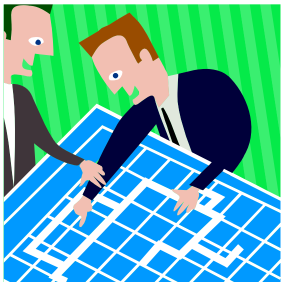
Routers & Routing

Joao.Neves@fe.up.pt

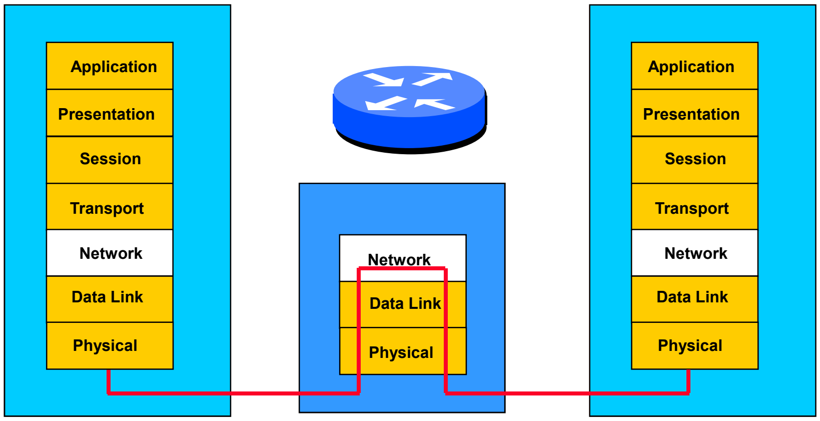
Router Main Functions



- Routing = build maps and choose routes

- Switching = move packets between interfaces, based on routing tables information

Router: working at layer 3

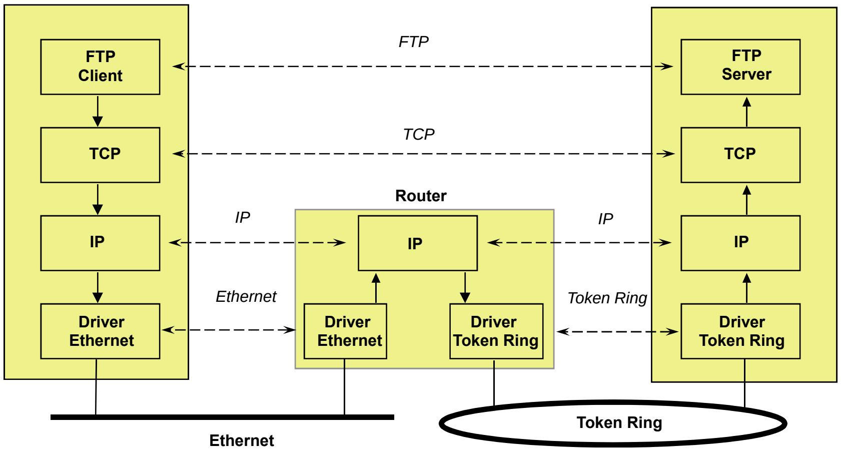


Application
Presentation
Session
Transport
Network
Data Link
Physical

Network
Data Link
Physical

Application
Presentation
Session
Transport
Network
Data Link
Physical

Router: working at layer 3

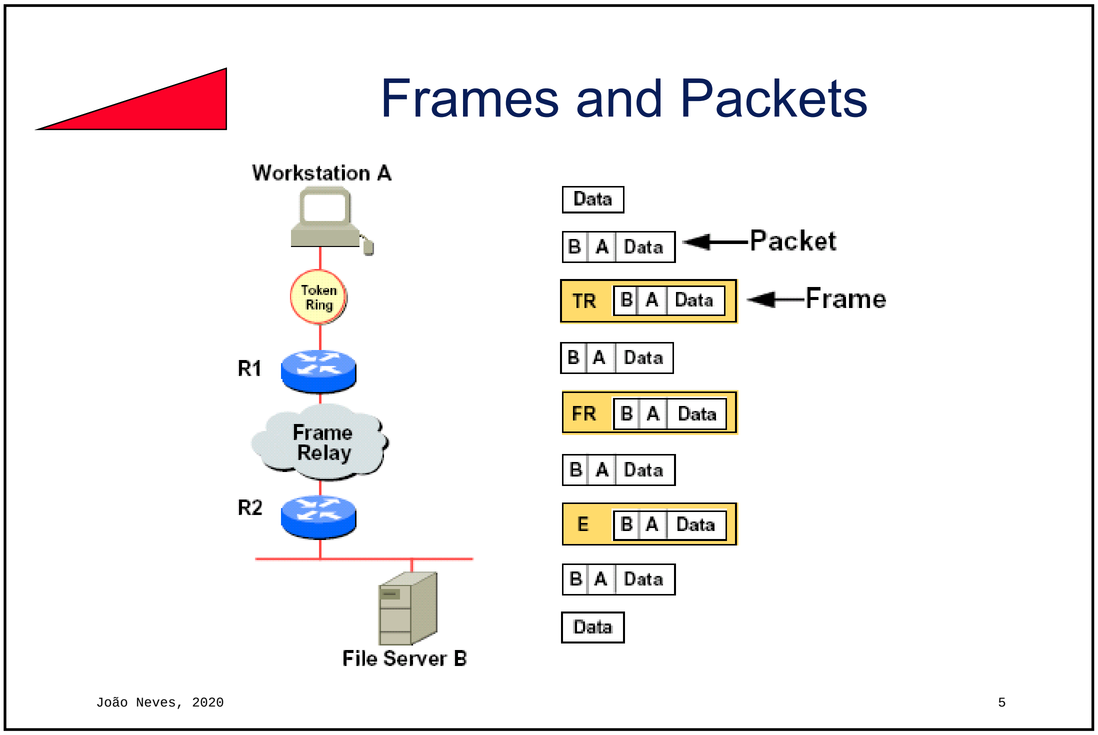


FTP Client
TCP
IP
Driver Ethernet

FTP
TCP
Router
IP
IP
IP
Ethernet
Driver Ethernet
Driver Token Ring
Token Ring

FTP Server
TCP
IP
Driver Token Ring

Ethernet
Token Ring



Frames and Packets

Workstation A

Token Ring

R1

Frame Relay

R2

File Server B

Data

B A Data ← Packet

TR B A Data ← Frame

B A Data

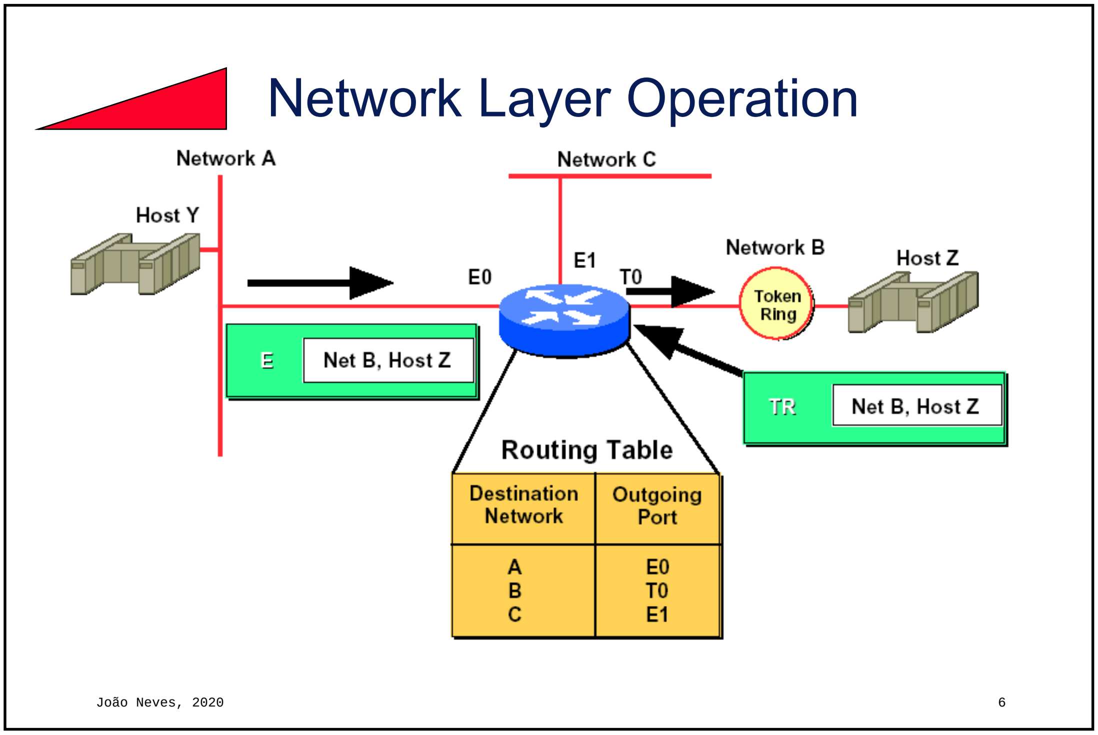
FR B A Data

B A Data

E B A Data

B A Data

Data

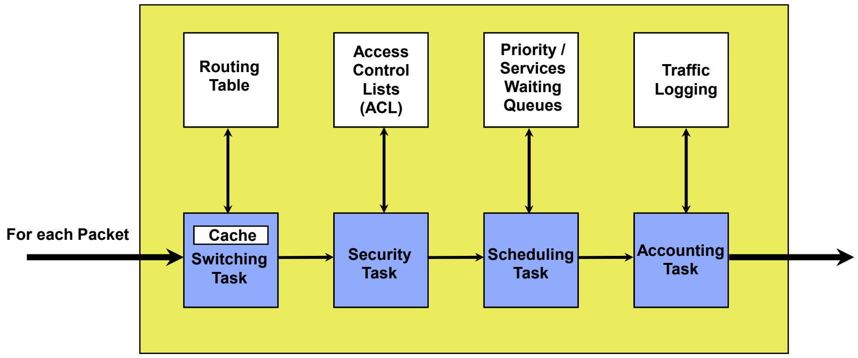


Network Layer Operation

Routing Table

Destination Network	Outgoing Port
A	E0
B	T0
C	E1

Layer 3 Processing



Routing Table

Access Control Lists (ACL)

Priority / Services Waiting Queues

Traffic Logging

For each Packet

Cache

Switching Task

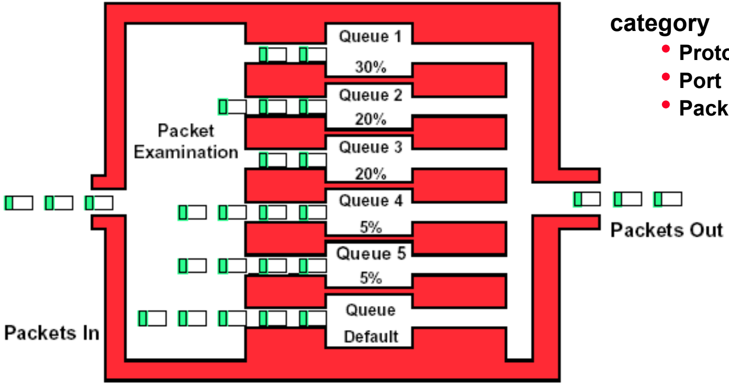
Security Task

Scheduling Task

Accounting Task

Each packet is processed in multiple modules, which has an impact on system performance.

