OSPF and BGP-4

KRNET 2000

June, 2000

AT&T Labs
Jeonghoon Mo

Overview

Routing and Internet Basic

- Internet Routing Architecture
- Interior Routing Protocol
- Exterior Routing Protocol
- Router, Routing Protocol and Routing Table
- Scalability Concerns
- CIDR

OSPF

BGP-4

Routing Example

Traceroute result from Berkeley to www.krnet.or.kr

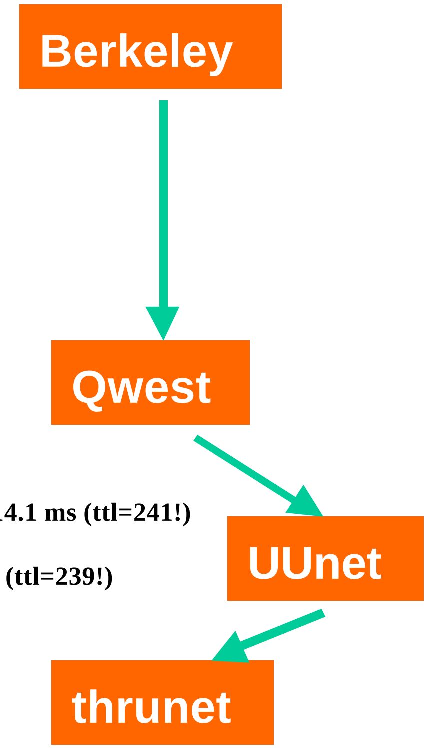
```
 1  cnr239net.EECS.Berkeley.EDU (128.32.239.1)  1.67 ms  0.585 ms  0.540 ms
 2  gigE5-0-0.inr-010-cory.Berkeley.EDU (169.229.1.45)  0.801 ms  0.768 ms  0.723 ms
 3  fast4-0-0.inr-002-eva.Berkeley.EDU (128.32.0.34)  1.31 ms  1.19 ms  1.20 ms
 4  pos0-2.inr-000-eva.Berkeley.EDU (128.32.0.73)  1.10 ms  1.34 ms  1.49 ms
 5  pos3-0.c2-berk-gsr.Berkeley.EDU (128.32.0.90)  1.78 ms  1.97 ms  1.44 ms

 6  SUNV--BERK.POS.calren2.net (198.32.249.14)  2.58 ms  2.66 ms  2.69 ms
 7  198.32.249.94 (198.32.249.94)  3.25 ms  2.64 ms  3.14 ms
 8  198.32.249.86 (198.32.249.86)  3.3 ms  3.51 ms  3.42 ms
 9  63.237.208.1 (63.237.208.1)  3.85 ms  3.91 ms  3.89 ms

10  svl-core-03.inet.qwest.net (205.171.14.85)  3.48 ms  3.98 ms  4.77 ms
11  sjo-core-03.inet.qwest.net (205.171.14.102)  3.66 ms  3.52 ms  3.16 ms
12  sjo-brdr-04.inet.qwest.net (205.171.22.121)  3.60 ms  4.3 ms  3.65 ms
13  205.171.4.98 (205.171.4.98)  14.3 ms (ttl=241!)  14.4 ms (ttl=241!)  14.6 ms (ttl=241!)

14  175.at-6-0-0.XR2.SAC1.ALTER.NET (152.63.50.190)  15.1 ms (ttl=241!)  14.1 ms (ttl=241!)  14.1 ms (ttl=241!)
15  184.ATM7-0.GW3.SAC1.ALTER.NET (152.63.51.117)  15.3 ms  14.8 ms  16.1 ms
16  thrunet2-gw.customer.alter.net (157.130.194.50)  152 ms (ttl=239!)  151 ms (ttl=239!)  153 ms (ttl=239!)

17  210.117.67.193 (210.117.67.193)  151 ms  151 ms  152 ms
18  210.117.67.124 (210.117.67.124)  152 ms (ttl=47!)  155 ms (ttl=47!)  153 ms (ttl=47!)
19  210.221.0.135 (210.221.0.135)  154 ms (ttl=46!)  153 ms (ttl=46!)  153 ms (ttl=46!)
20  s211-33-122-139.thrunet.ne.kr (211.33.122.139)  154 ms  153 ms  152 ms
```

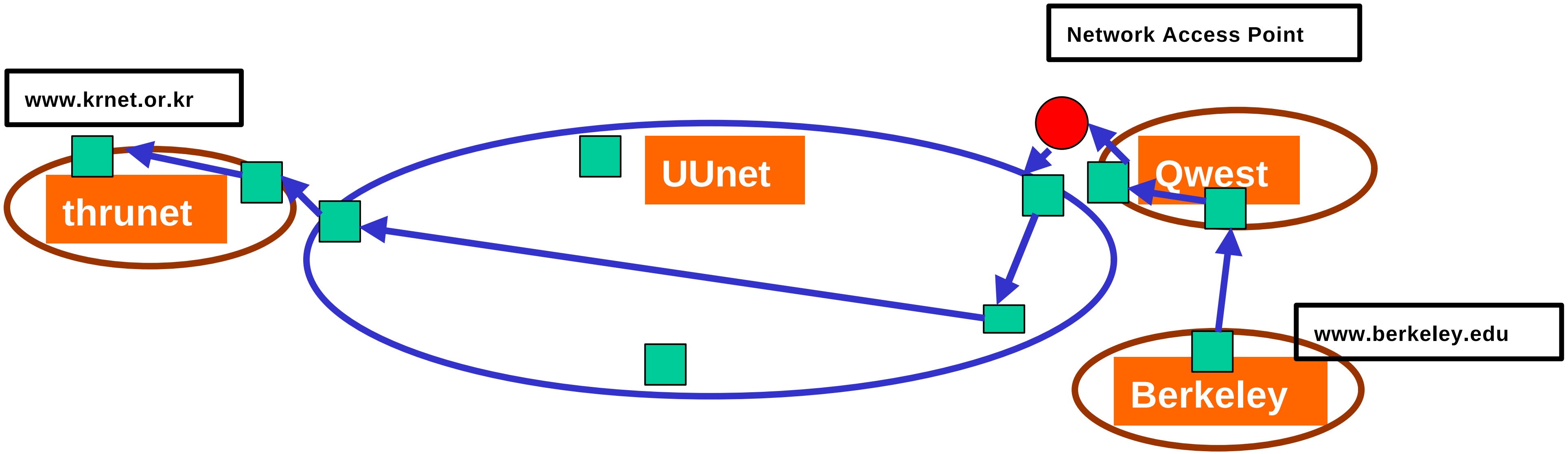


Passes 20 hops and 4 different ASes.

Routing Example

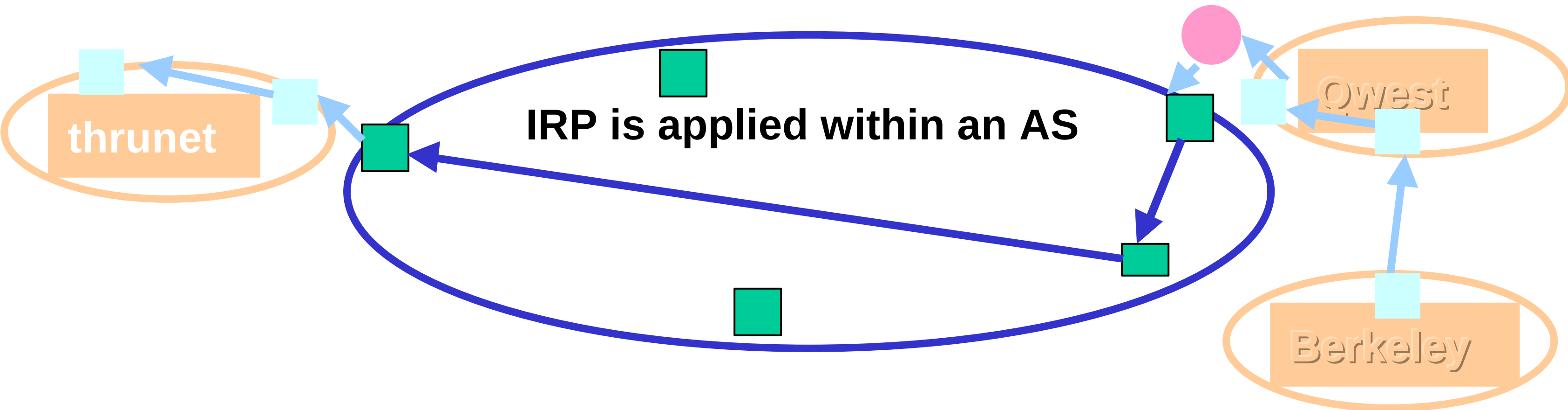
routers

AS is a network operated by an independent authority.



There are 4 ASes, Berkeley, Qwest, UUnet, and thrunet in this example.
The size of an AS varies from a small campus network to a global network.
Internet routing can be classified into Routing within an AS (IRP)
and Routing between ASes (ERP).

Interior Routing Protocol (IRP)



IRP or IGP (Interior Gateway Protocol)

Determine connectivity and routes within a single AS

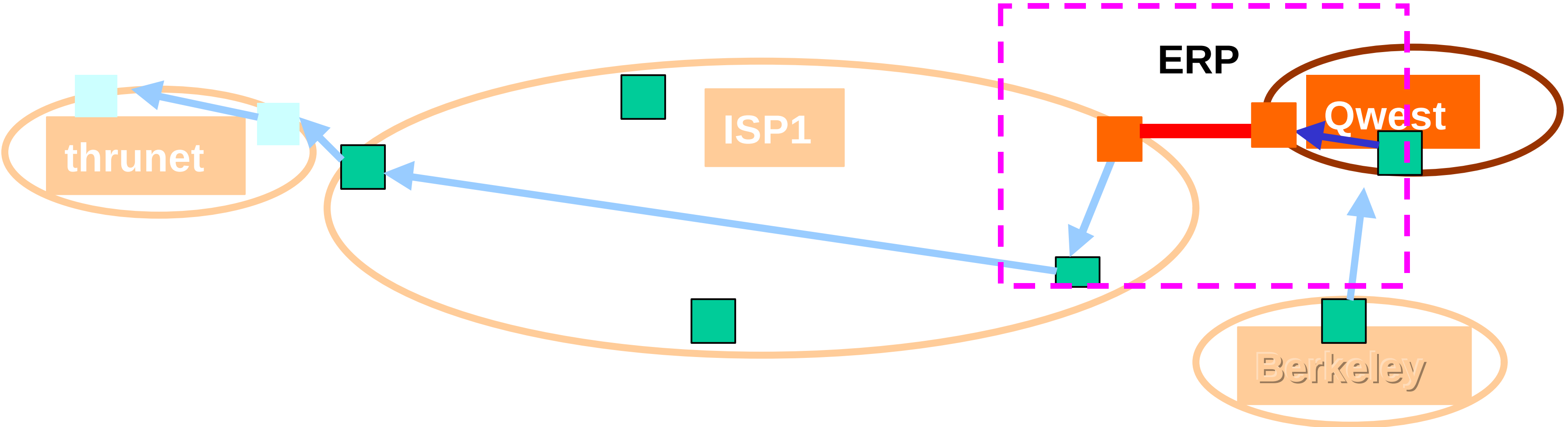
Dynamic route revision as connectivity changes

Distance Vector Algorithms and Link State Algorithms

OSPF, RIP, Integrated IS-IS, IGRP are IRPs.

Internet Service Provider (ISP)
AS in business of providing connectivity

Exterior Routing Protocol (ERP)



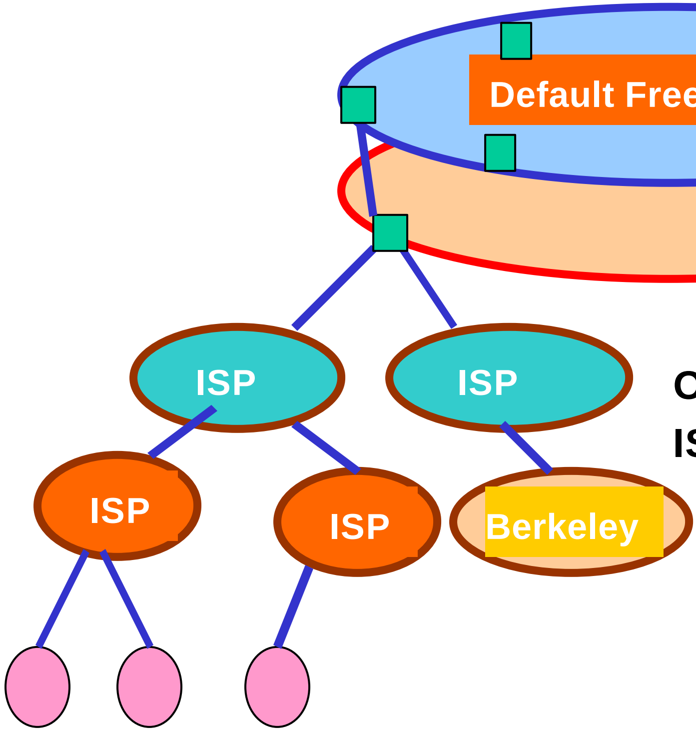
ERP or EGP (Interior Gateway Protocol) are used between ASes.

Determine connectivity and routes within between ASes

ISP1 and Qwest are peering each other and are peer to each other.

EGP, BGP, IDPR are example of ERPs.

Internet Routing Architecture



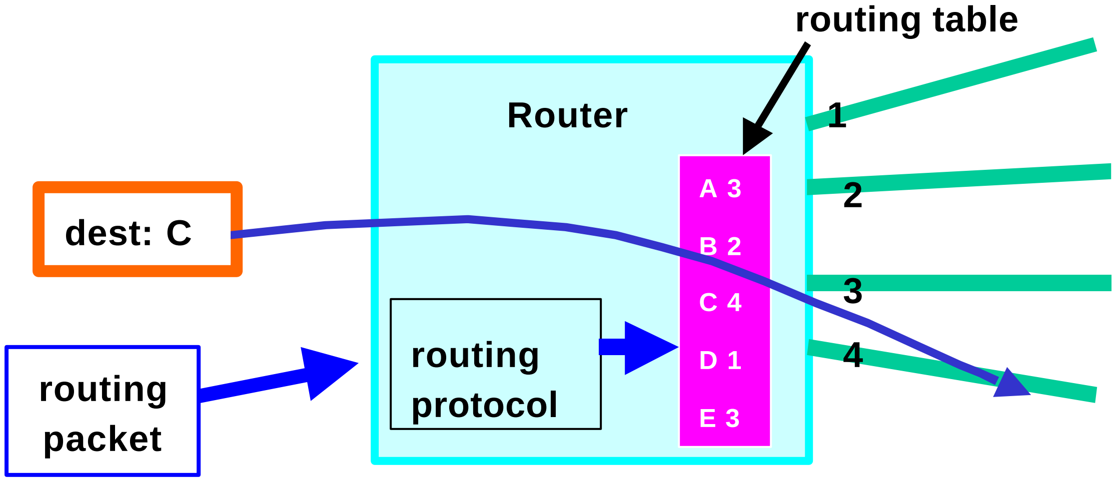
Organized as rough hierarchical ASes.
ISP is an AS that offers connectivity to the Internet.

