UNIT3 ROUTING

Unicast Routing

In unicast routing, a packet is routed, hop by hop, from its source to its destination bythe help of forwarding tables. The source host needs no forwarding table because itdelivers its packet to the default router in its local network.

The destination host needsno forwarding table because it receives the packet from its default router in itslocal network.

Routing a packet from its source to its destination means routing the packet from a source router to a destination router.

Least-Cost Routing

When an internet is modeled as a weighted graph, one of the ways to interpret the best route from the source router to the destination router is to find the least cost between the two.

That is, the source router chooses a route to the destination router insuch a way that the total cost for the route is the least cost among all possible routes.

In Figure (below) the best route between A and E is A-B-E, with the cost of 6.

This meansthat each router needs to find the least-cost route between itself and all the other routersto be able to route a packet towards the destination.

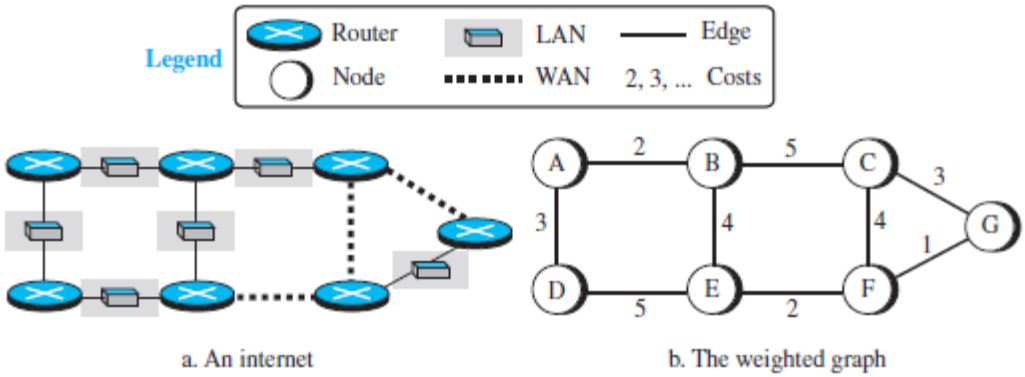


Fig:Internet with graph. [Source : Data Communications and Networking by Behrouz A. Forouzan]

Least-Cost Trees

If there are N routers in an internet, there are (N - 1) least-cost paths from each router to any other router.

This means we need N · (N - 1) least-cost paths for the whole internet.
For example, If we have only 10 routers in an internet, we need 90 least-cost paths.

A least-cost treeis a tree with the source router as the root that spans the whole graph (visits all other nodes) and in which the path between the root and any other node is the shortest.

In this way, we can have only one shortest-path tree for each node; we have N least-cost trees for the whole internet.

Figure (below) **shows the seven least-cost trees** for the internet

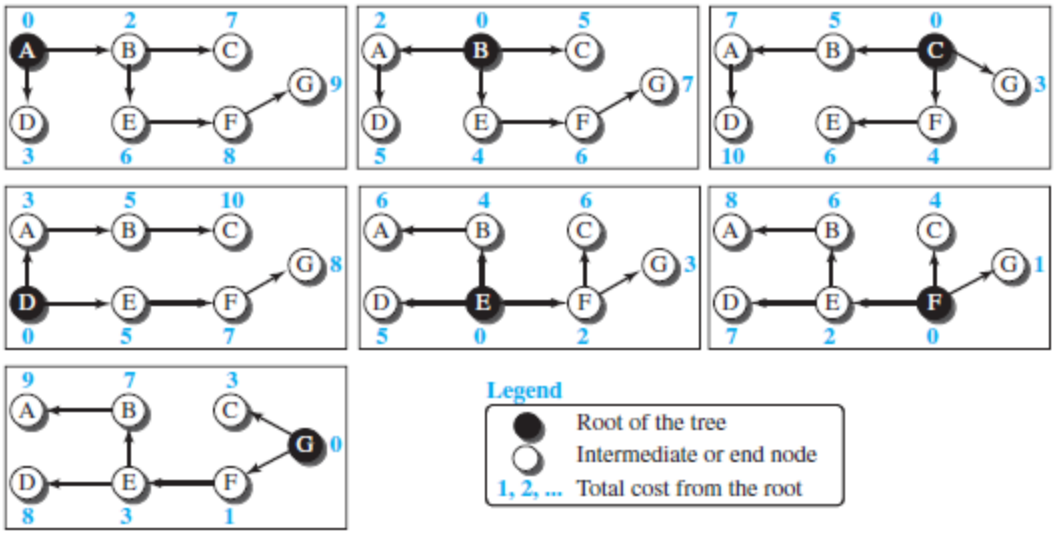


Fig: Seven least-cost trees for the internet .[Source : Data Communications and Networking by Behrouz A. Forouzan].

1. The least-cost route from X to Y in X's tree is the inverse of the least-cost route from Y to X in Y's tree; the cost in both directions is the same.

   For example, inFigure, the route from A to F in A's tree is (A - B- E- F), but the routefrom F to A in F's tree is (F - E- B- A), which is the inverse of the first route.The cost is 8 in each case.

2. Instead of travelling from X to Z using X's tree, we can travel from X to Y using X's tree and continue from Y to Z using Y's tree.