Queues



Packet Examination

Queue 1

30%

Queue 2

20%

Queue 3

20%

Queue 4

5%

Queue 5

5%

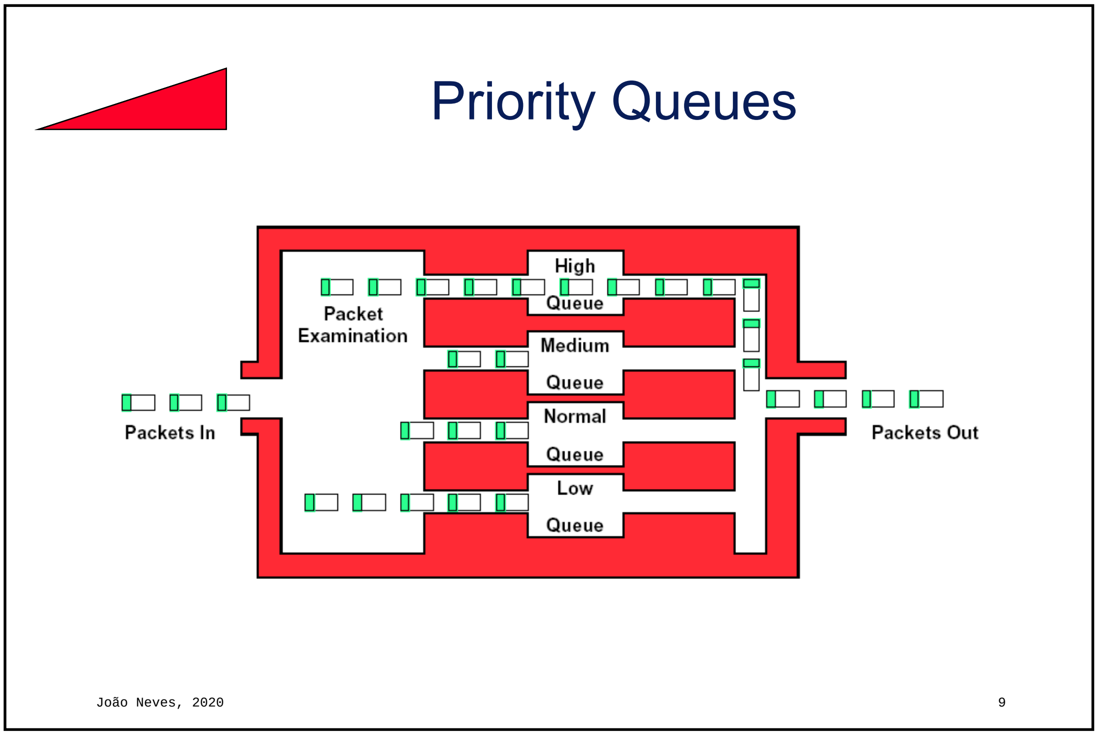
Queue Default

Packets In

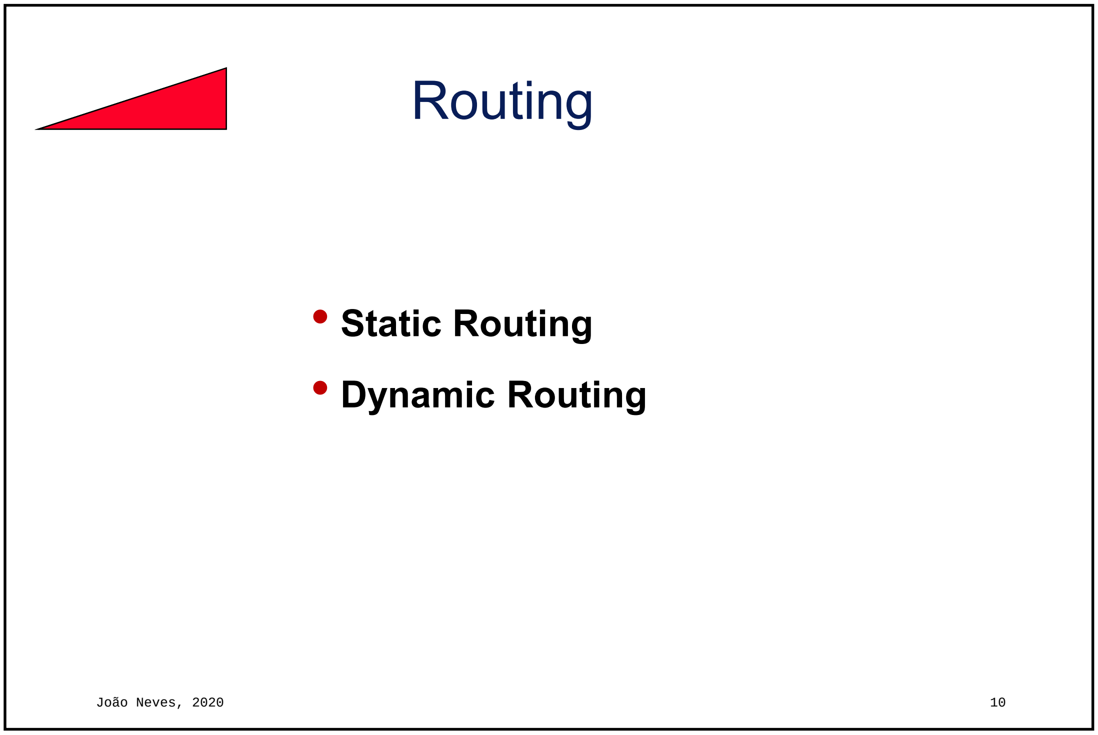
Packets Out

Bandwidth reservation by category

- Protocol
- Port
- Packet size

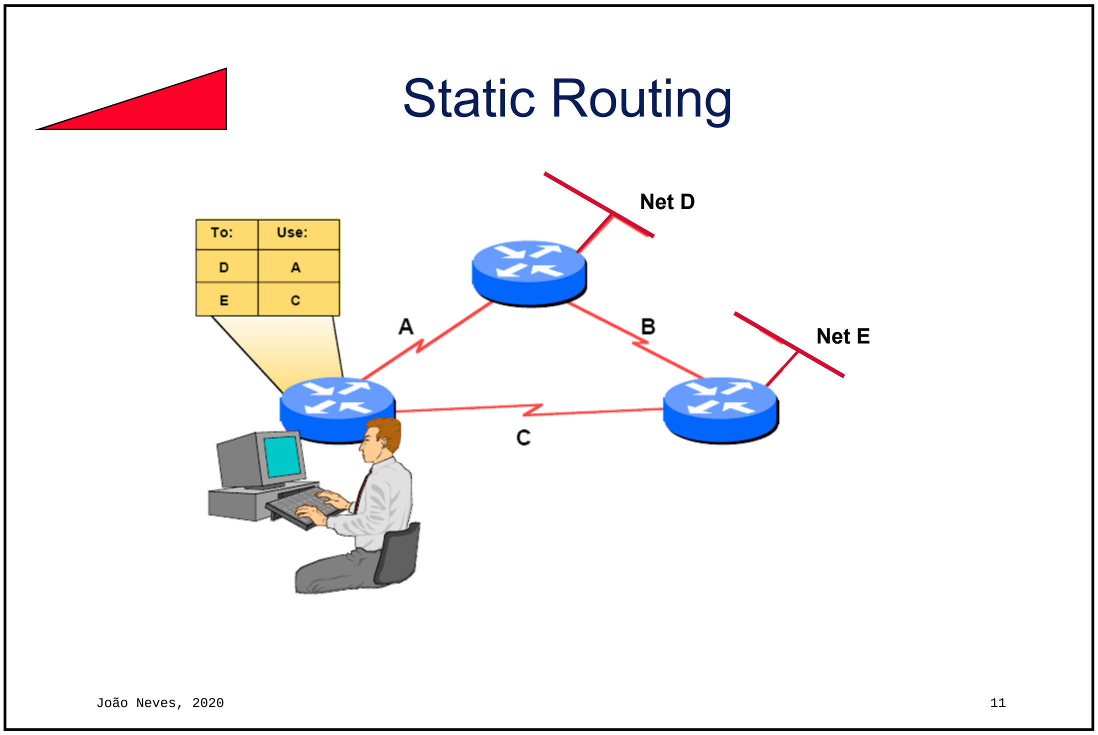


Priority Queues



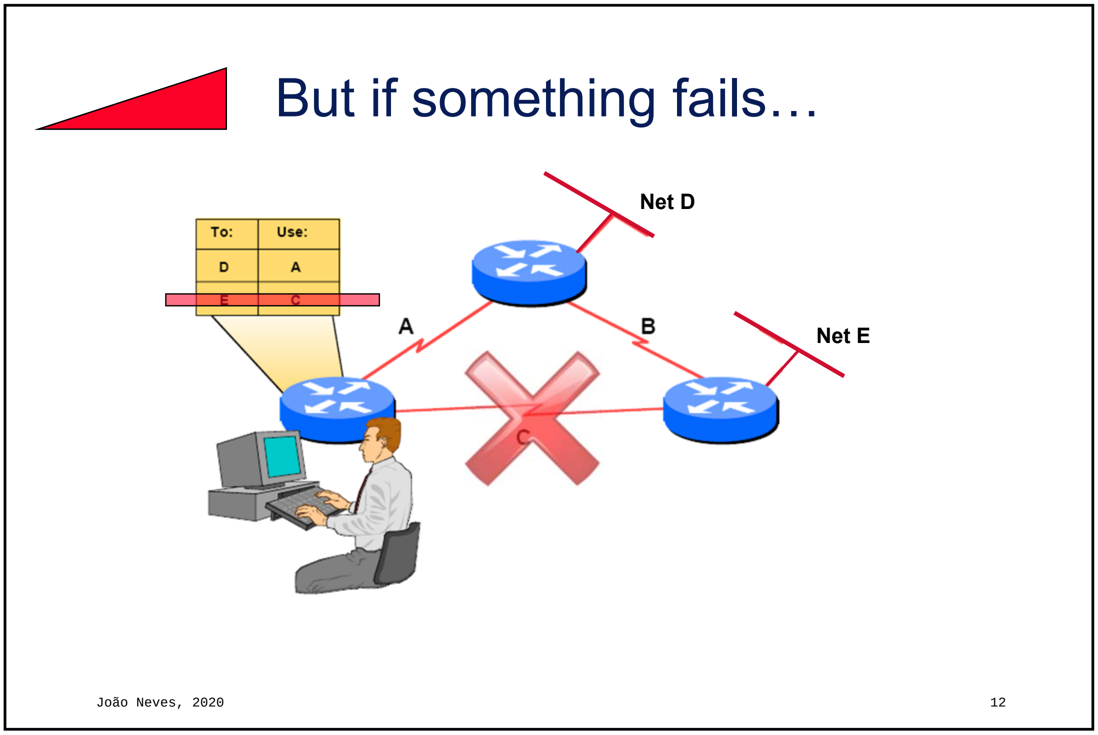
Routing

- **Static Routing**
- **Dynamic Routing**



Static Routing

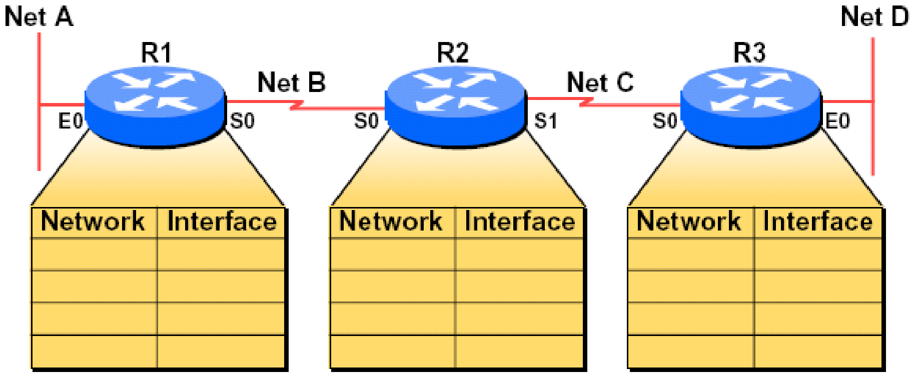
To:	Use:
D	A
E	C



But if something fails…

To:	Use:
D	A
~~E~~	~~C~~

Dynamic Routing



R1:

Network	Interface

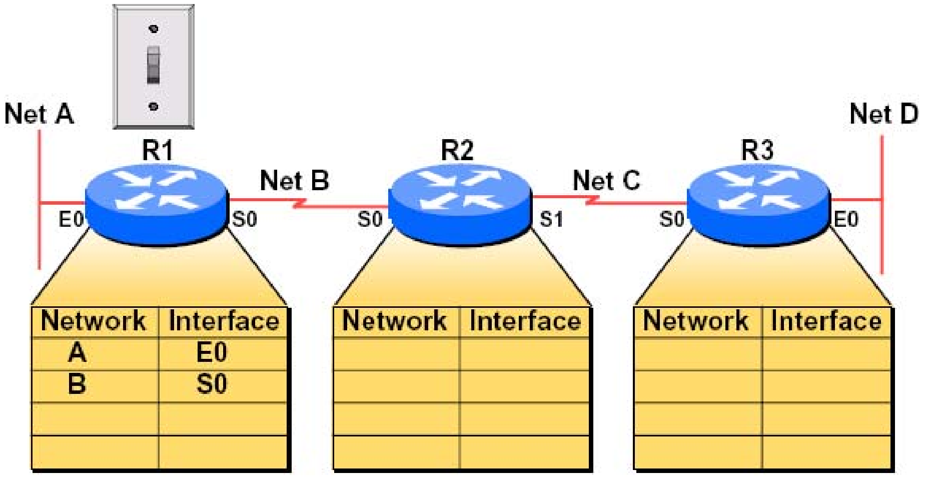
R2:

Network	Interface

R3:

Network	Interface

Dynamic Routing



R1:

Network	Interface
A	E0
B	S0

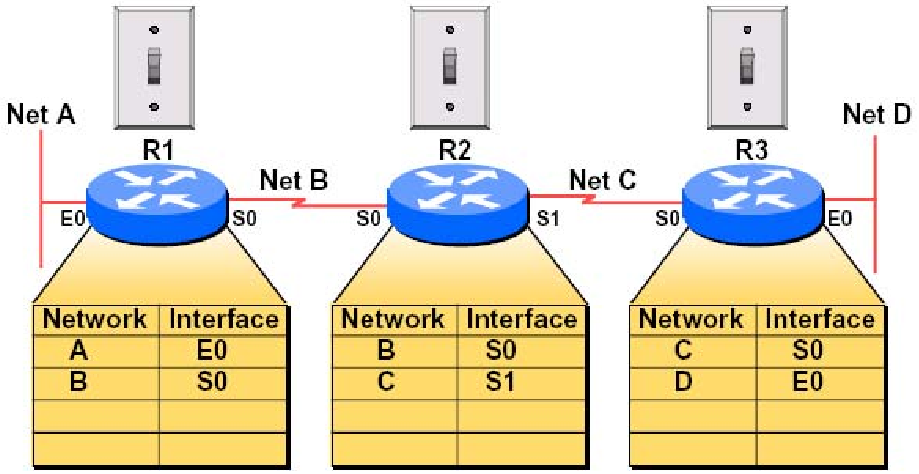
R2:

Network	Interface

R3:

Network	Interface

Dynamic Routing

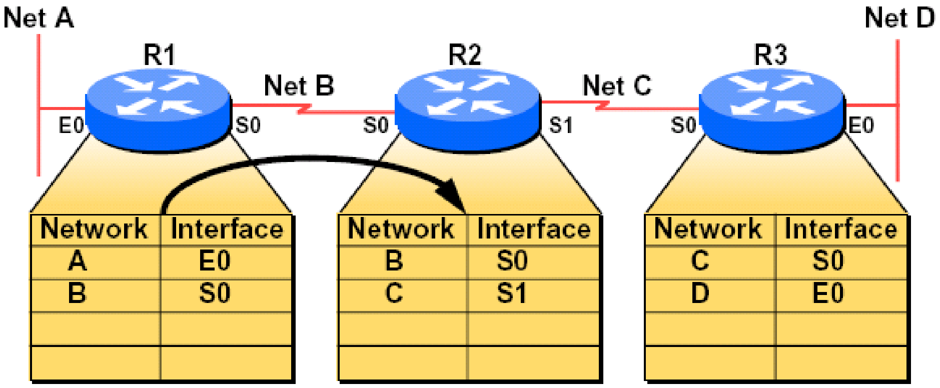


Network	Interface
A	E0
B	S0

Network	Interface
B	S0
C	S1

Network	Interface
C	S0
D	E0

Dynamic Routing

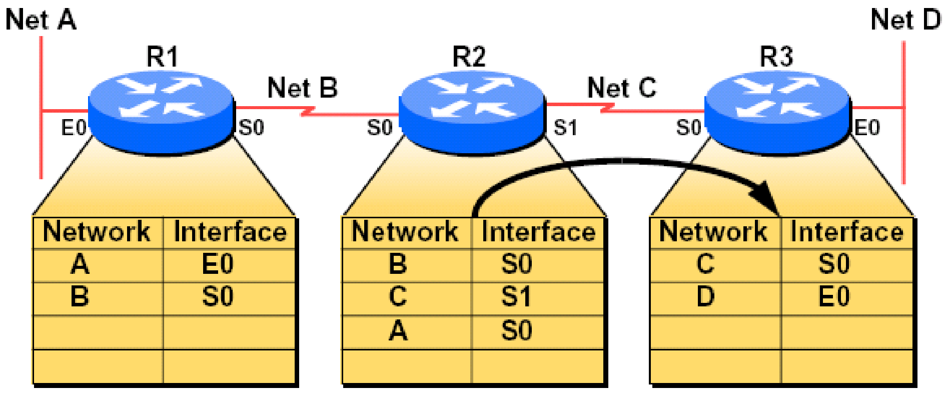


Network	Interface
A	E0
B	S0

Network	Interface
B	S0
C	S1

Network	Interface
C	S0
D	E0

Dynamic Routing

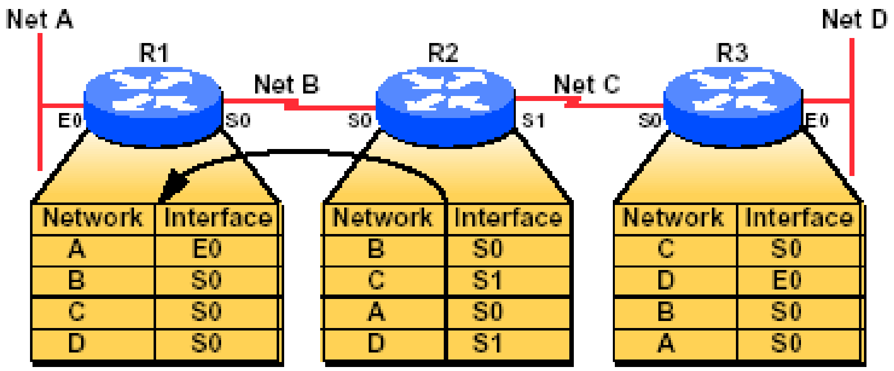


Network	Interface
A	E0
B	S0

Network	Interface
B	S0
C	S1
A	S0

Network	Interface
C	S0
D	E0

Dynamic Routing



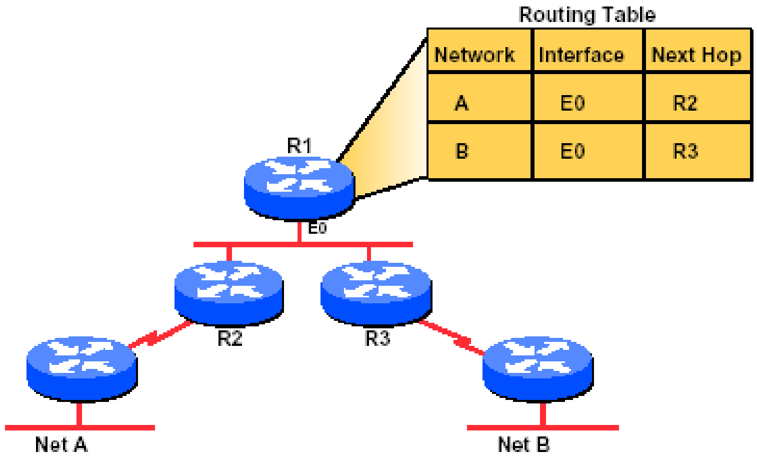
Network	Interface
A	E0
B	S0
C	S0
D	S0

Network	Interface
B	S0
C	S1
A	S0
D	S1

Network	Interface
C	S0
D	E0
B	S0
A	S0

Router Interfaces

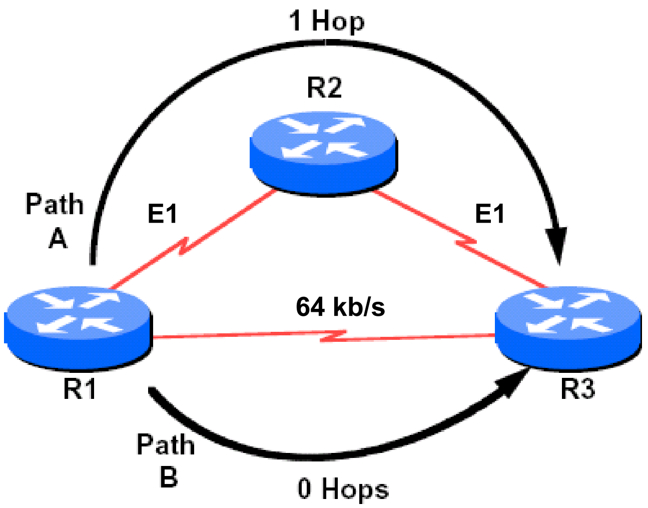
The Next Hop...