- Core (Default Free Zone)
  - No default routes
  - Complete routing table (60,000+ entries)

- Most ASes
  - Default route pointing up the hierarchy
  - Carry only subset of Internet Addresses

3500 ISPs in 1997 (20 major ones with 150 nodes each) [Walrand 1998]

Router, Routing Protocol and Routing Table



A router forwards packets based on its routing table.

- The router passes the packet destined to C to link 4.

Routing protocol

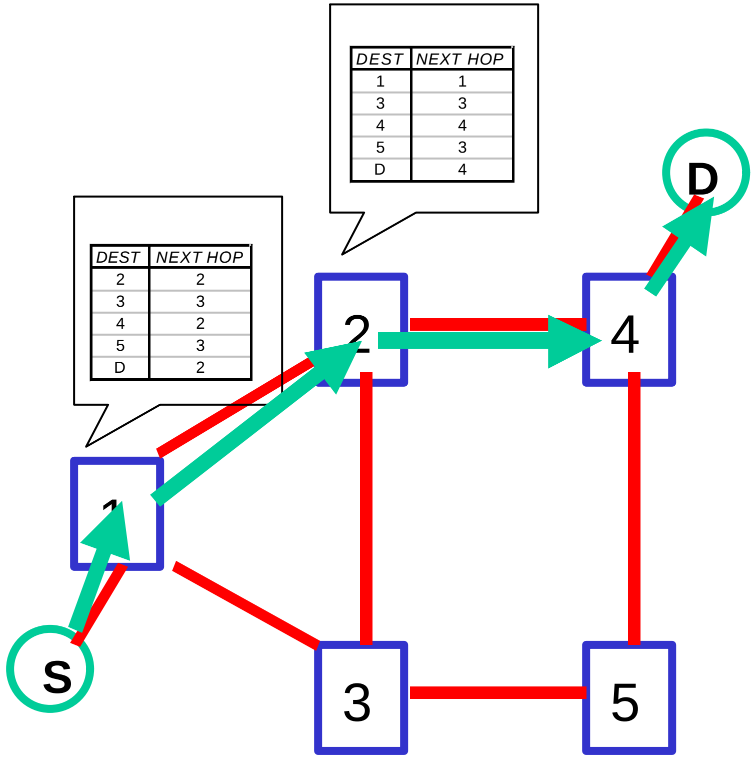
- advertises its routes to other routers (help other routers to generate their routing table)
- generates its routing table (computes routes).
- IRP is responsible for Intra-AS entries while ERP for Extra-AS entries.

Routing Table

Routing table is a mapping from destination address to an output link.

Upon receiving a packet, routers forward the packet to the next hop based on the routing table.

Routing protocol compares metrics (distances) of routes to determine a route.



Scalability becomes big concerns.

Exponential growth causes many scalability problems.

- Address Space Depletion
- Huge Routing Table
- Increased Control Traffic
- BGP Scaling Problem

.....

A lot of efforts have been put to address these concerns.

- IPv6, subnetting, CIDR
- Default Route, CIDR, Address Aggregation
- Multi-Area OSPF
- AS Confederation, Route Reflection

] **(later discussion)**

Internet Addressing Problem

32 bit class-based Internet address has the form network.host.
Class A network can have 16million hosts. However, Class B
and C network can have 64 thousand and 256 hosts, respectively.

Class A	0	network (7bits)	host (24bits)		
Class B	1	0	network (14bits)	host (16bits)	
Class C	1	1	0	network (21bits)	host (8bits)

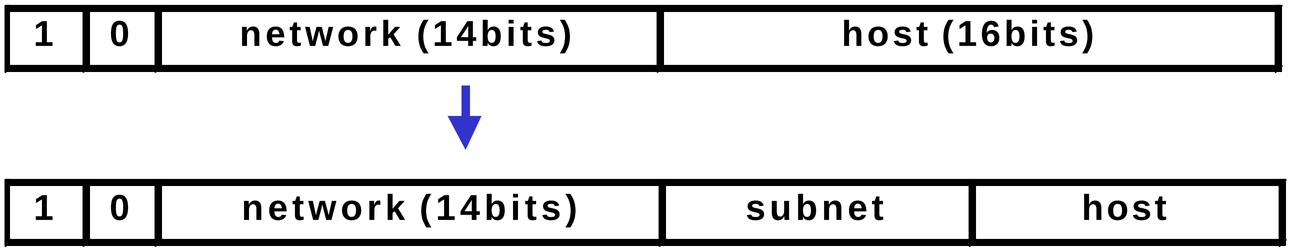
Class A network is too large and C is very small, that does not
meet what users need.

IPv6, subnetting and CIDR are suggested.

Solutions to Address Depletion Problem

IPv6: uses 128bit address rather than 32bit.

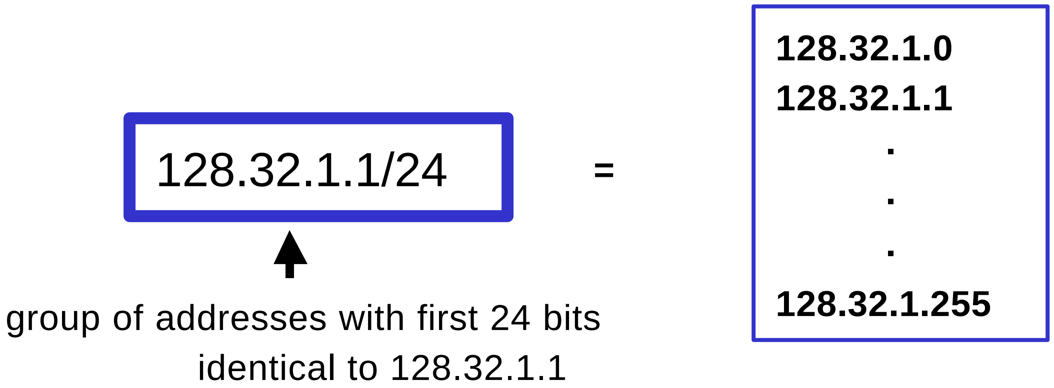
Subnetting: divide the host part into subnet.host.



Use 32 bit mask to specify subnet address.

CIDR (Classless InterDomain Routing)

CIDR addresses two scalability concerns on routing table and address space depletion.



By using address prefixes instead of IP address, CIDR helps to reduce routing table sizes.

Theoretically, 32bit address can specify 4 billion addresses.

BGP-4 supports CIDR.

Overview

Routing and Internet Basic

OSPF Protocol

BGP-4 Protocol

OSPF Basic: Hello Protocol, DB synch, Link State

...

OSPF Protocol

OSPF Network Types

Multi Area OSPF

OSPF Extensions

OSPF Basic

Popular Link State Routing protocol.

Packets called "LSAs" (link state advertisements) are flooded for each node to keep the state of the network.

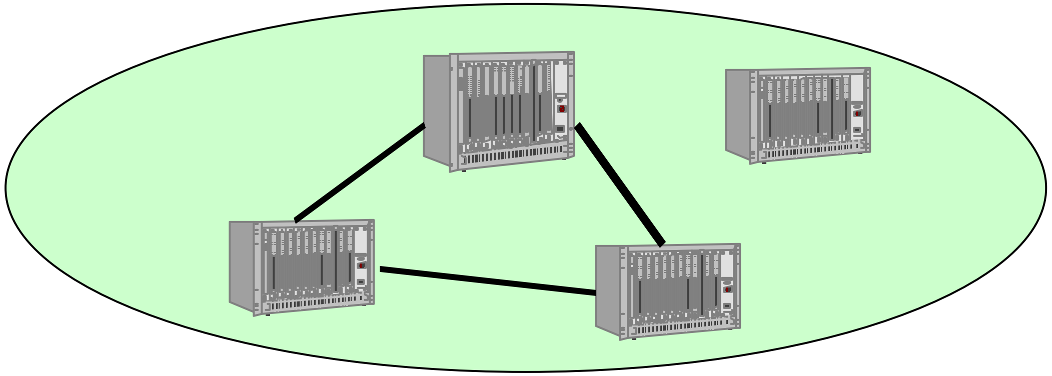
A method called "Reliable Flooding" is used.

Routing is computed using the Dijkstra algorithm.

Hello Protocol is used to find neighborhood.

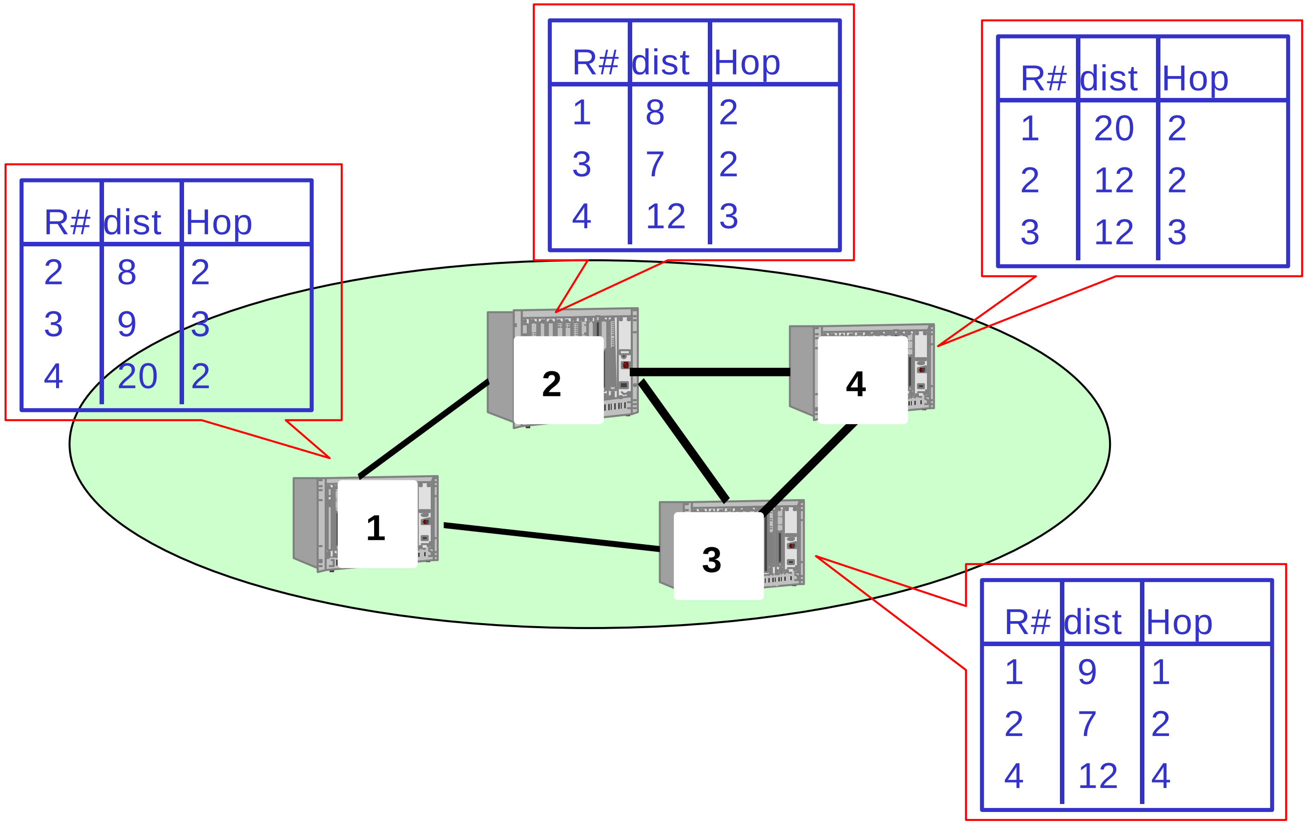
Open implies that the protocol is not proprietary.

OSPF is a Link State Routing Protocol.



Each router has full knowledge of network topology.

Distance Vector Algorithm



Router 1 does not know how router 4 is connected. It only knows it can arrive router 4 via router 2.

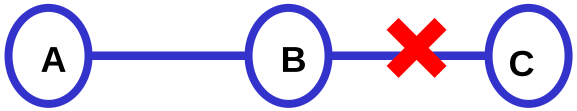
Link State vs. Distance Vector

Link State	Distance Vector
Whole Topology	Next Hop and Distance
All paths to destination	Single path to destination
Fast Convergence	Slow Convergence
	prone to loop
	Count to Infinity
OSPF, IS-IS	RIP, IGRP

Link State algorithms are more in use than DV now.

Counting to Infinity

A problem of DV protocols.



Well known problem of a Distance vector algorithm.

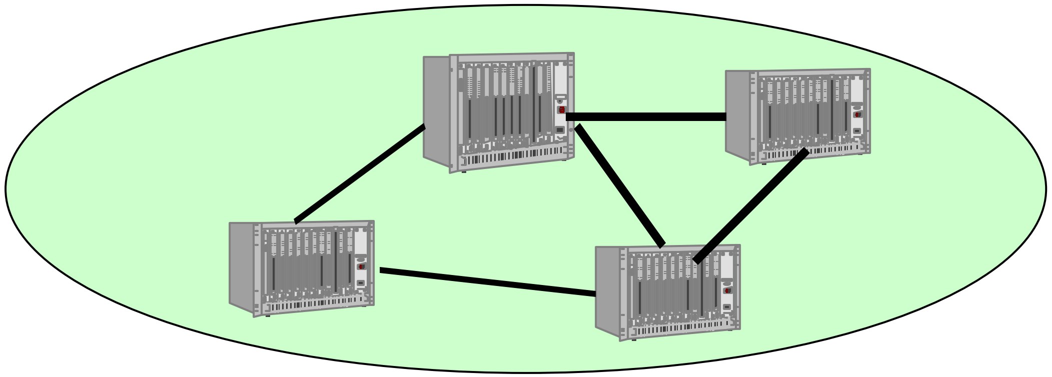
Assume that the link between B and C breaks down.
Instead of declaring that C is unreachable, B will try to reach C via A, which is impossible until the distance reaches infinity.

Solutions such as Hold down, split horizon, and triggered update are suggested.

It shows the slow convergence of DV algorithms

Link State Advertisement (LSA)

A packet called "LSA" is used to deliver link state information.
Each router generates LSAs of its links and sends it out regularly.
It also generated whenever a link state change happens.
A mechanism called "Reliable Flooding" is used to spread LSAs over whole network.



