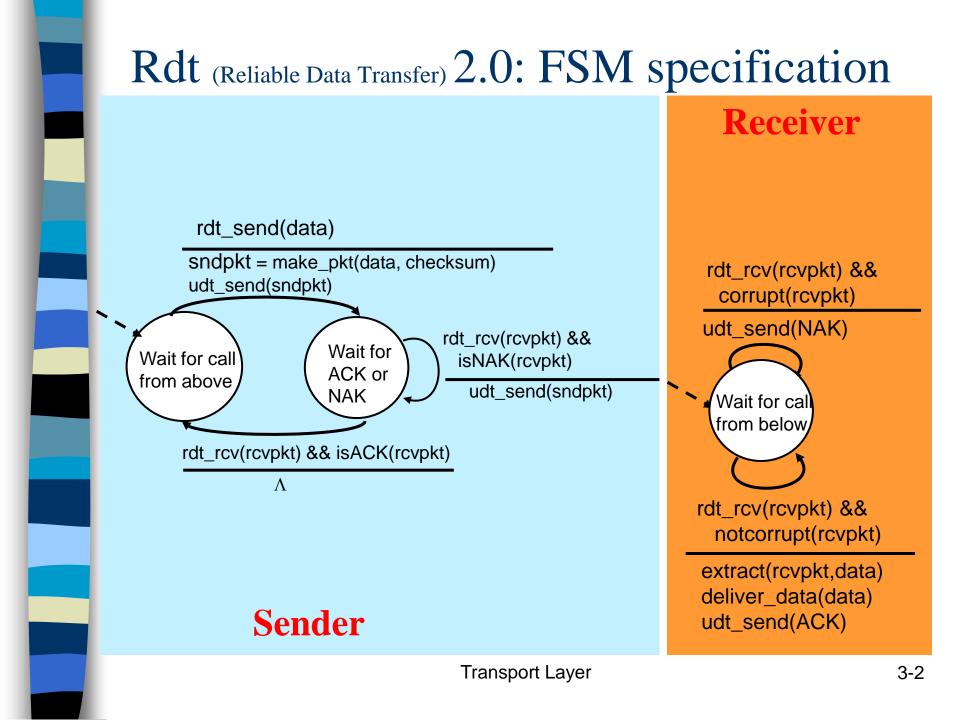
Protocoles et Interconnexions

Course Overview and Introduction

Dario Vieira Department of Computer Science EFREI



rdt (Reliable Data Transfer) 3.0: channels with errors and loss

New assumption:

underlying channel can time for ACK also lose packets (data or ACKs)

- checksum, seq. #, ACKs, retransmissions will be of lost): help, but not enough

Approach: sender waits

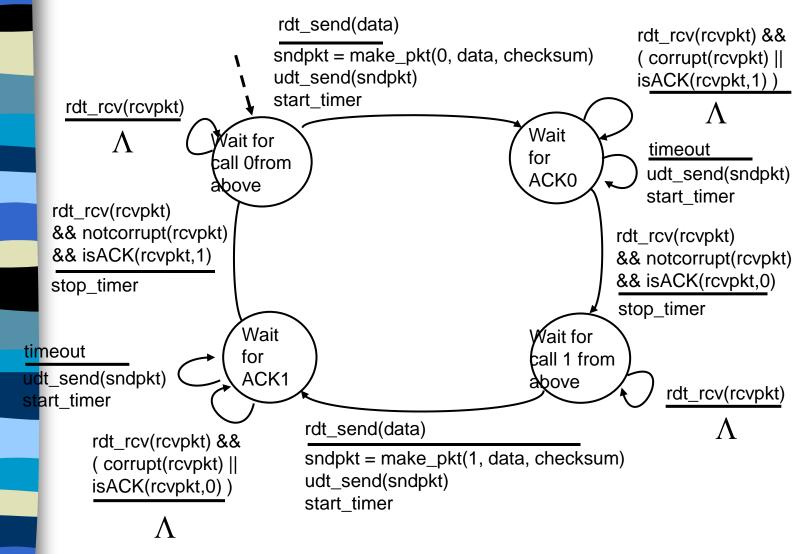
"reasonable" amount of

retransmits if no ACK received in this time

if pkt (or ACK) just delayed (not

- retransmission will be duplicate, but use of seq. #'s already handles this
- receiver must specify seq # of pkt being ACKed
- requires countdown timer

rat3.0 sender



Chapter 4: Network Layer

4. 1 Introduction 4.2 Virtual circuit and datagram networks 4.3 What's inside a router 4.4 IP: Internet Protocol Datagram format IPv4 addressing ICMP

– IPv6

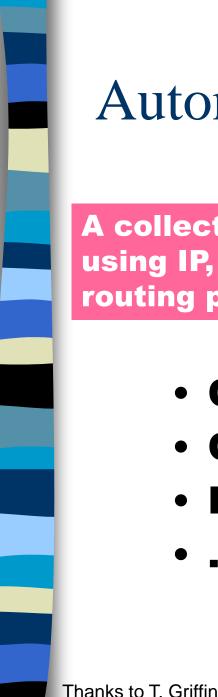
4.5 Routing algorithms

- Link state
- Distance Vector
- Hierarchical routing

4.6 Routing in the Internet

- RIP
- OSPF
- BGP

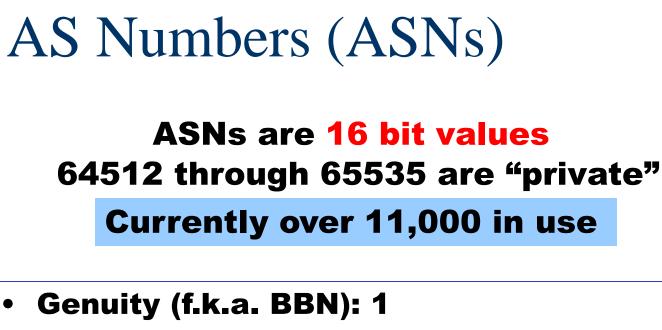
4.7 Broadcast and multicast routing



Autonomous Systems (ASes)

A collection of physical networks glued together using IP, that have a unified administrative routing policy.

- Campus networks
- Corporate networks
- ISP Internal networks

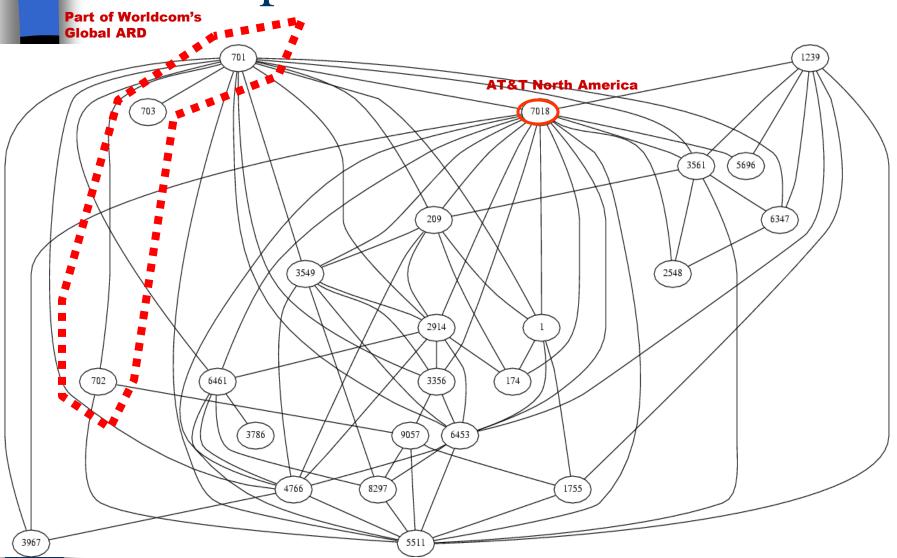


- MIT: 3
- Harvard: 11
- UC San Diego: 7377
- AT&T: 7018, 6341, 5074, ...
- UUNET: 701, 702, 284, 12199, ...
- Sprint: 1239, 1240, 6211, 6242, ...

ASNs represent units of routing policy

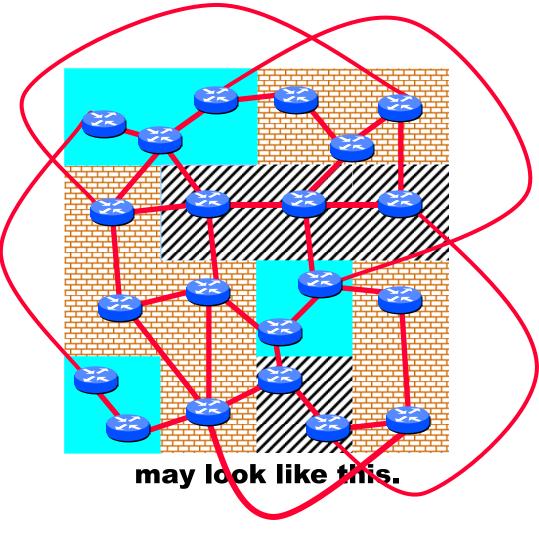
^{...}

AS Graphs Can Be Fun

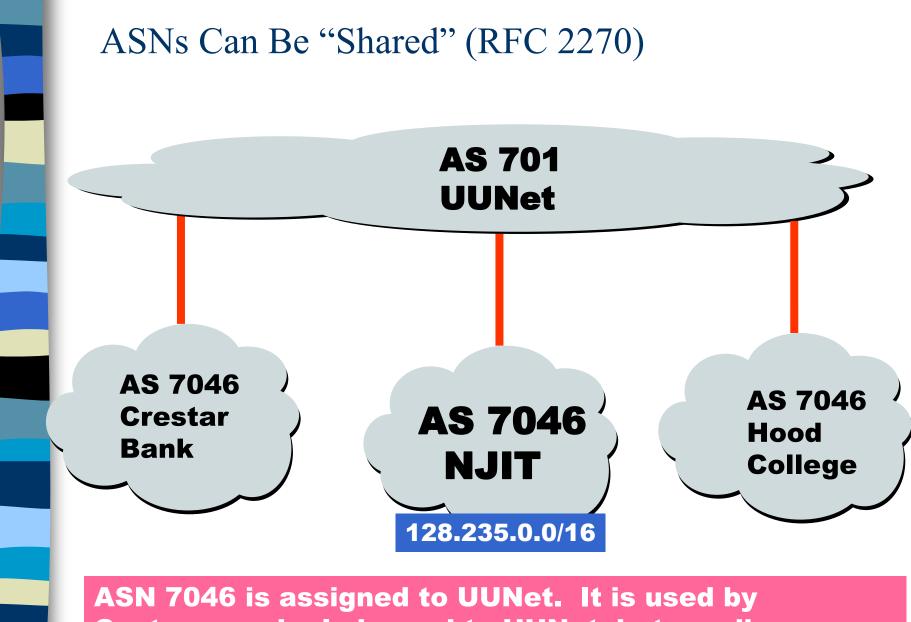


The <u>subgraph</u> showing all ASes that have more than 100 neighbors in full graph of 11,158 nodes. July 6, 2001. Point of view: AT&T route-server

AS Graph != Internet Topology

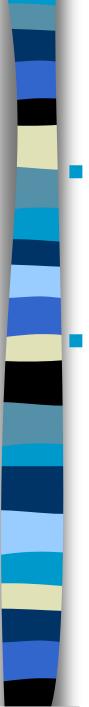


Reality may be closer to this...



