Routing Protocols

Basic Operation

- ❑ Routing updates
  - ❑ Neighbors
  - ❑ Routes & metrics
  - ❑ Selection criteria

 T04.2

Routing protocols specify the details of communication between routers. Routers exchange information with their neighbors. The information exchanged may include network reachability, network state and metrics.

A **metric** is a measure of how good a particular route is. Different protocols use different metrics and exchange different types of information.

Distance Vector vs. Link State

- Distance vector
  - Updates are routing tables.
  - Update neighbors
  - Propagation delay; slow convergence
  - Simple
- Link state
  - Updates are link states.
  - Update all nodes
  - Fast convergence
  - Complex

 T04.3

Distance vector protocols require that routers send all or a large part of the routing table to neighbor routers. This results in large amounts of routing traffic. Each neighbor must then pass any new information on to its neighbors and so on. This type of information exchange results in **propagation delay**. Before a router can learn of a routing change, one of its neighbors must learn of the change and pass this information along in the next routing update. However, distance vector protocols are relatively simple.

Link state protocols require that routers send updates to all routers in the same area. This allows for hierarchical routing by specifying different areas. The amount of routing traffic is small, but the operation of the routing protocol is more complex. Each router must calculate the state of the network based upon the **link state advertisements** (LSA) received from all other routers in the same area. LSAs contain information about network links, such as which are active and which are not.

Interior vs. Exterior

- ❑ Autonomous System (AS)
  - ❑ One administrative control
- ❑ Interior
  - ❑ Same AS
- ❑ Exterior
  - ❑ Different AS

 T04.4

An **Autonomous System**, or AS, is a group of networks under one administrative control. To everyone outside the AS, the AS appears to have a consistent routing implementation. Details of routing techniques and decisions are hidden outside of the AS.

An **interior routing protocol** is one used within an AS. This type of protocol is not used to exchange routing information with anyone outside the AS.

An **exterior routing protocol** is used between ASs. This type of protocol is used specifically to exchange routing information with other Autonomous Systems.

PSINet

CIDR

- CIDR blocks (Classless Inter-Domain Routing)
- Superblock announcements
- A superblock is specified via a netmask.
  - 256 class C networks: 199.100.0.0 - 199.100.255.0, netmask 255.255.255.0 -> 256 routes
  - 1 class B network: 199.100.0.0, netmask 255.255.0.0 -> 1 route
- Outside the AS there is a single routing entry for the entire block.

 T04.5

CIDR network numbers, or superblocks, allow the aggregation of multiple routing table entries into a single entry. This benefits the Internet by reducing the load on routers. It also improves performance since fewer routes need to be examined when deciding how to route a packet.

PSINet

RIP

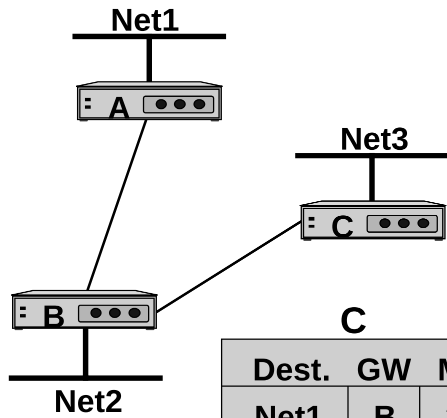
- ❑ Routing Information Protocol
- ❑ Distributed with Berkeley UNIX
- ❑ RFC 1058
- ❑ Distance vector
- ❑ Internal
- ❑ Hop count metric

Routing Information Protocol (RIP) is widely used due to its distribution with UNIX. RIP is an internal, distance vector protocol that uses a simple hop-count metric. The **hop count** is the number of intermediate gateways from the current device to the destination. Fewer gateways mean a better route.

RIP - Sample Flaw

A

Dest.	GW	M
Net2	B	1
Net3	B	2



B

Dest.	GW	M
Net1	A	1
Net3	C	1

C

Dest.	GW	M
Net1	B	2
Net2	B	1

 T04.7

Consider the above network.