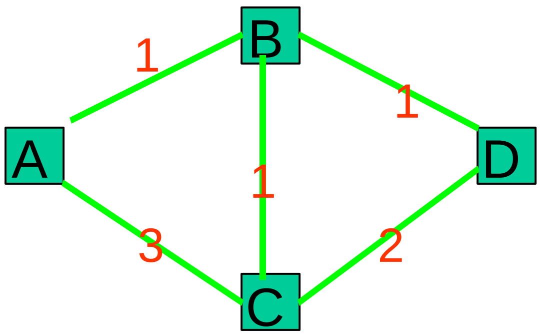
Example of LSA

R_LSA_B

links 3
(B,A): 1
(B,C): 1
(B,D): 1

R_LSA_A

links: 2
(A,B): 1
(A,C): 3



R_LSA_D

links: 2
(D,B): 1
(D,C): 3

R_LSA_C

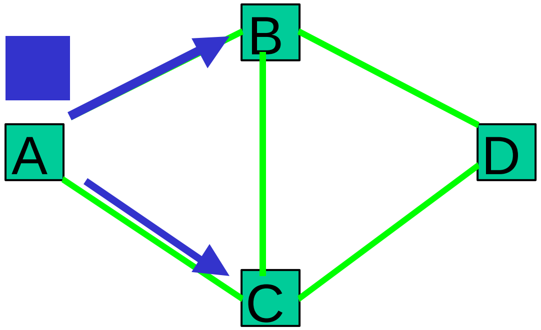
links: 3
(C,A): 2
(C,B): 1
(C,D): 2

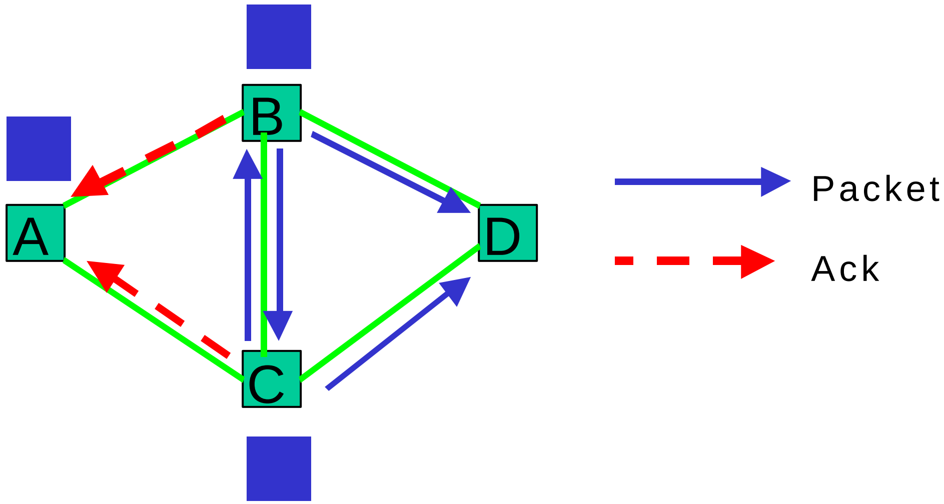
Collection of 4 LSAs forms LSA Database.

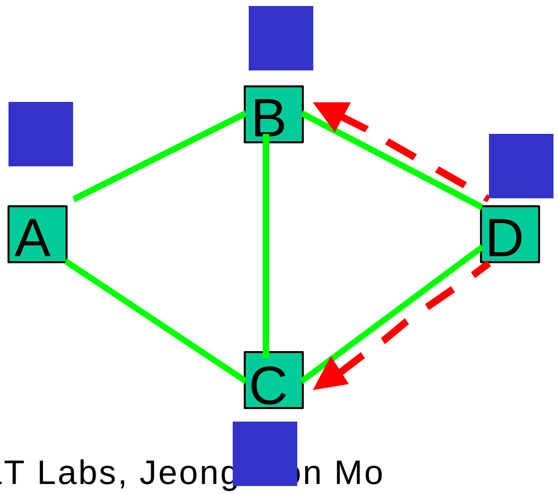
From the LSA DB, shortest paths can be computed.

Reliable Flooding

A robust way to spread packets over an entire network.







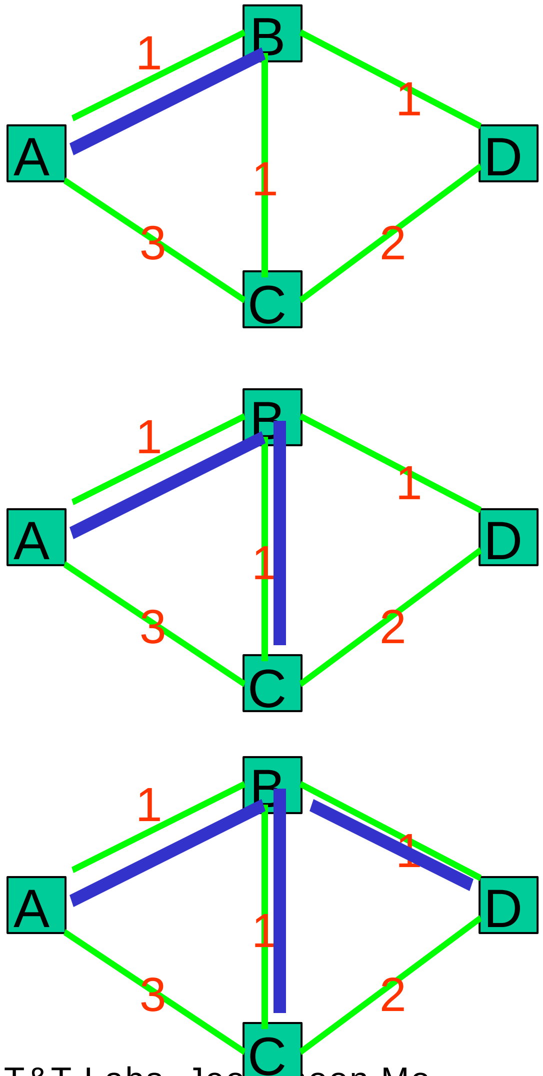
If a router fails to receive ACK before RETXN_INTERVAL (5sec), it retransmit the packet.

Dijkstra's Algorithm

Forms a shortest path spanning tree with an arbitrary node as a root.

The complexity of Dijkstra is $O(l \log n)$ where l is number of links and n is the number of nodes

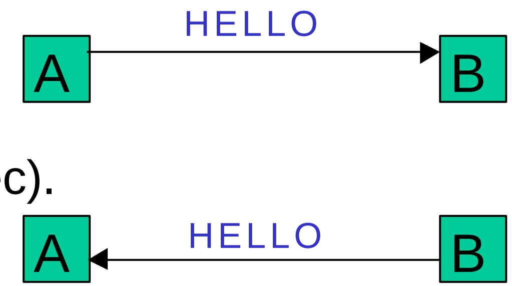
Dijkstra Illustrated



Iter.	Destination Added to a tree	Candidate List (cost, next hop)
1	A	B(1,A) C(3,A)
2	B	C(2,B) D(2,B)
3	C	D(2,B)
4	D	EMPTY

Neighborhood Discovery

Hello packet is used to find its neighborhood.



A send hello packet periodically out to
all its interfaces at every HELLO_INTERVAL (10sec).

If A hears hello back from its neighbor
A recognizes the existence of B.

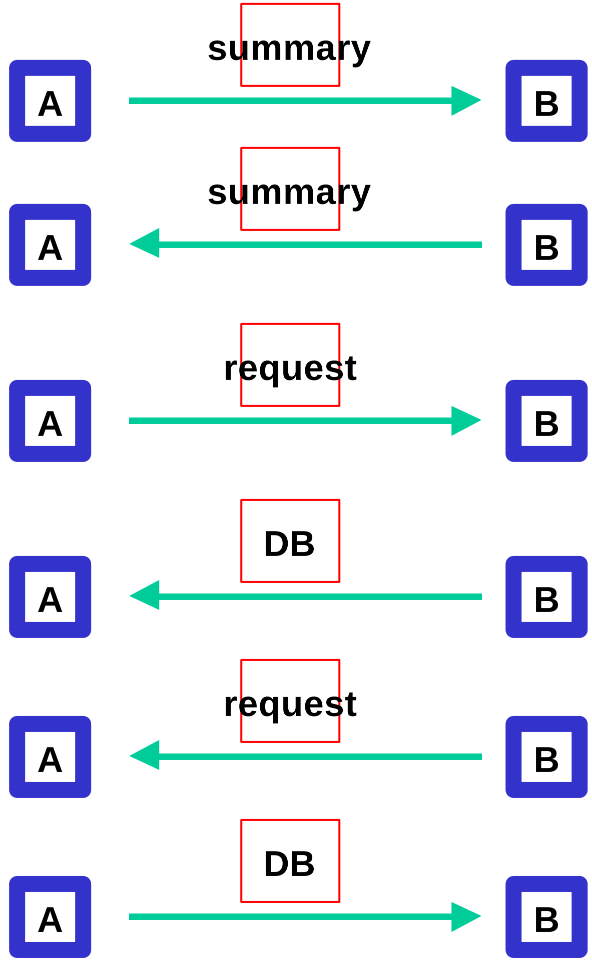
If A fails to hear 4 consecutive hellos
back from B, A declares the trunk is dead and floods new LSAs.

This part of protocol is called "HELLO PROTOCOL".

Initial Database Synchronization

A special mechanism to synchronize LSA database when a trunk added between two routers.

Instead of exchanging all LSAs directly,
1. Exchange description of database
2. And then, exchange missing Database.



Note that LSA database should be synchronized to avoid incompatible routing table

OSPF Protocol

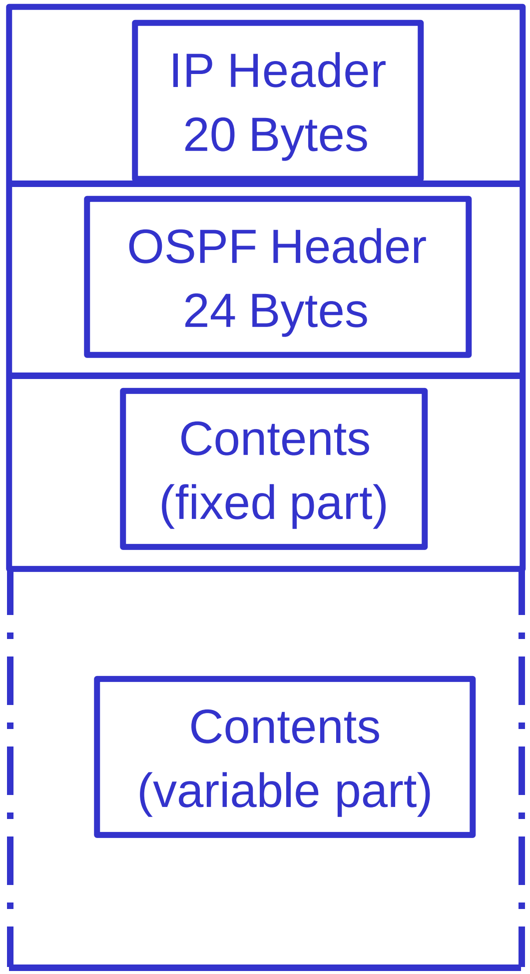
OSPF Packet Types

LSA Types

OSPF Packet

- OSPF Header
  - Version
  - Packet type
  - Length
  - Router id
  - Area id
  - Checksum
  - Authentication type & data

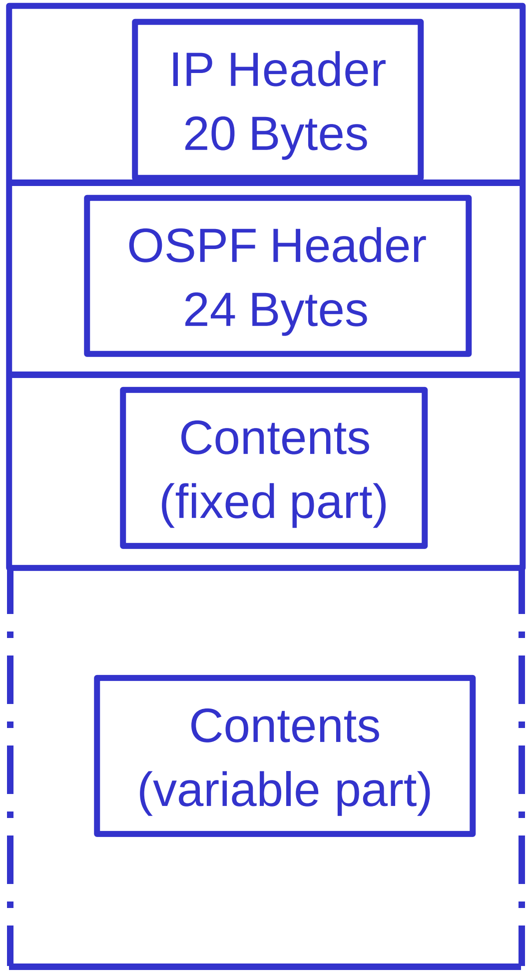
- Contents
  - Size depends on
    OSPF message type
    # neighbors listed and/or LSA type



Protocol field = 89 in IP header => OSPF packet

OSPF Packet Types

1. Hello
2. Database Description
3. Link State Request [LSR]
4. Link State Update   [LSU]
5. Link State Acknowledgement



Packet type 2, 3 are used in DB synch. Process.

A single LSU may contain many LSAs.

Type 5 packets are sent in response to LSAs.

LSA Packet

LSA Types

1. Router LSA
2. Network LSA
3. Network Summary LSA
4. ASBR Summary LSA
5. AS External LSA

There are 5 types of LSA in RFC 2178.
However, there are more proprietary vendor implemented LSAs to extend functionality of OSPF protocol.

Each router originates one router LSA that describes its active neighbors.
Type 3 and 4 LSAs are used in multi-area OSPF.

AS External LSA is used to introduce external prefixes into an AS.

OSPF Network Types

Point-to-Point subnets

Broadcast subnets

Ethernet, Token Ring, FDDI,

Nonbroadcast Multiaccess (NBMA) subnets

X.25, Frame Relay, ATM

Point-to-Multipoint subnets

OSPF in Broadcast Network

So far, we have discussed OSPF in point-to-point networks.
In broadcast networks such as Ethernet, OSPF can take advantage of broadcasting ability.

Hello Protocol: Instead of sending n hellos, each OSPF node broadcasts
one hello to the broadcast address AllSPFRouter (224.0.0.5).

Database Synchronization:

Pick a designated router (DR) and all the other routers keep their DB synchronized with the DR.