For example, in Figure , we can go from A to G in A's tree using the route (A - B- E - F - G). We can also go from A to E in A's tree (A - B- E) and then continue in E's tree using the route (E - F- G).

The combination of the two routes in the second case is the same route as in the first case. The cost in the first case is 9; the cost in the second case is also 9 (6 + 3).

ROUTING ALGORITHMS

Distance-Vector Routing

In distance-vector routing, a router continuously tells all of its neighbors what it knows about the whole internet.

Bellman-Ford Equation

In distance-vector routing **Bellman-Ford** equationis used to find the least cost (shortest distance) between a source node, ***x***, and a destinationnode, ***y***, through some intermediary nodes (**a, b, c,** . . .) when the costs between thesource and the intermediary nodes and the least costs between the intermediary nodes andthe destination are given.

The following shows the general case in which D***ij***is the shortest distance and c***ij***is the cost between nodes ***i*** and ***j***.

$$D_{xy} = \min\{(c_{xa} + D_{ay}), (c_{xb} + D_{by}), (c_{xc} + D_{cy}), \ldots\}$$

In distance-vector routing, we want to update an existing least cost with aleast cost through an intermediary node, such as ***z***, **ie,**if the intermediate node is shorter.

In this case,the equation can be written as:

$$D_{xy} = \min\{D_{xy}, (c_{xz} + D_{zy})\}$$

Graphical idea behind Bellman-Ford equation

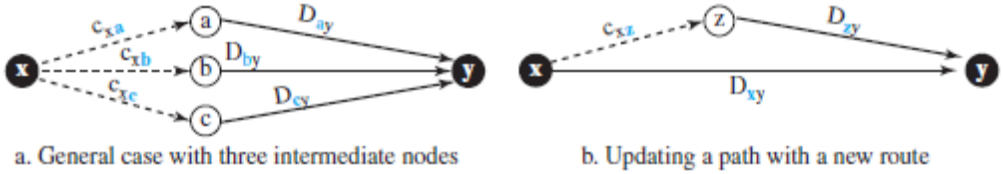


Fig: Graphical idea .[Source : Data Communications and Networking by Behrouz A. Forouzan].

Bellman-Ford equation help us to build a new least-cost pathfrom previously established least-cost paths.

In the Figure (above), we can think of (*a-y*),(*b-y*), and (*c-y*) as previously established least-cost paths and (*x-y*) as the newleast-cost path.

We can even think of this equation as the builder of a new least-cost treefrom previously established least-cost trees if we use the equation repeatedly.

Distance Vectors

The concept of a **distance vector** is the reason for the name distance-vector routing.A least-cost tree is a combination of least-cost paths from the root of the tree to all destinations.

Figure shows the tree for node A in the internet in Figure and the corresponding distance vector.

A distance vector does not give the path to the destinations as the least-cost tree does; itgives only the least costs to the destinations.

Note that the *name* of the distance vector defines the root, the *indexes* define the destinations,and the *value* of each cell defines the least cost from the root to the destination.

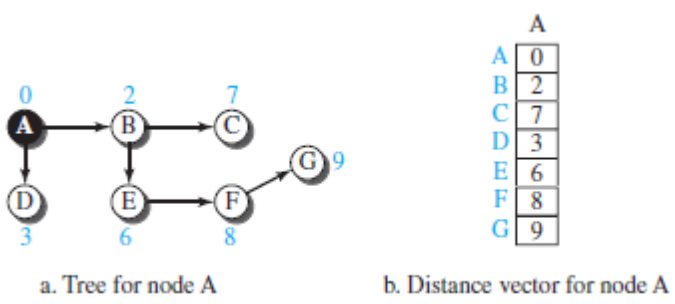


Fig: Seven least-cost trees for the internet . [Source : Data Communications and Networking by Behrouz A. Forouzan].

Each node in an internet, when it starts its function, creates a very basic distance vector with the minimum information the node can obtain from its neighborhood. The node sends some greeting messages out of its interfaces and discovers the identity ofthe immediate neighbors and the distance between itself and each neighbor.

It then makes a simple distance vector by inserting the discovered distances in the correspondingcells and leaves the value of other cells as infinity.

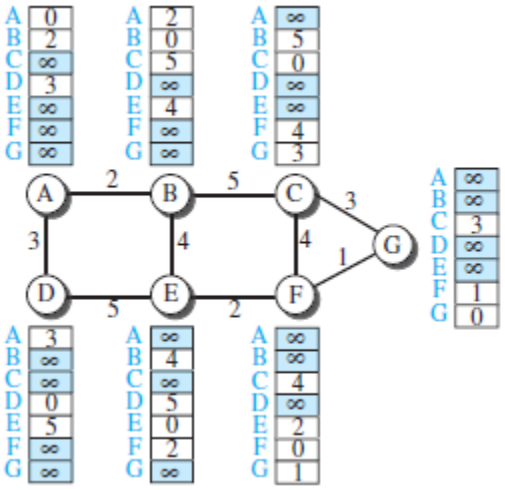


Fig: First distance vector for internet .[Source : Data Communications and Networking by Behrouz A. Forouzan].

Description of above diagram

Consider (For example), Node A thinks that it is not connected to node G because the correspondingcell shows the least cost of infinity.