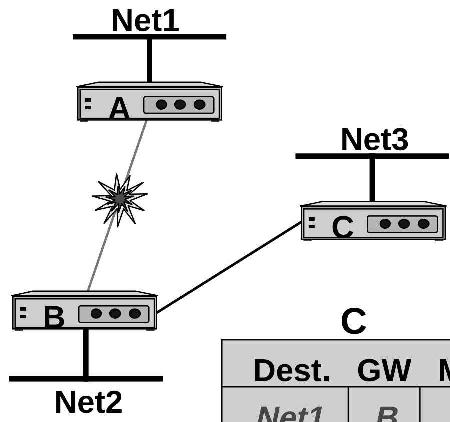
A can reach Net2 via B and it is 1 hop away. A can reach Net3 via B and it is 2 hops away.

B and C have similar reachability information. Direct routes are not included in the routing tables for this example. (B has a direct route to Net2 but this is not included in this example.)

RIP - Sample Flaw

A

Dest.	GW	M
~~Net2~~	~~B~~	~~1~~
~~Net3~~	~~B~~	~~2~~



B

Dest.	GW	M
~~Net1~~	~~A~~	~~1~~
Net3	C	1

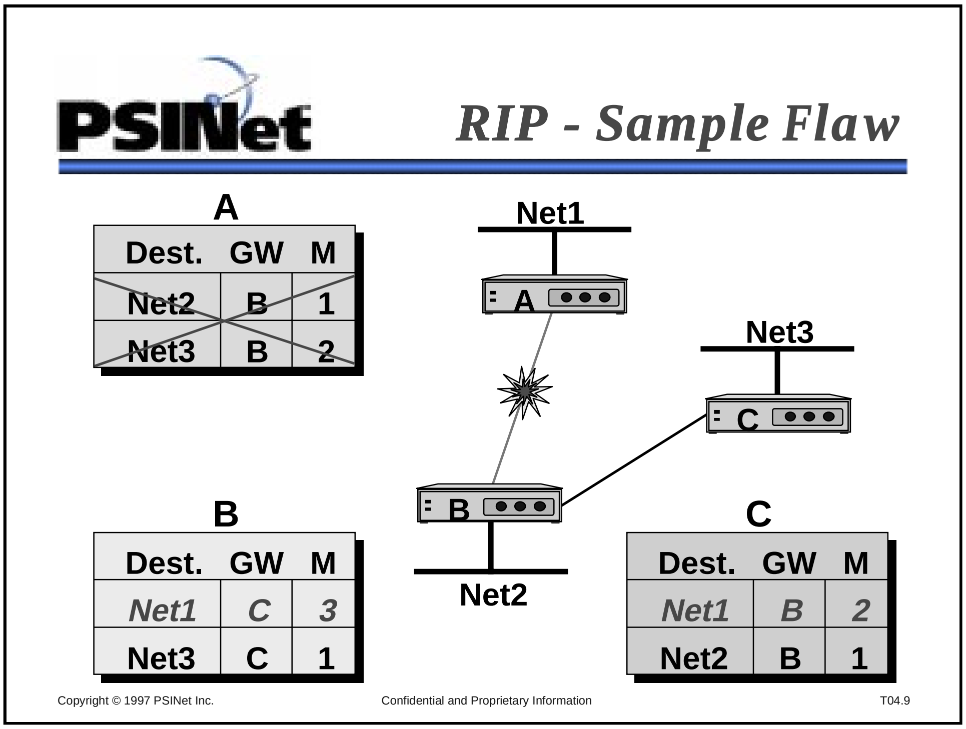
C

Dest.	GW	M
Net1	*B*	*2*
Net2	B	1

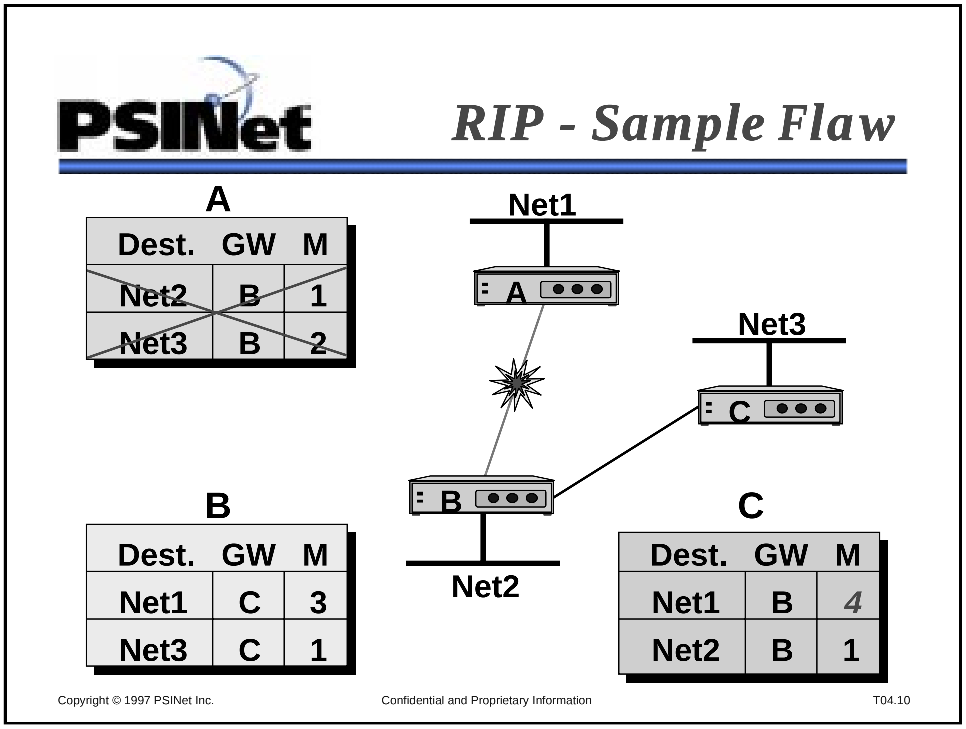
 T04.8

What happens if the link between A and B goes down?

The route for Net1 in B’s routing table is marked invalid. B then listens to its other neighbors (in this case, just C) for a route to Net1.



B will update its routing table with the information it has learned from C. Now what happens to a packet at router B that is destined for Net1?



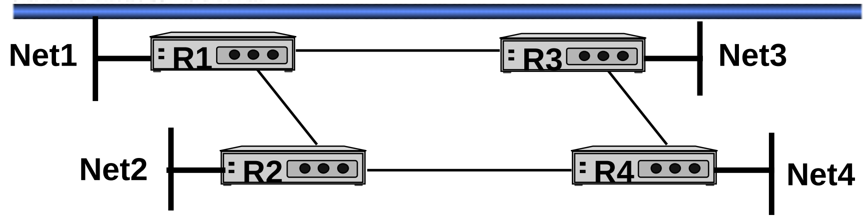
With the next routing update between B and C, C will update its metric for the Net1 route to reflect the current information C is receiving from B.