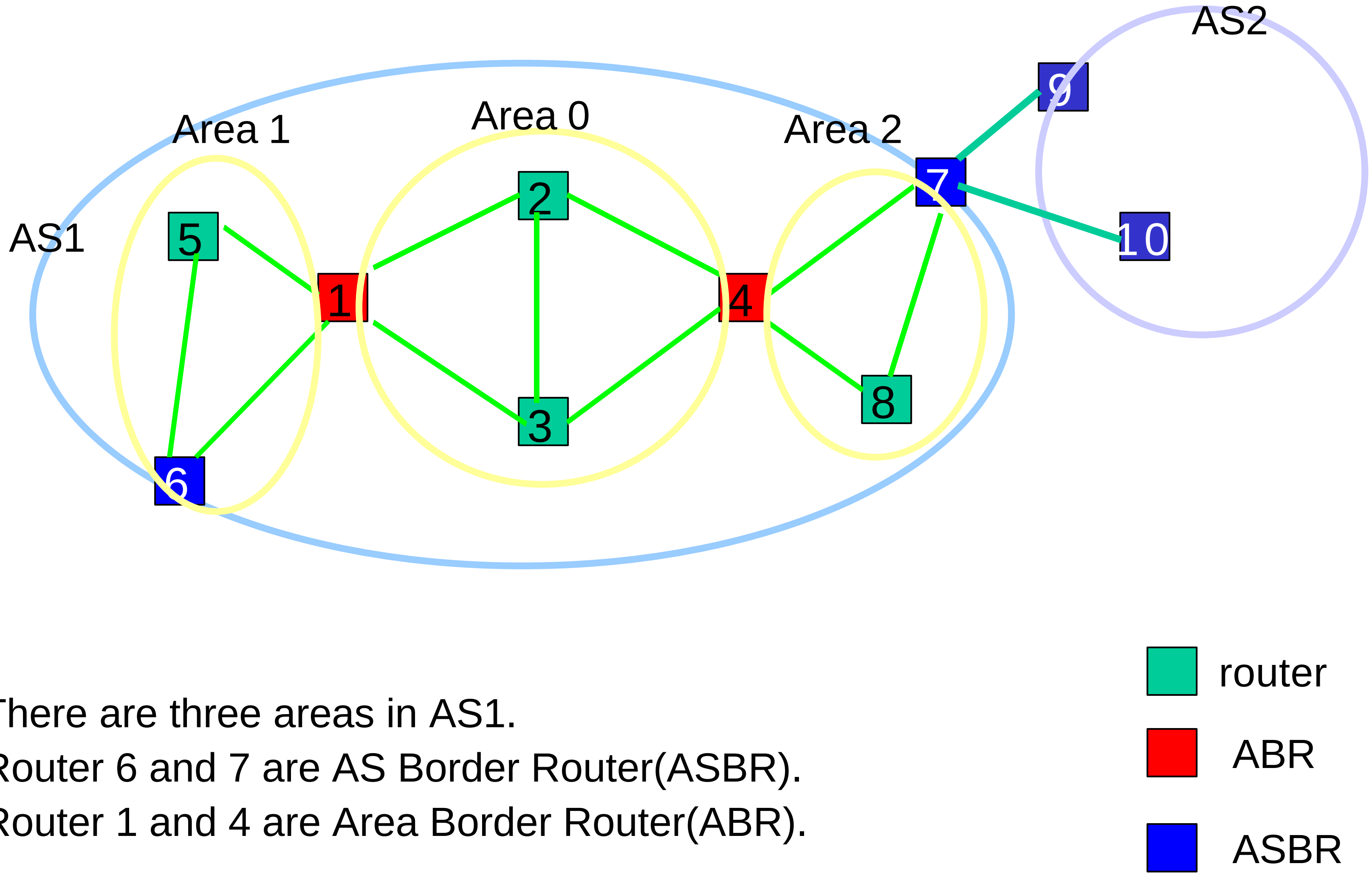
Instead of n(n-1)/2 adjacencies, they are now n.

Backup DR is elected just in case of DR failure.

Abstraction: reducing the number of LSAs.

Multi Area OSPF

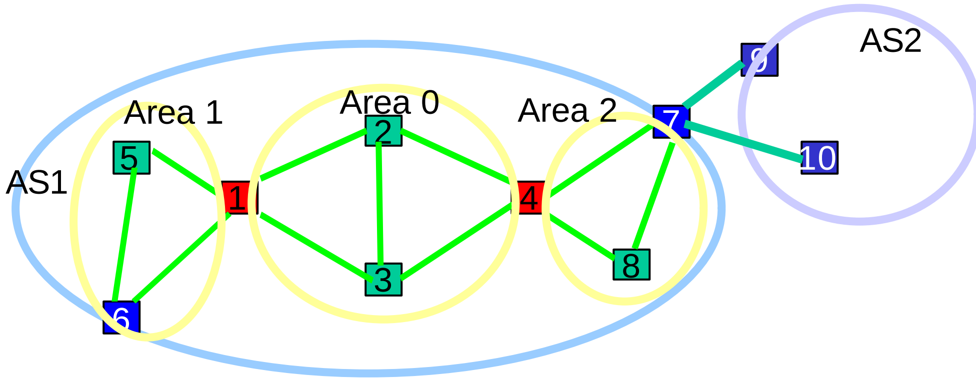
Dividing an AS into multiple Areas.



There are three areas in AS1.
Router 6 and 7 are AS Border Router(ASBR).
Router 1 and 4 are Area Border Router(ABR).

Why Multi-Area?

Scalability



As the network becomes big,

The size of routing table increases.

The router CPU overhead increases. [computing shortest path]

The control traffic increases.

Hierarchical routing reduces routing table size. (cf. flat routing)

The LSA Database size is reduced.

CPU overhead and control traffic is reduced.

MA-OSPF helps building larger networks.

router

ABR

ASBR

MA OSPF

Combination of Link State and Distance Vector Algorithm

within an area: Link State Algorithm

outside an area: Distance Vector Algorithm

Two level hierarchical routing

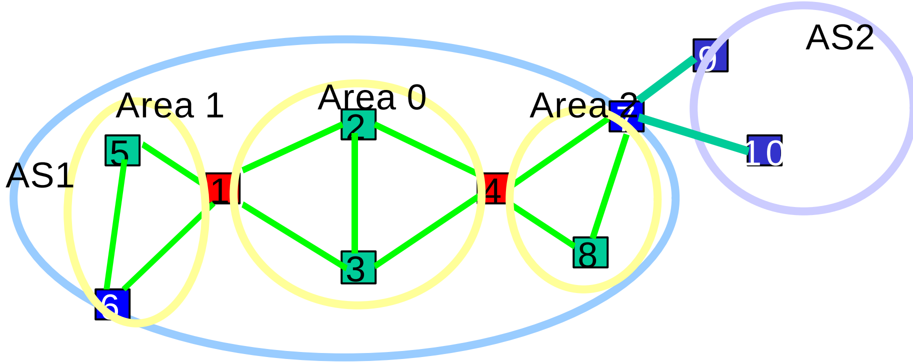
OSPF backbone area

Other areas directly attached to the backbone area.

Area border routers are between the backbone and other areas

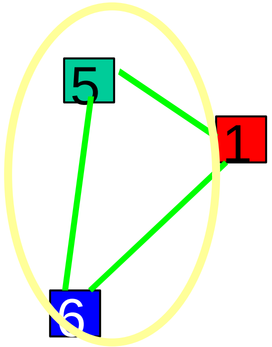
Multi Area may cause less efficient path.

LS + DV Algorithm



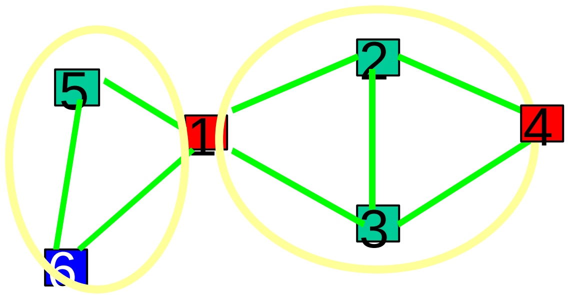
In a multi-area OSPF network, the view of network is is limited to areas to which routers belong.

Router 5 sees Area 1 only and have dist. table to other routers



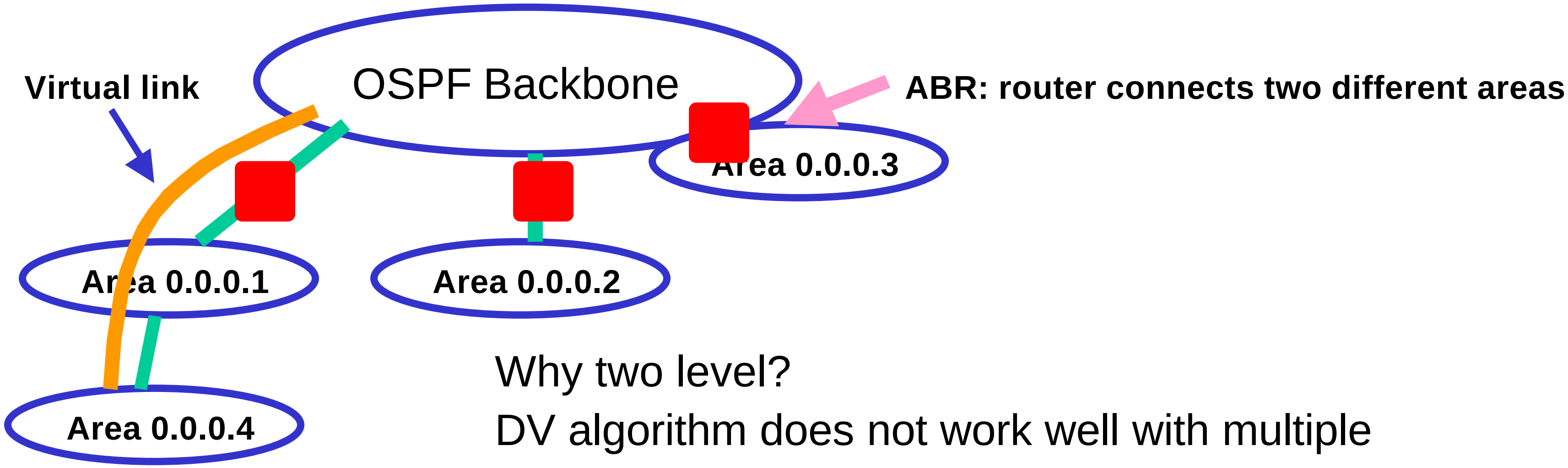
R#	dist
2	3
3	3
4	7
7	8
8	9
9	20
10	20

ABR 1 sees Area 1 and Area 0 have dist. table to other routers



R#	dist
7	8
8	9
9	20
10	20

Two level Hierarchy



Why two level?
DV algorithm does not work well with multiple paths (counting to infinity).

Is two level physical?
No. OSPF virtual links are used to connect them logically.

It is hard to configure where and when to configure virtual links.

More on LSA types

Network summary LSA (Type 3) and ASBR summary LSA (type 4) are used in MA OSPF to construct routing table for other areas.

Network Summary LSA (LSA Type 3)

- Originated by ABR.
- Summarize one area info and flood it to other area.
- Distance from ABR to routers in other areas.

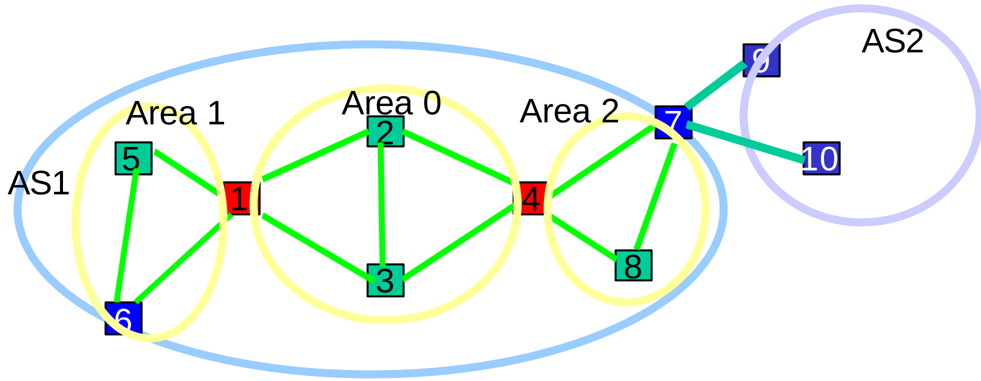
ASBR summary LSA (LSA type 4)

- Originated by ABR.
- Distance from the ABR to ASBRs

AS-External LSA (LSA type 5)

- Originated by ASBR.
- distance to routers outside AS from an ASBR.

Network Summary LSA
ASBR Summary LSA



How router 5 or 2 knows about other areas?

Network summary LSA (NSL) is the answer.

ABR 1 generates NSL for area 1 and floods it into Area 0.

ABR 1 generates NSL for area 0 and floods it into Area 1.

Flooding of NSL is limited to an area.

ABR 1 will not flood ABR 4 generated LSA into area 1.

ASBR summary LSA is describes the distance from ABR to ASBR.

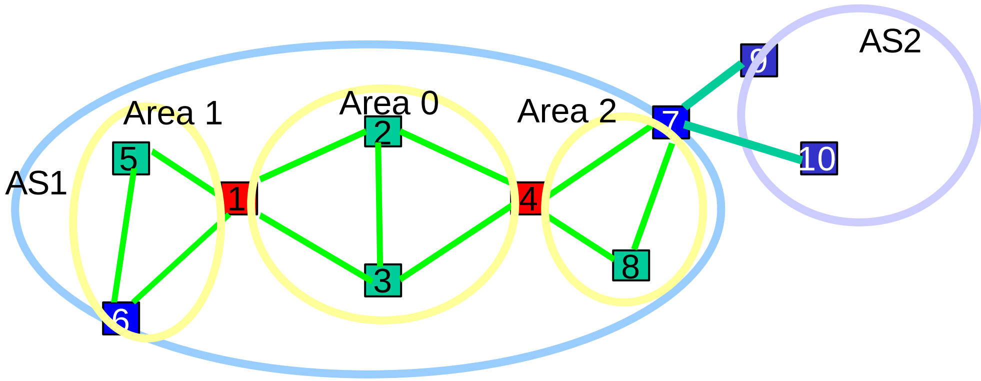
ABR 1 will generate one ASBR Summary LSA to ASBR 6 and another to ASBR 7.

How does ABR know which one is ASBR?

PRECONFIGURED.

Flooding of ASBR Summary LSA is also limited to an area.

AS External LSA



AS External LSA is originated by ASBR.

AS External LSA describes the distance from an ASBR to destination outside the AS.

The outside info. can be from BGP peering or other routing protocols.

ASE LSA is the only LSA that is flooded over the whole AS.
(note that summary LSA and router LSA is limited to an area.)

ASBR 6 and 7 generates AS External LSA

OSPF Extensions

Stub Areas

- One of the OSPF area types
- Area with limited resources
- AS-External LSA is not flooded, instead using default route.