To improve these vectors, the nodes in the internetneed to help each other by exchanging information. After each node has created its vector,it sends a copy of the vector to all its immediate neighbors. After a node receives adistance vector from a neighbor, it updates its distance vector using the Bellman-Fordequation (second case).

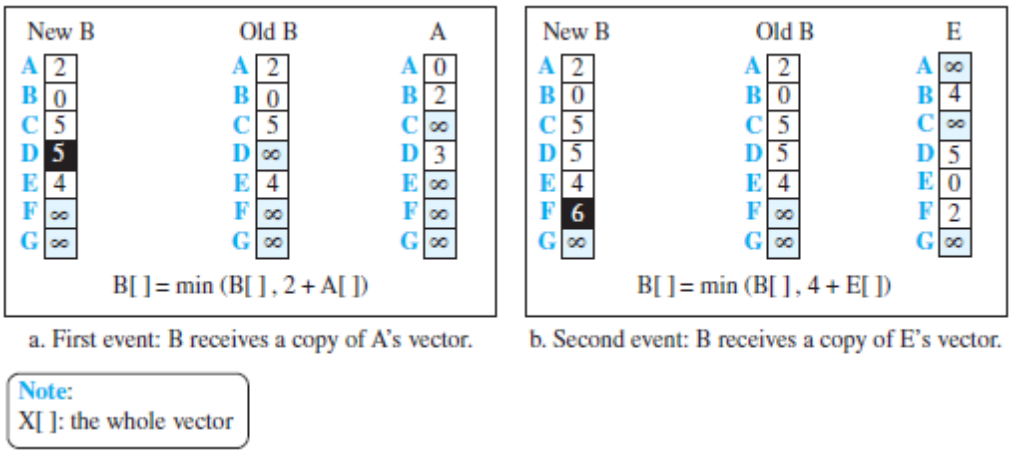
The figure(below) shows two asynchronous events, happening one after another with some time in between.

In the first event, node A has sent its vector to node B. Node B updates its vector using the cost cBA= 2. **In the second event**, node E has sent its vector to node B.Node B updates its vector using the cost cEA= 4.

After the first event, node B has one improvement in its vector: its least cost to node D has changed from infinity to 5 (via node A). After the second event, node B hasone more improvement in its vector; its least cost to node F has changed from infinityto 6 (via node E).

By exchanging the vectors,we can stabilize the system and allows all nodes to find the ultimate least costbetween themselves and any other node.

After updating a node, it immediately sends its updated vector to all neighbors.



a. First event: B receives a copy of A's vector.

b. Second event: B receives a copy of E's vector.

Fig:Updating distance vector.[Source : Data Communications and Networking by Behrouz A. Forouzan].

Count to Infinity

For a routingprotocol to work properly, if a link is broken (cost becomes infinity), every other routershould be aware of it immediately, but in distance-vector routing, this takes some time.
The problem is called count to infinity.

Two-Node Loop

Example of count to infinity is the two-node loop problem.

To understand the problem, consider the Figure (below).
The figure shows a system with three nodes.

Initially both nodes A and B know how to reach node X. But suddenly, the link between A and X fails. Node Achanges its table.

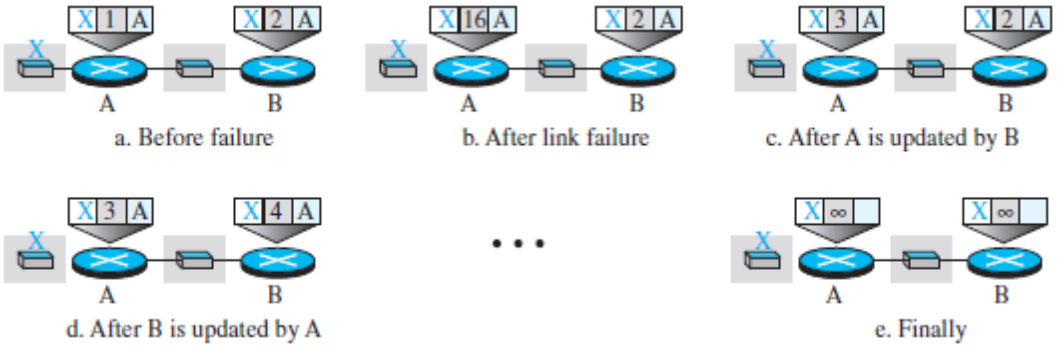


Fig:Two node instability.[Source : Data Communications and Networking by Behrouz A. Forouzan].

If A can send its table to B immediately, everything is fine. However, the system becomes unstable if B sends its forwarding table to A before receiving A'sforwarding table. Node A receives the update and, assuming that B has found a way toreach X, immediately updates its forwarding table. Now A sends its new update to B.

Now B thinks that something has been changed around A and updates its forwardingtable. The cost of reaching X increases gradually until it reaches infinity. At thismoment, both A and B know that X cannot be reached. However, during this time thesystem is not stable. Node A thinks that the route to X is via B; node B thinks that theroute to X is via A. If A receives a packet destined for X, the packet goes to B and thencomes back to A. Similarly, if B receives a packet destined for X, it goes to A andcomes back to B. Packets bounce between A and B, creating a two-node loop problem.