During the next routing update, B will increment the metric for the route to Net1 to reflect the information C is broadcasting.

This process will continue until the metrics reach 16 and the routes are marked invalid.

PSINet

RIP - Sample Update



R2

Dest.	GW	M
Net1	R1	1
Net2	Direct	
Net3	R1	2
Net4	R4	1

What will R1's routing table look like?

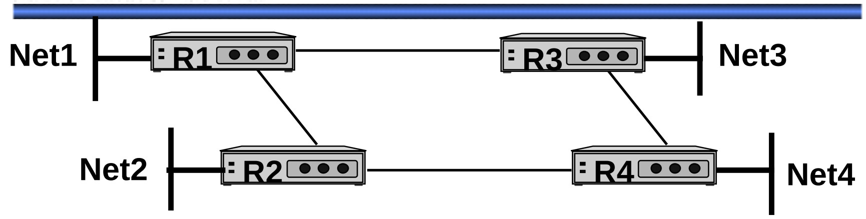
R3

Dest.	GW	M
Net1	R1	1
Net2	R4	2
Net3	Direct	
Net4	R4	1

 T04.11

Given that the routers in the above example are utilizing RIP, what will R1's routing table look like after receiving updates from R2 and R3?

RIP - Sample Update



R1

Dest.	GW	M
Net1	Direct	
Net2	R2	1
Net3	R3	1
Net4	R2	2

R1 will utilize the direct route to Net1

R1 will choose the shorter route to Net2 via R2

R1 will choose the shorter route to Net3 via R3

R1 has to choose the route to Net4. The route to Net4 via R2 is 2 hops away as is the route via R3.

RIP maintains a single entry for each destination network.

RIP - Stability Features

- ❑ Hop count limit
- ❑ Split horizons
- ❑ Poisoned reverse update

 Confidential and Proprietary Information T04.13

Hop count limit
Any destination that is more than 15 hops away is considered unreachable.