Customers single homed to UUNet, but needing BGP for some reason (load balancing, etc..) [RFC 2270]

Thanks to T. Griffin



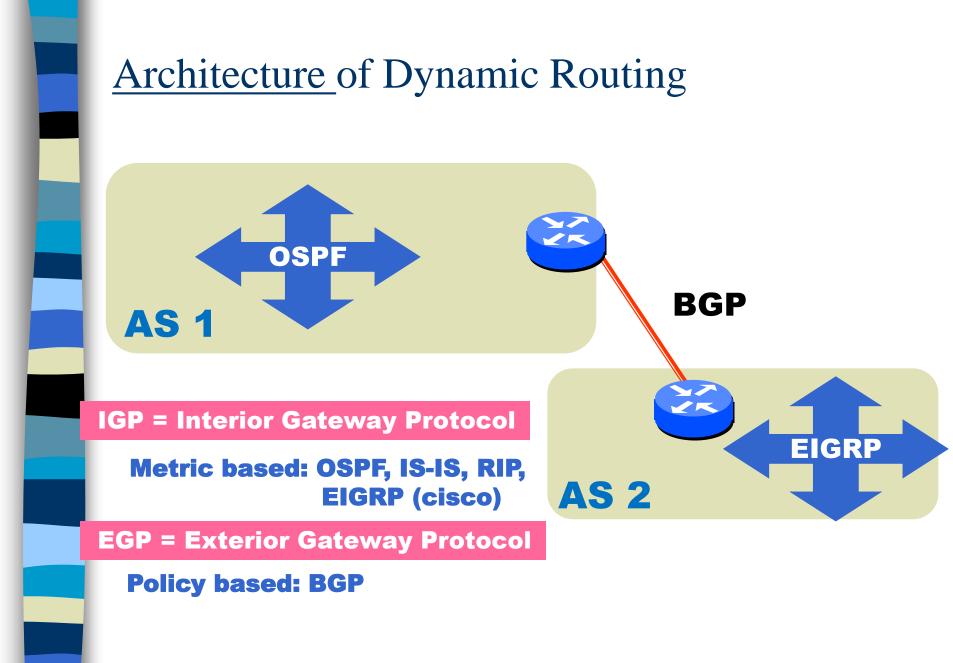
Hierarchical Routing

Aggregate routers into regions, "Autonomous Systems" (AS)

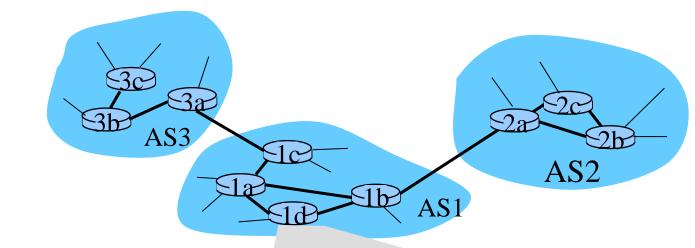
Routers in same AS run same routing protocol

- Intra-AS routing protocol
 - IGP: Interior Gateway Protocol
 - e.g. , RIP

- Gateway router
- At "edge" of its own AS
- Direct link to router in another AS
- All ASes run the same Inter-AS routing protocol
 - EGP: Exterior Gateway
 Protocol
 - e.g., Border Gateway Protocol (BGP)
- Routers in different AS can
 run different intra-AS routing
 protocol
 Network Layer

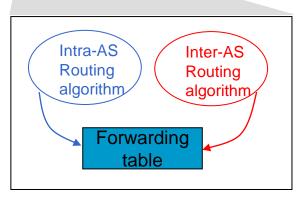


Interconnected ASes



Forwarding table configured by both intra- and inter-AS routing algorithm

- intra-AS sets entries for internal dests
- inter-AS & intra-As sets
 entries for <u>external dests</u>



Inter-AS tasks

Suppose router in AS1 receives datagram destined outside of AS1

 router should forward packet to gateway router, but which one?

-3a-

1c

-1a-

AS1

3c

AS3

-3b

Job of inter-AS routing!

AS1 must

- Learn which dests are reachable through AS2, which through AS3
- Propagate this reachability info to all routers in AS1

2c

AS2

other networks

h

other

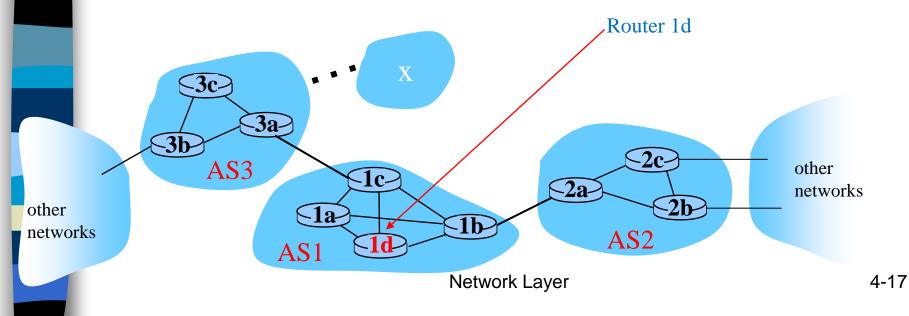
networks

Network Layer

Example: Setting forwarding table in router 1d

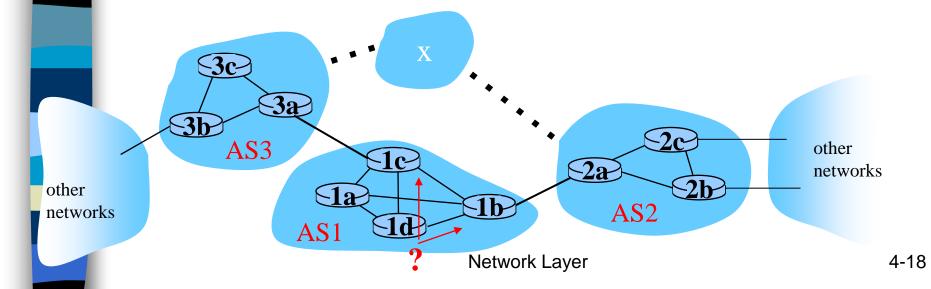
Suppose AS1 learns (via inter-AS protocol) that subnet *x* reachable via AS3 (gateway 1c) but not via AS2.

- inter-AS protocol propagates reachability info to all internal routers
- Router 1d determines from intra-AS routing info that its interface / is on the least cost path to 1c.
 - installs forwarding table entry (x, l)



Example: Choosing among multiple ASes

- Now suppose AS1 learns from inter-AS protocol that subnet *x* is reachable from AS3 and from AS2.
- To configure forwarding table, router 1d must determine which gateway it should forward packets towards for dest x
 - This is also job of inter-AS routing protocol!



Chapter 4: Network Layer

4. 1 Introduction 4.2 Virtual circuit and datagram networks 4.3 What's inside a router 4.4 IP: Internet Protocol Datagram format IPv4 addressing ICMP

– IPv6

4.5 Routing algorithms

- Link state
- Distance Vector
- Hierarchical routing

4.6 Routing in the Internet

- RIP
- OSPF
- BGP

4.7 Broadcast and multicast routing

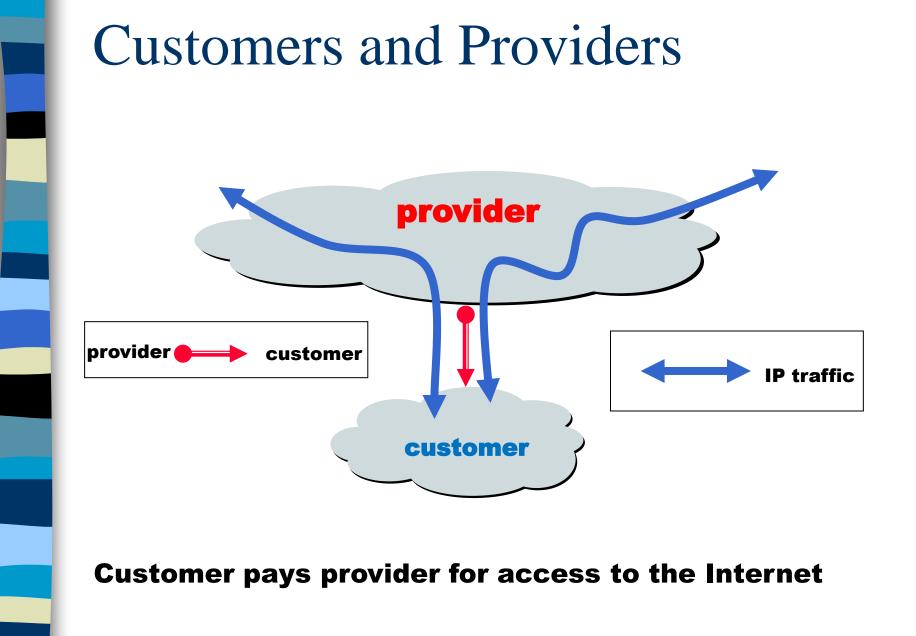


Intra-AS Routing

- 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)

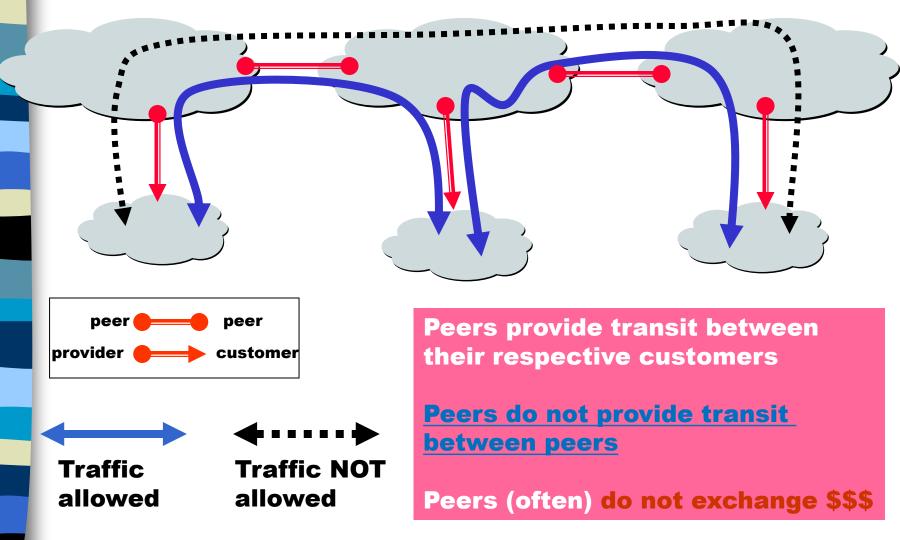
Relationships Between Networks

Thanks to T. Griffin



Thanks to T. Griffin

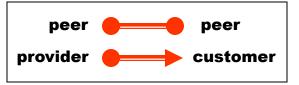
The "Peering" Relationship



Thanks T. Griffin

Peering Provides Shortcuts

Peering also allows connectivity between the customers of "Tier 1" providers.