Database overflow support

MOSPF

- Multicast extensions to OSPF
- add group-membership LSA
- add calculations for multicast paths

Summary of OSPF

- OSPF is one of the most popular IRP protocols.
- OSPF is link state protocol.
- OSPF uses hello protocol, DB synch process, reliable flooding.
- OSPF computes shortest path using Dijkstra algorithm.
- We explained 5 types of LSAs: router, network, network summary, AS summary, and AS external LSAs.
- LSUPD packet is used to deliver LSAs.
- Multi-Area OSPF

Overview

Routing and Internet Basic

OSPF

BGP-4

- Introduction
- Characteristics of BGP
- BGP Attributes
- Route Selection
- I-BGP and E-BGP
- BGP Extensions
  - IBGP scaling
  - Route Flap Dampening
  - BGP Communities
- Further Issues

BGP-4 (Border Gateway Protocol version 4)

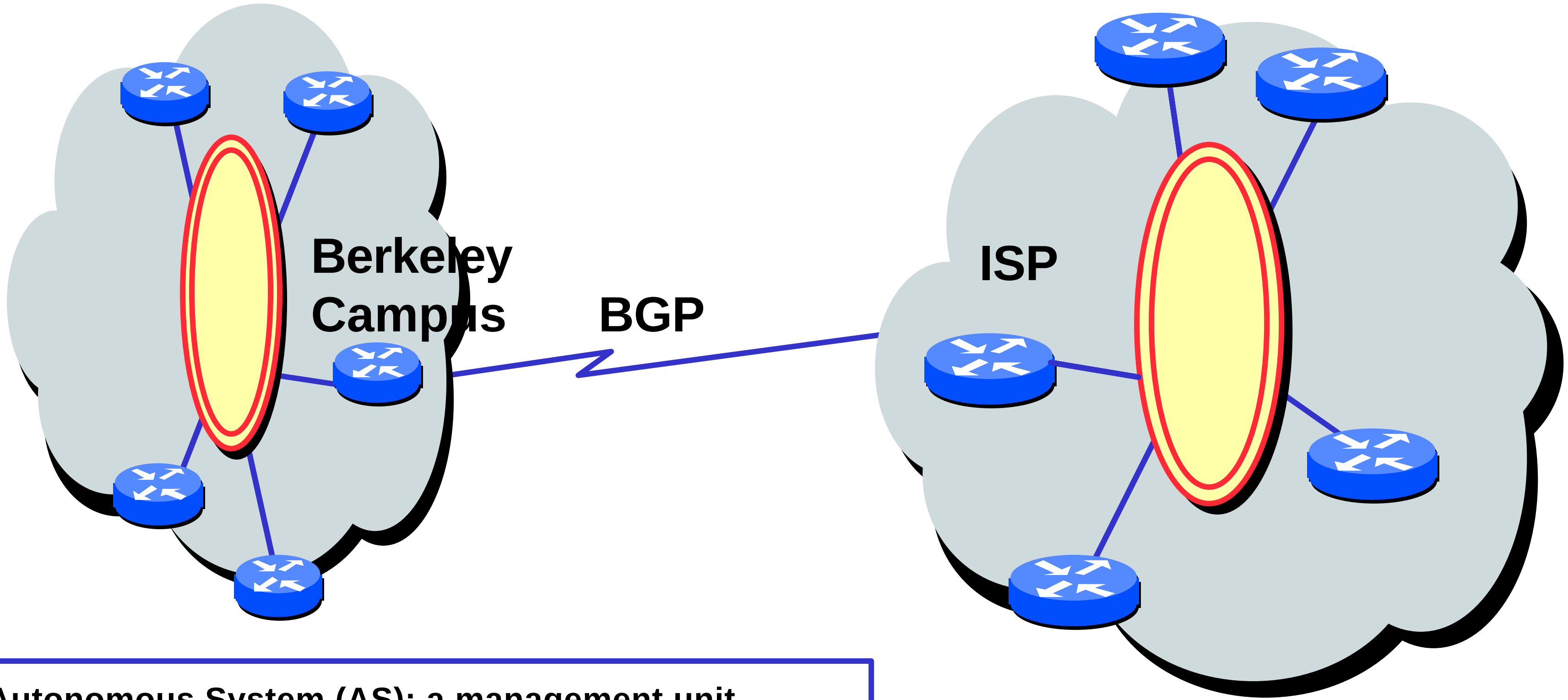
One of the most popular EGP protocols in the Internet

used to carrying routing information between ASes.

Current Internet Standard for Interdomain Routing

Version 4 supports CIDR

BGP used to share routing info. with other ASes.



Autonomous System (AS): a management unit

ISP: Internet Service Provider

Increasing Interests in BGP

Explosive increase in the number of ISPs

more than 3500 in US, 1977

Increase in Internet Based Companies

Internet advertising, Internet Sales companies

Mostly ISPs have great interests in BGP.

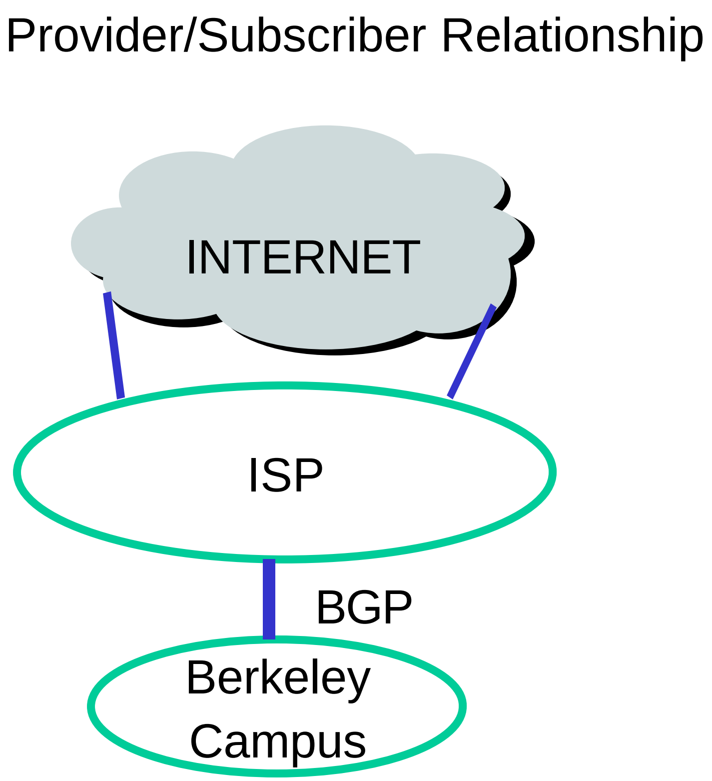
Routing Policies in BGP

BGP routing is rather political than optimal as in IGP.

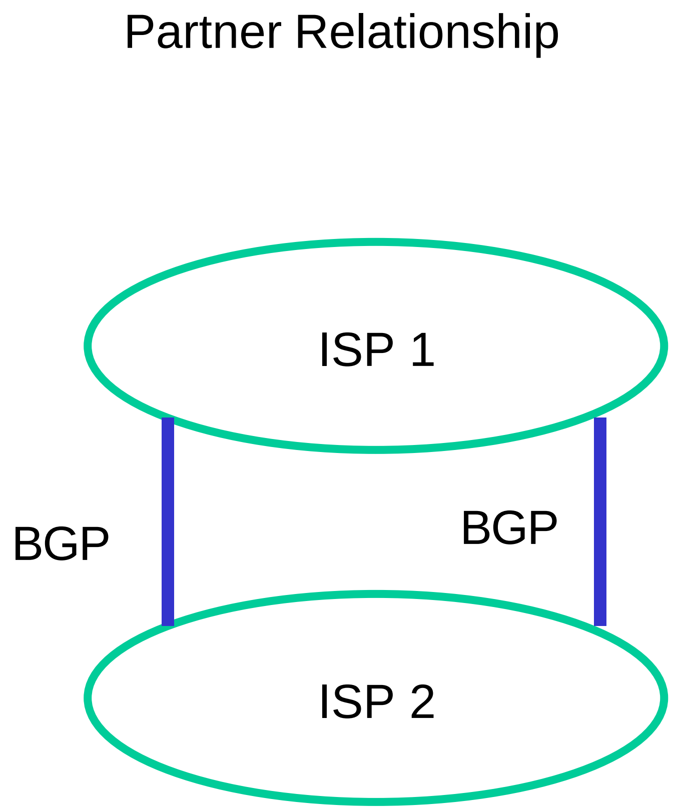
An ISP configures peering with other ISPs if NECESSARY.

When two routers start a BGP session, they are BGP peers of each other.

Peering Relationships

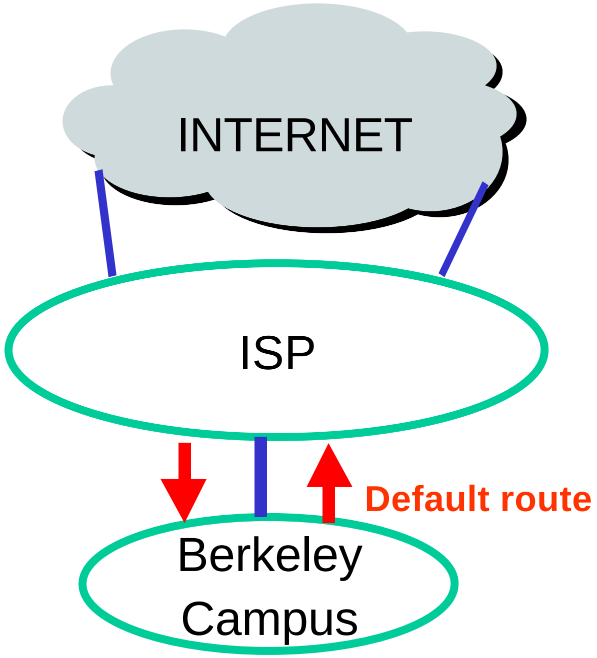


Berkeley Campus connected
to Internet via ISP.
ISP provides TRANSIT SERVICE.

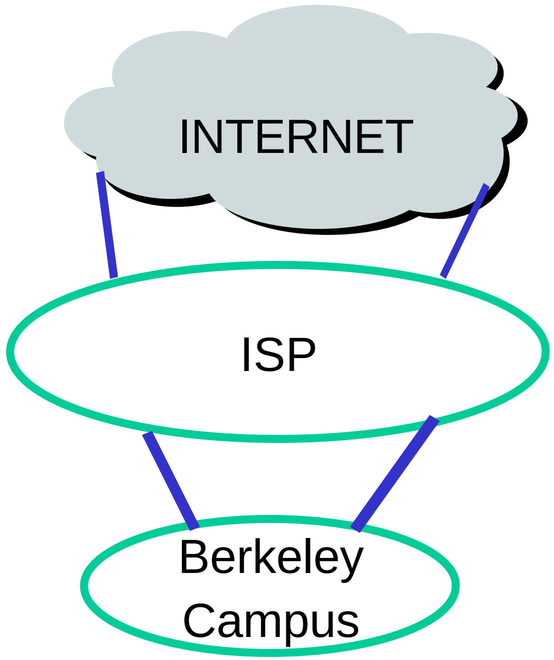


When is BGP needed?

Singly homed subscriber may not need BGP.

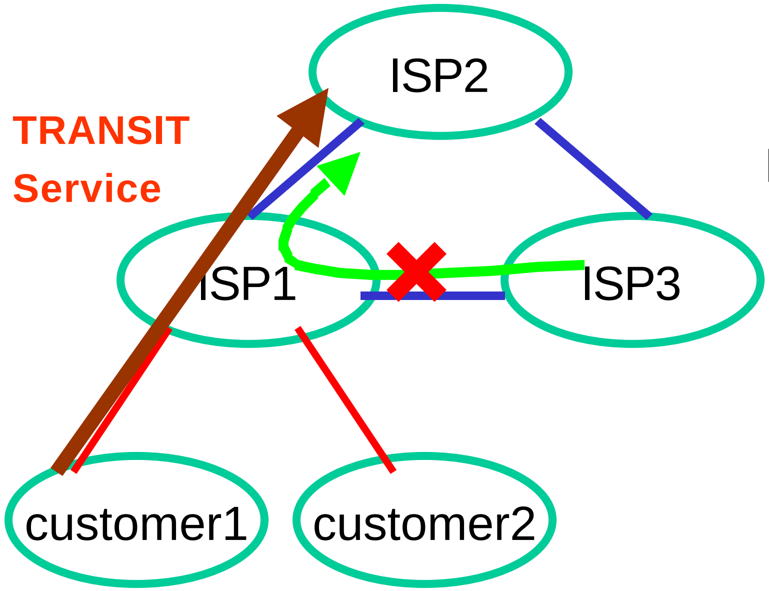


Doubly homed subscriber needs to choose which route to choose.



Upstream to ISP, use default route.
For downstream link, use static route.

When is BGP needed?

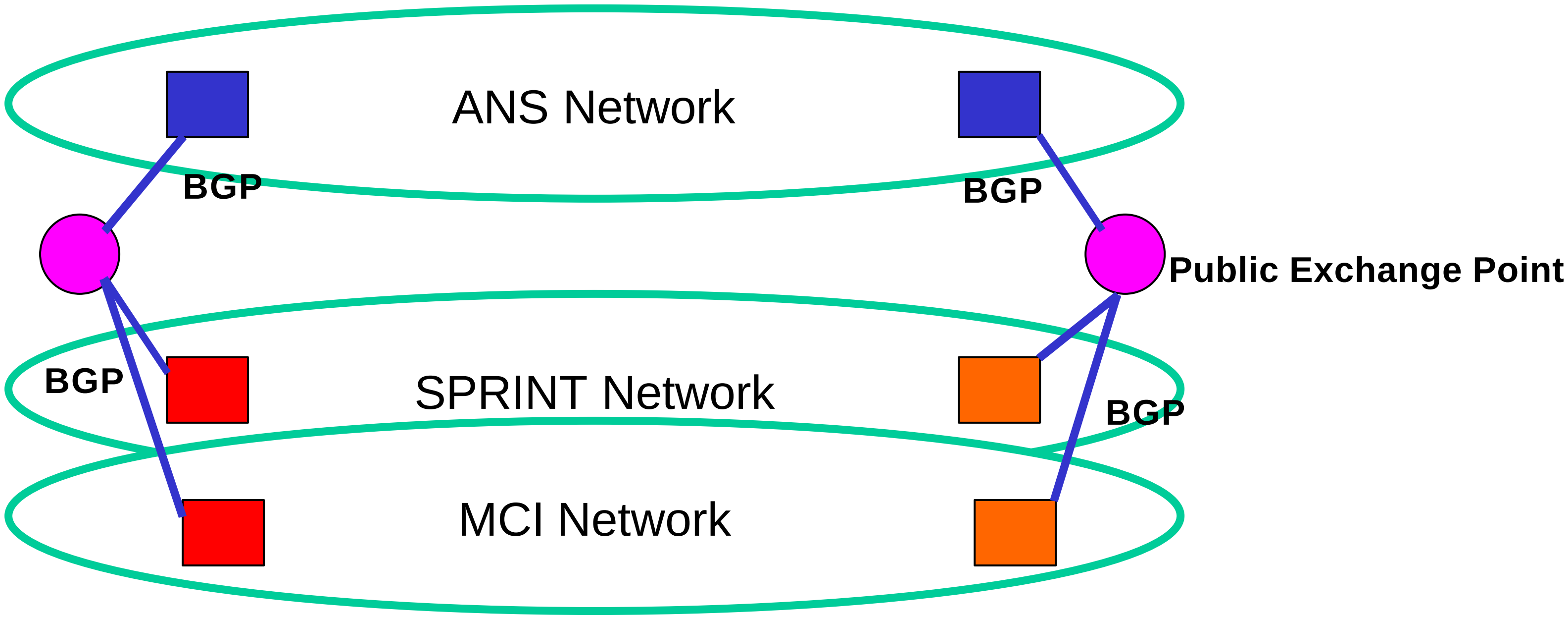


ISP may want to apply different policies to different ASes.