A few solutions have been proposed for instability of this kind.

Split Horizon

One solution to instability is called ***split horizon***. In this method, instead of floodingthe table through each interface, each node sends only part of its table through eachinterface.

If, according to its table, node B thinks that the optimum route to reach X isvia A, it does not need to advertise this piece of information to A; the information hascome from A (A already knows).

Taking information from node A, modifying it, andsending it back to node A is what creates the confusion. In thismethod, node B eliminatesthe last line of its forwarding table before it sends it to A. In this case, node Akeeps the value of infinity as the distance to X.

Later, when node A sends its forwardingtable to B, node B also corrects its forwarding table. The system becomes stableafter the first update: both node A and node B know that X is not reachable.

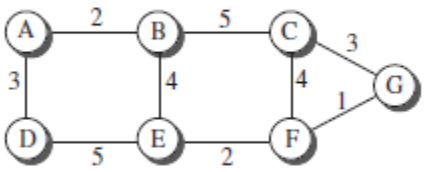
Link-State Routing

A routing algorithm that directly creates least-cost trees andforwarding tables is **link-state (LS) routing.** This method uses the term link-state to define the characteristic of a link (an edge) that represents a network in the internet.

Link-State Database (LSDB)

To create a least-cost tree with this method, each node needs to have a complete *map* ofthe network, which means it needs to know the state of each link. The collection of statesfor all links is called the **link-state database** (**LSDB).**

Example for LSDB



a. The weighted graph

	A	B	C	D	E	F	G
A	0	2	∞	3	∞	∞	∞
B	2	0	5	∞	4	∞	∞
C	∞	5	0	∞	∞	4	3
D	3	∞	∞	0	5	∞	∞
E	∞	4	∞	5	0	2	∞
F	∞	∞	4	∞	2	0	1
G	∞	∞	3	∞	∞	1	0

b. Link state database

Fig: Example Link state data base.[Source : Data Communications and Networking by Behrouz A. Forouzan].

This method is called **flooding.**

Each node cansend some greeting messages to all its immediate neighbors (those nodes to which it is connected directly) to collect two pieces of information for each neighboring node: theidentity of the node and the cost of the link.

The combination of these two pieces ofinformation is called the ***LS packet*** **(LSP**); the LSP is sent out of each interface, asshown in Figure .

When a node receives an LSPfrom one of its interfaces, it compares the LSP with the copy it may already have. If thenewly arrived LSP is older than the one it has (found by checking the sequence number),it discards the LSP.

If it is newer or the first one received, the node discards the oldLSP (if there is one) and keeps the received one. It then sends a copy of it out of each interface except the one from which the packet arrived. This guarantees that floodingstops somewhere in the network (where a node has only one interface).

After receiving all new LSPs, each node creates the comprehensiveLSDB as shown in Figure(below). This LSDB is the same for each node and shows thewhole map of the internet.

In other words, a node can make the whole map if it needsto, using this LSDB.

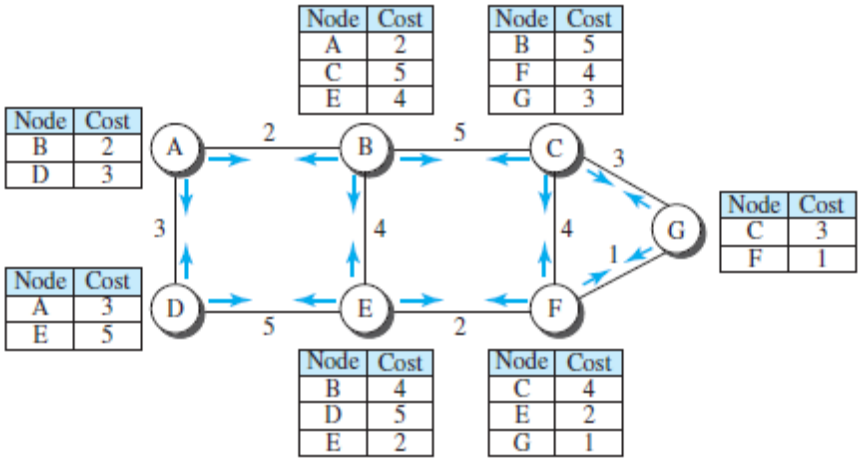


Fig:Creation of LSP.[Source : Data Communications and Networking by Behrouz A. Forouzan].

Note:

In the **distance-vector routing algorithm**, each router tells its neighbors what it knows about the whole internet; in the **link-state routing algorithm**, each router tellsthe whole internet what it knows about its neighbors.

Formation of Least-Cost Trees

To create a least-cost tree for itself, using the shared LSDB, each node needs to run thefamous **Dijkstra Algorithm.**

This algorithm uses the following steps:

1. The node chooses itself as the root of the tree, creating a tree with a single node, and sets the total cost of each node based on the information in the LSDB.

2. The node selects one node, among all nodes not in the tree, which is closest to the root, and adds this to the tree. After this node is added to the tree, the cost of all other nodes not in the tree needs to be updated because the paths may have been changed.

3. The node repeats step 2 until all nodes are added to the tree.

UNICAST ROUTING PROTOCOLS

Hierarchical Routing in Internet

The Internet today is made of a huge number of networks and routers that connectthem.