Peering Wars



- Reduces upstream transit costs
- Can increase end-to-end performance
 - May be the only way to connect your customers to some part of the Internet ("Tier 1")

Don't Peer

- You would rather have customers
- Peers are usually your competition
- Peering relationships may require periodic renegotiation

Peering struggles are by far the most contentious issues in the ISP world!

Peering agreements are often confidential.

Chapter 4: Network Layer

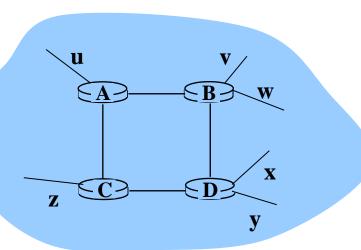
- 4. 1 Introduction
 - 4.3 What's inside a router
 - 4.4 IP: Internet Protocol
 - Datagram format
 - IPv4 addressing

- 4.5 Routing algorithms
 - Link state
 - Distance Vector
 - Hierarchical routing
- 4.6 Routing in the Internet
 - RIP
 - OSPF
 - BGP
- 4.7 Broadcast and multicast routing

RIP (Routing Information Protocol)

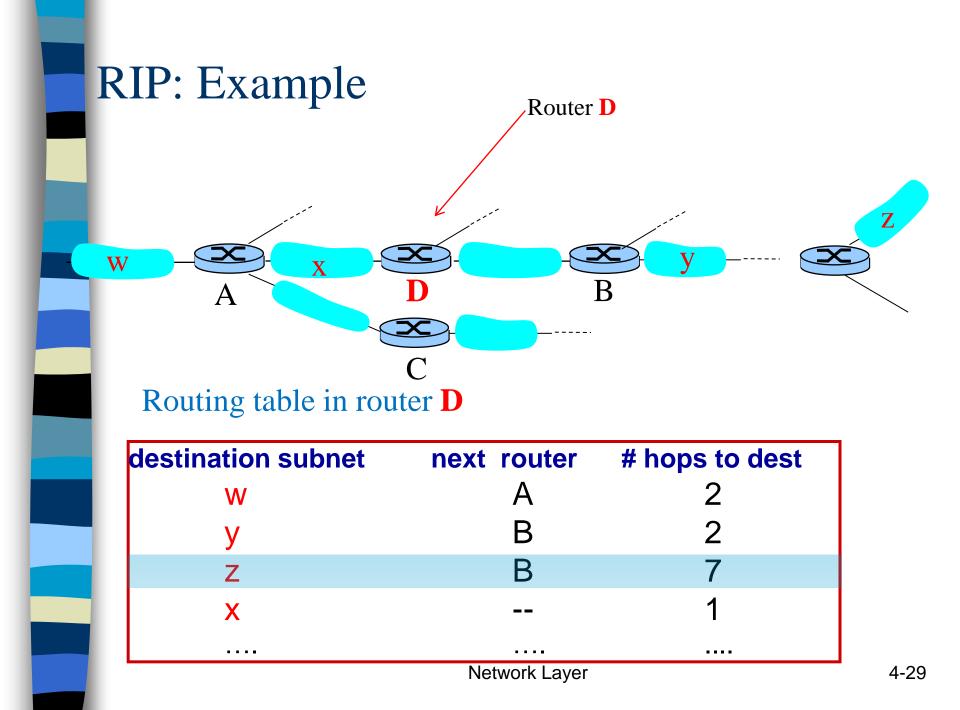
Included in BSD-UNIX distribution in 1982 Distance vector algorithm

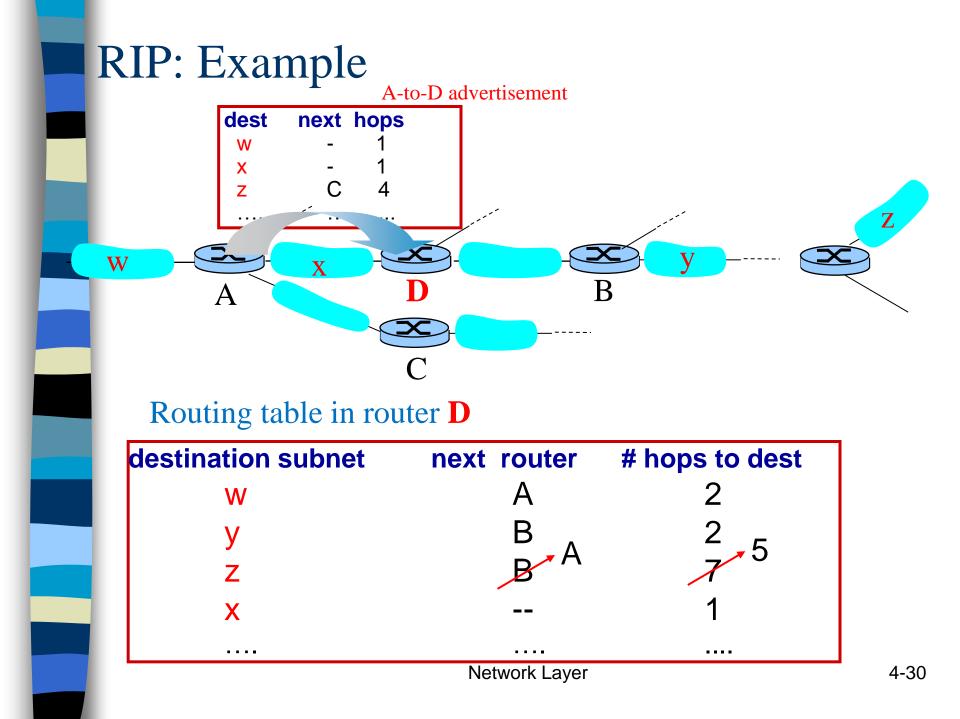
- distance metric: # hops (max = 15 hops), each link has cost 1
- DVs exchanged with neighbors every 30 sec in response message (advertisement)
- each advertisement: list of up to 25 destination subnets (in IP addressing sense)



from router A to destination subnets

<u>subnet</u>	<u>hops</u>
u	1
V	2
W	2
Х	3
У	3
Z	2





RIP: Link Failure and Recovery

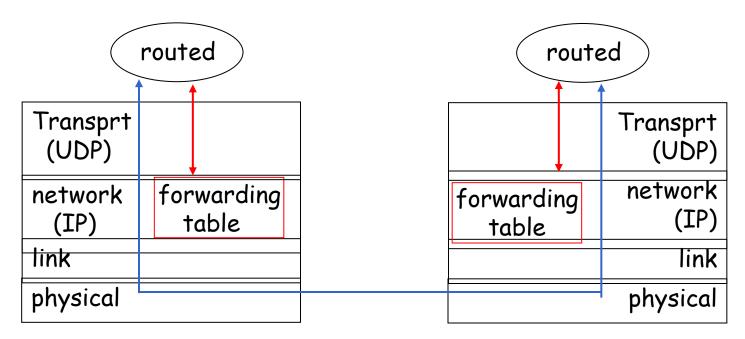
If no advertisement heard after 180 sec --> neighbor/link declared dead

- Routes via 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
- Poison reverse used to prevent ping-pong loops (infinite distance = 16 hops)



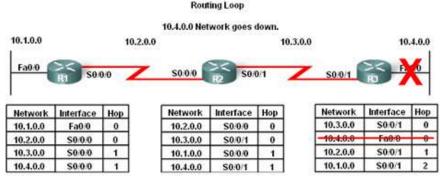
RIP Table Processing

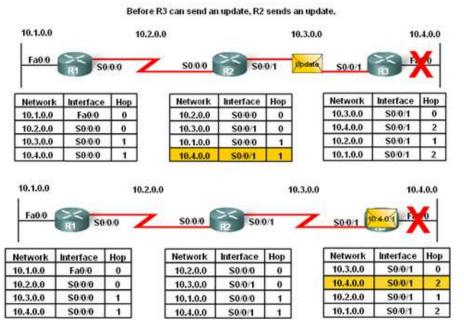
- RIP routing tables managed by application-level process called route-d (daemon)
- Advertisements sent in UDP packets, periodically repeated



Routing loops are

- A condition in which a packet is continuously transmitted within a series of routers
- without ever reaching its destination.





Count to Infinity

 This is a routing loop whereby packets bounce infinitely around a network

Count to Infinity

Each round of updates continues to increase hop count.



Network	Interface	Нор
10.1.0.0	Fa0/0	0
10.2.0.0	S0/0/0	0
10.3.0.0	S0/0/0	1
10.4.0.0	S0/0/0	24

Network	Interface	Нор
10.2.0.0	S0/0/0	0
10.3.0.0	S0/0/1	0
10.1.0.0	S0/0/0	1
10.4.0.0	S0/0/1	23

Network	Interface	Нор
10.3.0.0	S0/0/1	0
10.4.0.0	S0/0/1	22
10.2.0.0	S0/0/1	1
10.1.0.0	S0/0/1	2

Setting a maximum

 Distance Vector routing protocols set a specified metric value to indicate infinity

- Once a router "counts to infinity" it marks the route as unreachable



10.4.0.0 is unreachable. Hop count is 16.

Network	Interface	Нор	Network	Interface	Нор	Network	Interface
10.1.0.0	Fa0/0	0	10.2.0.0	S0/0/0	0	10.3.0.0	S0/0/1
10.2.0.0	S0/0/0	0	10.3.0.0	S0/0/1	0	10.4.0.0	S0/0/1
10.3.0.0	S0/0/0	1	10.1.0.0	S0/0/0	1	10.2.0.0	S0/0/1
10.4.0.0	S0/0/0	16	10.4.0.0	S0/0/1	16	10.1.0.0	S0/0/1

The Split Horizon Rule is used to prevent routing loops

Split Horizon rule:

 A router should not advertise a network through the interface from which the update came

Split Horizon Rule for 10.4.0.0

R2 only advertises 10.3.0.0 and 10.4.0.0 to R1. R2 only advertises 10.2.0.0 and 10.1.0.0 to R3.

R1 only advertises 10.1.0.0 to R2.

R3 only advertises 10.4.0.0 to R2.