ISP1 offers TRANSIT service to customer1. However, it does not to other ISPs.

Peering in Practice

There are number of Exchange Points such as MAEs, CIXs, FIXs. MAE-east, one of them, located near washing,D.C. hosts more than 60 ISPs.



Characteristics of BGP-4

Path Vector Protocol

Runs over TCP (port 179)

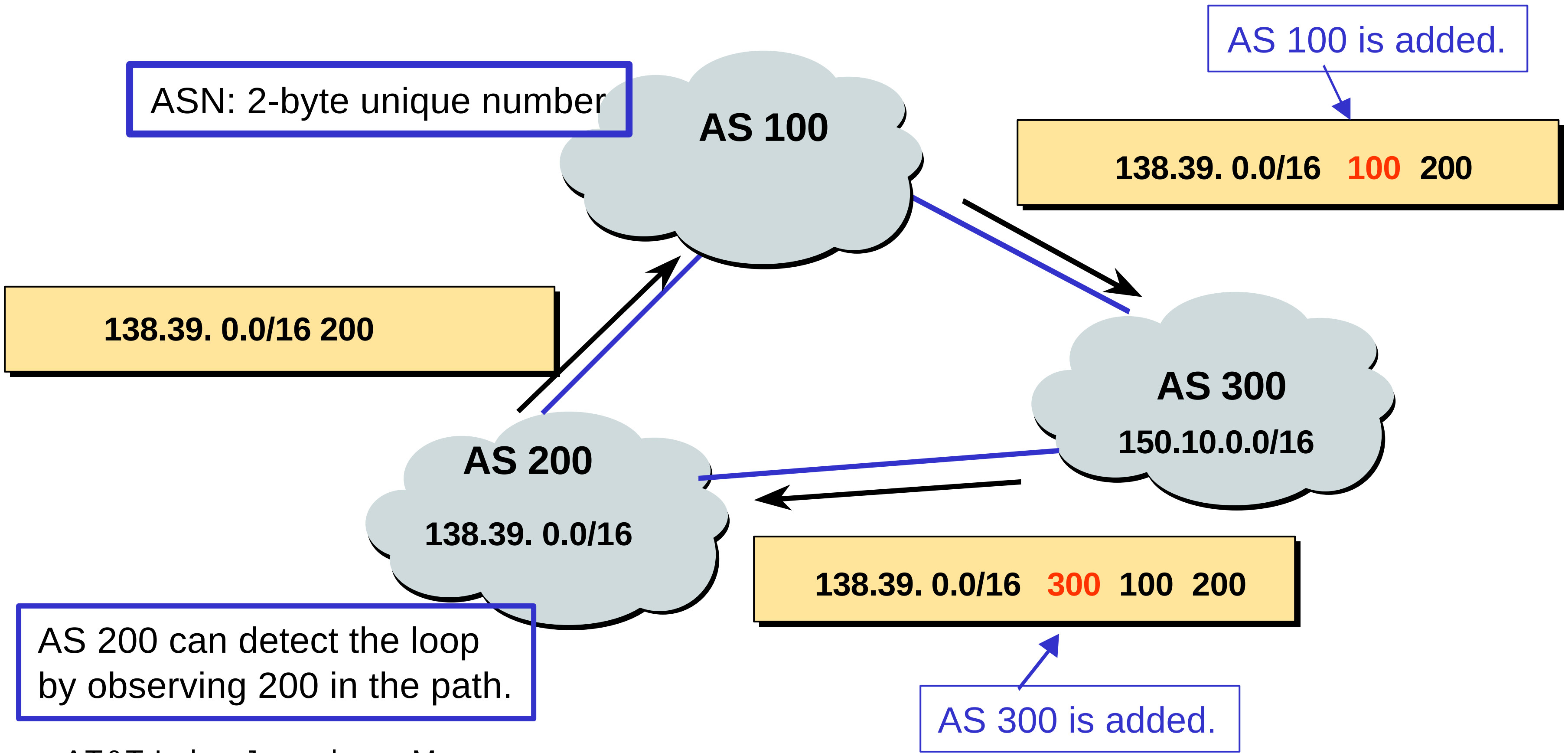
Incremental

Peering is configured manually.

Prefix : CIDR term
aggregate of addresses

BGP is a path vector protocol.

When advertising prefixes, AS path is distributed to prevent looping.



BGP achieves reliability by TCP.

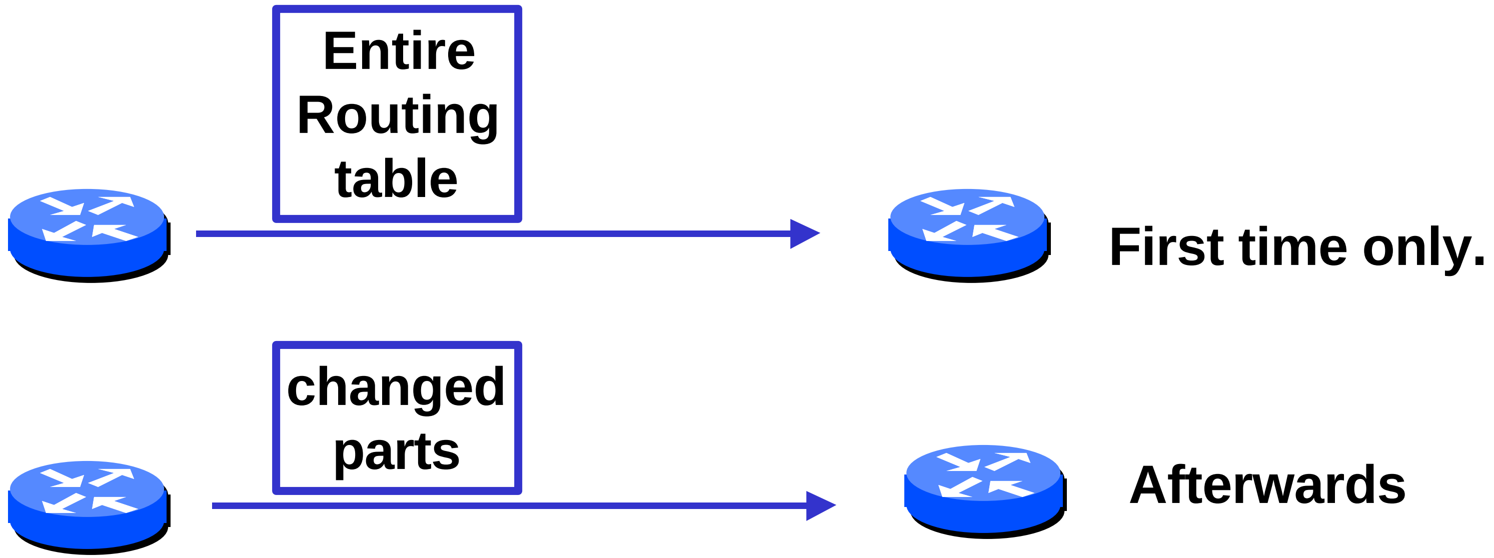
Instead of establishing its own retransmission protocol, BGP runs over TCP with port 179.

TCP is a reliable transmission protocol.

BGP is incremental.

Initial data exchange between two routers are the entire BPG routing table.

Incremental updates are sent out as the routing table changes.



BGP has 4 packet types.

OPEN
UPDATE
NOTIFICATION
KEEPALIVE

OPEN is the first message sent out establishing TCP connection. It is used to identify each other.
NOTIFICATION is for error handling.
KEEPALIVE is used to maintain TCP connection.

UPDATE is the primary message used to communicate routing information.

BGP UPDATE MESSAGE

Used to advertise a single feasible route or to withdraw multiple routes.

Common Header (19)
Withdrawn Routes Len (2)
Withdrawn Routes (var)
Path Attributes Length (2)
Path Attributes (var)
Network Layer (var) Reachability Info.

BGP Overview

Characteristics of BGP

BGP Attributes

Route Selection

I-BGP and E-BGP

BGP Extensions

- IBGP scaling
- Route Flap Dampening
- BGP Communities

Further Issues

BGP Attributes

- One of the most important features.
- Attributes describes prefixes
  - how the prefix came to be routed
  - the path of ASes
  - metrics used in path selection

- There are 7 attributes in RFC 1771.
- Some are mandatory.
- Attributes can be used to extend functionality of BGP

Common Header (19)
Withdrawn Routes Len (2)
Withdrawn Routes (var)
Path Attributes Length (2)
Path Attributes (var)
Network Layer (var) Reachability Info.

ORIGIN (1)
AS_PATH (2)
NEXT_HOP (3)
MULTI_EXIT_DESC (4)
LOCAL_PREF (5)
ATOMIC_AGGREGATE (6)
AGGREGATOR (7)
COMMUNITY(8)

BGP Attributes

ORIGIN (1) origin type, whether it is from within the AS, outside AS or else
AS_PATH (2) AS paths of the prefix
NEXT_HOP (3) next hop router to reach the prefix
MULTI_EXIT_DESC (4) used for route selection
LOCAL_PREF (5) used for route selection
ATOMIC_AGGREGATE (6)
AGGREGATOR (7)
COMMUNITY(8) used for policy control
ORIGINATOR_ID (9)
CLUSTER_LIST (10) — used for router reflector, Cisco defined
DEST_PREF (11) MCI defined
Advertiser (12) Baynet defined
rcid_path (13) Baynet defined
MP-REACH-NLRI (14)
MP-UNREACH-NLRI (15) — used for multiprotocol extension

BGP Overview

Characteristics of BGP

BGP Attributes

Route Selection

I-BGP and E-BGP

BGP Extensions

- IBGP scaling
- Route Flap Dampening
- BGP Communities

Further Issues

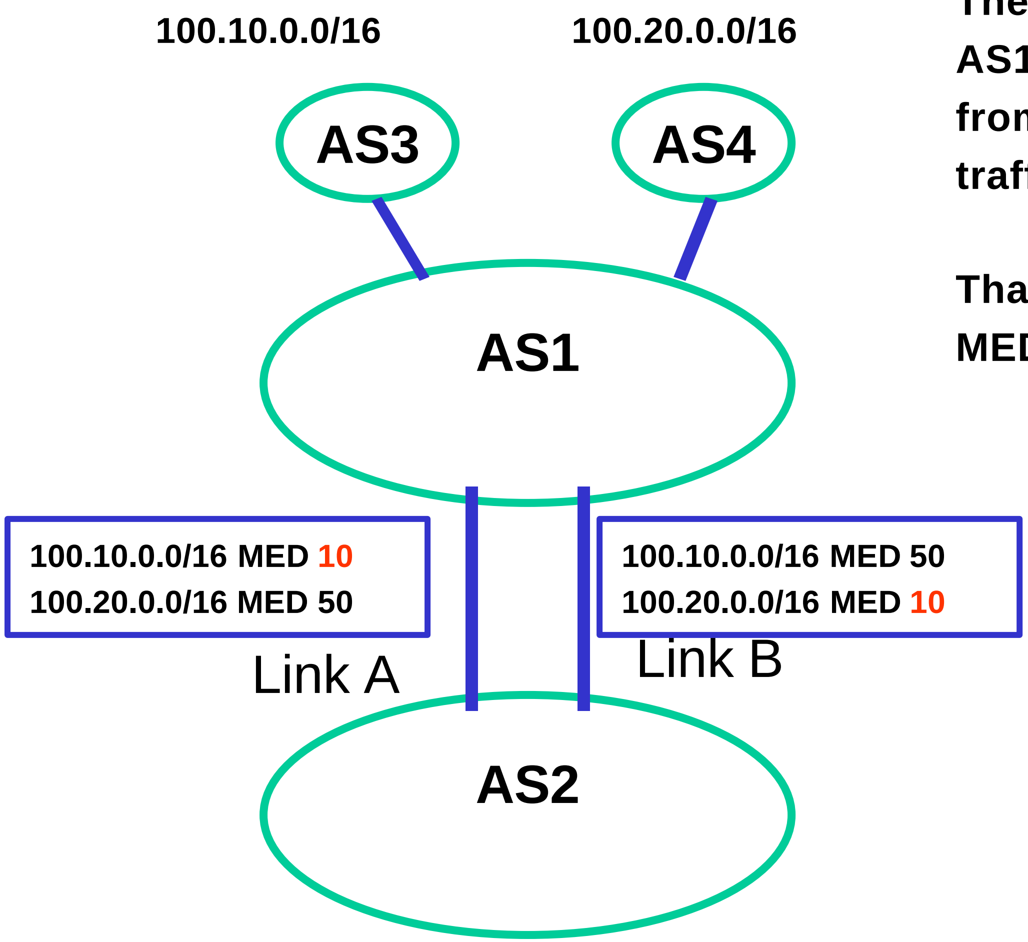
Route Selection in BGP

ORIGIN (1)
AS_PATH (2)
NEXT_HOP (3)
MULTI_EXIT_DESC (4)
LOCAL_PREF (5)
ATOMIC_AGGREGATE (6)
AGGREGATOR (7)
COMMUNITY(8)

Route selection is important since it has profound impact on the network usage.
If there is only one route for a prefix, select it otherwise, the following tie-breaking rule is applied.

Highest LOCAL_PREF
↓
Shortest AS_PATH
↓
Lowest MULTI_EXIT_DESC
↓
Shortest distance to NEXT_HOP
↓
Route with lowest id from EBGP neighbor
↓
Route with lowest id from IBGP neighbor

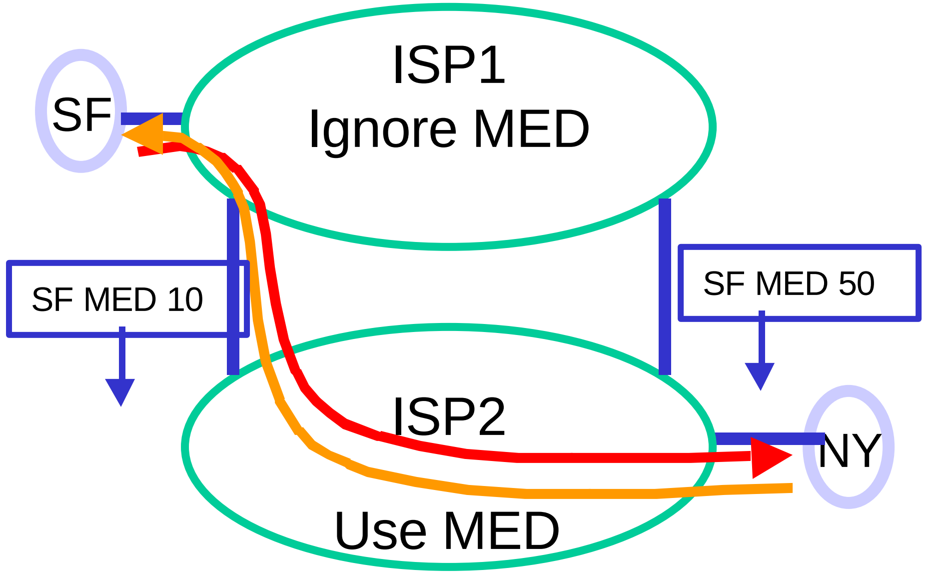
MULTI EXIT DESC (MED)



There are two links bet. AS1 and AS2. AS1 wants to utilize link A for traffic from AS2 to AS3 and link B for traffic from AS2 to AS4.