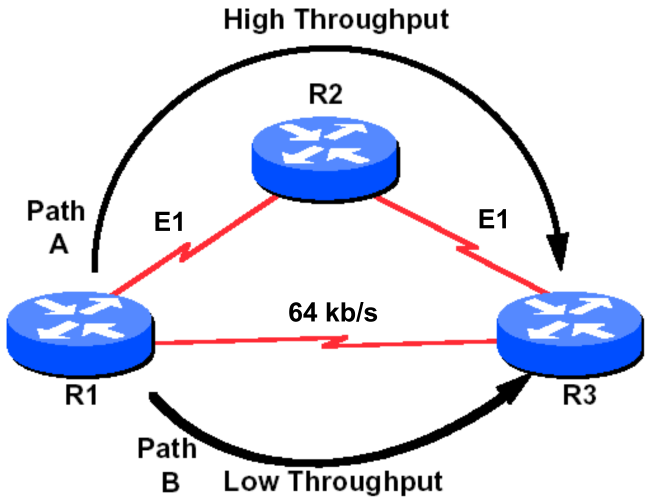
Routing Table

Network	Interface	Next Hop
A	E0	R2
B	E0	R3

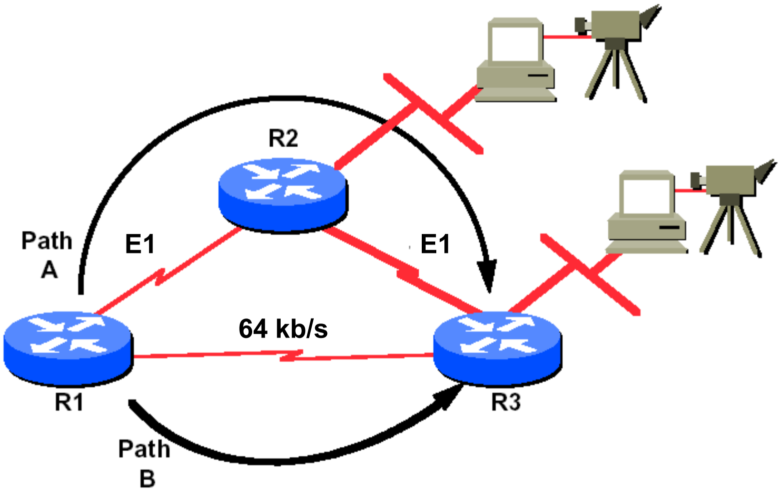
Metrics: hop count



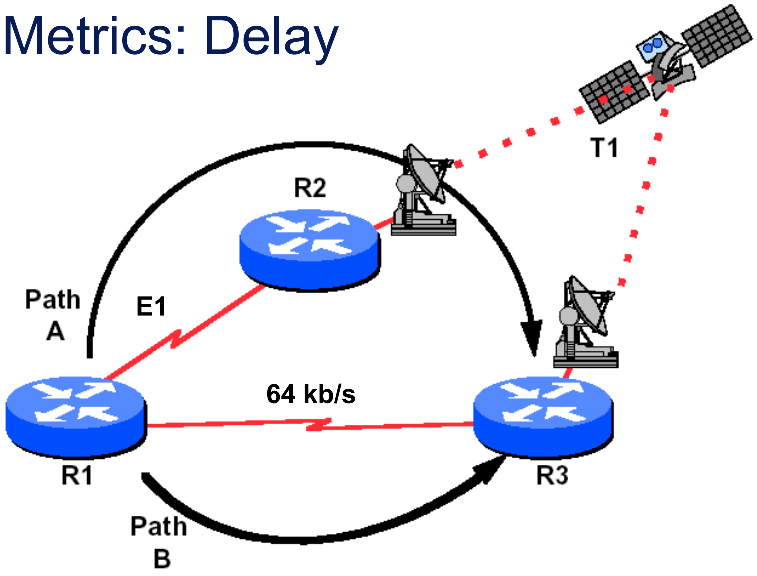
Metrics: Throughput



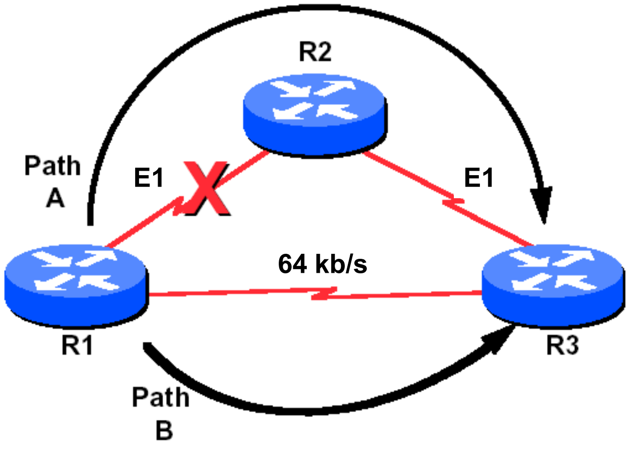
Metrics: Load



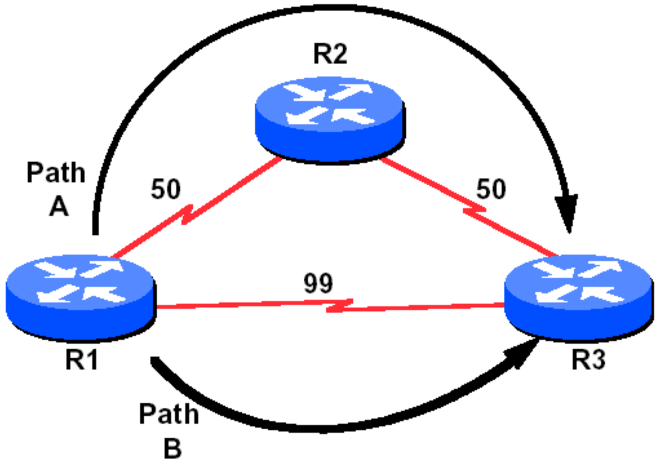
Metrics: Delay



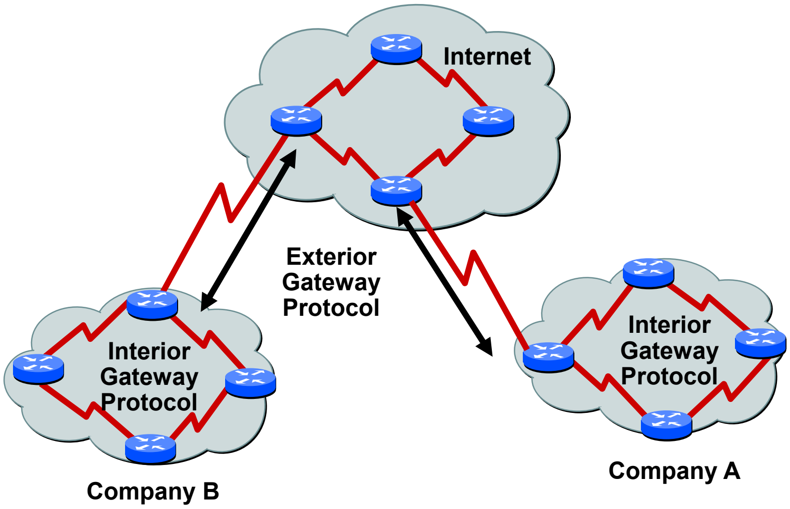
Metrics: Reliability



Metrics: Administrative Cost



Interior and Exterior Routing



Internet

Exterior Gateway Protocol

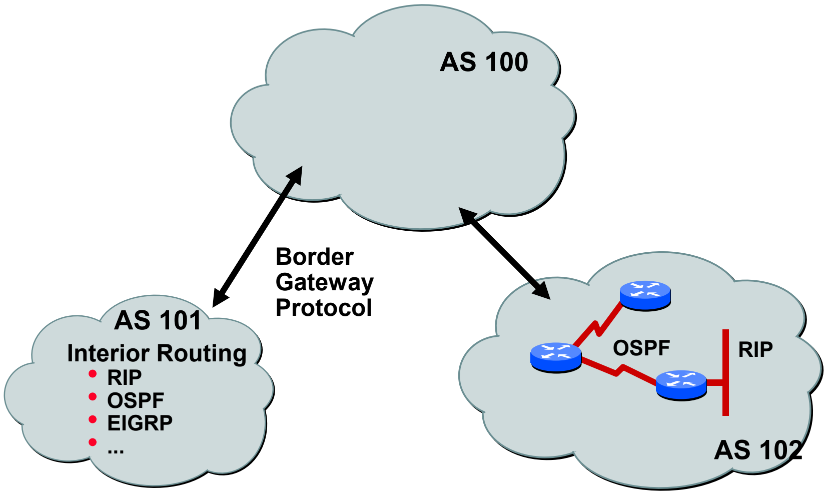
Interior Gateway Protocol

Interior Gateway Protocol

Company B

Company A

Autonomous System (AS)



AS 100

Border Gateway Protocol

AS 101

Interior Routing

- RIP
- OSPF
- EIGRP
- ...

OSPF

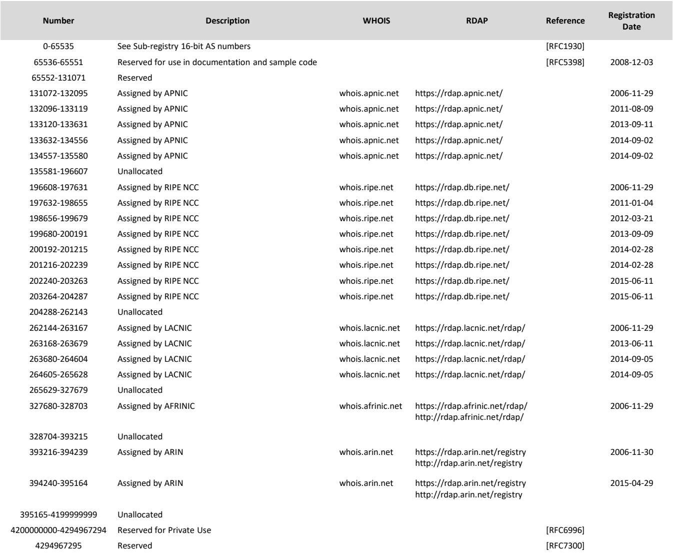
RIP

AS 102

Characteristics of an AS

- An AS is a set of networks subordinated to a single Technical Management and that share the same Routing policy (Unique Administrative Management);
- An AS# is a number in the range [1, 65535];
- AS# of the range [64512, 65535] are reserved for private use;
- Interior Gateway Protocol (IGP) operate within an AS to ensure IP connectivity within it;
- Exterior Gateway Protocol (EGP) operate between AS's to allow routing and policies between them.

32-bit Autonomous System Numbers

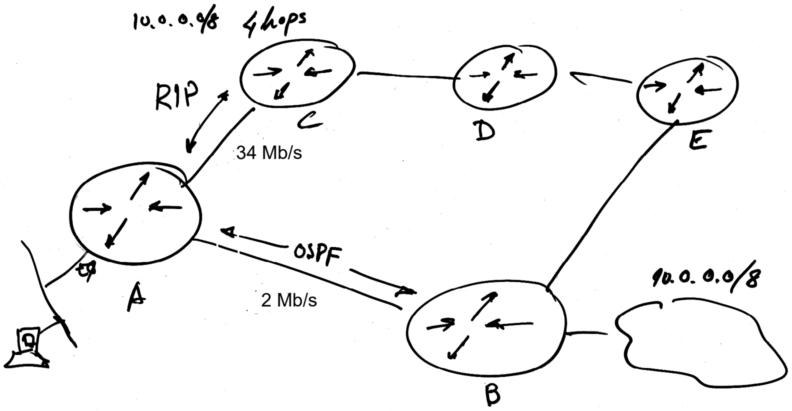


Number	Description	WHOIS	RDAP	Reference	Registration Date
0-65535	See Sub-registry 16-bit AS numbers			[RFC1930]	
65536-65551	Reserved for use in documentation and sample code			[RFC5398]	2008-12-03
65552-131071	Reserved				
131072-132095	Assigned by APNIC	whois.apnic.net	https://rdap.apnic.net/		2006-11-29
132096-133119	Assigned by APNIC	whois.apnic.net	https://rdap.apnic.net/		2011-08-09
133120-133631	Assigned by APNIC	whois.apnic.net	https://rdap.apnic.net/		2013-09-11
133632-134556	Assigned by APNIC	whois.apnic.net	https://rdap.apnic.net/		2014-09-02
134557-135580	Assigned by APNIC	whois.apnic.net	https://rdap.apnic.net/		2014-09-02
135581-196607	Unallocated				
196608-197631	Assigned by RIPE NCC	whois.ripe.net	https://rdap.db.ripe.net/		2006-11-29
197632-198655	Assigned by RIPE NCC	whois.ripe.net	https://rdap.db.ripe.net/		2011-01-04
198656-199679	Assigned by RIPE NCC	whois.ripe.net	https://rdap.db.ripe.net/		2012-03-21
199680-200191	Assigned by RIPE NCC	whois.ripe.net	https://rdap.db.ripe.net/		2013-09-09
200192-201215	Assigned by RIPE NCC	whois.ripe.net	https://rdap.db.ripe.net/		2014-02-28
201216-202239	Assigned by RIPE NCC	whois.ripe.net	https://rdap.db.ripe.net/		2014-02-28
202240-203263	Assigned by RIPE NCC	whois.ripe.net	https://rdap.db.ripe.net/		2015-06-11
203264-204287	Assigned by RIPE NCC	whois.ripe.net	https://rdap.db.ripe.net/		2015-06-11
204288-262143	Unallocated				
262144-263167	Assigned by LACNIC	whois.lacnic.net	https://rdap.lacnic.net/rdap/		2006-11-29
263168-263679	Assigned by LACNIC	whois.lacnic.net	https://rdap.lacnic.net/rdap/		2013-06-11
263680-264604	Assigned by LACNIC	whois.lacnic.net	https://rdap.lacnic.net/rdap/		2014-09-05
264605-265628	Assigned by LACNIC	whois.lacnic.net	https://rdap.lacnic.net/rdap/		2014-09-05
265629-327679	Unallocated				
327680-328703	Assigned by AFRINIC	whois.afrinic.net	https://rdap.afrinic.net/rdap/ http://rdap.afrinic.net/rdap/		2006-11-29
328704-393215	Unallocated				
393216-394239	Assigned by ARIN	whois.arin.net	https://rdap.arin.net/registry http://rdap.arin.net/registry		2006-11-30
394240-395164	Assigned by ARIN	whois.arin.net	https://rdap.arin.net/registry http://rdap.arin.net/registry		2015-04-29
395165-4199999999	Unallocated				
4200000000-4294967294	Reserved for Private Use			[RFC6996]	
4294967295	Reserved			[RFC7300]	

Interior and Exterior Routing

- One router may run one or more Interior Gateway Protocol (IGP) simultaneous;
- One router may run only one Exterior Gateway Protocol (EGP);
- One router may run one or more IGPs for exchanging routes within its AS, simultaneously may run one EGP for routes exchange with others AS.

Route Redistribution



34 Mb/s

2 Mb/s

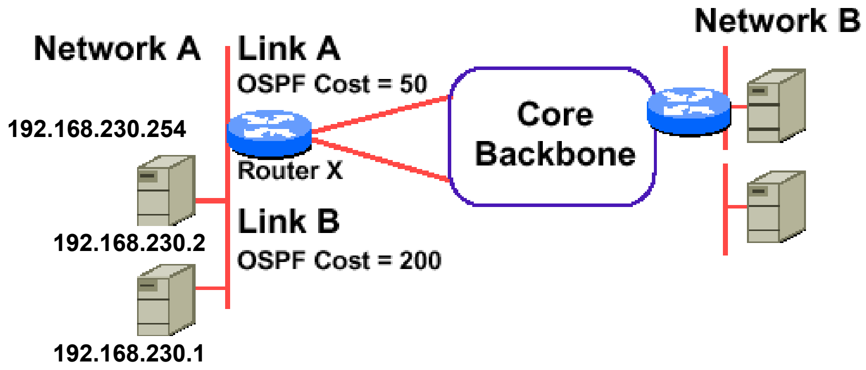
Administrative Distance

- Cisco routers use the routing concept of administrative distance;
- The administrative distance is a value ∈ [0, 255];
- When a router speaks multiple routing protocols, it will have a problem deciding the best path based on different metric structures and incompatible algorithms;
- Administrative distance rates the trustability of a routing protocol.

Route Source	Default Distance
Directly connected interface	0
Static route	1
EIGRP summary route	5
External BGP	20
Internal EIGRP	90
IGRP	100
OSPF	110
IS-IS (IP)	115
RIPv1, RIPv2	120
EGP	140
External EIGRP	170
Internal BGP	200
Unknown	255

Policy-based Routing Example

- Router X has two E1 links
  - Link A was assigned a lower OSPF cost
  - Router X is forced to forward IP traffic on Link A

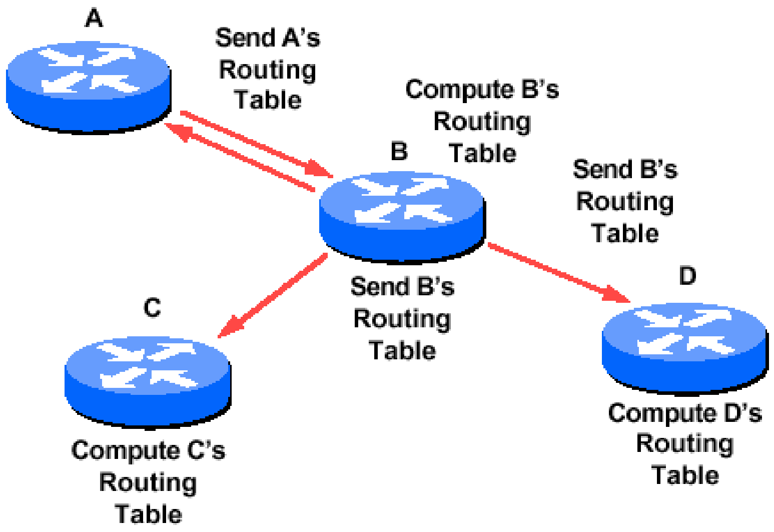


Characterization of Routing Protocols

Which information is exchanged?

- ***Distance-vector***
  Based on "rumors" (I heard that ...)
- ***Path-vector***
  Advertised de distance and the path to destination
- ***Link-state***
  Based on "advertising" (we only have exactly what you are looking for ...)