10.1.0.0		-	10.2.0.0	\sim	\times	10	0.3.0.0		10.4	.0.0
Fa0/0	Update S0/	0/0	Ż	S0/0 Upda	ate Update	0/1	Z	50/0/1)/0
Network	Interface	Нор	1>	Network	Interface	Нор	\geq	Network	Interface	Нор
	Interface Fa0/0	Hop 0	\geq	Network 10.2.0.0	Interface S0/0/0	Hop 0	\geq			Hop 0
10.1.0.0			R		The second states of	N	R	Network	Interface	0
Network 10.1.0.0 10.2.0.0 10.3.0.0	Fa0/0	0	R	10.2.0.0	S0/0/0	Ò	\geq	Network 10.3.0.0	Interface S0/0/1	

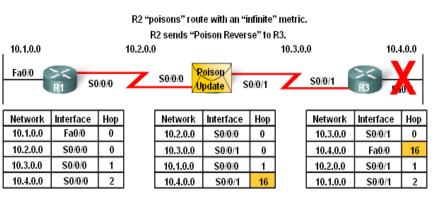


Split horizon with poison reverse

- A router learns of an unreachable route through an interface
- Advertise it as unreachable back through the same interface

Poison Reverse Network 10.4.0.0 goes down. R3 "poisons" route with an "infinite" metric. R3 sends triggered Poison Update to R2. 10.1.0.0 10.2.0.0 10.3.0.0 10.4.0.0 Fa0/0 Poisop S0/0/0 S0/0/1 S0/0/0 \$0/0/1 pdate Network Interface Hop Network Interface Hop Interface Network Нор 10.1.0.0 Fa0/0 0 S0/0/0 10.2.0.0 0 10.3.0.0 S0/0/1 0 10.2.0.0 S0/0/0 0 10.3.0.0 S0/0/1 0 10.4.0.0 Fa0/0 16 10.3.0.0 S0/0/0 1 10.1.0.0 S0/0/0 1 S0/0/1 10.2.0.0 1 S0/0/0 10.4.0.0 2 S0/0/1 10.4.0.0 1 10.1.0.0 S0/0/1

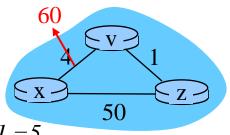
Poison Reverse



Distance Vector: link cost changes

Link cost changes:

- Good news travels fast
- Bad news travels slow "count to infinity" problem!
- ✤ 44 iterations before algorithm stabilizes



-Bef: $D_v(x) = 4$, $D_v(z) = 1$, $D_z(v) = 1$, $D_z(x) = 4+1=5$

- -Aft: $D_{v}(x) = min\{60, c(v,z) + D_{z}(x)\} = 6$
- -Next: $D_z(x) = ... = c(z, v) + D_v(x) = 1 + 6 = 7$
- Then: $D_v(x) = c(v,z) + D_z(x) = 1 + 7 = 8 \iff$

Poisoned reverse:

- If z routes through v to get to x :
 - z tells v: $D_z(x) = \infty$ (so v won't route to x via z)

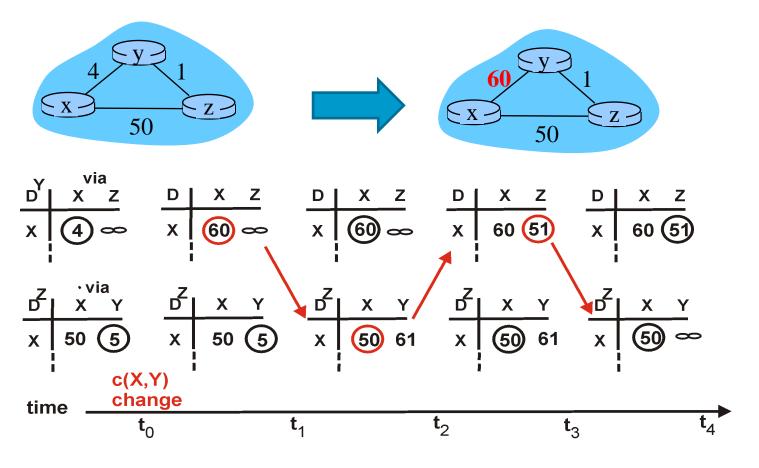
$$\rightarrow D_{\nu}(x) = 60 \rightarrow D_{z}(x) = 50 \rightarrow D_{\nu}(x) = 51$$

Will this completely solve count to infinity problem?

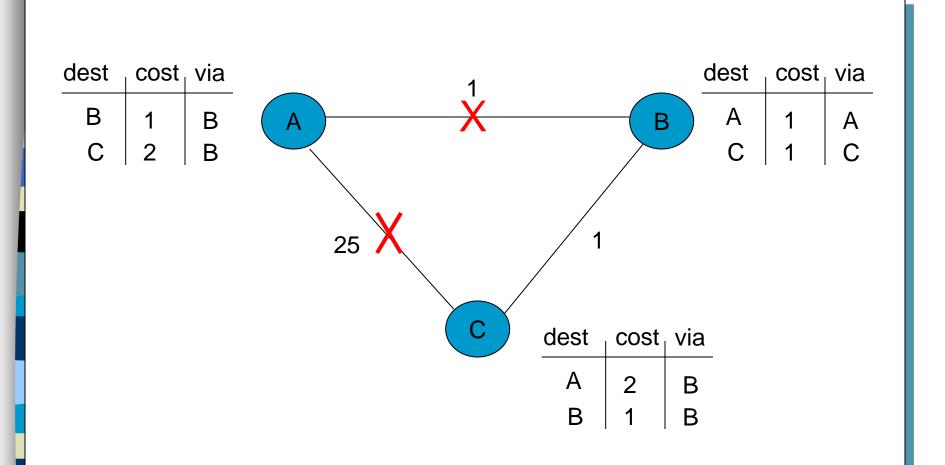
Distance Vector: link cost changes

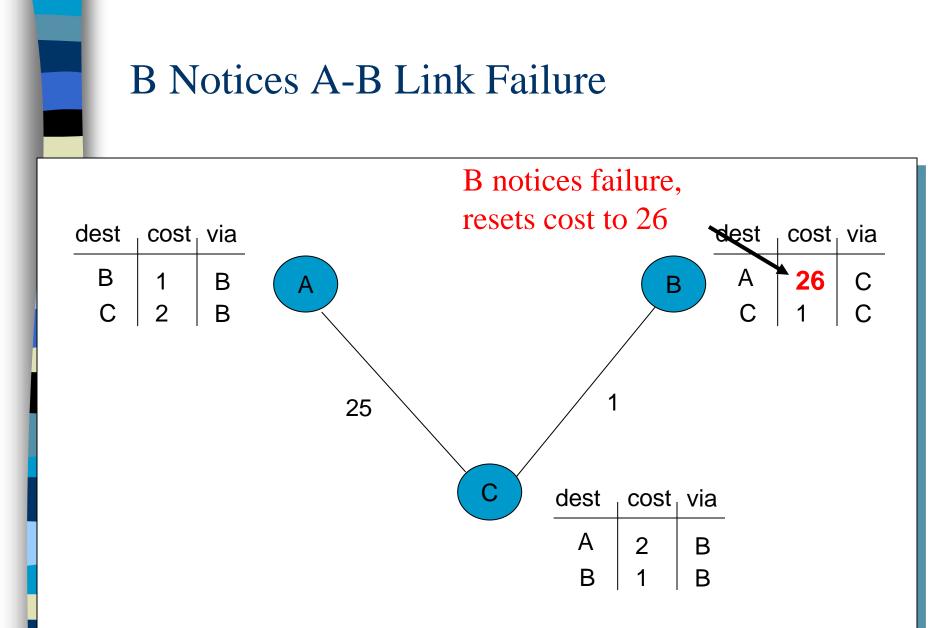
Poisoned reverse:

- If z routes through y to get to x:
 - z tells y: $D_z(x) = \infty$ (so y won't route to x via z)