Routing in the Internet cannot be done using a single protocolfor two reasons:

A scalability problem and an administrative issue.

Scalability problem means that the size of the forwarding tables becomes huge, searching for a destination in a forwarding table becomes time-consuming, and updating creates a huge amount of traffic.

The administrative issue is related to the Internet structure .

Each ISP is run by an administrative authority. The administrator needs to have control in its system.

Hierarchical routing means considering each ISP as an **autonomous system (AS).**

Each AS can run a routing protocol that meets its needs, **but the global Internet** runs a global protocol to join and connect all ASs together.

The routing protocol run in each AS is referred to as intra-AS routing protocol, intradomain routing protocol, or interior gateway protocol (IGP);

The global routing protocol is referred to as inter-AS routing protocol, interdomain routing protocol, or exterior gateway protocol (EGP).

Routing Information Protocol (RIP)

The **Routing Information Protocol** (**RIP**) is one of the most widely used intradomainrouting protocols based on the distance-vector routing algorithm.

Hop Count

A router in this protocol implements the distance-vector routing algorithm. **First,** since a router in an AS needs to know how to forward a packet to different networks(subnets) in an AS, RIP routers advertise the cost of reaching differentnetworks instead of reaching other nodes in a theoretical graph.

The cost is defined between a router and the network in which the destination host islocated.

Second, to make the implementation of the cost simpler (independent fromperformance factors of the routers and links, such as delay, bandwidth, and so on**), the cost is defined as**the number of hops, which means the number of networks (subnets)a packet needs to travel through from the source router to the final destinationhost.

Note that the network in which the source host is connected is not counted inthis calculation because the source host does not use a forwarding table; the packet isdelivered to the default router.

Figure (below) shows the concept of hop count advertisedby three routers from a source host to a destination host.

In RIP, the maximumcost of a path can be 15, which means 16 is considered as infinity (no connection).

For this reason, RIP can be used only in autonomous systems in which the diameterof the AS is not more than 15 hops.

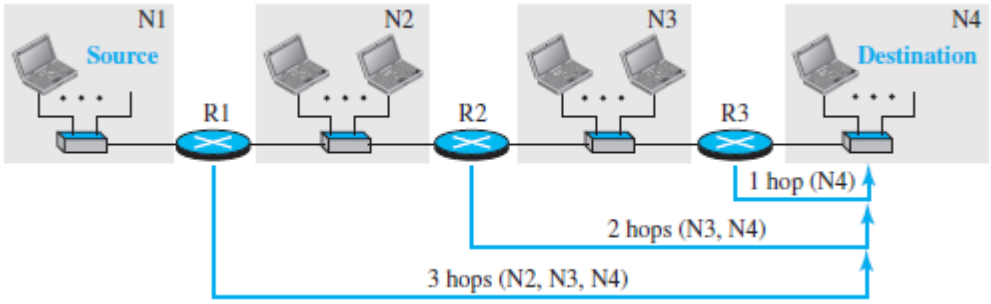


Fig: Hop counts in RIP .[Source : Data Communications and Networking by Behrouz A. Forouzan].

Forwarding Table

A forwarding table in RIP is a three-column table in which the first column is the address of the destination network, the second column is the address of thenext router to which the packet should be forwarded, and the third column is the cost(the number of hops) to reach the destination network.

Figure shows the three forwarding tables for the routers in Figure (above).

Note that the first and the third columnstogether convey the same information as does a distance vector, but the costshows the number of hops to the destination networks.

Forwarding table for R1

Destination network	Next router	Cost in hops
N1	——	1
N2	——	1
N3	R2	2
N4	R2	3

Forwarding table for R2

Destination network	Next router	Cost in hops
N1	R1	2
N2	——	1
N3	——	1
N4	R3	2

Forwarding table for R3

Destination network	Next router	Cost in hops
N1	R2	3
N2	R2	2
N3	——	1
N4	——	1

Fig:Forwarding tables in RIP.[Source : Data Communications and Networking by Behrouz A. Forouzan].

For example, R1 definesthat the next router for the path to N4 is R2; R2 defines that the next router to N4 isR3; R3 defines that there is no next router for this path. The tree is then R1 - R2- R3- N4.

What is the use of the third column in the forwarding table?.

The third column is not needed for forwarding the packet, but it is needed for updating the forwarding table when there is a change in the route.

RIP Implementation

RIP is implemented as a process that uses the service of UDP on the portnumber 520.

RIP is a routing protocol to help IP route its datagrams through the AS,the RIP messages are encapsulated inside UDP user datagrams, which in turn areencapsulated inside IP datagrams.

That is, RIP runs at the application layer, butcreates forwarding tables for IP at the network layer.

RIP Messages

Two RIP processes, a client and a server, need to exchange messages.

RIP-2 defines the format of the message, as shown in Figure .
The message Entry, can be repeated as needed in a message. Each entry carries the information related to one line in the forwarding table of the router that sends the message.