Distance-Vector Routing



A

Send A's Routing Table

Compute B's Routing Table

B

Send B's Routing Table

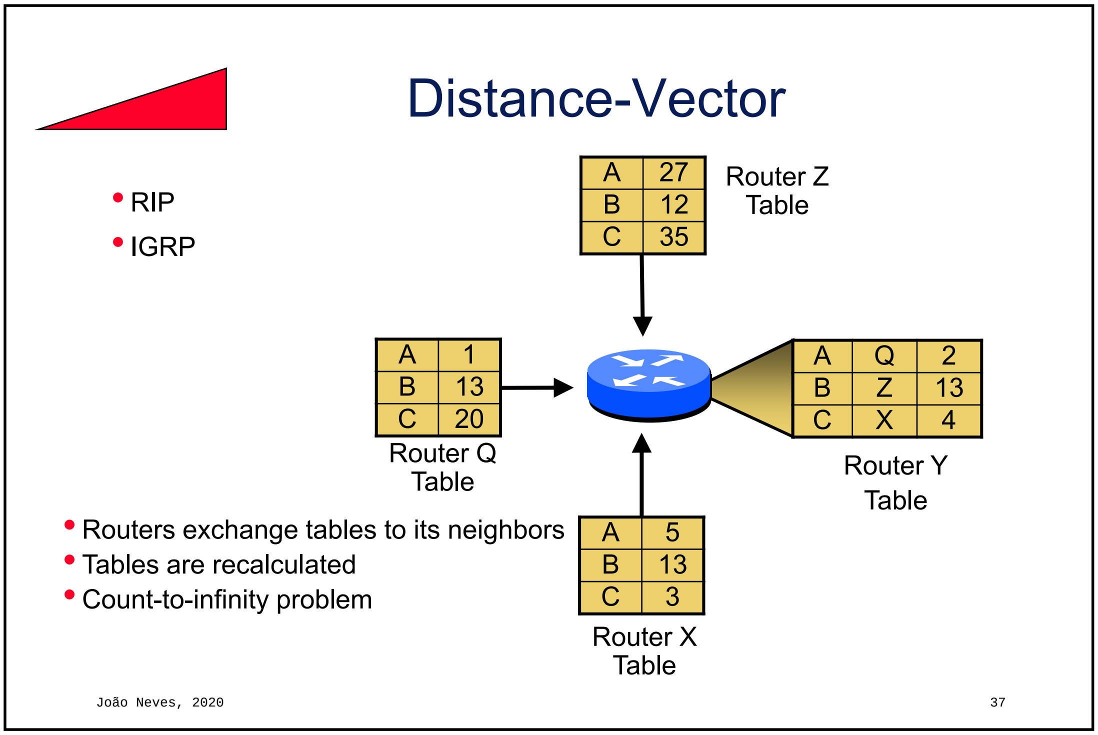
C

Send B's Routing Table

D

Compute C's Routing Table

Compute D's Routing Table



Distance-Vector

- RIP
- IGRP

Router Z Table

A	27
B	12
C	35

Router Q Table

A	1
B	13
C	20

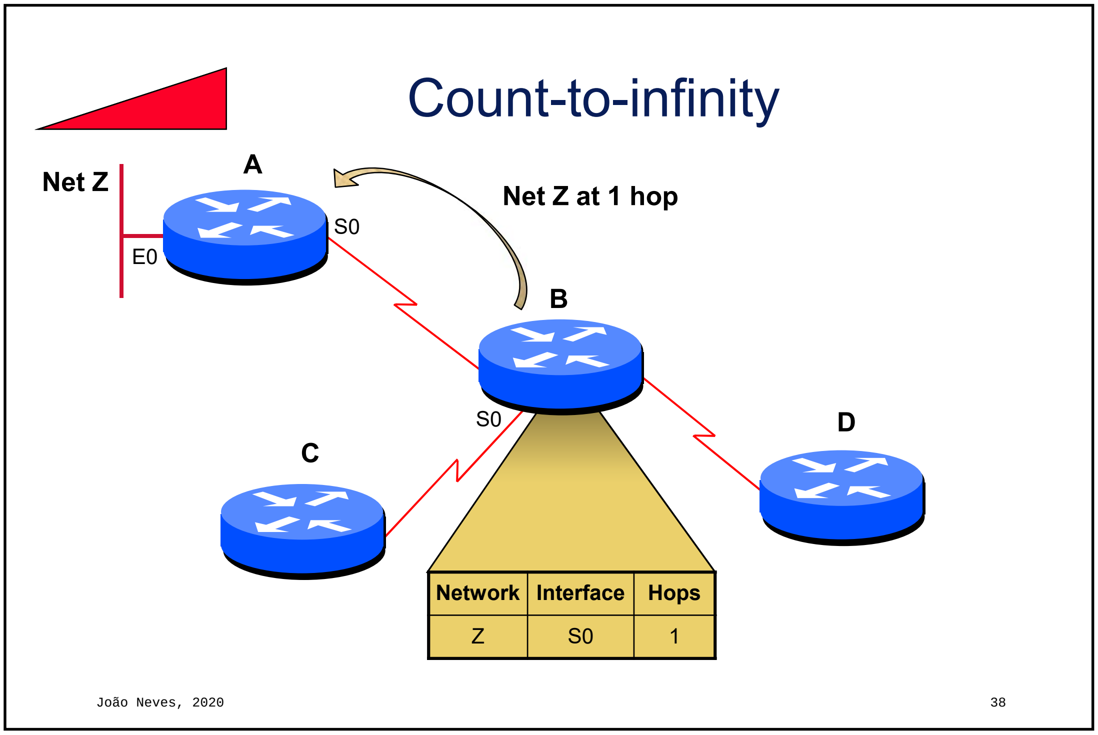
Router Y Table

A	Q	2
B	Z	13
C	X	4

Router X Table

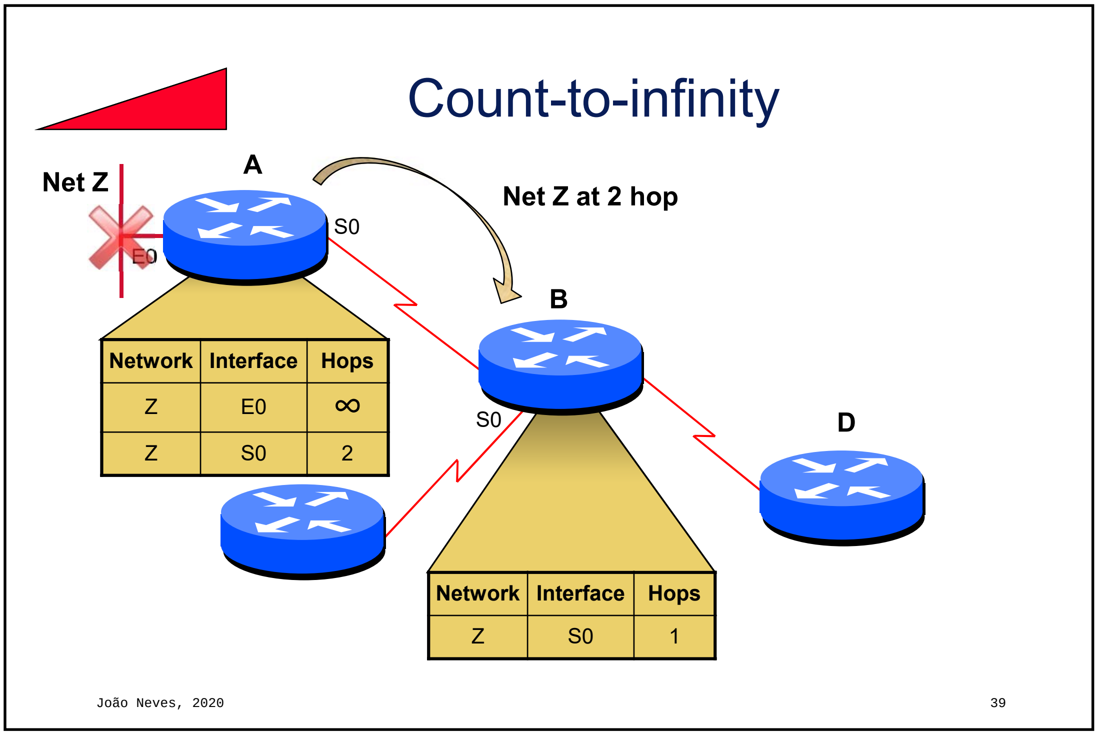
A	5
B	13
C	3

- Routers exchange tables to its neighbors
- Tables are recalculated
- Count-to-infinity problem



Count-to-infinity

Network	Interface	Hops
Z	S0	1

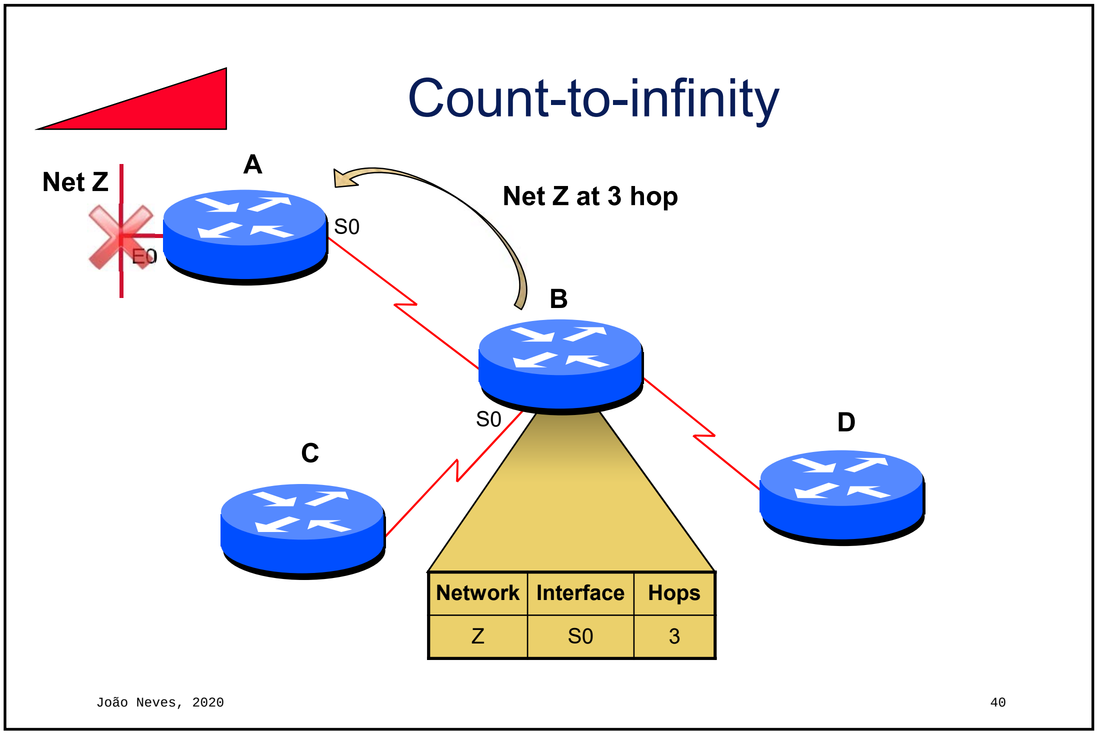


Count-to-infinity

Net Z at 2 hop

Network	Interface	Hops
Z	E0	∞
Z	S0	2

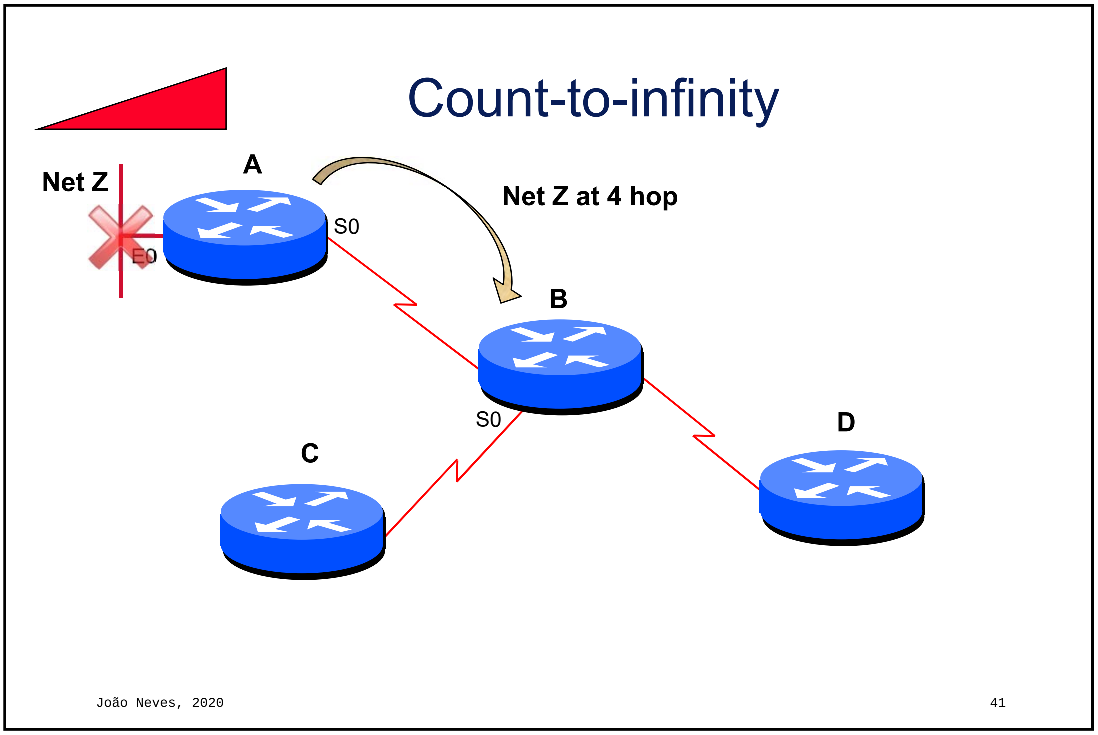
Network	Interface	Hops
Z	S0	1



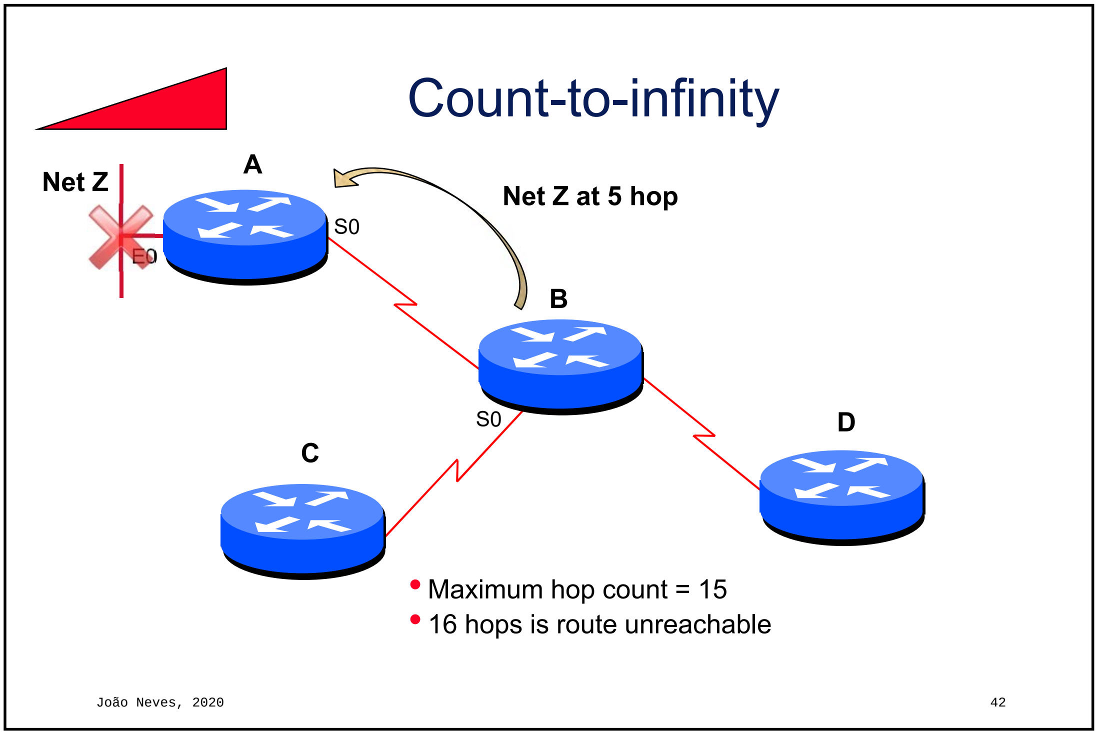
Count-to-infinity

Net Z at 3 hop

Network	Interface	Hops
Z	S0	3



Count-to-infinity

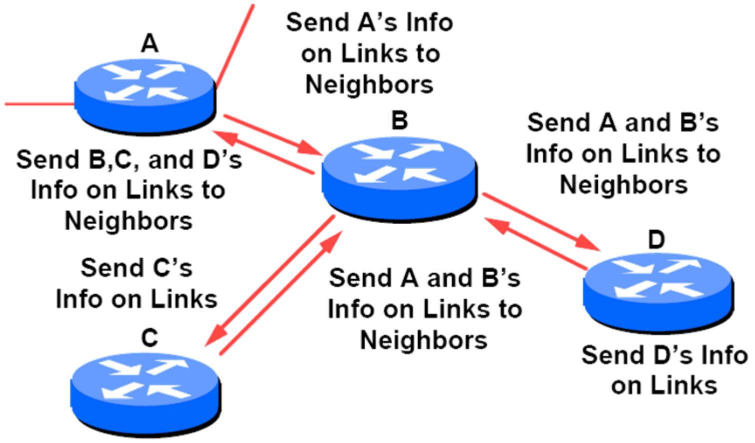


Count-to-infinity

- Maximum hop count = 15
- 16 hops is route unreachable

Link-State Routing

- All the routers compute the shortest paths using Dijkstra SPF algorithm



A

Send A's Info on Links to Neighbors

B

Send A and B's Info on Links to Neighbors

Send B,C, and D's Info on Links to Neighbors

D

Send C's Info on Links

Send A and B's Info on Links to Neighbors

C

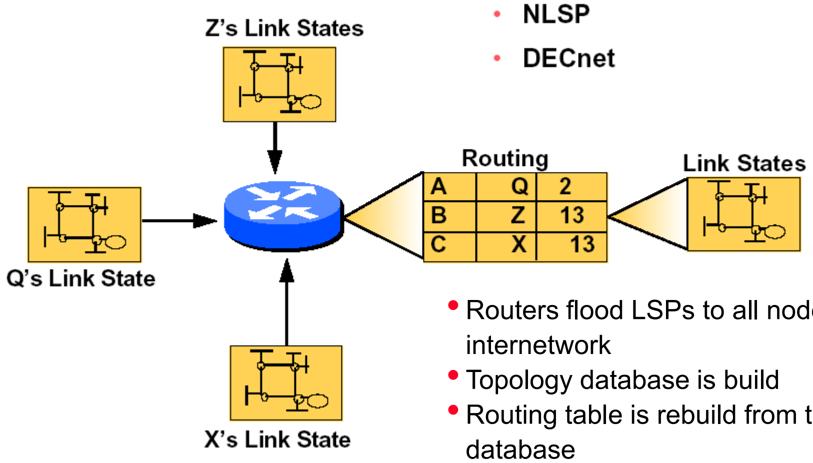
Send D's Info on Links

Link-State Routing

- Build Link-State Packet (LSP) that includes information learned from local discoveries
- Flood LSP for all routers
- Use the received LSPs to build a global network model
- Build the routing table

Link-State Routing

- OSPF
- NLSP
- DECnet



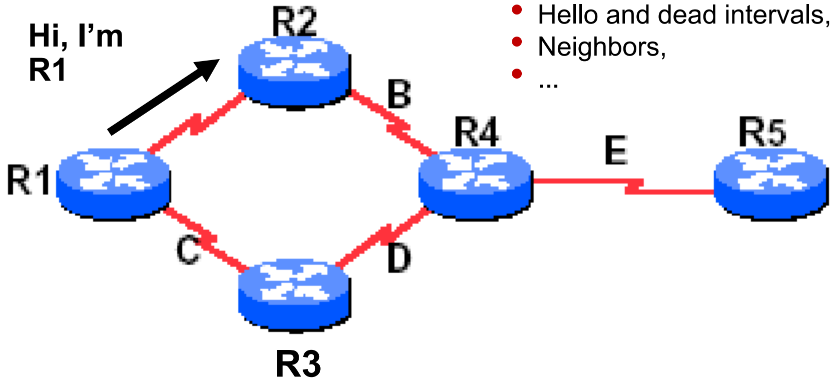
A	Q	2
B	Z	13
C	X	13

- Routers flood LSPs to all nodes in the internetwork
- Topology database is build
- Routing table is rebuild from topology database

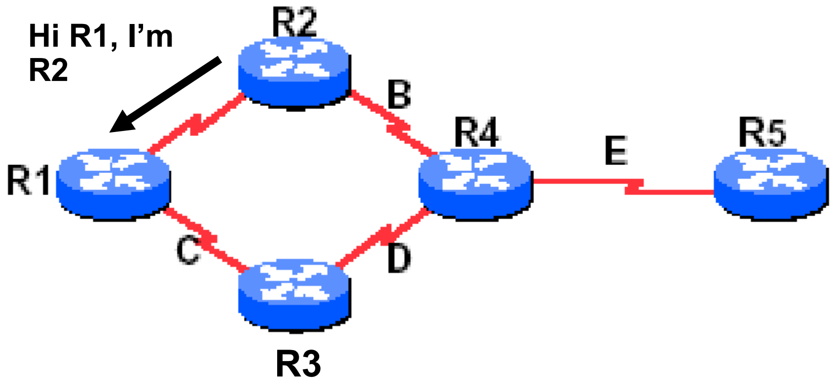
Neighbor Discovery

Hello packets include:

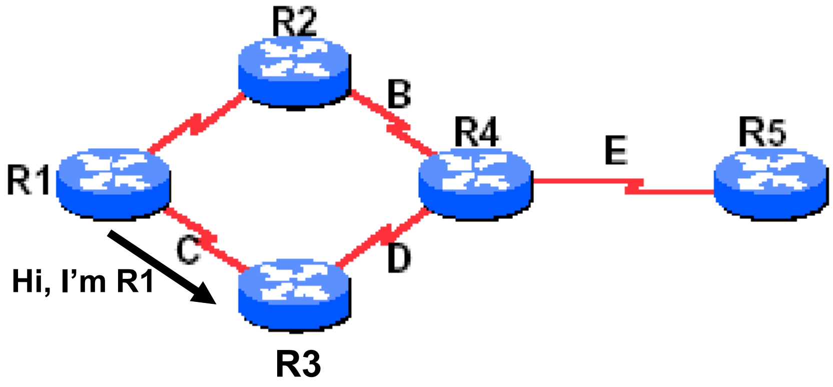
- Router ID,
- Hello and dead intervals,
- Neighbors,
- ...