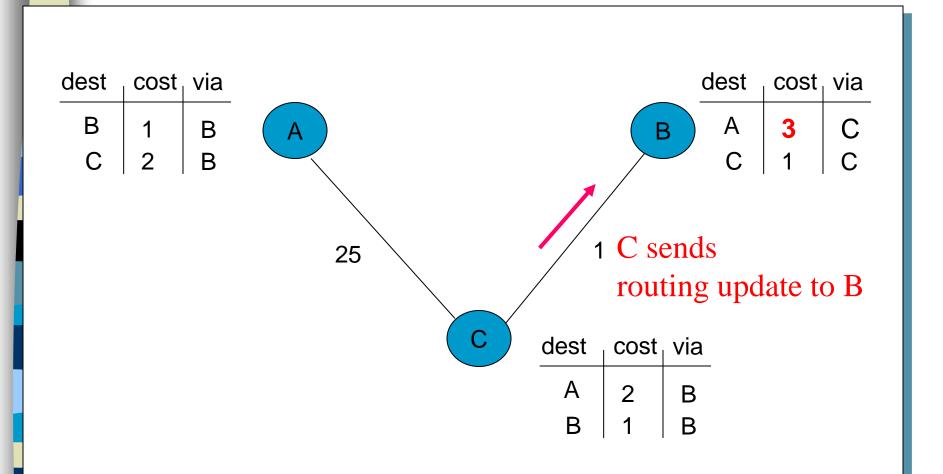


Link Failure Causes "Counting to Infinity" Effect

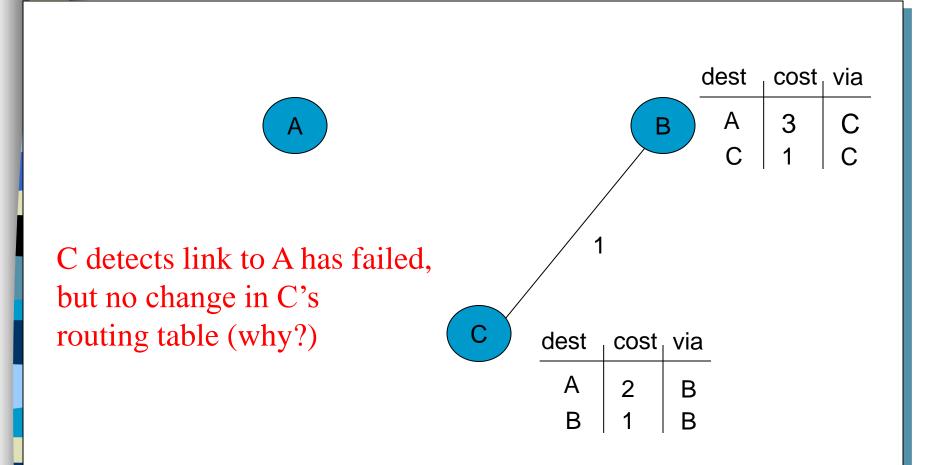




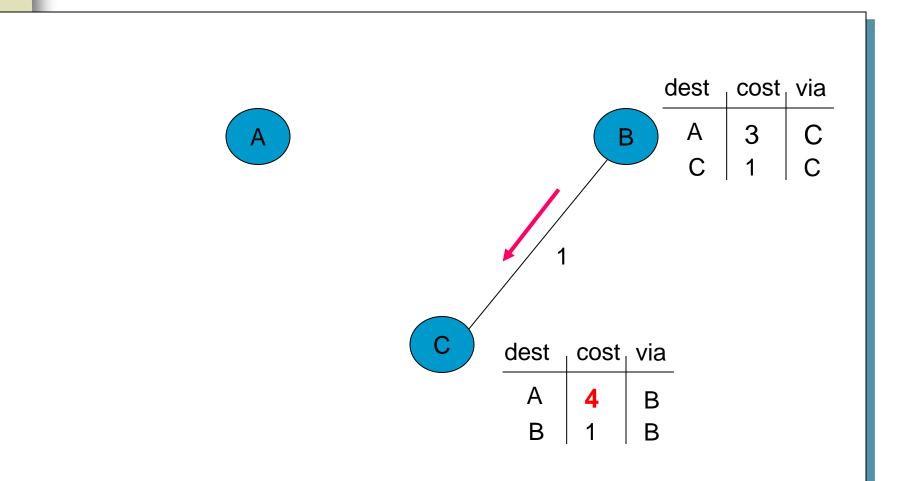
C Sends Dist. Vector to B



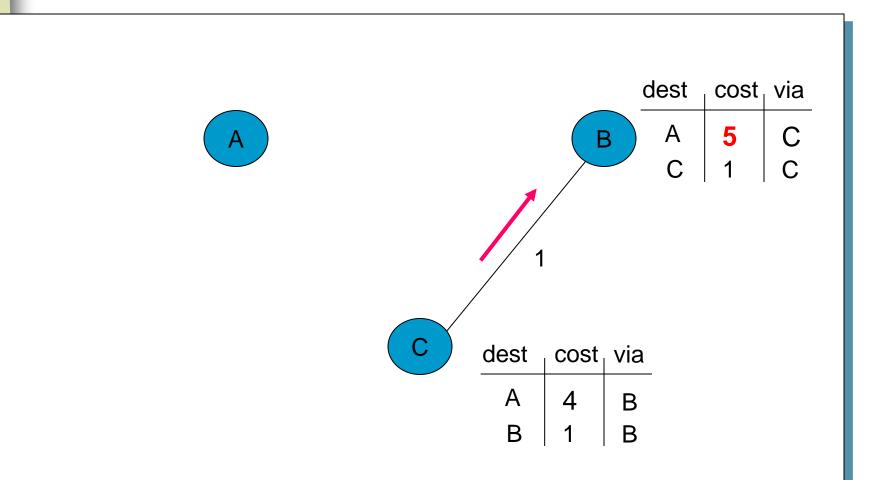
A-C Link Fails



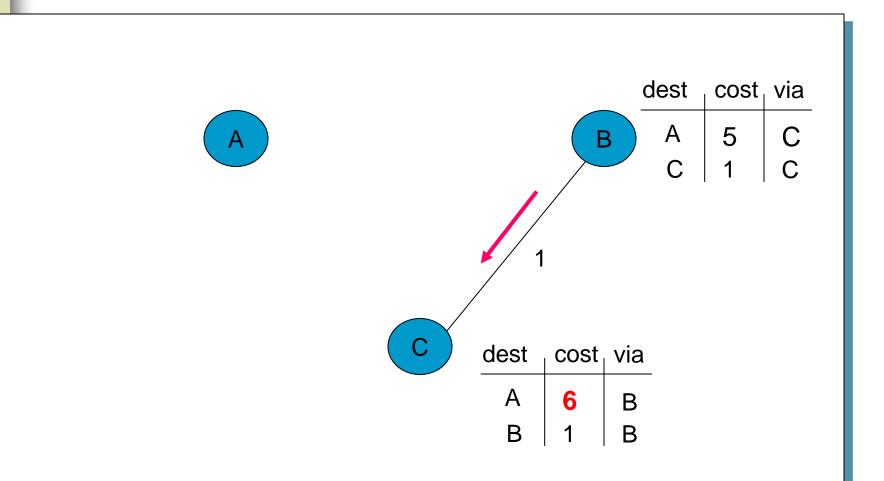
Now, B and C Count to Infinity



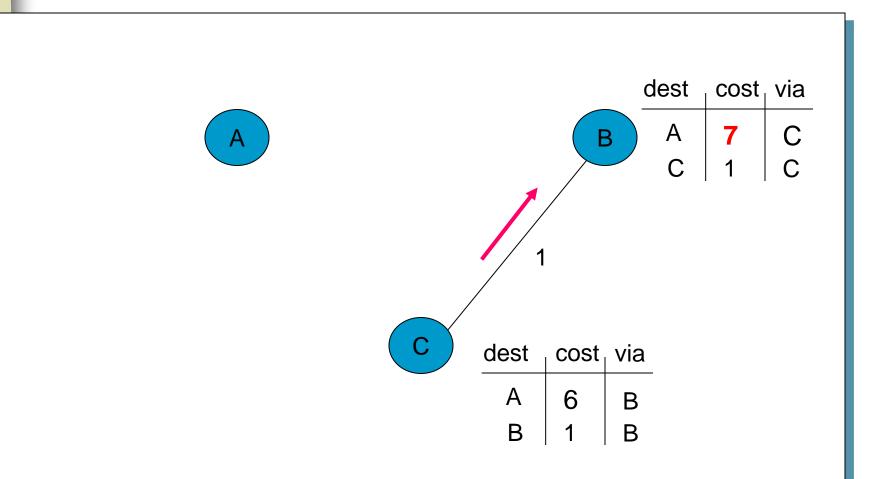
B and C Count to Infinity (2)



B and C Count to Infinity (3)



B and C Count to Infinity (4)





Some "Solutions"

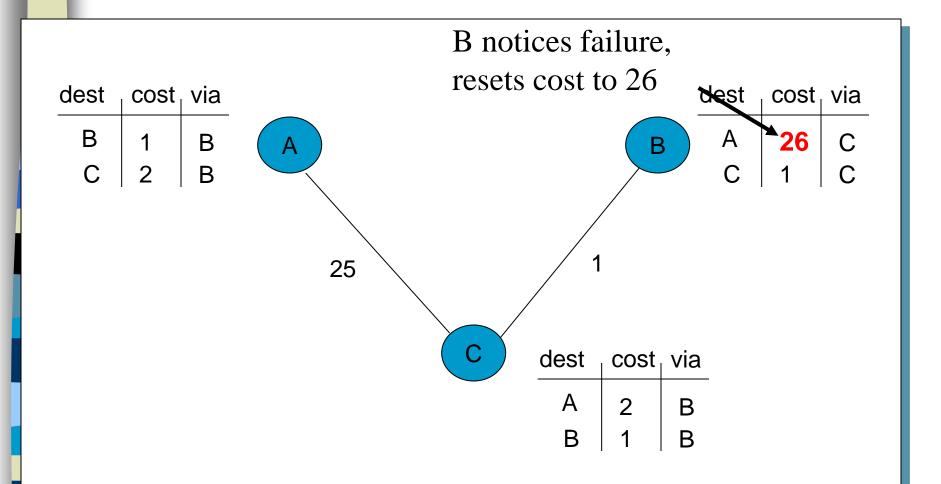
Split horizon

- C does not advertise route to B when it sends its distance vector
- A router should not advertise a network through the interface from which the update came

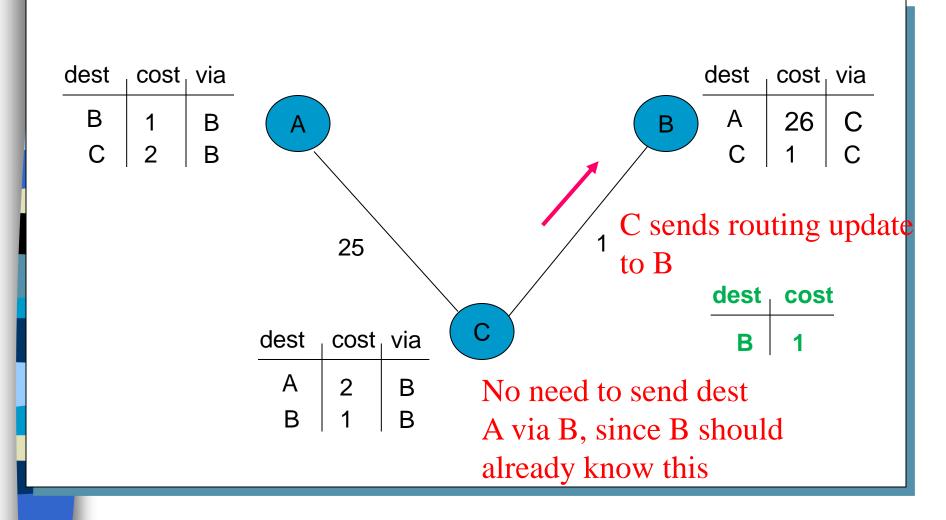
Poisoned reverse

- C advertises route to B with infinite distance in its distance vector

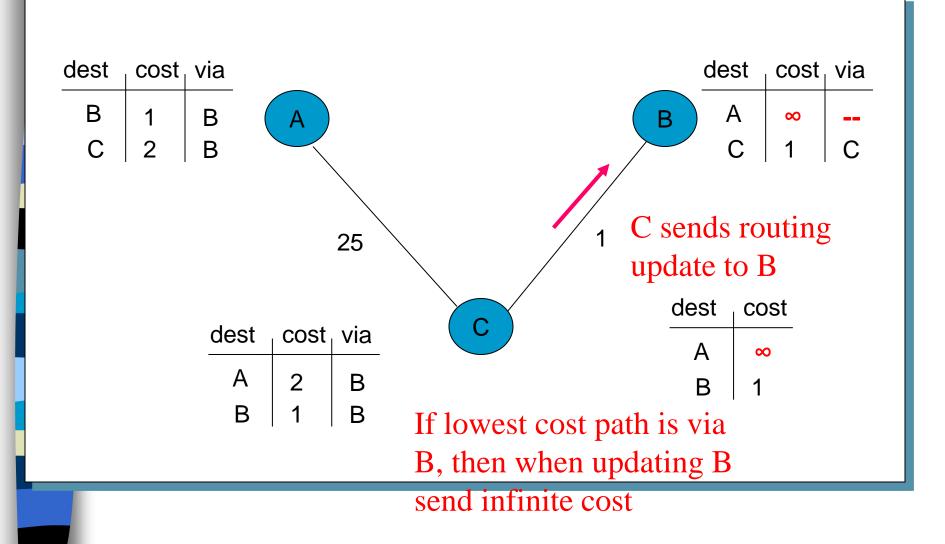
B Notices A-B Link Failure



Split Horizon



Split Horizon With Poisoned Reverse





PART II

Implementing Intra-Network Relationships with OSPF

Protocoles et Interconnexions

Course Overview and Introduction

Dario Vieira Department of Computer Science EFREI

Chapter 4: Network Layer

4. 1 Introduction 4.2 Virtual circuit and datagram networks 4.3 What's inside a router 4.4 IP: Internet Protocol Datagram format IPv4 addressing <lu> ICMP

– IPv6

4.5 Routing algorithms

- Link state
- Distance Vector
- Hierarchical routing

4.6 Routing in the Internet

- RIP
- OSPF
- BGP

4.7 Broadcast and multicast routing

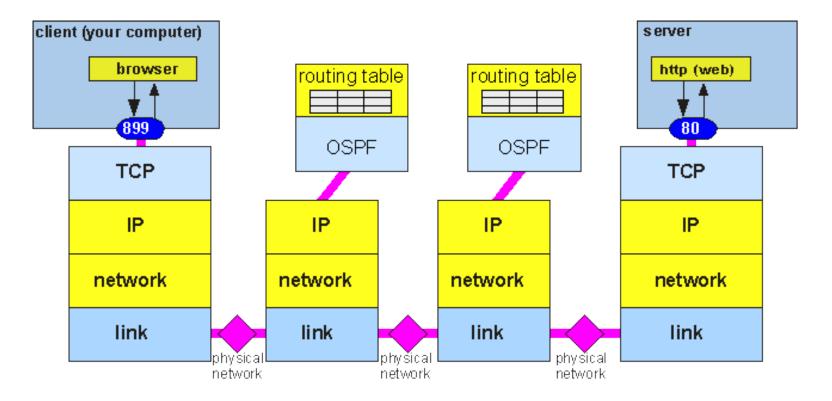


OSPF (Open Shortest Path First)

- "Open": publicly available
 - Uses Link State algorithm
 - LS packet dissemination
 - 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 over IP