Split horizons
Routers do not broadcast information back to the source. In other words, if router A learns something from router B, router A will not include this route in updates to router B.

Poisoned reverse update
If router A learns something from router B, when router A sends an update to router B, the route will be included but the metric will be 16 indicating unreachable.

PSINet

RIP - Summary

- ❑ Advantages
  - ❑ Simple to implement
  - ❑ Widely used
- ❑ Disadvantages
  - ❑ Propagation delay; slow convergence
  - ❑ Instability
  - ❑ Not classless
  - ❑ Single route only
  - ❑ 16 hops is unreachable.

 T04.14

To summarize, RIP is widely used due to its distribution with UNIX. It is a simple protocol to understand and implement.

Along with simplicity come some disadvantages such as no load balancing and an inability to support classless or CIDR routing. Since RIP is a distance vector protocol, it also has the disadvantage of slow convergence; when changes in the routing topology arise, RIP is slow to stabilize. Also, any network that is more than 15 hops away from the source is unreachable. This prevents RIP from scaling to meet the needs of the current Internet growth.

IGRP

- Interior Gateway Routing Protocol
- Cisco Systems' proprietary routing protocol
- "An Introduction to IGRP"
  - Distance vector
  - Internal

 Confidential and Proprietary Information T04.15

IGRP is similar to RIP in that it is a distance vector protocol. An IGRP router periodically exchanges network reachability information with its neighbors. IGRP is proprietary to Cisco Systems, so IGRP can only be utilized between two or more Cisco routers.

PSINet

IGRP - Metric

- ❑ Combination of metrics
  - ❑ Topological delay time
  - ❑ Bandwidth - bits/sec
  - ❑ Load - fraction
  - ❑ Reliability (error rate) - %
  - ❑ Maximum Transmission Unit (MTU) *
  - ❑ Hop count *

* MTU and hop count are not currently used when determining the overall path metric.

 T04.16

IGRP uses a combination of values to determine the metric or measure of a particular route.

Topological delay time
Time to get to the destination assuming no network traffic.

Bandwidth
In bits per sec of the "narrowest" segment of the path.

Load
How much of the bandwidth is currently in use.

Reliability
Current error rate. The fraction of packets that arrive at the destination undamaged.

MTU*
Maximum packet size that can travel along the path without undergoing fragmentation.

Hop count*
Number of intermediate gateways along the path.

* MTU and hop count are not currently used when determining the overall path metric.

PSINet

IGRP - Metric

- Magic formula

[(K1 / Be) + (K2 * Dc)] * r

K1	bandwidth weight
K2	delay weight
Be	bandwidth * (1 - load)
Dc	topological delay
r	reliability

 T04.17

This is the formula run over the individual metric values in order to create the composite metric value. The composite value is the one used for routing table entry comparison.

The two constants allow the network administrator to adjust performance of the protocol.

PSINet

IGRP - Stability Features

- Loop suppressing features
  - Split horizon
  - Metric checking
  - Maximum metric
  - Flash updates

 T04.18

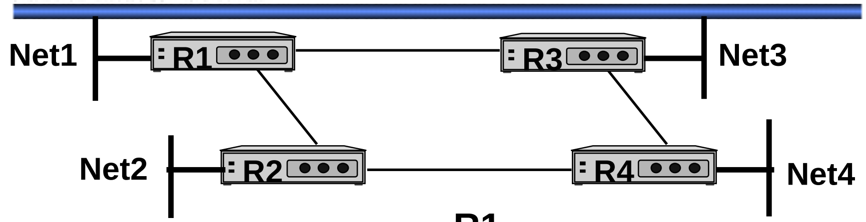
Split horizon
Information is not passed back to the source. In other words, if router A learns a route to Net1 from router B, router A will not include the route to Net1 when providing router B with a routing update.

Metric checking
If the composite metric at the next hop is greater than at the current hop, no traffic is sent to the next hop. The metric should be decreasing at each step along the path.

Maximum metric
IGRP’s maximum metric is 16 million.

Flash updates
When the router realizes there has been a change in network, routing updates can be sent immediately without waiting for the next scheduled update time.

IGRP Selection Criteria



R1

Dest.	GW	M
Net1	Direct	0
Net2	R2	1321
Net3	R3	1321
Net4	R2 R3	2005 2005

 T04.19

R1 will choose to route traffic directly to Net1.