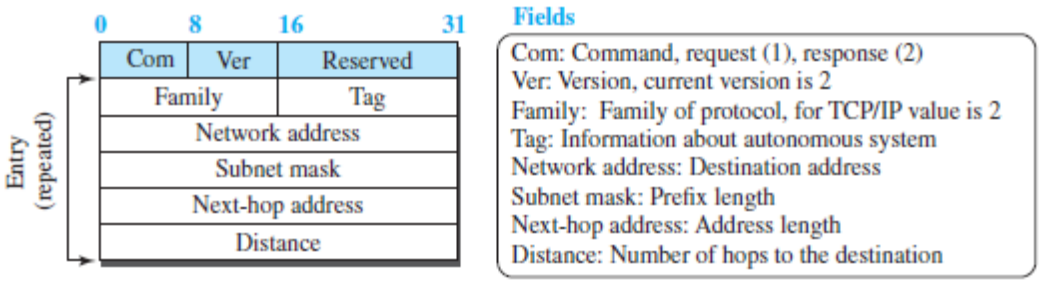


Fig: RIP message format.[Source : Data Communications and Networking by Behrouz A. Forouzan].

RIP has two types of messages:

Request and response. A request message is sentby a router that has just come up or by a router that has some time-out entries.
Arequest message can ask about specific entries or all entries.

A response (or update)message can be either solicited or unsolicited. A solicited response message is sentonly in answer to a request message. It contains information about the destinationspecified in the corresponding request message.

RIP Algorithm

RIP implements the same algorithm as the distance-vector routing algorithm.

- Instead of sending only distance vectors, a router needs to send the whole contents of its forwarding table in a response message.
- The receiver adds one hop to each cost and changes the next router field to the address of the sending router.
- The received router selects the old routes as the new ones except in the followingthree cases:

1. If the received route does not exist in the old forwarding table, it should be addedto the route.
2. If the cost of the received route is lower than the cost of the old one, the receivedroute should be selected as the new one.
3. If the cost of the received route is higher than the cost of the old one, but the value of the next router is the same in both routes, the received route should be selected as the new one.

Timers in RIP

RIP uses **three timers** to support its operation.

The periodic timer controls the advertisingof regular update messages. Each router has one periodic timer that is randomly setto a number between 25 and 35 seconds (to prevent all routers sending their messagesat the same time and creating excess traffic). The timer counts down; when zero isreached, the update message is sent, and the timer is randomly set once again.

The expiration timer governs the validity of a route. When a router receives update informationfor a route, the expiration timer is set to 180 seconds for that particular route. Everytime a new update for the route is received, the timer is reset.

If there is a problem on aninternet and no update is received within the allotted 180 seconds, the route is consideredexpired and the hop count of the route is set to 16, which means the destination isunreachable.

Every route has its own expiration timer. The garbage collection timer isused to purge a route from the forwarding table.

The garbage collection timer isused to purge a route from the forwarding table. When the information about a routebecomes invalid, the router does not immediately purge that route from its table.

Instead, it continues to advertise the route with a metric value of 16. At the same time,a garbage collection timer is set to 120 seconds for that route. When the count reacheszero, the route is purged from the table.

Open Shortest Path First (OSPF)

Open Shortest Path First (**OSPF**) is an intradomain routing protocol like RIP.It is based on the link-state routing protocol.

Metric

In OSPF, like RIP, the cost of reaching a destination from the host is calculated fromthe source router to the destination network.

However, each link (network) can beassigned a weight based on the throughput, round-trip time, reliability, and so on

Figure (below) shows the idea of the cost from a router to the destination hostnetwork.

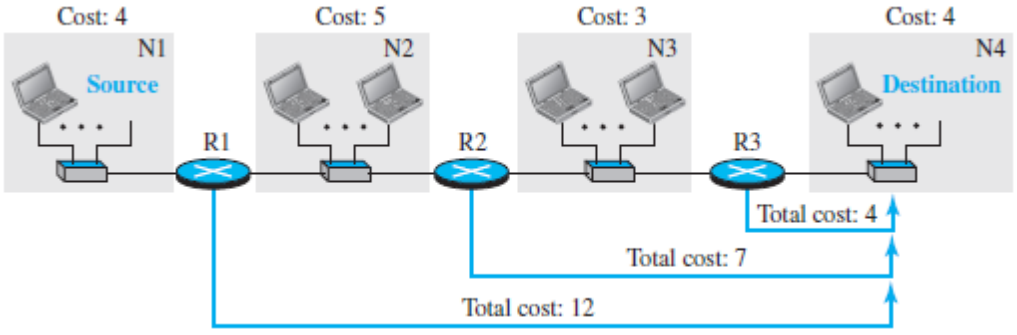


Fig:Metric in OSPF.[Source : Data Communications and Networking by Behrouz A. Forouzan].

Forwarding Tables

Each OSPF router can create a forwarding table after finding the shortest-path treebetween itself and the destination using Dijkstra's algorithm.

Forwarding table for R1

Destination network	Next router	Cost
N1	——	4
N2	——	5
N3	R2	8
N4	R2	12

Forwarding table for R2

Destination network	Next router	Cost
N1	R1	9
N2	——	5
N3	——	3
N4	R3	7

Forwarding table for R3

Destination network	Next router	Cost
N1	R2	12
N2	R2	8
N3	——	3
N4	——	4

Fig:Forwading table in OSPF.[Source : Data Communications and Networking by Behrouz A. Forouzan].

Areas

OSPF was designed to handle routing in a small or large autonomous system.