That can be achieved by setting lower MED value to the link of preference.

MED is used in provider/subscriber situation.



MED is set by one AS and used by a different AS.

Some ISP can use MED unfairly.

ISP1 is configured to ignores MED and to send lower MED to ISP2. Customer of ISP1 in SF wants to exchange traffic with a customer of ISP2 in NY.

As a result, traffic between SF to NY utilizes mostly ISP2' s network.

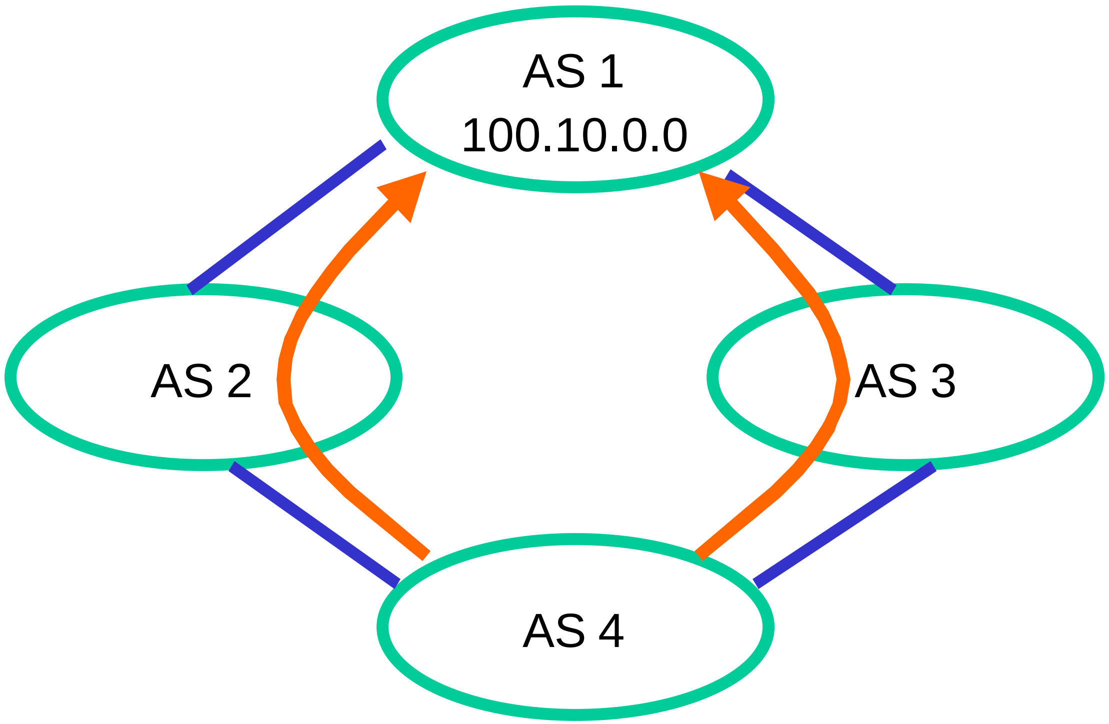
LOCAL PREF

Route selection using MED is not sufficient.

- **Selection is dictated by other AS.**
- **Can be used when there are multiple links between two ASes.**

Highest LOCAL_PREF value wins

LOCAL PREF



Two paths to AS1 from AS4.

If AS3 gives better service than AS2, AS4 may want to send traffic via AS3.

AS4 configure the value of prefix heard from AS3 to be higher.

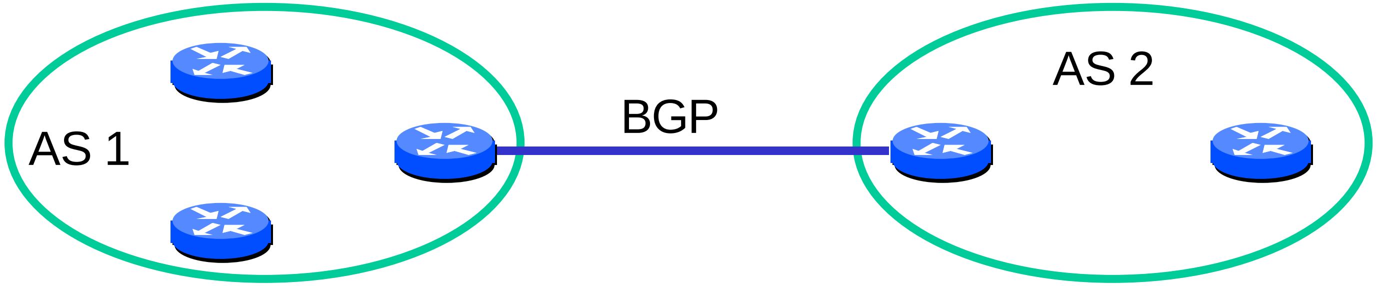
The decision of LOCAL_PREF value is local to AS.

BGP Overview

- Characteristics of BGP
- BGP Attributes
- Route Selection
- I-BGP and E-BGP
- BGP Extensions
  - I-BGP scaling
  - Route Flap Dampening
  - BGP Communities
- Further Issues

Internal BGP (I-BGP)

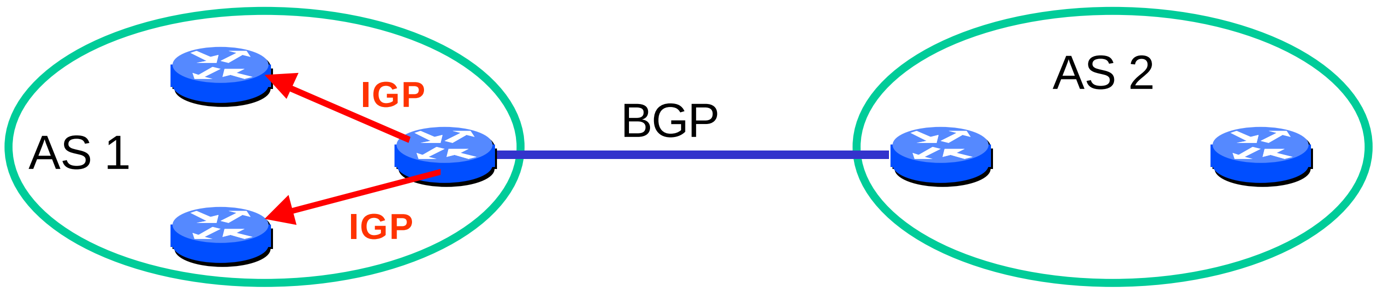
So far, we have talked about BGP as an EGP between ASes. This is called External-BGP (E-BGP) to distinguish it from Internal-BGP.



However, it is only one of two usage of BGP, E-BGP and I-BGP.
In the above figure, two routers connected by BGP can learn the routes from each other. But how about the other three routers?

Needs for I-BGP

We may inject routes to IGP such as OSPF.



This may work for smaller network but does not scale well.

- Computing shortest paths is burdensome.
- Tremendous increase in LSA traffic.

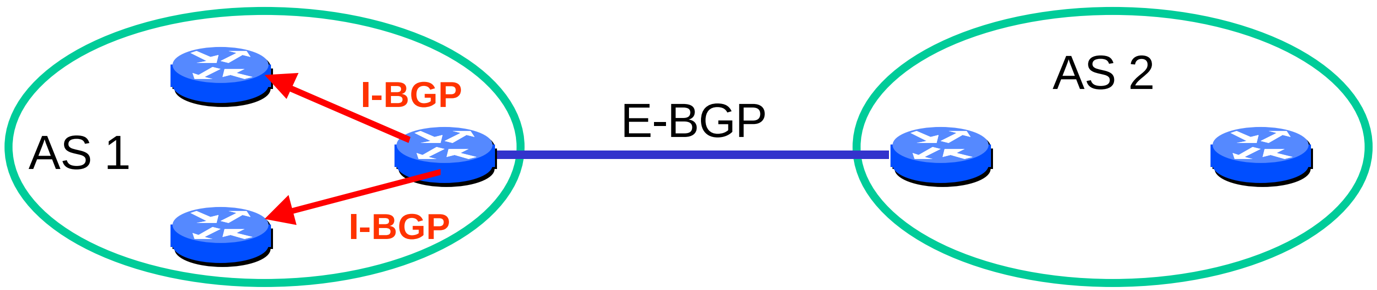
Another approach can be using Internal-BGP (I-BGP).

I-BGP vs. E-BGP

I-BGP is the same as E-BGP in that they share the same packet types, the same attribute types. However, they have different advertising rules.

Prefixes learned from E-BGP neighbor can be advertised to I-BGP neighbors.
But,
Prefixes learned from an I-BGP neighbor cannot be to other I-BGP neighbors.

Why?
To prevent LOOPING.

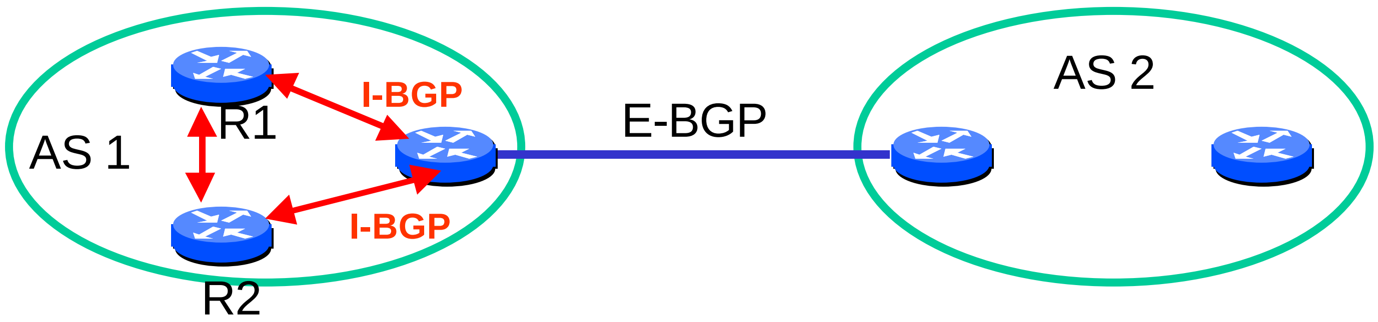


I-BGP connections forms a full-mesh.

The major result of the following rule

Prefixes learned from an I-BGP neighbor cannot be to other I-BGP neighbors.

is that there needs to be connection between R1 and R2 to advertise the prefixes between R1 and R2.



Note that full mesh is I-BGP connections is independent of physical connectivity.

BGP Overview

Characteristics of BGP

BGP Attributes

Route Selection

I-BGP and E-BGP

BGP Extensions

- IBGP scaling
- Route Flap Dampening
- BGP Communities

Further Issues

BGP problems and its Extensions

I-BGP Scaling Problem: # of I-BGP connections increases too fast.

Route Reflection, RFC1966

AS Confederation, RFC1965

Instability Issue:

Route Flap Damping

Complex Policy Control:

BGP Community Attributes RFC 1997

Security Issue:

TCP MD5 Authentication, IETF draft (1998)

Multiprotocol Extensions

RFC 2283 (1998)

Capabilities Negotiation

Two approaches to IBGP scaling

Route Reflection: hierarchical Approach
AS Confederation: Divide and Conquer Approach

Problem:

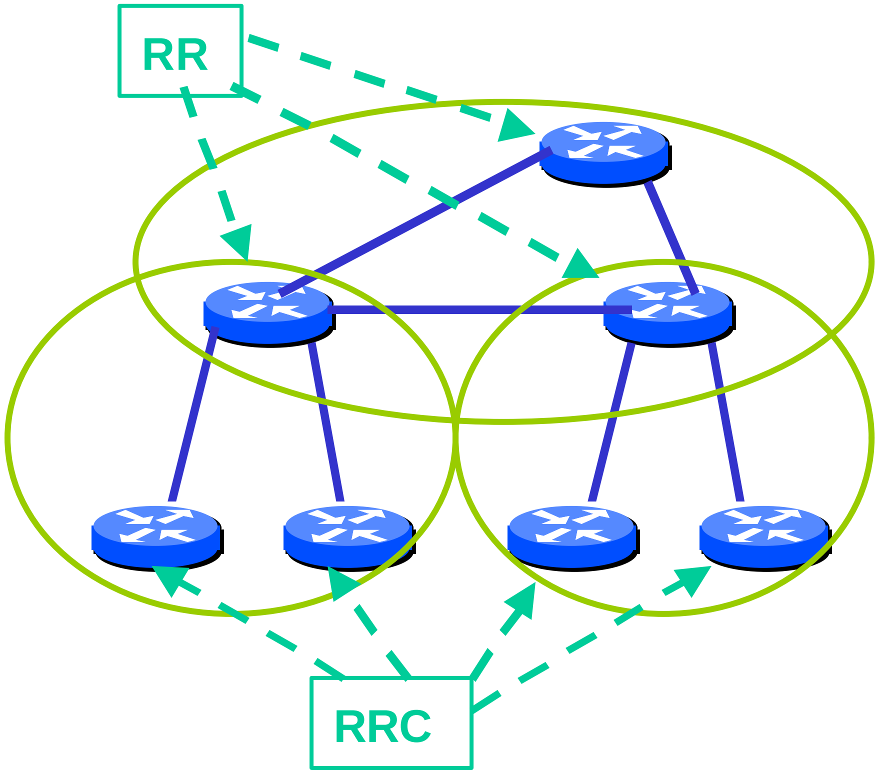
How to deliver all routing info without full-mesh?

subject to:

No Looping.

Route Reflection

RFC1966 (1996)



Hierarchical Approach
Consists of Route Reflectors (RR)
Route Reflector Clients (RRC).

RRC depends on RR in learning and advertising routes.