R2 advertises the best route to Net2 and R3 advertises the best route to Net3.

R2 and R3 have equally good routes to Net4, so R1 will maintain both path choices and load balance traffic through both R2 and R3.

Notice that the metrics are dependent upon many values and cannot be determined by adding one to the previous metric as is done with RIP. Thus, the actual metric values maintained by R1 will be larger than those advertised by R2 and R3 by some amount that is dependent upon the load, delay, bandwidth and reliability of the links between R1-R2 and R1-R3..

IGRP - Summary

- Advantages
  - Relatively simple to implement
  - Control over metrics
  - Flash updates reduce propagation delay
  - Load balancing
- Disadvantages
  - Classfull
  - Still some propagation delay

 T04.20

IGRP has some advantages over RIP. The metric is more sophisticated and thus allows for greater control over the routing choices made. If a network administrator wants to favor bandwidth over delay, that decision can be implemented using IGRP. Flash updates reduce propagation delay of network changes, but they do not eliminate this delay.

A disadvantage of IGRP is that it is classfull and does not support CIDR routing.

OSPF

- ❑ Open Shortest Path First
- ❑ RFC 1583 version 2
- ❑ Link state
- ❑ Internal
- ❑ Metric - interface cost

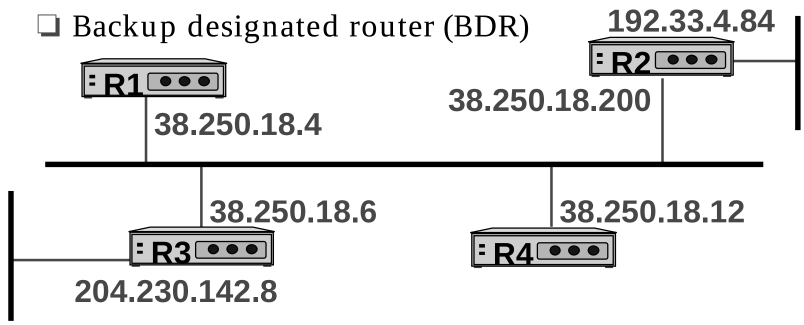
 T04.21

OSPF is an internal link state routing protocol. The state of the network is known by each node. The network information is used to create a shortest path tree at each node.

A cost is associated with leaving each OSPF interface. The sum of all costs along a path is the metric for that path. The costs are generally assigned manually.

OSPF Network

- ❑ Router ID
- ❑ Multi-access network
  - ❑ Designated router (DR)
  - ❑ Backup designated router (BDR)



 Confidential and Proprietary Information T04.22

Each OSPF router has a **router ID**. This value is determined by taking the maximum IP address of all the router's interfaces, regardless of whether or not OSPF is currently running on the interface.

The **designated router** is the router with the highest Router ID. Each router also has a priority associated with it to better control which router becomes the Designated Router - highest priority and highest Router ID.

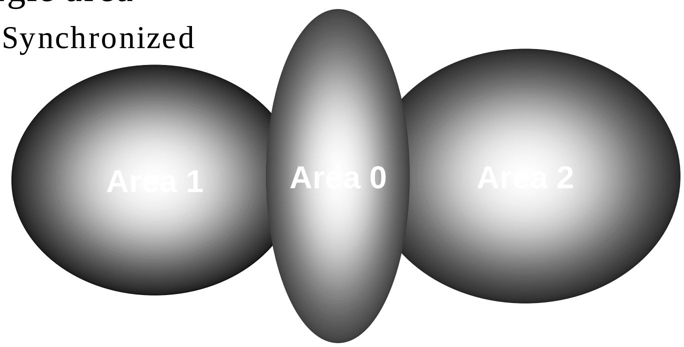
The **backup designated router** has the next highest priority/Router ID in the area.

The process of choosing the DR and BDR is slow and involves much network traffic. Thus, once an election is complete, it is not challenged until the DR and BDR are taken down and brought back up. The DR and BDR cannot be "overthrown" without being removed from operation.

With the exception of keep alive packets, all routers communicate with the DR and the BDR. Updates are not broadcast to all OSPF routers on a particular network.