The formationof shortest-path trees in OSPF requires that all routers flood the whole AS with theirLSPs to create the global LSDB.

This may not create a problem in a small AS, but create traffic in large AS.

To prevent this, the ASneeds to be divided into small sections called ***areas.***

Each area acts as a small independentdomain for flooding .

Each router in an area needs to know the information about the link states not only in its area but also in other areas.

For this reason, one of the areas in the AS isdesignated as the *backbone area*, responsible for gluing the areas together.

The routers in the backbone area are responsible for passing the information collected by each areato all other areas. In this way, a router in an area can receive all LSPs generated in other
areas. For the purpose of communication, each area has an area identification.

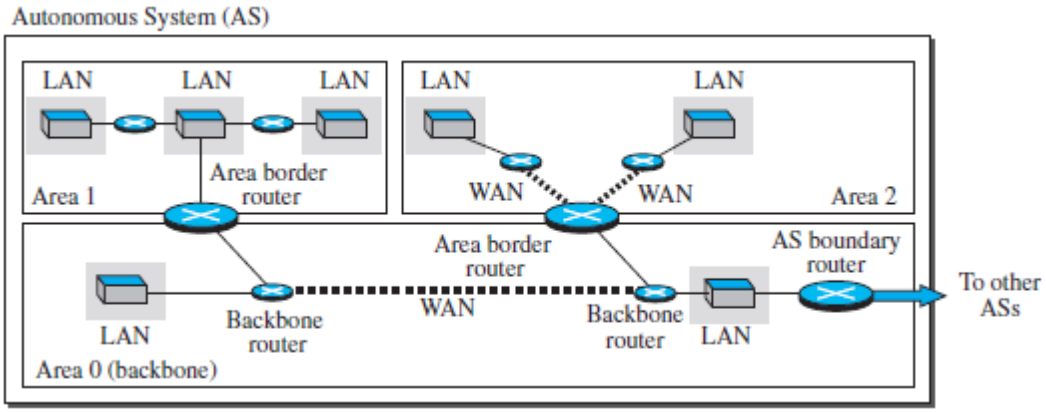


Fig:Areas in AS.[Source : Data Communications and Networking by Behrouz A. Forouzan].

OSPF Implementation

OSPF is implemented as a program in the network layer, using the service of the IP forpropagation. An IP datagram that carries a message from OSPF sets the value of theprotocol field to 89. This means that, the OSPF messages are encapsulated inside datagrams.

OSPF has two versions: version 1 and version 2.

OSPF Messages

OSPF is a very complex protocol; it has five different types of messages.