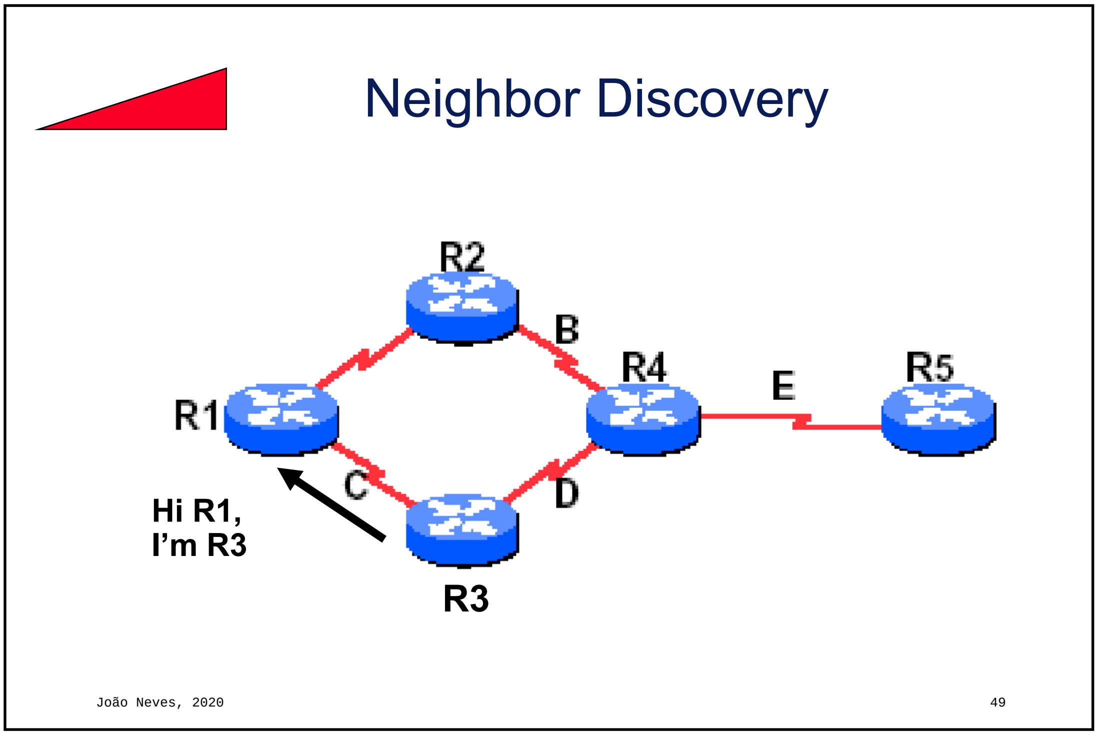


Neighbor Discovery



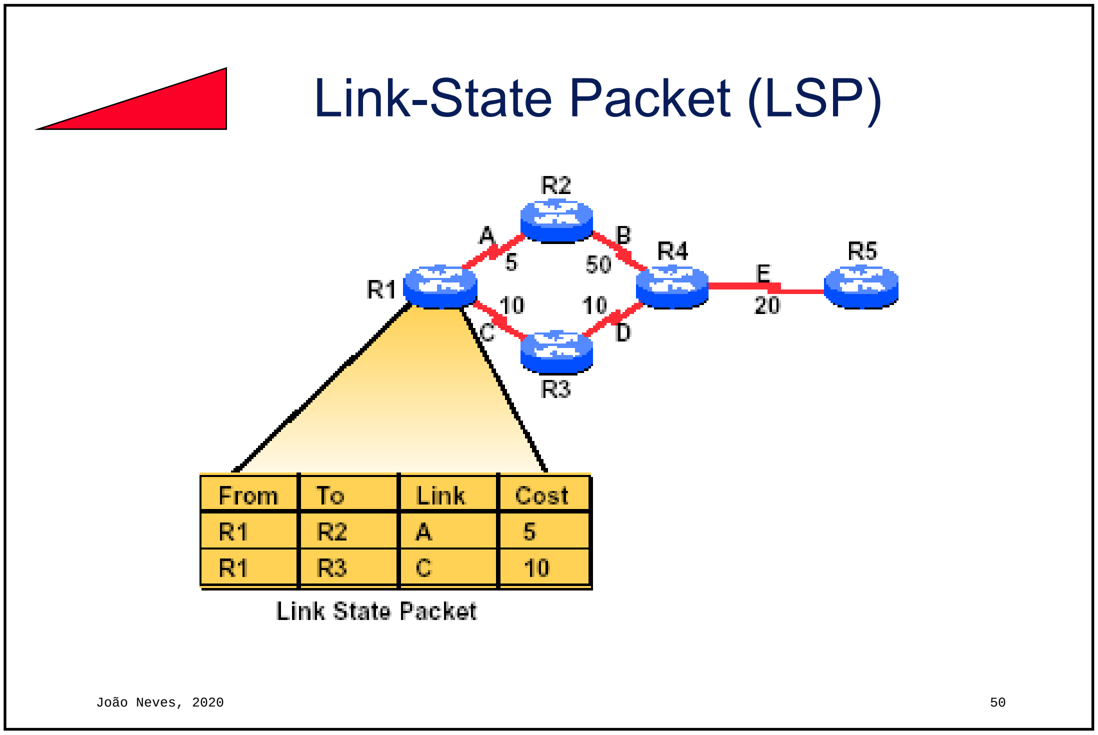
Neighbor Discovery





Neighbor Discovery

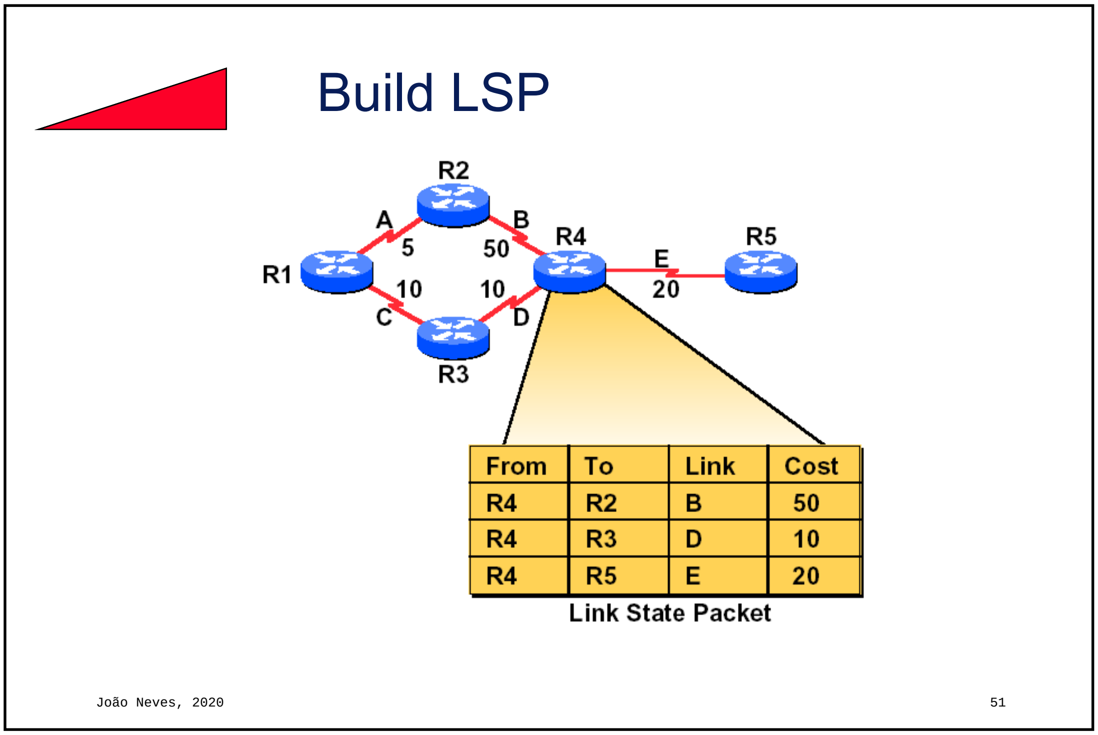
Hi R1,
I'm R3



Link-State Packet (LSP)

From	To	Link	Cost
R1	R2	A	5
R1	R3	C	10

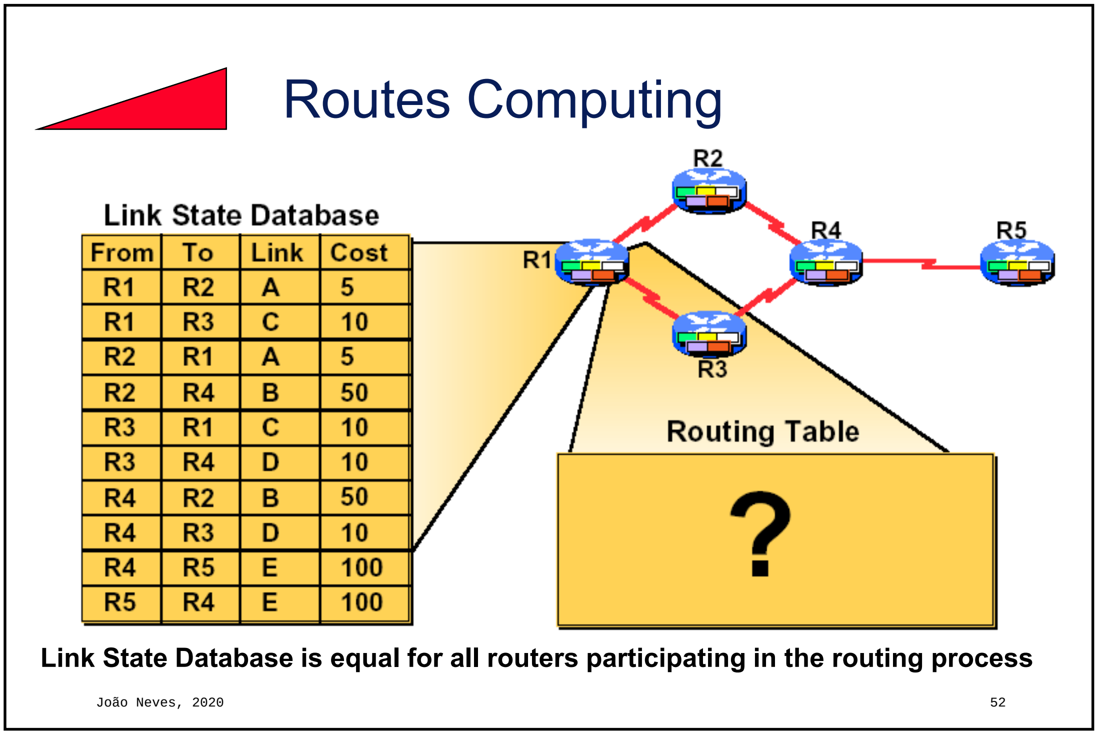
Link State Packet



Build LSP

From	To	Link	Cost
R4	R2	B	50
R4	R3	D	10
R4	R5	E	20

Link State Packet



Routes Computing

Link State Database

From	To	Link	Cost
R1	R2	A	5
R1	R3	C	10
R2	R1	A	5
R2	R4	B	50
R3	R1	C	10
R3	R4	D	10
R4	R2	B	50
R4	R3	D	10
R4	R5	E	100
R5	R4	E	100

Routing Table

?

Link State Database is equal for all routers participating in the routing process

Routing Protocols

IP classless Routing Algorithm

```
Route_IP_Datagram(datagram, routing_table):

Extract destination IP address, $I_D$, from datagram;
**If** prefix of $I_D$ matches address of any directly connected network **then**
    Send datagram to destination over that network
    ($I_D$ is mapped to a physical address and the datagram is
    encapsulated in a frame.)
**else**
    **foreach** routing table entry **do**
        Let N be the bitwise-and of $I_D$ and the net mask;
        **if** N equals the network address field of the entry **then**
            forward the datagrama to the specified next hop
    **end**
**If** the table contains a default route **then**
    send datagram to the default router
**else**
    return forwarding error;
```

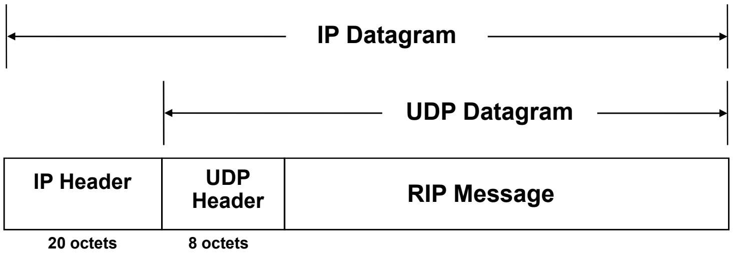
Routing Table

Network	Interface	Next Hop	Distance / Metric	Age	Status
194.117.30.0/24	Ethernet1	192.135.129.35	[90/307200]	02:03:50	D
198.113.178.0/24	Ethernet0	192.150.42.177	[110/20]	04:10:22	O
193.136.77.0/24	Ethernet0	192.150.42.177	[110/20]	03:36:50	O
193.136.32.0/21		192.150.42.178	[20/0]	1d20	B
192.135.129.160/30	Serial0				C
192.135.129.32/28	Ethernet1				C
192.150.42.0/24	Ethernet0				C
127.0.0.0/8	Null0				C
0.0.0.0/0		192.150.42.177	[1/0]		S
193.137.32.128/0		192.150.42.176	[1/0]		S

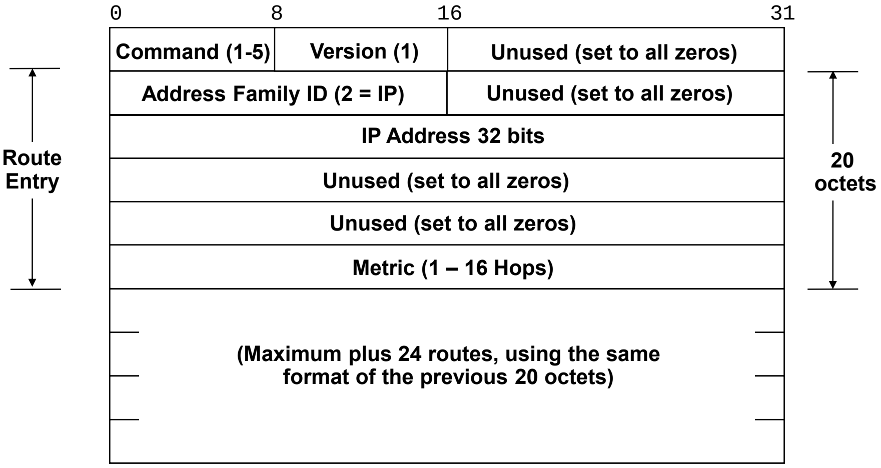
Routing Information Protocol (RIP)

- RIP is a standard and the oldest IP routing protocol
- It’s a Distance-Vector protocol
- Routers exchange full Routing tables
- Metric is Hop Count
- Mainly for it’s simplicity still is widely used

RIP encapsulated in UDP



RIP Message



Command (1-5)	Version (1)	Unused (set to all zeros)
Address Family ID (2 = IP)		Unused (set to all zeros)
IP Address 32 bits		
Unused (set to all zeros)		
Unused (set to all zeros)		
Metric (1 – 16 Hops)		
(Maximum plus 24 routes, using the same format of the previous 20 octets)		

RIP Message

0 8 16 31

Command (1-5)	Version (1)	Unused (set to all zeros)
Address Family ID (2 = IP)		Unused (set to all zeros)
IP Address 32 bits		
Unused (set to all zeros)		
Unused (set to all zeros)		
Metric (1 – 16 Hops)		
(Maximum plus 24 routes, using the same format of the previous 20 octets)		

20 octets

1 – REQUEST
2 – RESPONSE
3 – TRACE ON (obsolet)
4 – TRACE OFF (obsolet)
5 – RESERVED (Sun)

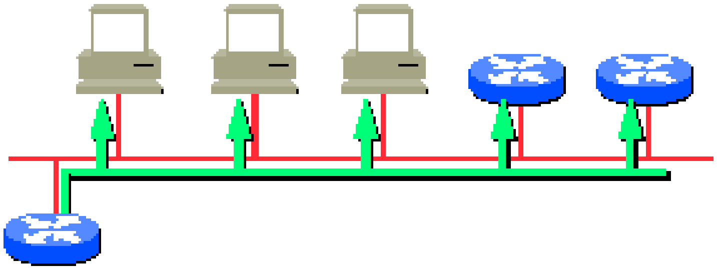
RIP Message

0 8 16 31

Command (1-5)	Version (1)	(must be zero)
Address Family ID (2 = IP)		(must be zero)
IP Address 32 bits		
(must be zero)		
(must be zero)		
Metric 1 .. 15, infinite = 16		
(Maximum plus 24 routes, using the same format of the previous 20 octets)		

20 octets

Broadcast Routing Updates



- RIP V1

- The routing updates are broadcasted every 30 second

RIP v2

- As RIPv1 is Distance-Vector
- Variable Length Subnet Mask
- Routing Updates by Multicast
- Additional Metrics (*throughput*)