OSPF - Areas

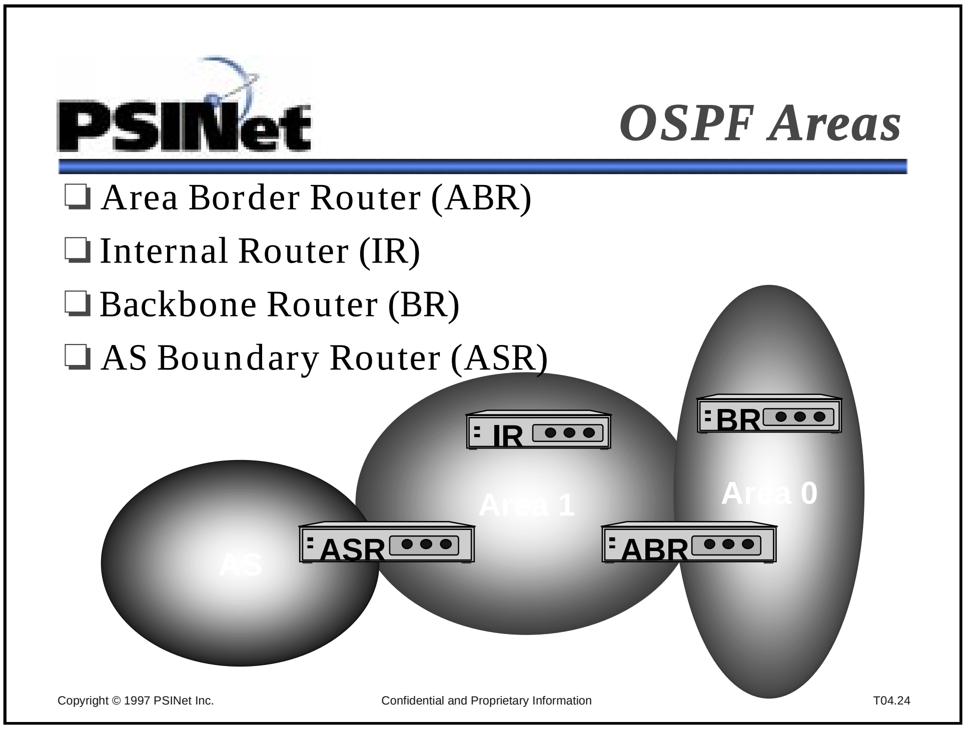
- Hierarchical routing
- Backbone area - 0
- Single area
  - Synchronized



 T04.23

OSPF allows a network to be broken into areas connected together by a backbone. The backbone is always area 0.

All routers within a particular area are synchronized with respect to link state knowledge. Creating a hierarchical routing structure allows for greater organizational control and reduces the effect changes in one area have on other areas.



OSPF router types are based upon functionality of the routers.

Area border routers (ABR)
Routers that join two or more areas. ABRs are responsible for distributing information into and out of the area. These routers are also backbone routers since they belong to the backbone area.

Internal routers (IR)
Routers that have no connections with other OSPF areas.

Backbone routers (BR)
Routers that are part of the backbone area. This includes all ABRs.

AS boundary routers (ASR)
Routers that exchange information with one or more routers belonging to other Autonomous Systems. AS Routers can be internal, area border or backbone routers.

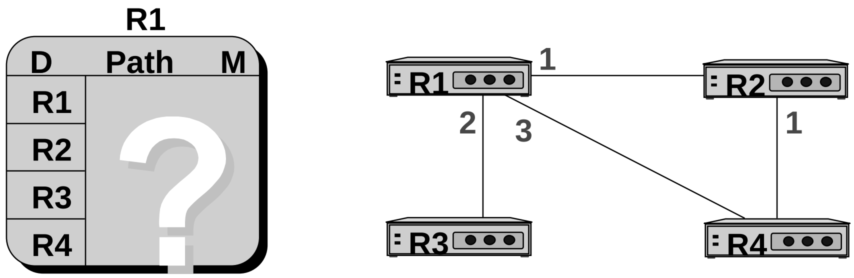
OSPF Operation

- ❑ Link state information
- ❑ Dijkstra’s shortest path algorithm
- ❑ Shortest path tree

Each OSPF router has a picture of the state of an area. The router runs Dijkstra’s shortest path algorithm over this information creating a shortest path tree. The router is the root of the tree and the tree contains the shortest path to each router on the network.

OSPF Example

❑ What will R1’s routing table look like?