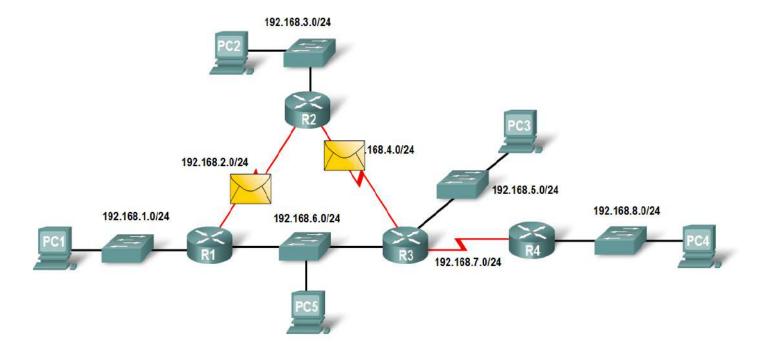
OSPF "advanced" features (not in RIP)

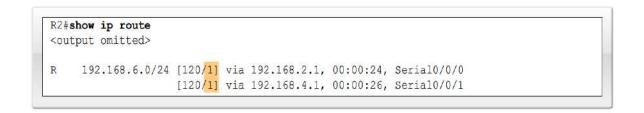
Security

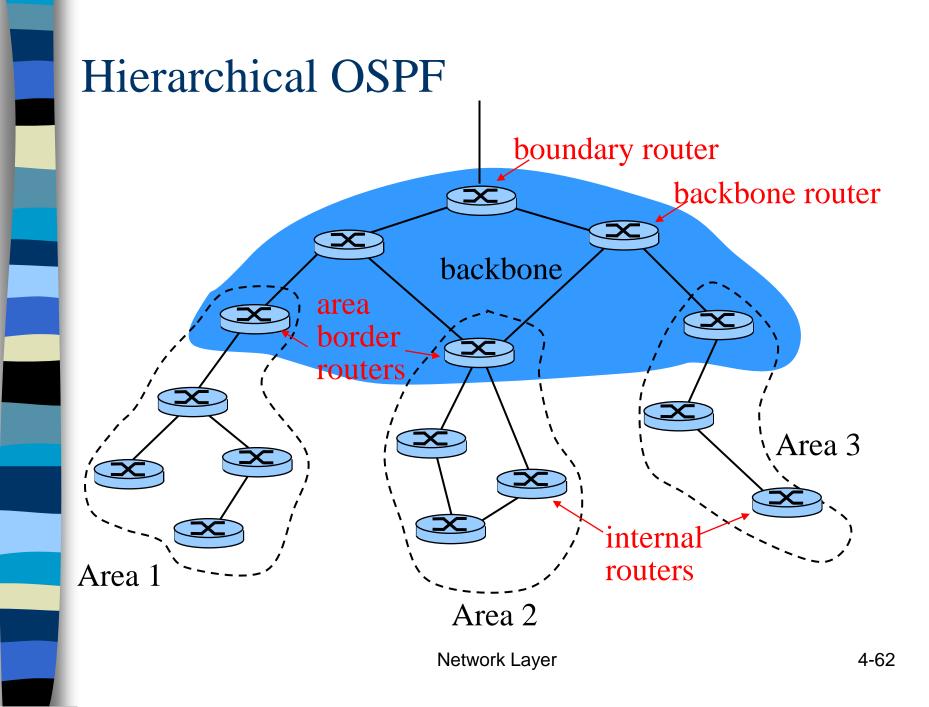
- all OSPF messages authenticated (to prevent malicious intrusion)
- Multiple same-cost paths allowed
 - only one path in RIP
- For each link, multiple cost metrics for different TOS (Type Of Service)
 - e.g., satellite link cost set "low" for best effort ToS; high for real time ToS
- Integrated uni-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

Network Layer

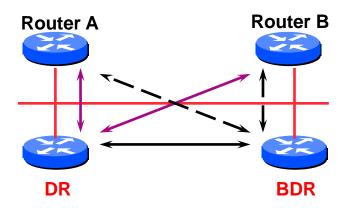
Multiple same-cost paths







Broadcast Network: Designated Router (DR)

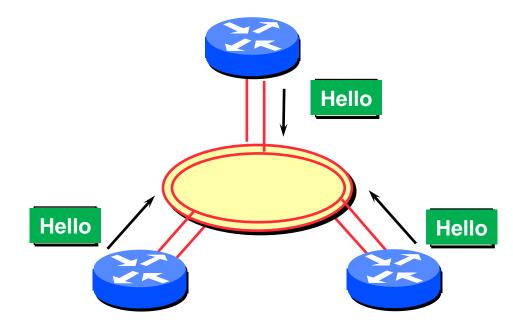


- Adjacencies only formed with DR (Routeurs Désignés) and (Backup Designated router) BDR
- All router exchanges routing information with the DR and BDR
- DR updates the database of all its neighbours
- This scales: 2n problem rather than having an nsquared problem

OSPF: How it works

Hello Protocol

- Responsible for establishing and maintaining neighbour relationships
- Elects Designated Router on broadcast networks

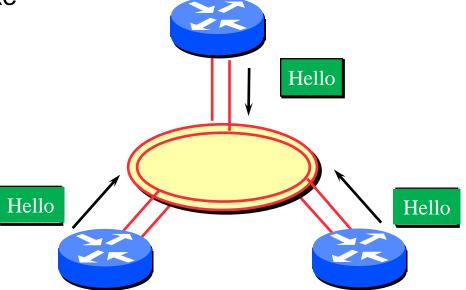




The Hello Packet

Contains information like

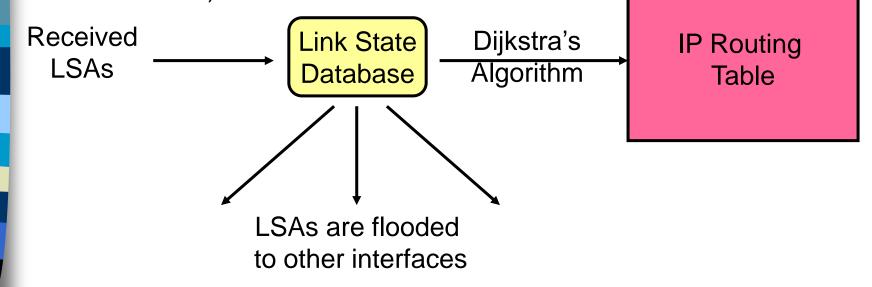
- Router Priority
- Hello Interval
- Router dead interval
- Network mask
- List of neighbors



Hello Packets sent periodically on all OSPF enabled interfaces

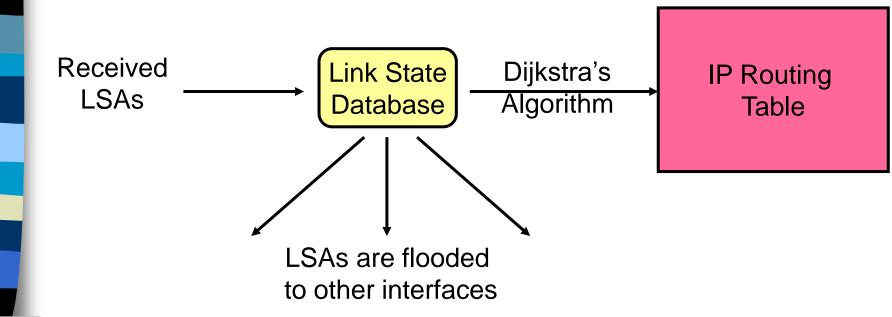
OSPF: How it works

- 1. Each router establishes a relationship ("adjacency") with its neighbors
- 2. Each router constructs a *Link State Packet* (*LSP*) which are distributed to all neighbours
 - Using an link state advertisements (LSAs)
 - LSA = (link id, state of the link, cost, neighbors of the link)

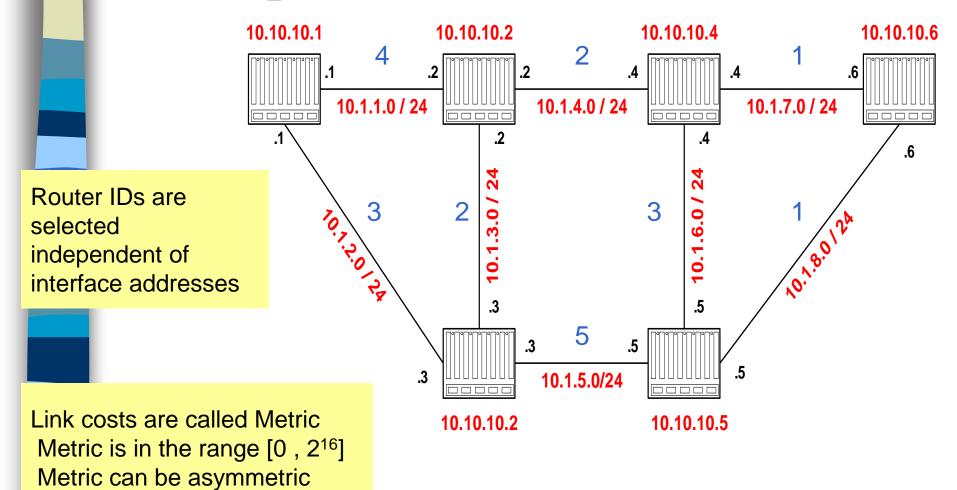


OSPF: How it works

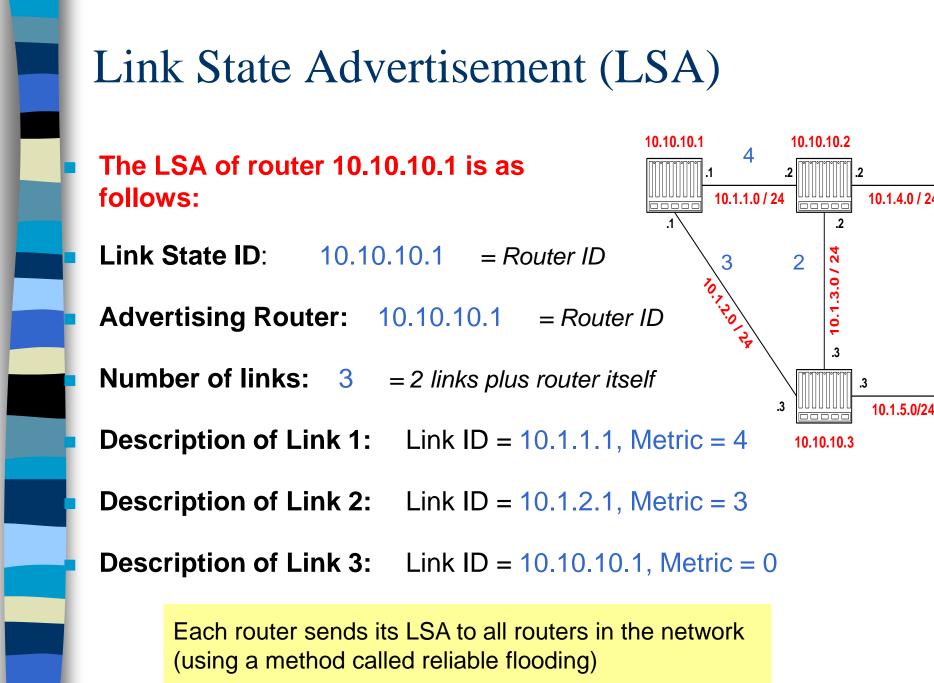
- 3. Each router maintains a database of all received LSAs (*topological database* or *link state database*), which describes the network has a graph with weighted edges
- 4. Each router uses its link state database to run a shortest path algorithm (Dijikstra's algorithm) to produce the shortest path to each network



