be taken? (of several advertised)
- Sending route advertisements to neighbors

BGP messages

- BGP messages exchanged using TCP.
- BGP messages:
  - OPEN: opens TCP connection to peer and authenticates sender
  - UPDATE: advertises new path (or withdraws old)
  - KEEPALIVE keeps connection alive in absence of UPDATES; also ACKs OPEN request
  - NOTIFICATION: reports errors in previous msg; also used to close connection

Why different Intra- and Inter-AS routing ?

Policy:

- Inter-AS: admin wants control over how its traffic is routed, who routes through its net.
- Intra-AS: single admin, so no policy decisions needed
- Examples of policy decisions: traffic starting or ending at IBM should not pass through Microsoft; Never put Iraq on a route starting at the Pentagon; Only use Bangladesh if no other route

Scale:

- hierarchical routing saves table size, reduced update traffic

Performance:

- Intra-AS: can focus on performance
- Inter-AS: policy may dominate over performance