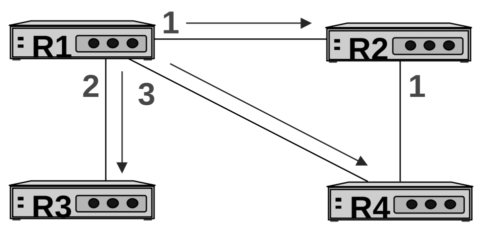
 Confidential and Proprietary Information T04.26

Assuming the routers are running OSPF and the metrics have been assigned as indicated, what will R1’s routing table look like?

PSINet — *OSPF Example*

R1

D	Path	M
R1	direct	0
R2	*R1-R2*	*1*
R3	*R1-R3*	*2*
R4	*R1-R4*	*3*



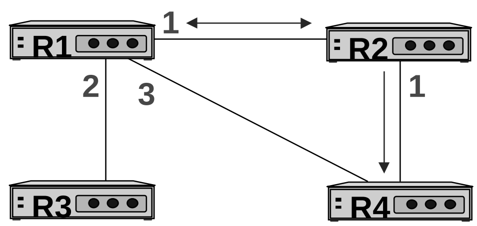
 Confidential and Proprietary Information T04.27

First, R1 examines each router it is connected to. Each such path is considered to be a prospect for entry into the table. The potential route with the smallest metric is added to the table.

OSPF Example

R1

D	Path	M
R1	direct	0
R2	R1-R2	1
R3	*R1-R3*	*2*
R4	*R1-R4*	*3*



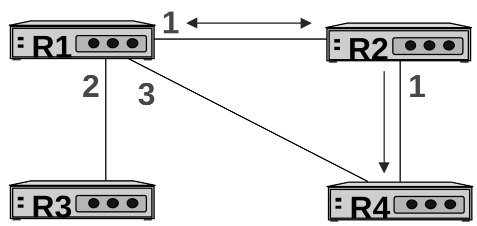
 Confidential and Proprietary Information T04.28

Now that the route to R2 has been decided upon, paths to all routers connected to R2 are analyzed and the paths and metrics are adjusted as necessary.

OSPF Example

R1

D	Path	M
R1	direct	0
R2	R1-R2	1
R3	*R1-R3*	*2*
R4	*R1-R2-R4*	*2*



 Confidential and Proprietary Information T04.29

While analyzing paths through R2, it is discovered that there is a better route to R4 through R2. The routing table is updated accordingly. The entry with the smallest metric is added to the routing table. Since both entries have the same metric, one router is chosen.

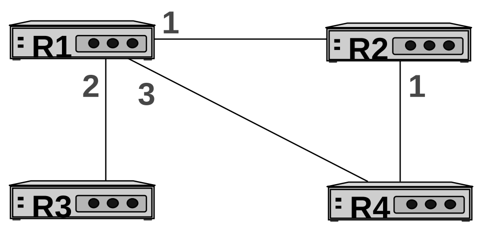
With this simple example, there are no more changes to the routing table.

PSINet

OSPF Example

R1

D	Path	M
R1	direct	0
R2	R1-R2	1
R3	R1-R3	2
R4	R1-R2-R4	2



 Confidential and Proprietary Information T04.30

This is the final state of R1’s routing table.

OSPF - Summary

- Advantages
  - Quick convergence
  - Load balancing
  - Classless
- Disadvantages
  - Computationally intensive

 Confidential and Proprietary Information T04.31

To summarize, OSPF converges quickly in the face of changes. Changes do not have to propagate gradually, but are flooded throughout the network. In addition, OSPF can maintain multiple routes for a single destination and supports CIDR addressing.

The main disadvantage of OSPF is the computation involved each time the network changes. Each node must recompute the shortest path tree. Thus, OSPF is not suited for networks with frequent changes.

BGP

- ❑ Border Gateway Protocol
- ❑ Successor to Exterior Gateway Protocol (EGP)
- ❑ RFC 1654 version 4
- ❑ External

BGP cannot really be compared with RIP, IGRP and OSPF since it is an exterior protocol.

BGP is used to exchange information between Autonomous Systems. Recall that an Autonomous System, or AS, is a collection of networks under one administrative control. To anybody outside an AS, the AS projects a consistent network state and routing policy.