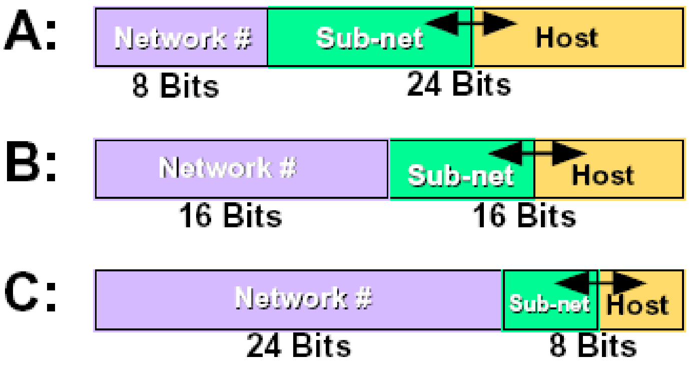
RIPv2 Message

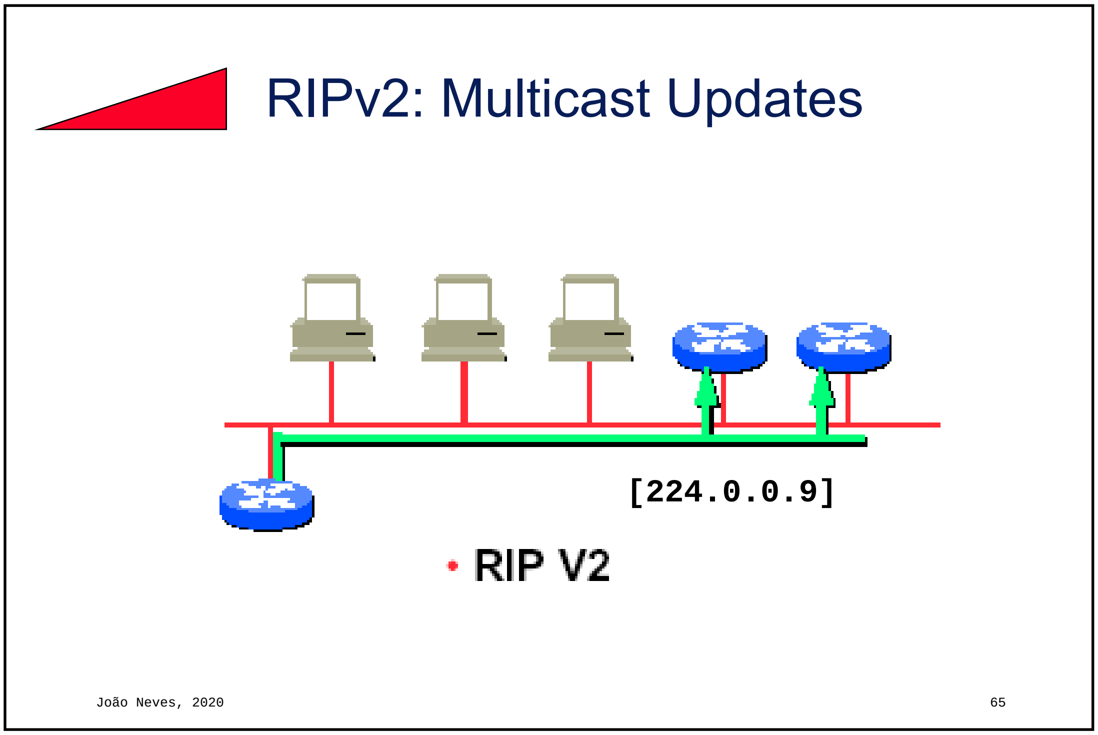
0	8	16	31
Command (1-5)	Version (2)	Routing Domain	
Address Family ID (2 = IP)		Route Tag (AS)	
IP Address			
32 bit Subnet Mask			
Next-Hop IP Address			
Metrics (1 – 16 Hops)			
(Maximum plus 24 routes, using the same format of the previous 20 octets)			

20 octets

Variable Mask

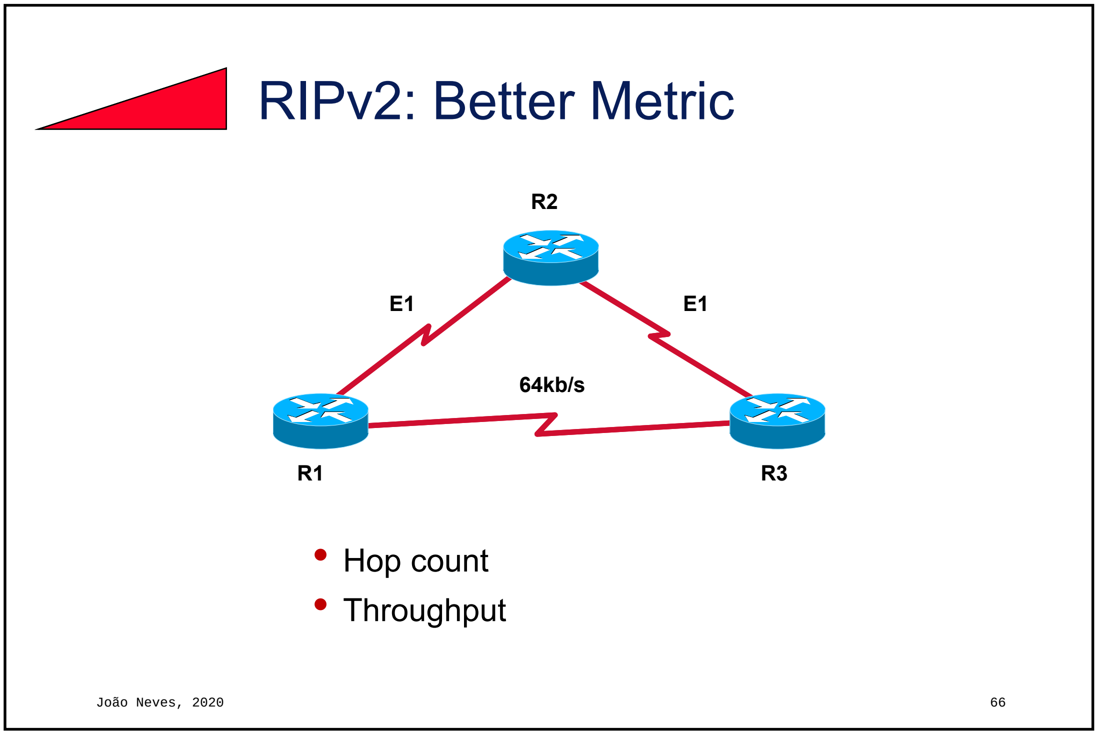
Variable Length Subnet Mask (VLSM)





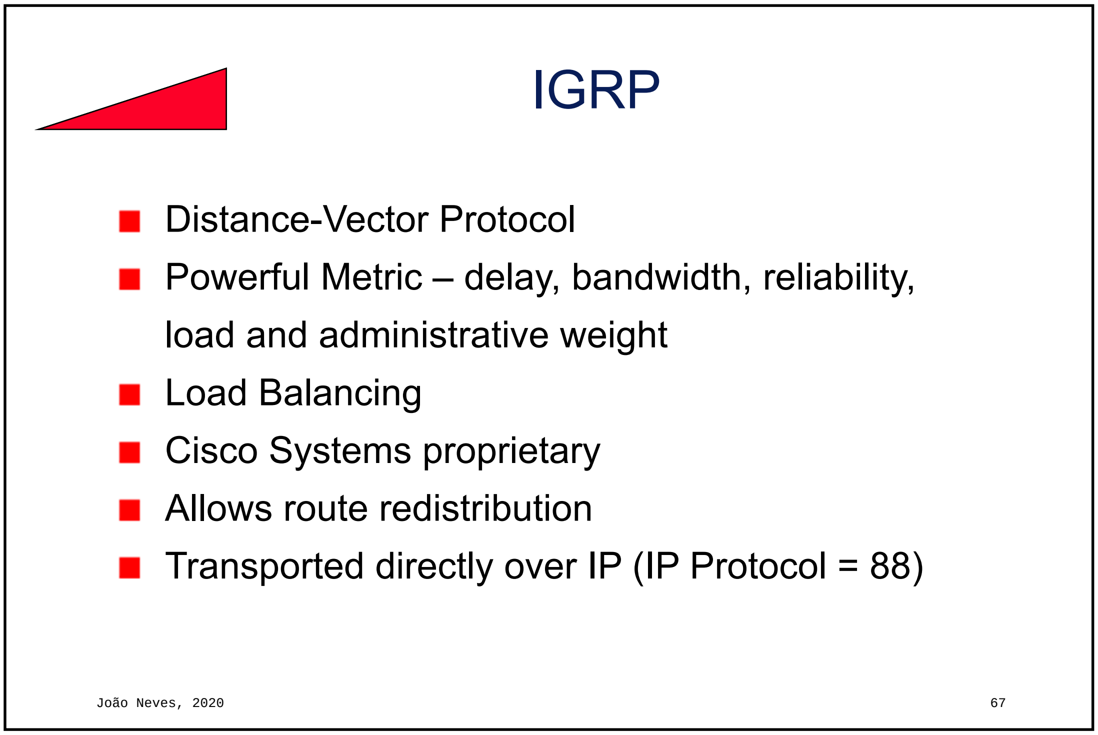
RIPv2: Multicast Updates

- RIP V2



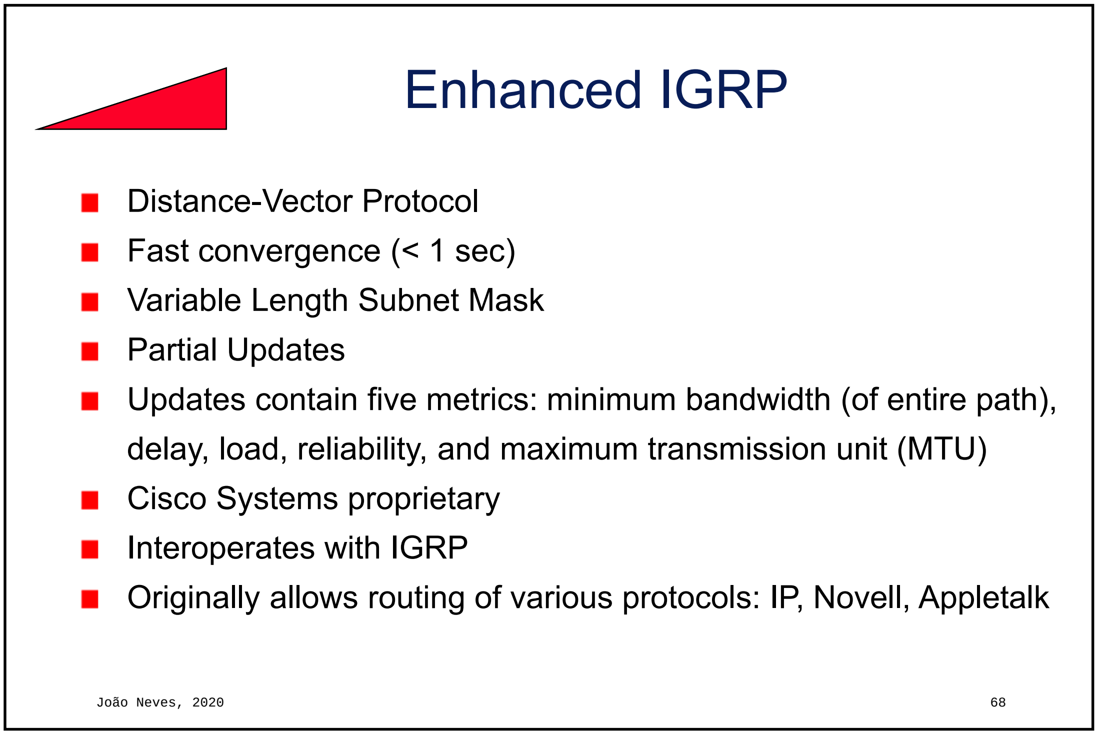
RIPv2: Better Metric

- Hop count
- Throughput



IGRP

- Distance-Vector Protocol
- Powerful Metric – delay, bandwidth, reliability, load and administrative weight
- Load Balancing
- Cisco Systems proprietary
- Allows route redistribution
- Transported directly over IP (IP Protocol = 88)



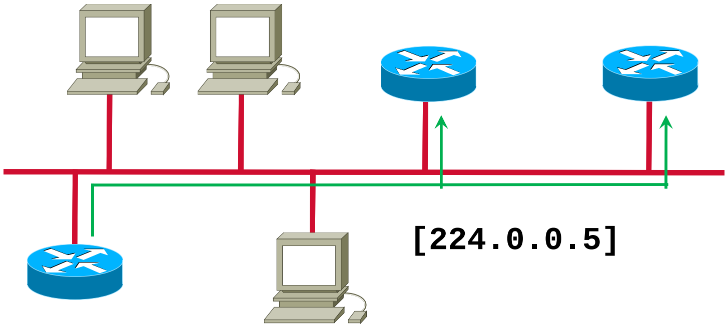
Enhanced IGRP

- Distance-Vector Protocol
- Fast convergence (< 1 sec)
- Variable Length Subnet Mask
- Partial Updates
- Updates contain five metrics: minimum bandwidth (of entire path), delay, load, reliability, and maximum transmission unit (MTU)
- Cisco Systems proprietary
- Interoperates with IGRP
- Originally allows routing of various protocols: IP, Novell, Appletalk

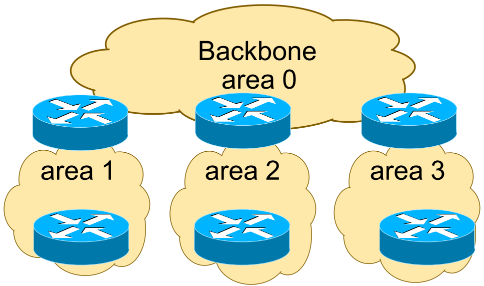
OSPF

- IETF Standard
- Link-State Protocol
- Accepts VLSM
- Small overhead
- Load Sharing
- Fast convergence
- Recognize areas
- Transport is done directly over IP (IP Protocol = 89)

Multicast Hello packets

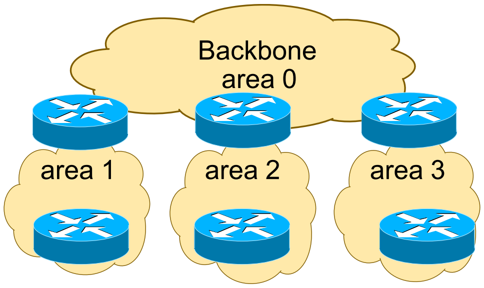


OSPF Areas



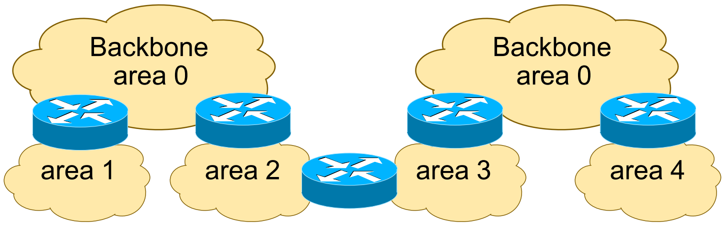
- OSPF recognizes one or more areas
- At least the backbone (area 0) must be configured
- The router may have interfaces in different areas
- All areas must be connected to the backbone
- Backbone must be contiguous

OSPF Areas Characteristics



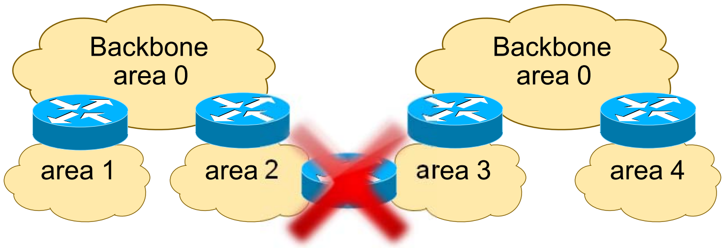
- Minimizes routing tables
- Topology of an area is invisible from outside
- Localizes impact of a topology change within an area
- Detailed LSA floods stops at the area boundary
- Hierarchical network design

OSPF Areas (cont.)



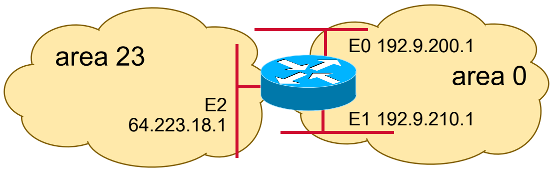
- How can we expand the areas coverage?

OSPF Areas (cont.)



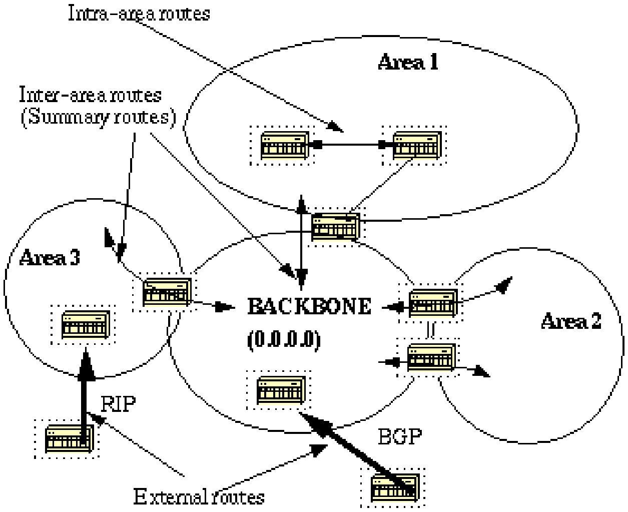
- The backbone must be contiguous!