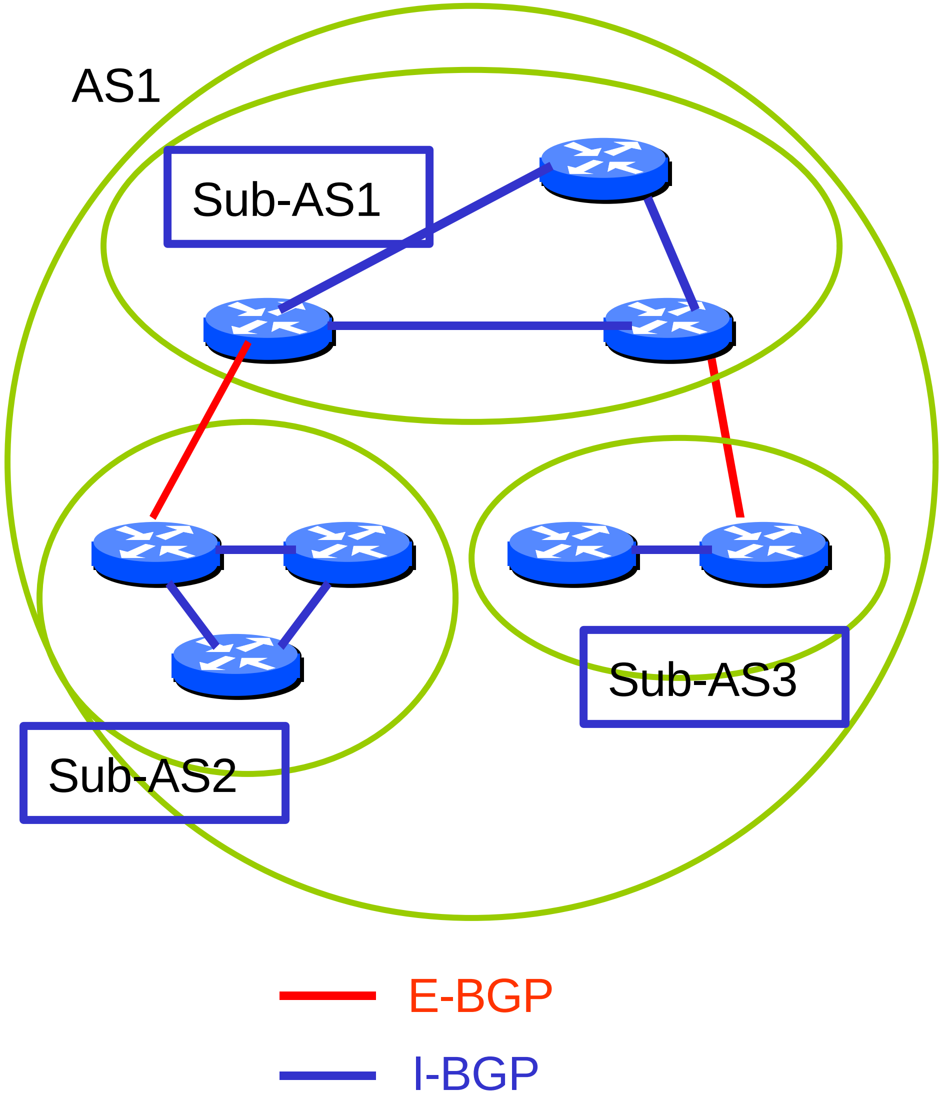
RRs are connected in a full mesh.

RR do not readvertise prefixes between non-clients.

Two more attributes are added.
ORIGINATOR_ID(9)
CLUSTER_LIST(10)

AS Confederation

RFC1965 (1996)

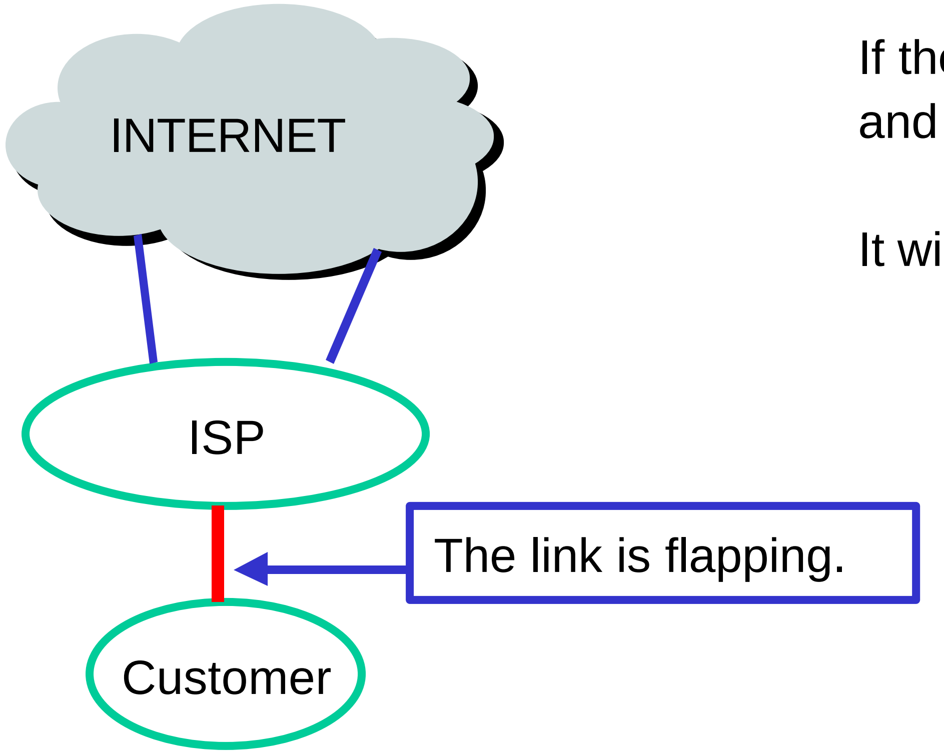


Divide and conquer
Divide a AS into sub-ASes.

Within a sub-AS, full mesh I-BGP
Between sub-ASes, E-BGP

Route Flapping

Route Flap : Constant up and down of a link



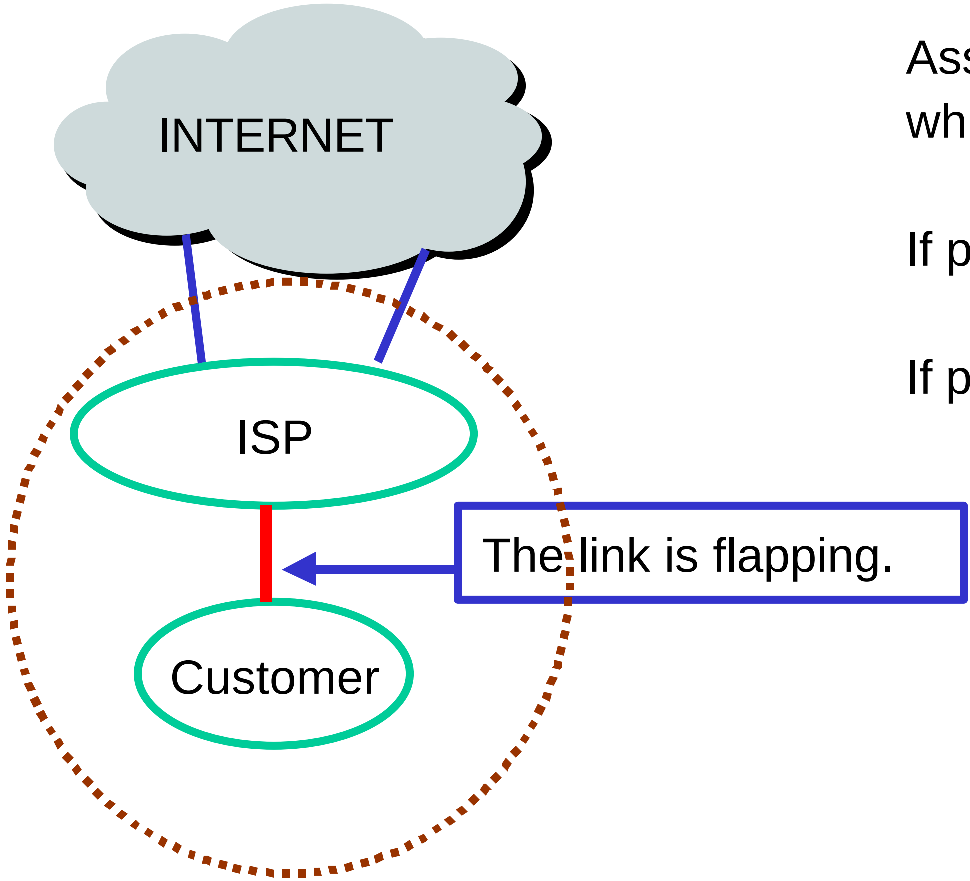
If the link is flapping, the ISP withdraw and establish routes for the customer.

It will affect the whole INTERNET.

Route Flap Damping

IETF draft (1998)

Idea : Localize the route advertisement of flapping routes.



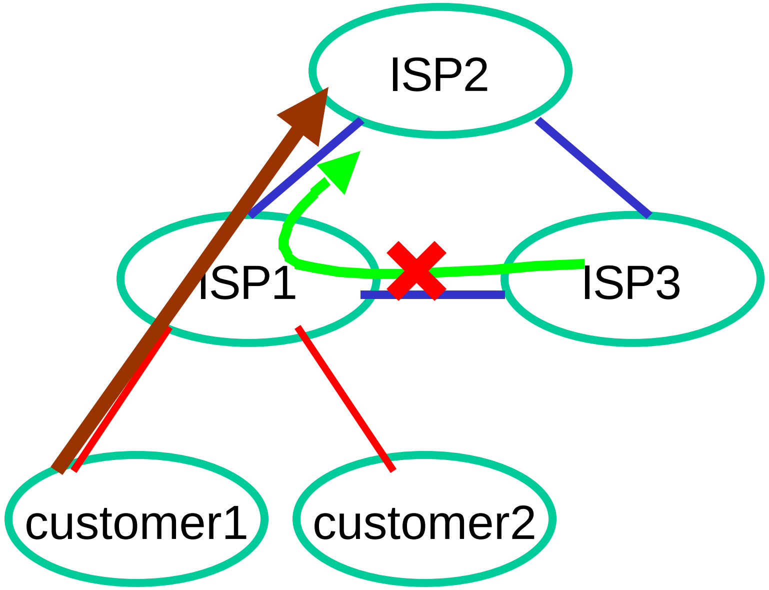
Assign penalty for every flapping, which decays exponentially.

If penalty is above a limit,
do not advertise.
If penalty is below a limit,
advertise.

Community Attribute

RFC 1997,RFC1998 (1996)

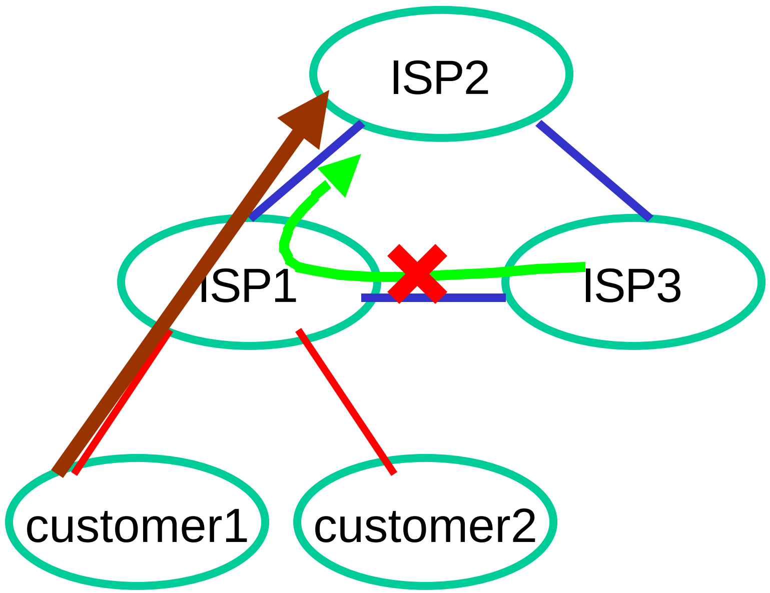
A BGP attribute added to simplify configuration of complex routing policies.



ISP1 offers transit services to its customers. But does not offer them to ISP2 and ISP3. That is, customers can pass ISP1's network, to ISP2 or ISP3. However, ISP2 (ISP3) cannot pass ISP1's network to ISP3 (ISP2).

ISP1 applies different policies to its customer and to different ISPs.

Community Attribute



Without community attributes, ISP1 should configure its router with a list of its customer prefixes to perform filtering.

However, if prefixes were several thousand, this would be very difficult.

BGP community attribute provide categorization of routes.

Summary of BGP

BGP is used in sharing routing information between different ASes.

BGP gives flexibility and control of traffics. However, load balancing is difficult problem in a multi-homed case.

We explained Important BGP attributes.

I-BGP is used within an AS to deliver BGP routing information.

Problems such as I-BGP scaling, Route Flapping and security issues are discussed and their possible solutions are discussed.

References

BGP4: Inter-domain routing in the Internet, J. Stewart III, Addison-Wesley, 1999

Interconnections, Bridges, Routers, Switches, and Internetworking Protocols, R. Perlman, 1st and 2nd editions, Addison-Wesley

BGP gives flexibility and control of traffics. However, load balancing is difficult problem in a multi-homed case.

We explained Important BGP attributes.

I-BGP is used within an AS to deliver BGP routing information.

Problems such as I-BGP scaling, Route Flapping and security issues are discussed and their possible solutions are discussed.

References

Books

OSPF: Anatomy of an Internet Routing Protocol, J. Moy, Addison-Wesley, 1998

BGP4: Inter-domain routing in the Internet, J. Stewart III, Addison-Wesley, 1999

Interconnections, Bridges, Routers, Switches, and Internetworking Protocols, R. Perlman, 1st and 2nd editions, Addison-Wesley

Internet Routing Architectures, Halabi, Cisco Press, 1997

Communication Networks: A first course, J. Walrand, McGraw Hill, 1998

Slides

Intro to BGP, P. Ferguson, Winter'97 NANOG, San Francisco, CA, February 9, 1997. Available at http://www.employees.org/~ferguson

OSPF SEMINAR with special emphasis on Multiple Area OSPF Issues Part I and Part II, P. Johri, 2000, AT&T proprietary

Internet Routing Technology, KRNET '99 Tutorial, Y. Shin, 1999

Available at http://www.krnet.or.kr

IP Routing Techniques, KRNET '98 Tutorial, W. Chang, 1998

Available at http://www.krnet.or.kr

References

RFCs and Internet Drafts

RFC 2328: OSPF version 2

RFC 1771: A Border Gateway Protocol 4

RFC 1965: Autonomous System Confederations for BGP

RFC 1996: BGP Route Reflection: An Alternative to Full Mesh I-BGP,

RFC 1997: BGP Community Attributes

RFC 1998: An Application of the BGP Community Attribute in Multihome Routing

RFC 2283: Multiprotocol Extensions for BGP-4

IETF draft: Capabilities Negotiation with BGP-4

Protection of GP Sessions via the TCP MD5 Signature Option

Use of BGP-4 Multiportocol Extensions for IPv6 Interdomain Routing

BGP-Route Flap Damping