PSINet

BGP

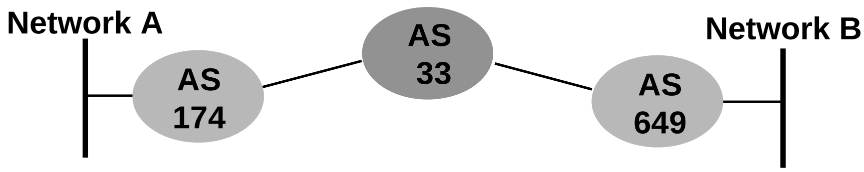
- Update messages
  - TCP
  - Network numbers / AS path
- Initial exchange - entire routing table
- Subsequent exchanges - changes
- Advertises only optimal path.
- Metric - shortest AS path

BGP is similar to other protocols in that routing information is exchanged with neighbors. However, BGP has some considerable differences:

- Initially the entire routing table is exchanged
- Subsequent updates are only sent when necessary, that is, when something in the routing table has changed.
- Keep alive packets are exchanged periodically to keep the BGP connection active.
- A BGP router can maintain multiple routes for a single destination. However, when advertising routes to neighbors, only the best route is advertised. The other routes are kept as a backup.
- The shortest AS path is generally chosen. There are parameters, such as local preference, to customize performance of routing. Manual control allows decisions to be based upon “political” distinctions such as avoiding an AUP network.

BGP

- Networks in AS 174 will have a path to Network B via AS path: AS 33, AS 649
- No knowledge of routing within AS 33 or AS 649

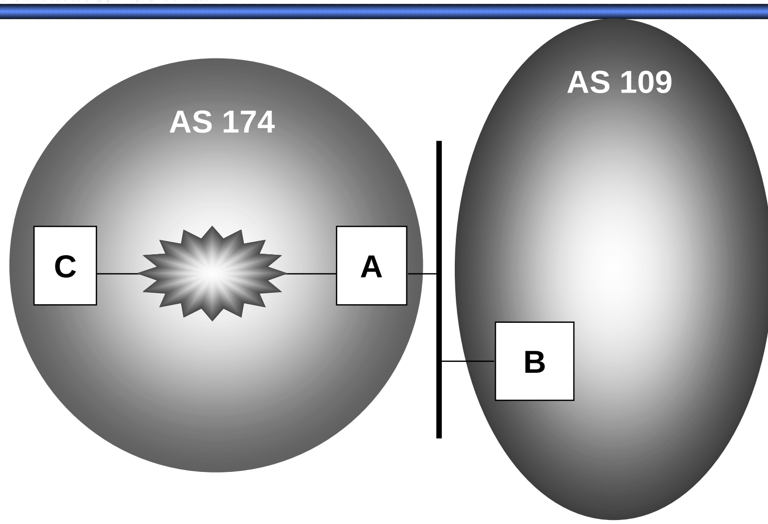


 Confidential and Proprietary Information T04.34

A network in AS 174, say Network A, knows the path to Network B is through AS 33 and then AS 649. Network A does not need to know anything about the path that must be taken to traverse AS 33.

PSINet

BGP



 Confidential and Proprietary Information T04.35

A and B are exterior BGP neighbors so they should share a common network. They are directly connected.

A and C are interior BGP neighbors so they do not need to share a common network. They do not have to be directly connected in order to exchange BGP updates.

Comparison

	RIP	IGRP	OSPF	BGP4
DV or LS	DV	DV	LS	Path Vec
TCP/UDP & Port	U - 520	IP - 9	T - 89	T - 179
Classless	No	No	Yes	Yes
Updates	Per.	Per.	Both	Trig.
Load Balance	No	Yes	Yes	No
Internal / External	Int.	Int.	Int.	Ext.
Metric	Hop Count	Load Errors Delay Bdwth	Sum of Int. Cost	Short. AS Path

 T04.36

DV or LS
Is the routing protocol a Distance Vector or a Link State protocol?

TCP/UDP & Port
Does the protocol specify information exchange via TCP or UDP and what port is used?

Classless
Does the protocol support CIDR or classless routing? Is netmask information contained in the routing updates?

Updates
Are updates periodic and occur based upon a time interval (Per.)? Or are they triggered and occur based upon information change (Trig.)?

Load Balance
Does the protocol load balance? That is, does the protocol allow multiple entries with the same destination?

Internal/External
Is the routing protocol used within an Autonomous System (internal) or between Autonomous Systems (external)?

Metric
What does the protocol use as a measure of how good a route is?

Questions?