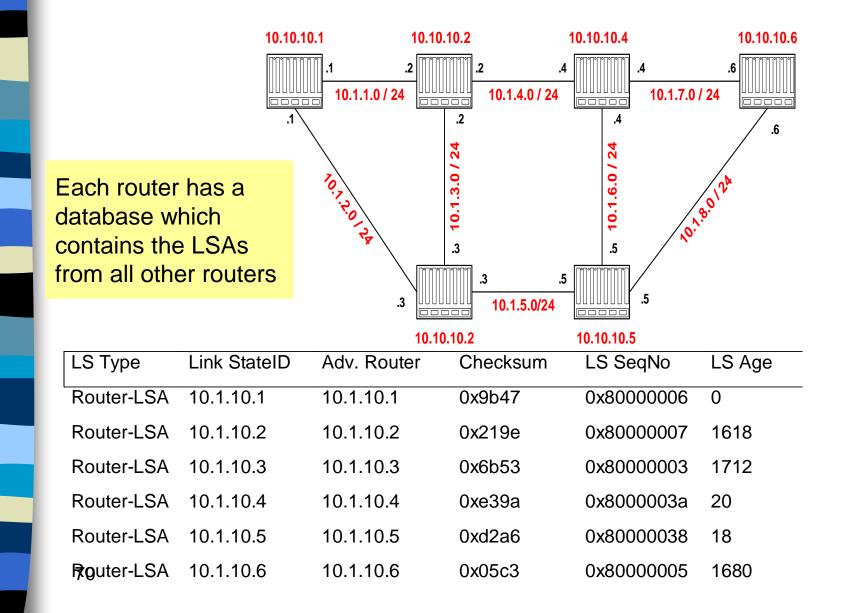
Example Network



68



Network and Link State Database

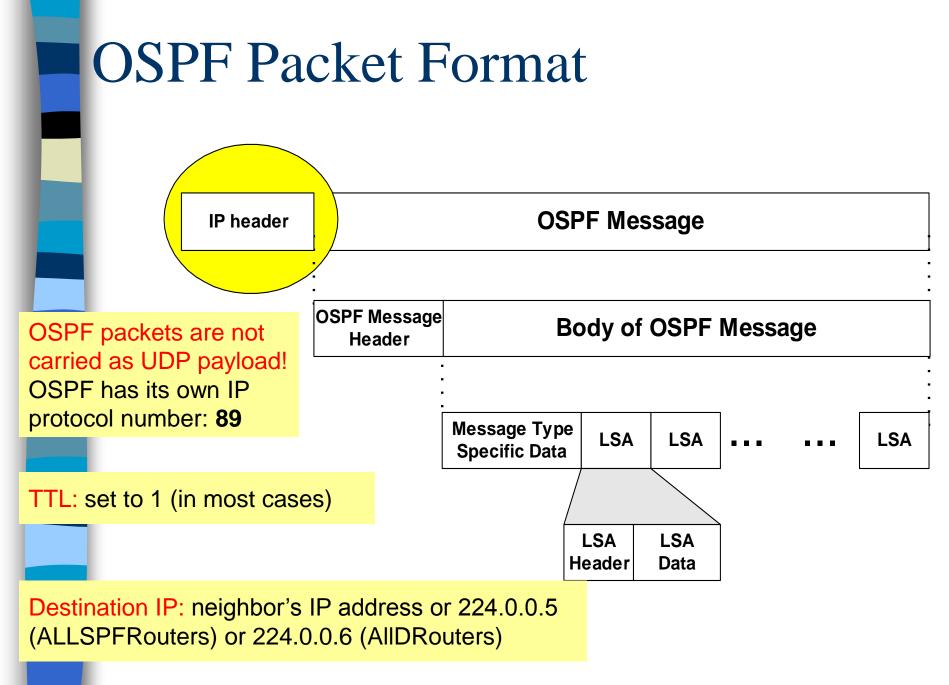


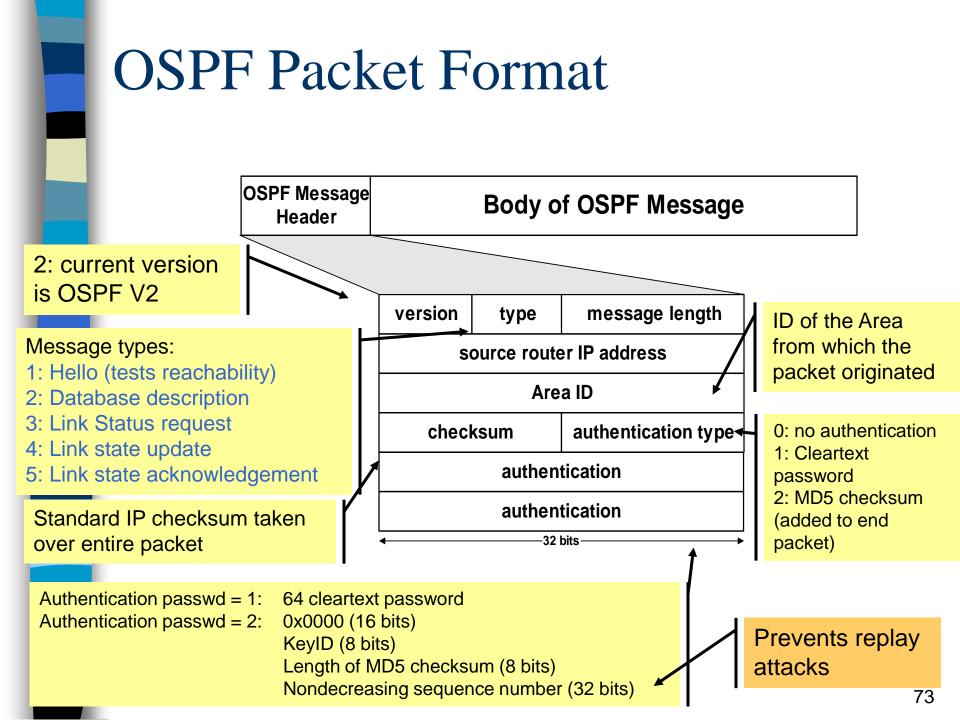


71

Link State Database

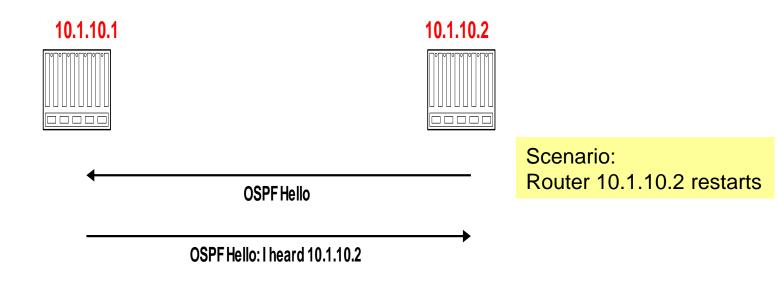
- The collection of all LSAs is called the link-state database
- Each router has an identical link-state database
 - Useful for debugging: Each router has a complete description of the network
- If neighboring routers discover each other for the first time, they will exchange their link-state databases
- The link-state databases are synchronized using reliable flooding



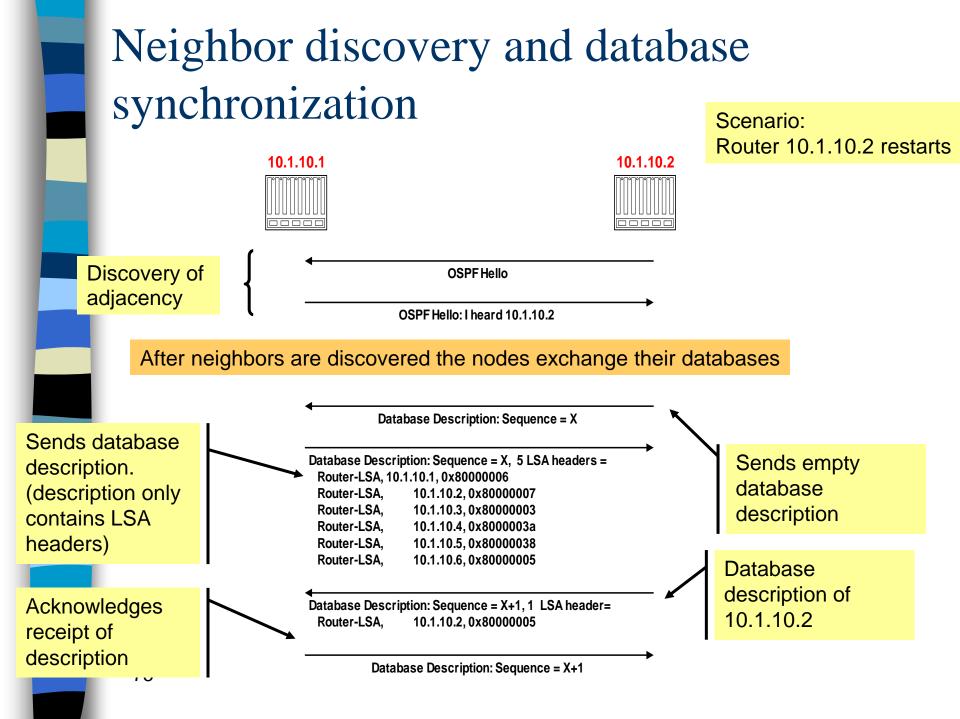


Discovery of Neighbors

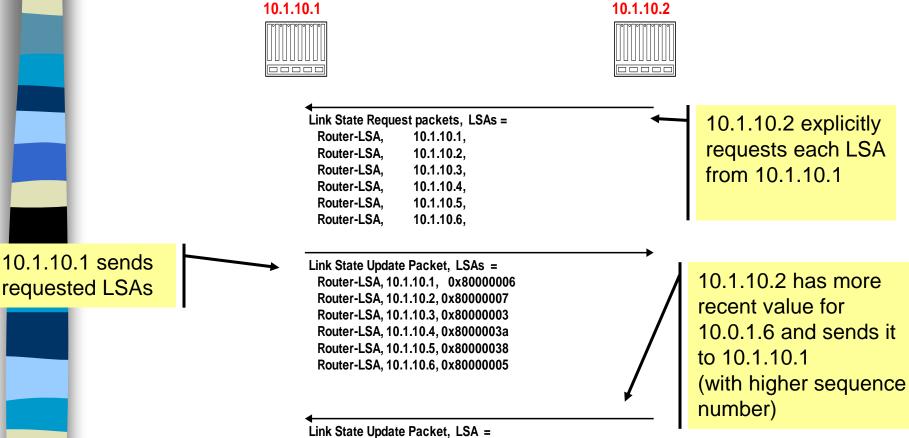
- Routers multicasts OSPF Hello packets on all OSPF-enabled interfaces
- If two routers share a link, they can become neighbors, and establish an adjacency