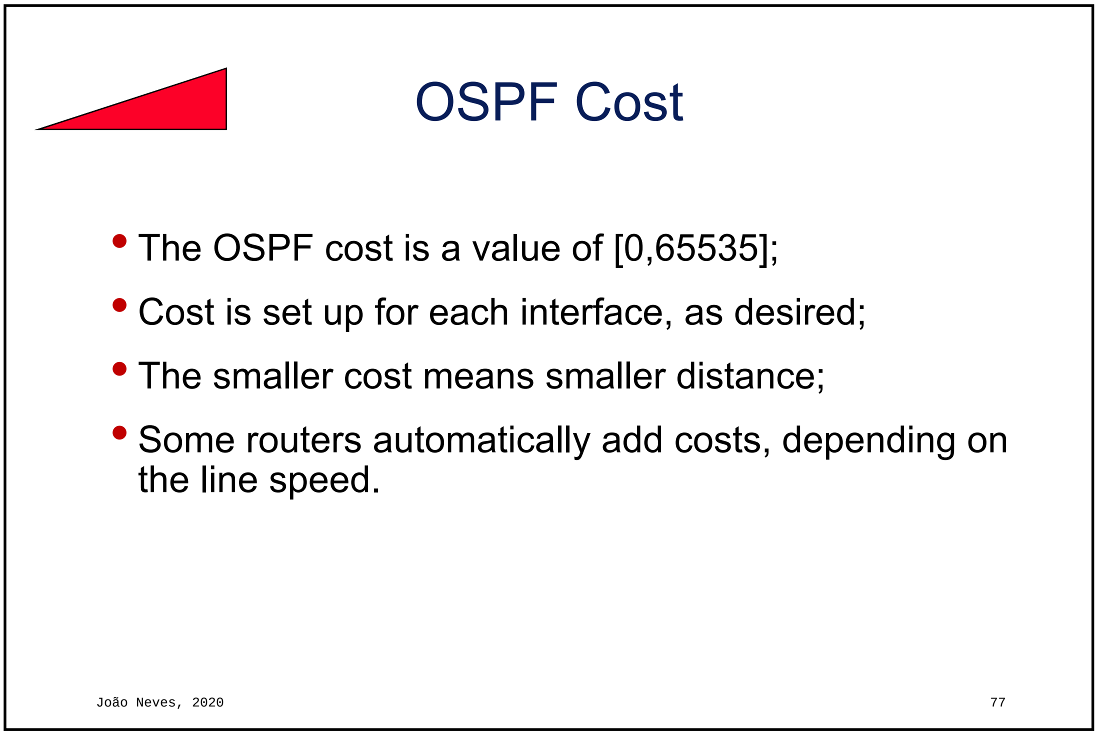
Example



```
Router#
interface Ethernet0
ip address 192.9.200.1 255.255.255.0

interface Ethernet1
ip address 192.9.210.1 255.255.255.0

interface Ethernet2
ip address 64.223.18.1 255.255.255.0

router ospf 100
network 192.9.0.0 0.0.255.255 area 0
network 64.223.18.1 0.0.0.0 area 23
```

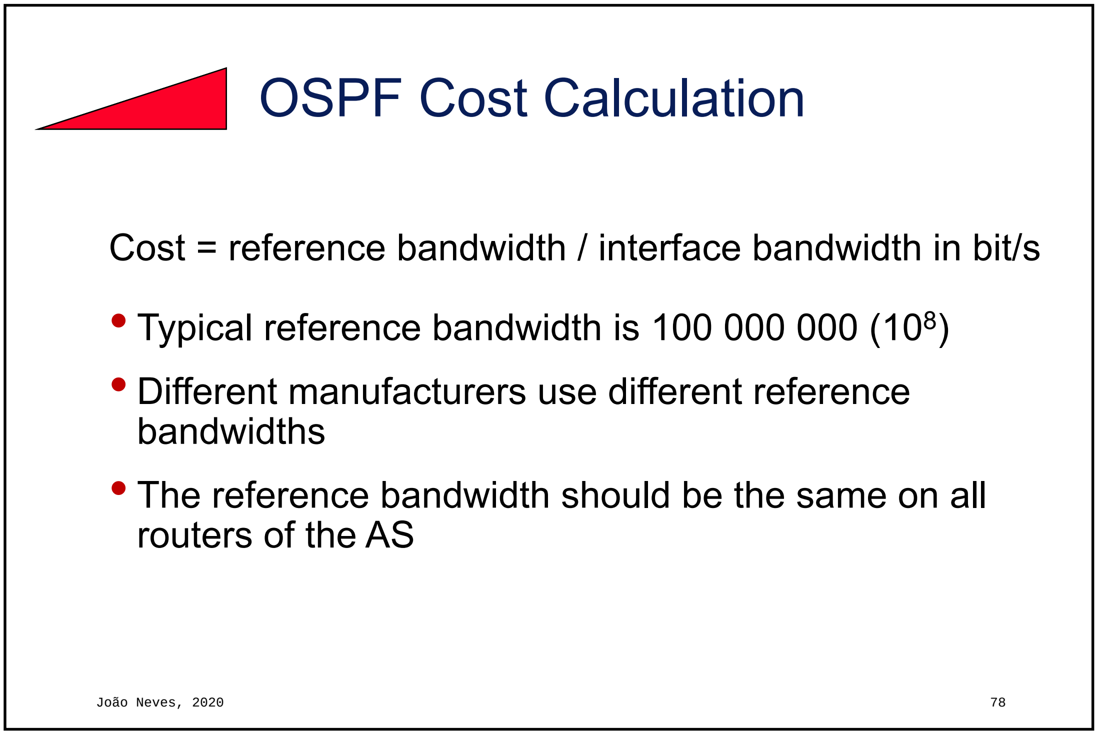
Area 0





OSPF Cost

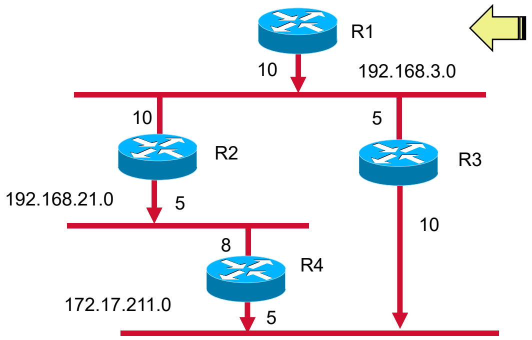
- The OSPF cost is a value of [0,65535];
- Cost is set up for each interface, as desired;
- The smaller cost means smaller distance;
- Some routers automatically add costs, depending on the line speed.



OSPF Cost Calculation

Cost = reference bandwidth / interface bandwidth in bit/s

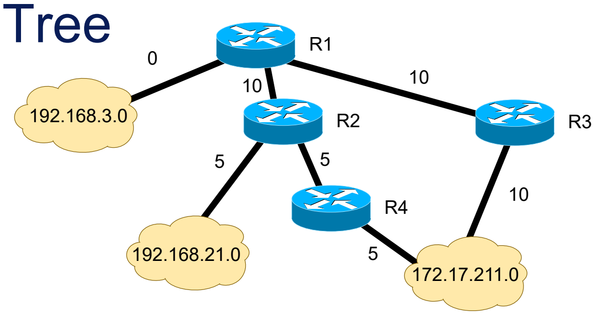
- Typical reference bandwidth is 100 000 000 (10^8)
- Different manufacturers use different reference bandwidths
- The reference bandwidth should be the same on all routers of the AS



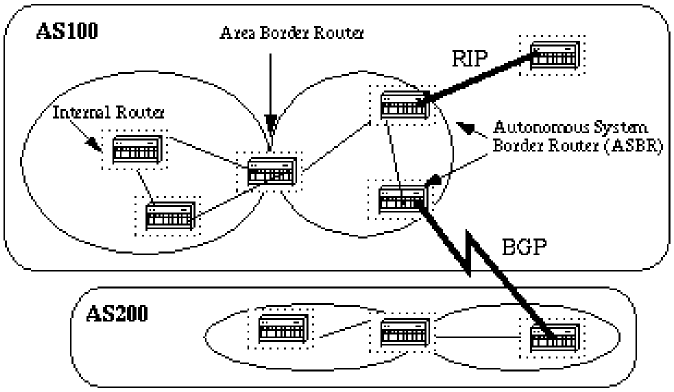
For the R1 shortest path tree computation, it will be the root of the tree

Shortest Path Tree

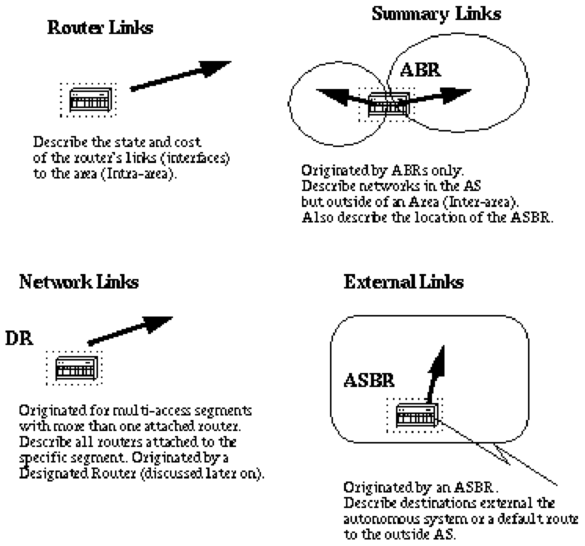
Directly connected networks will be reached with a cost of 0 and other networks will be reached according to the cost calculated in the tree



Area Border Router



Link-State Packets



Router Links

Describe the state and cost of the router's links (interfaces) to the area (Intra-area).

Summary Links

Originated by ABRs only. Describe networks in the AS but outside of an Area (Inter-area). Also describe the location of the ASBR.

Network Links

Originated for multi-access segments with more than one attached router. Describe all routers attached to the specific segment. Originated by a Designated Router (discussed later on).

External Links

Originated by an ASBR. Describe destinations external the autonomous system or a default route to the outside AS.

Default Convergence Time

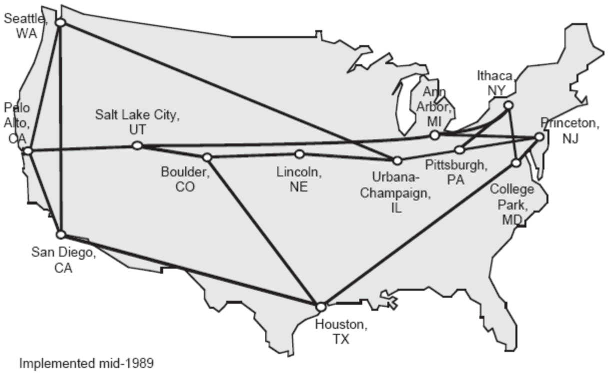
Routing Protocol	Time to Converge
• RIP	90s
• IGRP	270s
• EIGRP	~1s
• OSPF	~1s

Interior Routing

Traditional Distance Vector	Link-State	Advanced Distance Vector
RIP IGRP	OSPF	EIGRP

Exterior Routing

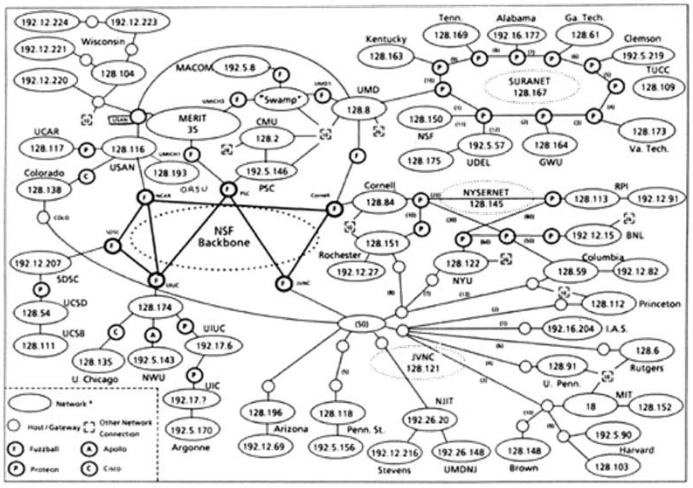
In 1990 the Internet was the National Science Foundation network (NSFNET) Backbone and connected networks [RFC1192]



Implemented mid-1989

NSFNET Backbone in 1987

NSFNet Physical Connectivity -- April 87



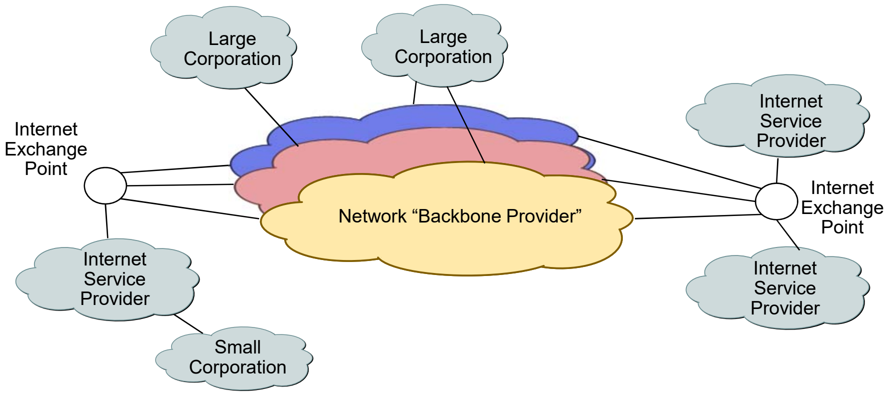
* For some networks internal structure (e.g. subnets) is suppressed.

ISI 4/17/87

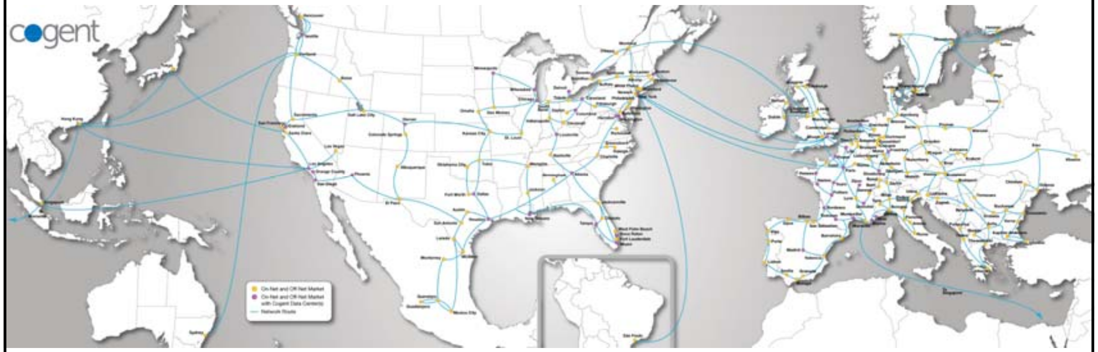
http://www.computerhistory.org/internet_history/internet_history_90s.shtml

Exterior Routing

Today, the Internet “is” a group of Network Backbone Providers (NBP) and peers.

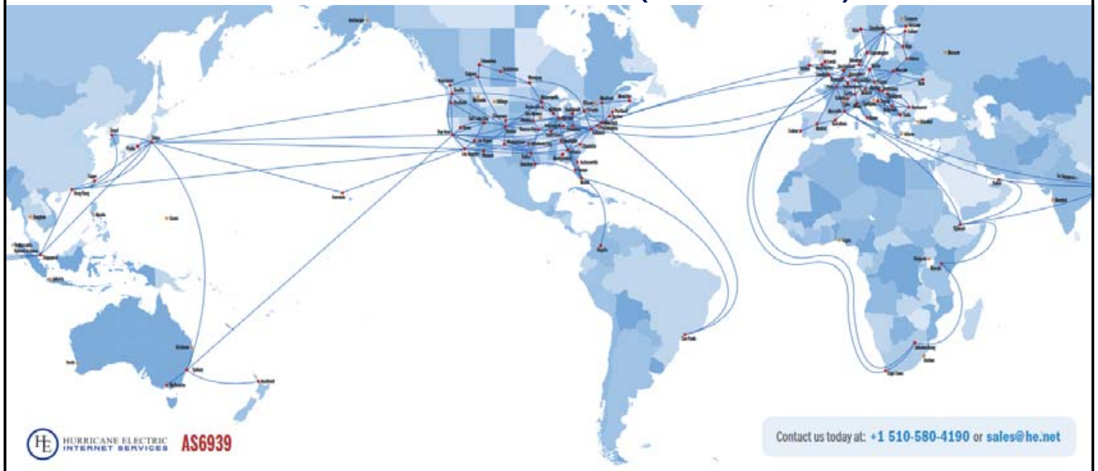

Large Corporation
Large Corporation
Internet Service Provider
Internet Exchange Point
Network “Backbone Provider”
Internet Exchange Point
Internet Service Provider
Internet Service Provider
Small Corporation

Backbone Provider AS174



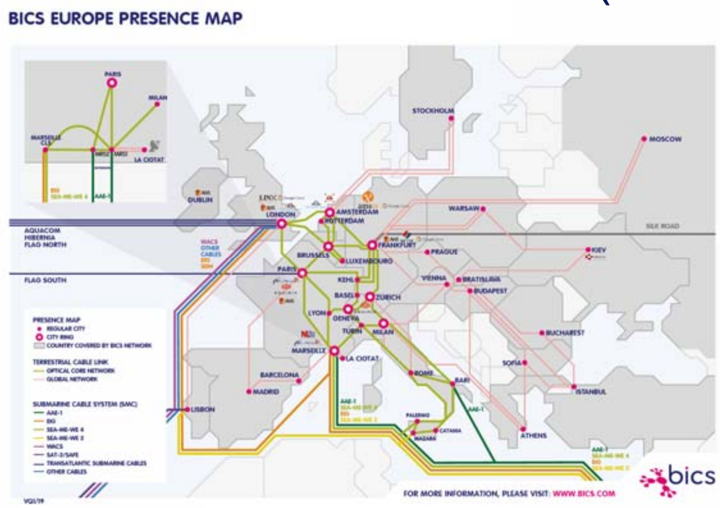
Source: March 2019, http://www.cogentco.com/en/network/network-map

Hurricane Electric IP Transit service (AS6939)


HURRICANE ELECTRIC INTERNET SERVICES
AS6939
Contact us today at: +1 510-580-4190 or sales@he.net

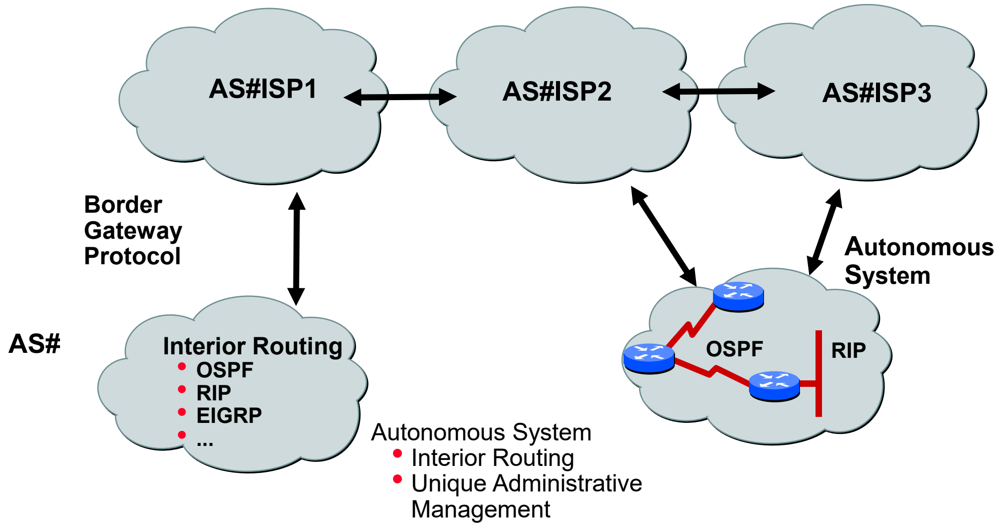
Source: March 2019, http://he.net/HurricaneElectricNetworkMap.pdf

BICS - Belgacom International Carrier Services SA (AS6774)



Source: https://bics.com/wp-content/uploads/2019/03/BICS-Europe-Map-Feb2019.png

Border Gateway Protocol



AS#ISP1

AS#ISP2

AS#ISP3

Border Gateway Protocol

AS#

Interior Routing
- OSPF
- RIP
- EIGRP
- ...