After becoming a neighbor, routers exchange their link state databases



Regular LSA exchanges

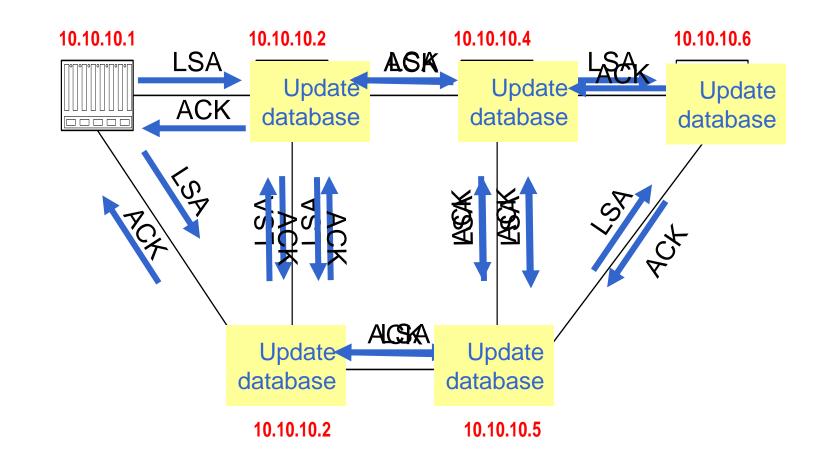


Link State Update Packet, LSA = Router-LSA, 10.1.1.6, 0x80000006

76

Routing Data Distribution

- LSA-Updates are distributed to all other routers via Reliable Flooding
- **Example:** Flooding of LSA from 10.10.10.1



Dissemination of LSA-Update

- A router sends and refloods LSA-Updates, whenever the topology or link cost changes (If a received LSA does not contain new information, the router will not flood the packet)
- Exception: Infrequently (every 30 minutes), a router will flood
 LSAs even if there are not new changes
- Acknowledgements of LSA-updates:
 - explicit ACK, or
 - implicit via reception of an LSA-Update



OSPF Areas

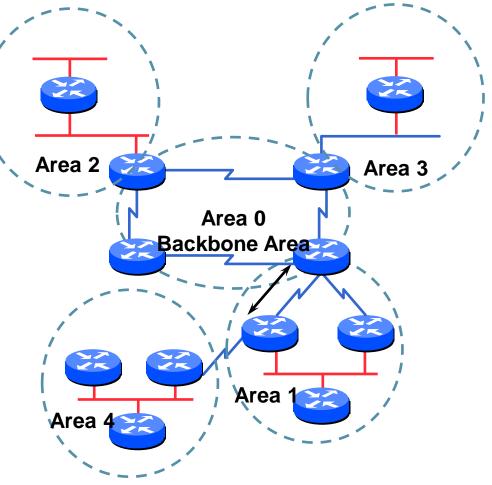
Group of contiguous hosts and networks

Per area topological database

- Invisible outside the area
- Reduction in routing traffic

Backbone area contiguous

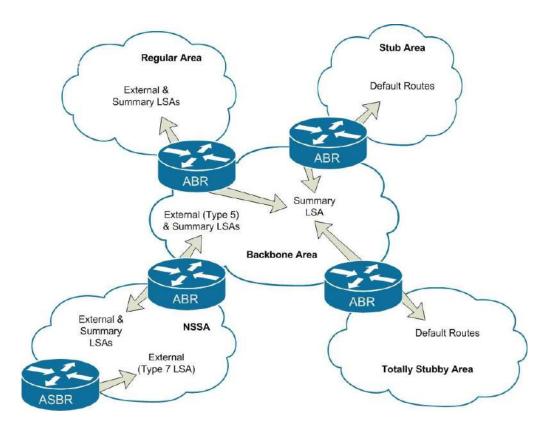
 All other areas must be connected to the backbone





Exemplo

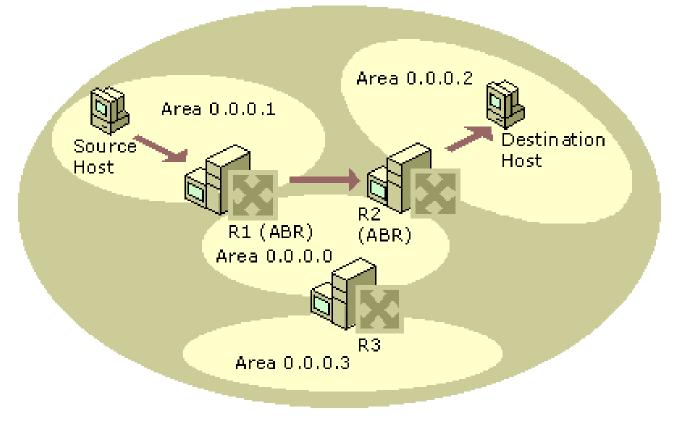
R1(config)# router ospf 10 R1(config-router)# **network 10.0.0 0.255.255.255 area 0** R1(config-router)# **network 192.168.100.0 0.0.3.255 area 0** R5(config)# router ospf 10 R5(config-router)# **area 2 stub**



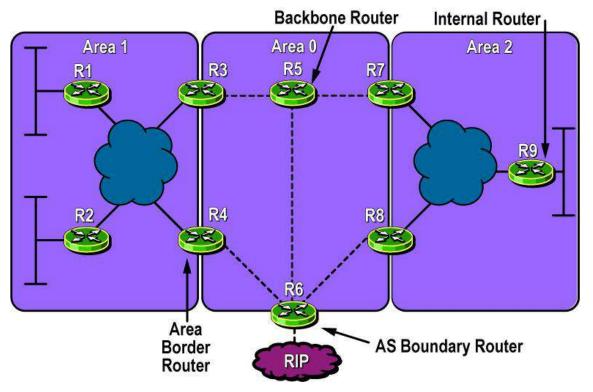


Exemplo

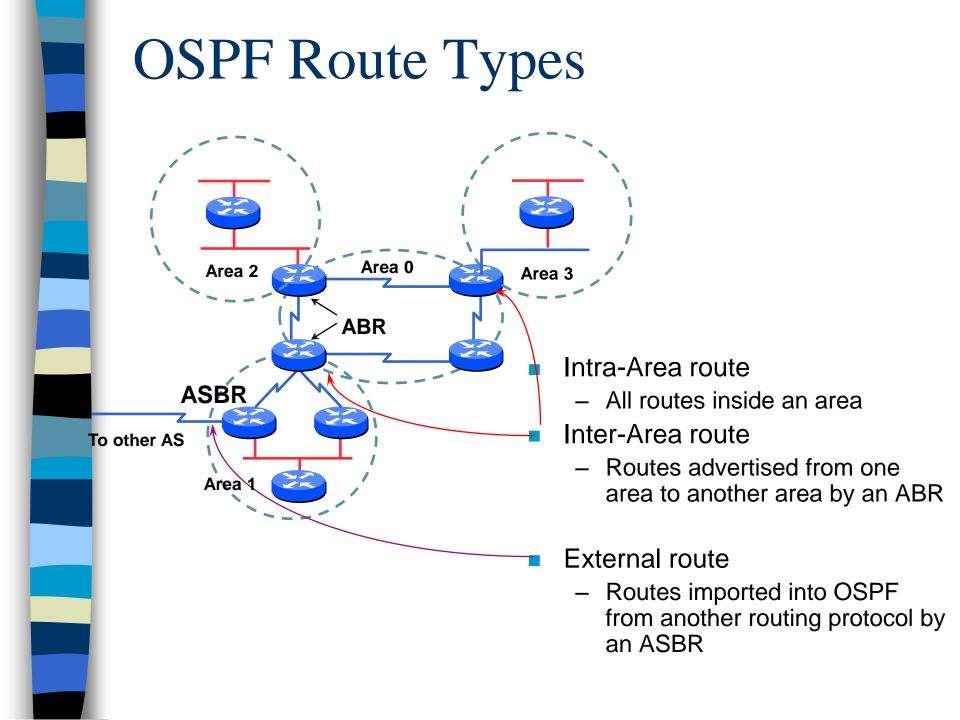
Autonomous System (AS)

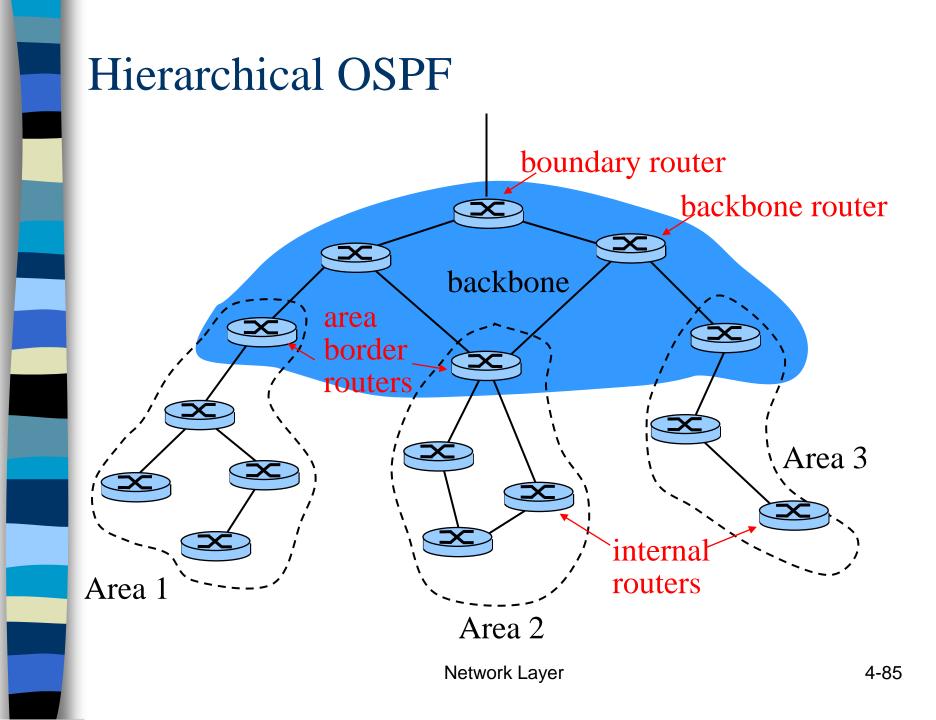


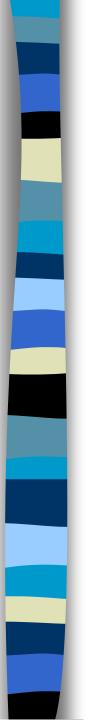
Router Roles



- Internal: All interfaces in a single area (routers 1,2,5 and 9 in diagram above)
- Backbone: At least one interface assigned to area 0 (routers 3,4, 5, 6, 7, 8, in diagram above)
- Area Border Router (ABR): Have interfaces in two or more areas (routers 1,4,5 in diagram above)
- Autonomous System Boundary Router (ASBR): Has at least one interface in an OSPF area and at least one interface outside of an OSPF area.







PART III

Implementing Inter-Network Relationships with BGP

Chapter 4: Network Layer

4. 1 Introduction 4.2 Virtual circuit and datagram networks 4.3 What's inside a router 4.4 IP: Internet Protocol Datagram format IPv4 addressing ICMP

– IPv6

4.5 Routing algorithms

- Link state
- Distance Vector
- Hierarchical routing

4.6 Routing in the Internet

- RIP
- OSPF
- BGP