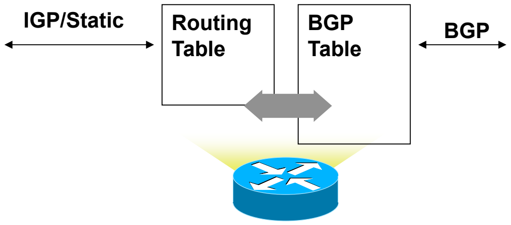
Autonomous System

OSPF

RIP

Autonomous System
- Interior Routing
- Unique Administrative Management

BGP and IP Routing Table



- The BGP process builds and manage a specific table, and the router uses the routing table for IP forwarding;
- The information from both tables may be redistributed, and usually is, but in a limited fashion;
- BGP selects a single best path to a destination, and inserts it in the IP routing table. IP forwarding decision is based on the routes in the IP table, NOT by the routes in the BGP table.

BGP4

- Path-Vector Protocol
- Inter-Autonomous System Communication
- Announces Reach ability Information
- Next Hop Paradigm
- Distributes routes information
- Supports Routing Politics
- Fast Convergence

BGP4 (cont.)

- Transport based on Transmission Control Protocol and sessions are established on port 179
- Incremental updates
- CIDR addresses
- Understands routes aggregation
- Authentication (is possible to verify who is sending routing updates)

TCP and Timers

- All communications between BGP peers are based on TCP;
- An IP connection must be established between the peers before a relationship can be set up;
- eBGP is designed to run only between directly connected neighbors;
- eBGP normally sets the time to live (TTL) in all packets to 1;
- You can you tell BGP to set the TTL to other value than 1 by declaring a multihop connection (**`ebgp-multihop`**);

TCP and Timers

- Keepalive interval indicates the time between successive Keepalive messages, used to maintain the session established in the absence of Updates;
- Hold time indicates the time that a router will wait without receiving an Update or Keepalive and before declaring a neighbor down;
- Each BGP speaker advertises its Hold time;
- Each BGP speaker compares its locally configured Hold time with the Hold time it receives from its peer, and chooses the lower of these two values;
- Keepalive timer is always set to one third of the Hold time.

Cisco BGP-4 Route selection criteria

1. **Prefer the path with the highest WEIGHT.**

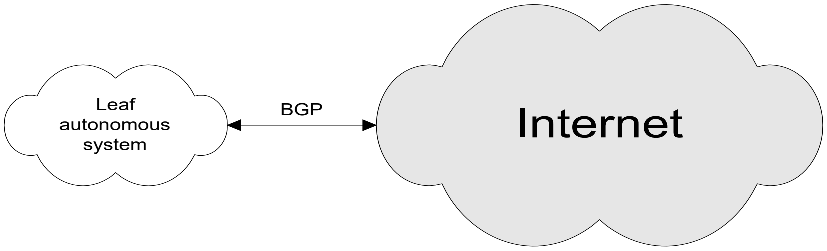
   Note: WEIGHT is a Cisco-specific parameter. It is local to the router on which it is configured.
2. **Prefer the path with the highest LOCAL_PREF (global within AS).**
3. **Prefer the path that was locally originated via a network or aggregate BGP subcommand or through redistribution from an IGP (next hop = 0.0.0.0).**
4. **Prefer the path with the shortest AS_PATH.**
5. **Prefer the path with the lowest origin type** (IGP < EGP < INCOMPLETE).
6. **Prefer the path with the lowest multi-exit discriminator (MED) –** from other AS**.**
7. **Prefer eBGP over iBGP paths.**
8. **Prefer the path with the lowest IGP metric to the BGP next hop** (closest IGP neighbor)**.**
9. **When both paths are external, prefer the path that was received first** (the oldest one).
10. **Prefer the route that comes from the BGP router with the lowest router ID.**

    The router ID is the highest IP address on the router, with preference given to loopback addresses. Also, you can use the bgp router-id command to manually set the router ID.
11. **Prefer the path that comes from the lowest neighbor address.**

    This address is the IP address that is used in the BGP neighbor configuration. The address corresponds to the remote peer that is used in the TCP connection with the local router.

Example No.1

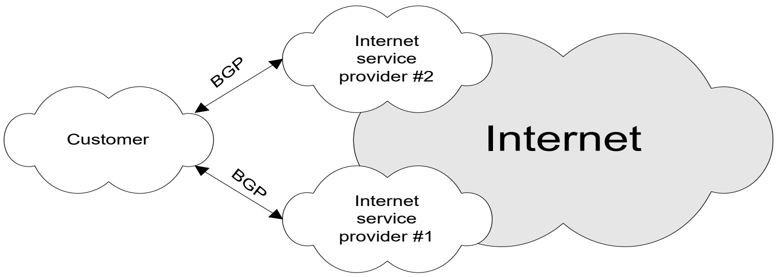
A large costumer or a small ISP connecting to the Internet



Example No.2

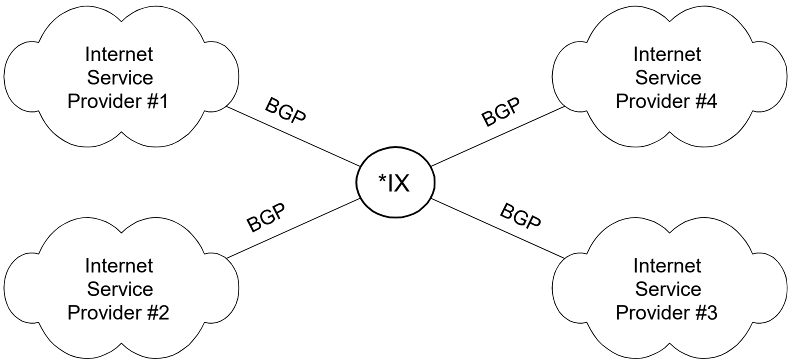
Customer with connectivity to more than one ISP (*multi-homed customer*)

- The use of BGP is mandatory
- Customer needs a unique public AS#
- Customer address space is independent of an ISP
- Large BGP Routing tables ➡ https://twitter.com/bgp4_table

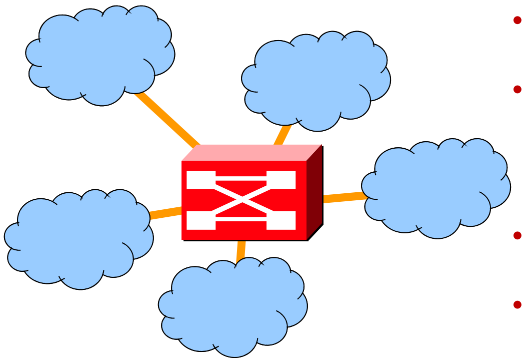


Example No.3

ISPs’ traffic exchange on an *Internet eXchange Point* (IXP)


Internet Service Provider #1
Internet Service Provider #4
BGP
BGP
*IX
BGP
BGP
Internet Service Provider #2
Internet Service Provider #3

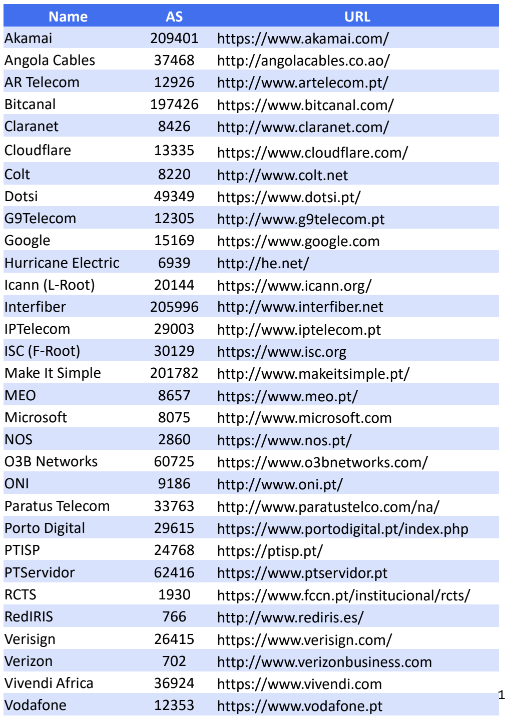
GigaPIX



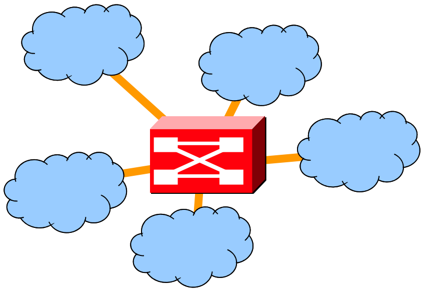
- GigaPIX is a not-for-profit Portuguese Internet Exchange Point managed and operated by FCT|FCCN;
- GigaPIX is present in four locations: three in Lisbon (two of the rooms are inside the Campus of Laboratório Nacional de Engenharia Civil, and the other at Equinix, in Prior Velho) and one in Porto (in Faculdade de Engenharia da Universidade do Porto);
- GigaPIX architecture covers the technical aspects of Physical and Data Link layers (layers 1 and 2);
- At the Physical Layer, GigaPIX provides each GigaPIX Member with a copper or fiber individual physical port;
- At the Data Link Layer the GigaPIX provides 100 Mbit/s, 1 Gbit/s and 10 Gbit/s Ethernet accesses.

Source: http://gigapix.pt

Who is connected to GigaPIX…



Name	AS	URL
Akamai	209401	https://www.akamai.com/
Angola Cables	37468	http://angolacables.co.ao/
AR Telecom	12926	http://www.artelecom.pt/
Bitcanal	197426	https://www.bitcanal.com/
Claranet	8426	http://www.claranet.com/
Cloudflare	13335	https://www.cloudflare.com/
Colt	8220	http://www.colt.net
Dotsi	49349	https://www.dotsi.pt/
G9Telecom	12305	http://www.g9telecom.pt
Google	15169	https://www.google.com
Hurricane Electric	6939	http://he.net/
Icann (L-Root)	20144	https://www.icann.org/
Interfiber	205996	http://www.interfiber.net
IPTelecom	29003	http://www.iptelecom.pt
ISC (F-Root)	30129	https://www.isc.org
Make It Simple	201782	http://www.makeitsimple.pt/
MEO	8657	https://www.meo.pt/
Microsoft	8075	http://www.microsoft.com
NOS	2860	https://www.nos.pt/
O3B Networks	60725	https://www.o3bnetworks.com/
ONI	9186	http://www.oni.pt/
Paratus Telecom	33763	http://www.paratustelco.com/na/
Porto Digital	29615	https://www.portodigital.pt/index.php
PTISP	24768	https://ptisp.pt/
PTServidor	62416	https://www.ptservidor.pt
RCTS	1930	https://www.fccn.pt/institucional/rcts/
RedIRIS	766	http://www.rediris.es/
Verisign	26415	https://www.verisign.com/
Verizon	702	http://www.verizonbusiness.com
Vivendi Africa	36924	https://www.vivendi.com
Vodafone	12353	https://www.vodafone.pt



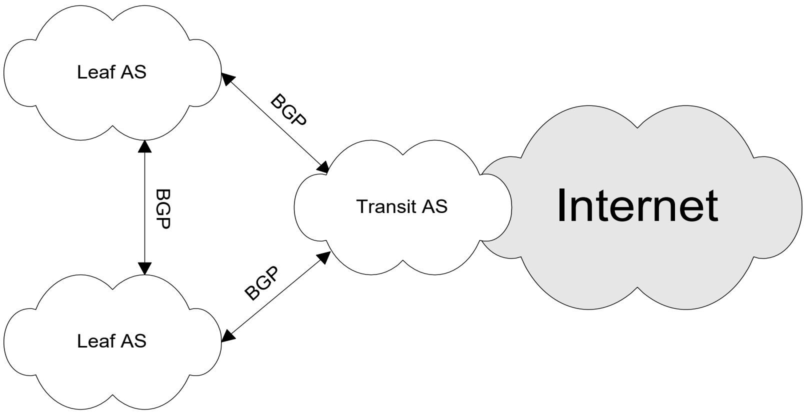
Source: https://gigapix.pt/en/members/

traceroute

```
# trace www.telepac.pt
Translating "www.telepac.pt"...domain server (193.136.192.40)
Translating "www.telepac.pt"...domain server (193.136.192.40) [OK]
Type escape sequence to abort.
Tracing the route to home.telepac.pt (194.65.62.76)

1 ROUTER11.GE.Porto.fccn.pt (193.137.4.18) 0 msec
  ROUTER11.GE.Porto.fccn.pt (193.137.4.2) 0 msec
  ROUTER11.GE.Porto.fccn.pt (193.137.4.18) 0 msec
2 ROUTER8.GE.Lambda.Lisboa.fccn.pt (193.137.1.241) 4 msec 8 msec 4 msec
3 ROUTER1.GE0-2-0.5.Lisboa.fccn.pt (193.137.0.11) 8 msec 8 msec 4 msec
4 ROUTER16.FE4-1.Lisboa.fccn.pt (193.137.0.21) 8 msec 8 msec 4 msec
5 GIGAPIX.telepac.pt (193.136.250.30) 4 msec 8 msec 12 msec
6 halley.telepac.net (194.65.12.161) 8 msec 8 msec 8 msec
7 lcatrt1.telepac.net (213.13.135.137) 12 msec 12 msec 12 msec
8 katrt4.telepac.net (213.13.135.202) 8 msec 8 msec 12 msec
9 diogo.sbd.telepac.pt (194.65.62.75) 8 msec * 8 msec
```

Example No.4

Use of an transit AS to transport traffic from other AS's



BGP hijacks

- BGP hijacks take place when an ISP announces the wrong Internet route to a specific destination;
- In most cases, BGP hijacks are accidents, such as typos;
- But in some cases, malicious ISPs intentionally announce a wrong BGP route in order to hijack traffic meant for particular targets, such as critical DNS servers, financial services, government sites, etc...
- These malicious BGP hijacks make traffic to those targets to flow through the malicious ISP's network, where it can sniff its content or perform Man-in-the-Middle attacks;
- Bitcanal (AS3266) *was the biggest BGP hijack offender in recent years, earning the nickname of "BGP hijack factory“.*

BGP hijacks

- BGP hija rong Internet r
- In most ca ;
- But in so ce a wrong BG cular targets, such as c NS servers, financial services, government sites, etc...
- These malicious BGP hija e traffic to those targets to flow through the maliciou etwork, where it can sniff its content or perform Man-i dle attacks;
- Bitcanal (AS3266) *was the GP hijack offender in recent years, earning the nick GP hijack factory“.*

Bitcanal (AS3266)
aka
Ebonyhorizon Telecomunicacoes S.A.
(AS42229)
Vila Nova de Gaia, Portugal…

BGP hijack factory - Case Study

- Hijack Internet address ranges that have gone unused for periods of time and “unannounced”;
- Announce to the Internet that their hosting facilities was the authorized location for those IP address blocks;
- After obtaining a chunk of IP addresses, Bitcanal would apparently sell or lease the space to spammers, who would then begin sending junk email from those addresses;
- Much of the hijacked address space routed by Bitcanal and used by its customers, was once assigned to business entities that no longer exist;
- But some were assigned to active organizations, such as the Texas State Attorney General’s office, as well as addresses managed by the U.S. Department of Defense!...

BGP hijack factory - Case Study

- Hijack Internet address ranges that have gone unused for periods of t
- A the
- nd used by business entities that n
- But some were assigned to active organizations, such as the Texas State Attorney General's office, as well as addresses managed by the U.S. Department of Defense!...

"Shutting down the BGP Hijack Factory"
Doug Madory, ORACLE Dyn,
Jul 10, 2018

https://dyn.com/blog/shutting-down-the-bgp-hijack-factory/

Routes Symmetry Validation

- To validate the symmetry of our traffic, in result of the routes propagation, the best approach is to consult a looking glass...

Looking Glass Sites

- **Route Views Project**
  http://www.routeviews.org/
- **BGP4.net**
  http://www.bgp4.net/rs
- **CERN**
  http://lg.cern.ch/

Route Views

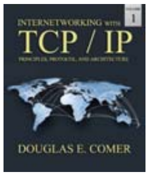
- Telnet to the server and make queries using a CLI *a la* IOS or query a Cisco box...

```
# telnet route-views2.routeviews.org
Trying 128.223.51.102...
Connected to route-views2.routeviews.org.
Escape character is '^]'.

Hello, this is zebra (version 0.95a).
Copyright 1996-2004 Kunihiro Ishiguro.

route-views2.routeviews.org> show ip bgp regexp AS#
```

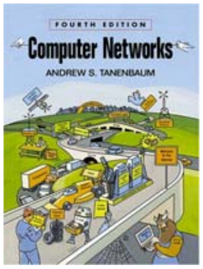
Routing Protocols

Protocols	RIP	HELLO	IGRP	OSPF	EIGRP	IS-IS	EGP	BGP4
Type	IGP	IGP	IGP	IGP	IGP	IGP	EGP	EGP
Algorithm	DV	DV	DV	SPF	DUAL	SPF	DV	PV
Metric	Hop count	Delay	Speed	Arb.	Speed	Arb.	Policy	Policy
Convergence	Slow	Unstable	Medium	Fast	Fast	Fast	Slow	Fast
Standard	IETF	No	No	IETF	No	OSI	Hist.	IETF
Complex	No	No	No	Yes	Yes	Yes	No	Yes
VLSM	No	No	No	Yes	Yes	Yes	No	Yes

Bibliography



- **Comer, Douglas E.**
  Internetworking with TCP/IP (VOL I)
  Pearson, 6th Edition (2014)
  ISBN-10: 0-13-608530-X
  ISBN-13: 978-0-13-608530-0



- **Tanenbaum, Andrew S.**
  Computer Networks
  Prentice Hall International Editions
  4th edition (August 9, 2002)
  ISBN 0-13-066102-3

- “A Border Gateway Protocol 4 (BGP-4)”, RFC 4271, January 2006