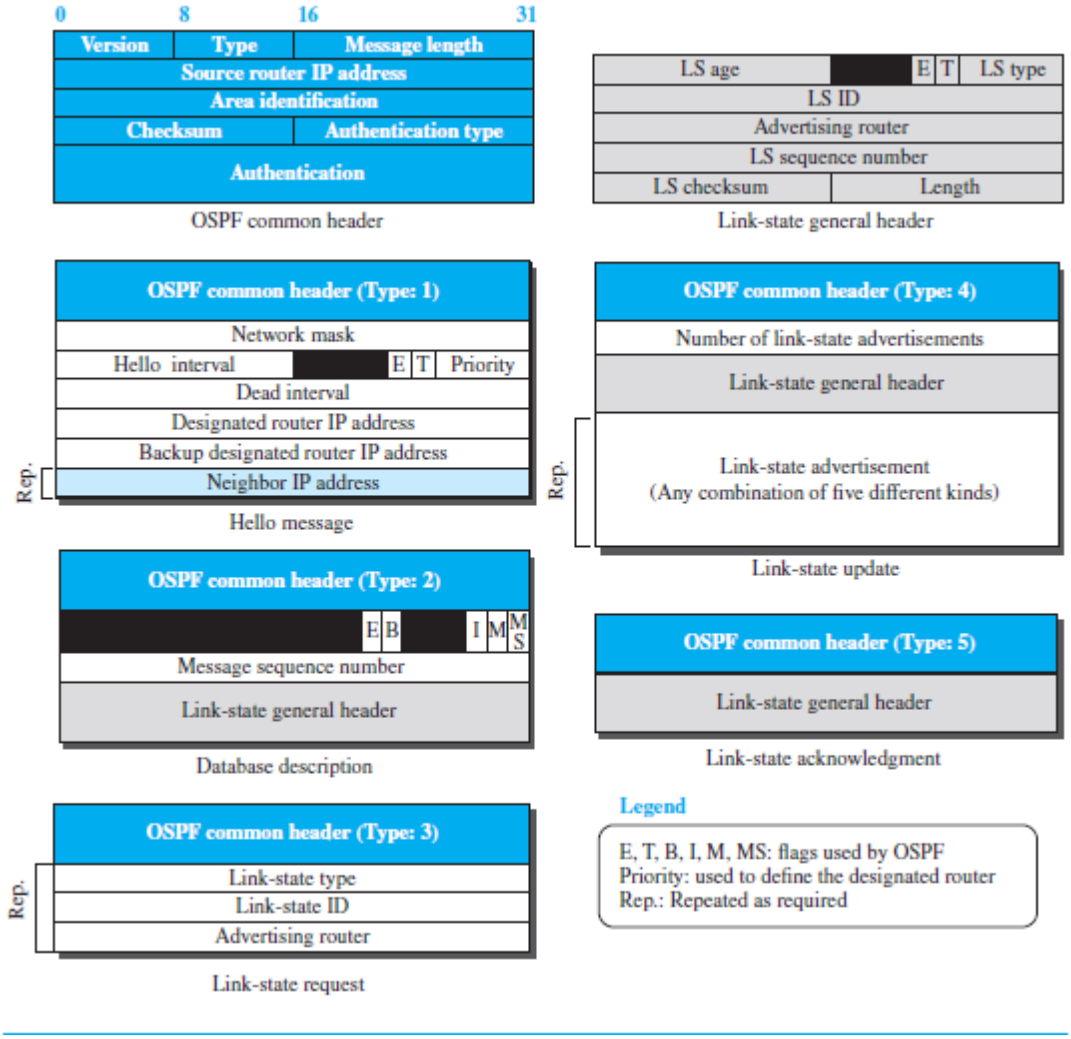


Fig: OSPF message format .[Source : Data Communications and Networking by Behrouz A. Forouzan].

The hello message (type 1) isused by a router to introduce itself to the neighbors.

The database description message (type 2) is sent in responseto the hello message to allow a newly joined router to acquire the full LSDB.

The link state request message (type 3) is sent by a router that needs information about a specificLS.

The link-state update message (type 4) is the main OSPF message used for buildingthe LSDB. This message, has five different versions (router link, networklink, summary link to network, summary link to AS border router, and external link).

The link-state acknowledgment message (type 5) is used to createreliability in OSPF; each router that receives a link-state update message needs toacknowledge it.

OSPF Algorithm

OSPF implements the link-state routing algorithm .

After each router has created the shortest-path tree, the algorithm needs to use it tocreate the corresponding routing algorithm.

The algorithm needs to be augmented to handle sending and receiving all five types of messages.

Border Gateway Protocol Version 4 (BGP4)

The **Border Gateway Protocol version 4 (BGP4)** is the only inter domain routing protocol used in the Internet today.

Consider an example of an internet with four autonomous systems. AS2, AS3, and AS4 are *stub* autonomous systems; AS1 is a *transient* one.

Here,data exchange between AS2, AS3, and AS4 should pass through AS1.

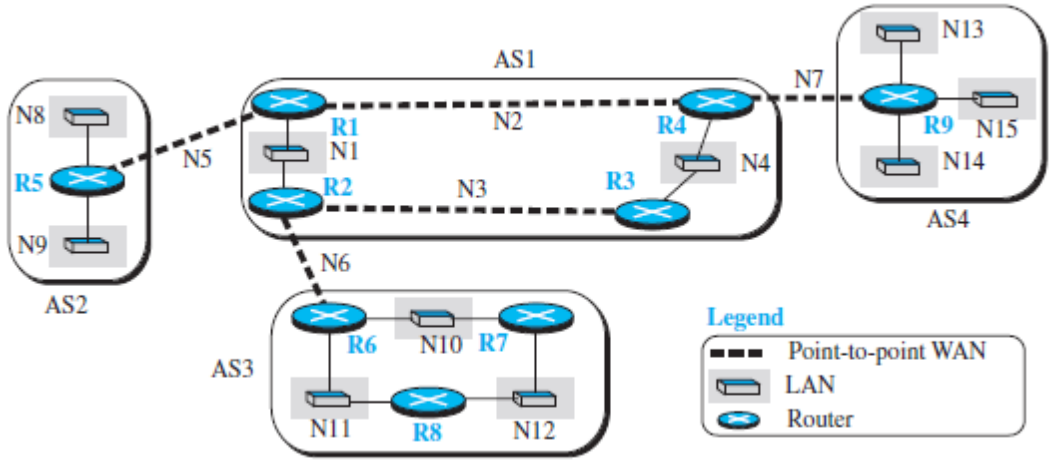


Fig:Sample internet with four AS.[Source : Data Communications and Networking by Behrouz A. Forouzan].

Each router in each AS knows how to reach a network that isin its own AS, but it does not know how to reach a network in another AS.

To enable each router to route a packet to any network in the internet, we first install a variation of BGP4, called external BGP (eBGP), on each border router (theone at the edge of each AS which is connected to a router at another AS).

We theninstall the second variation of BGP, called internal BGP (iBGP), on all routers.

The border routers will be running three routing protocols (intradomain, eBGP, and iBGP), but other routers are running two protocols (intradomain and iBGP).

Operation of External BGP (eBGP)

BGP is a point-to-point protocol. When the software is installed on two routers, they try to create a TCP connection using the well-known port 179.

The two routers that run the BGP processes are calledBGP peers or BGP speakers.

The eBGP variation of **BGP allows two physically connected border routers** in twodifferent ASs to form pairs of eBGP speakers and exchange messages.

The routers that we use in Figure has three pairs: R1-R5, R2-R6, and R4-R9.

The connection between these pairs is established over three physical WANs (N5,N6, and N7). There is a need for a logical TCP connection to be created overthe physical connection to make the exchange of information possible.

Each logical connection in BGP **is referred to as a session**. This means that we need threesessions, as shown in Figure (below).

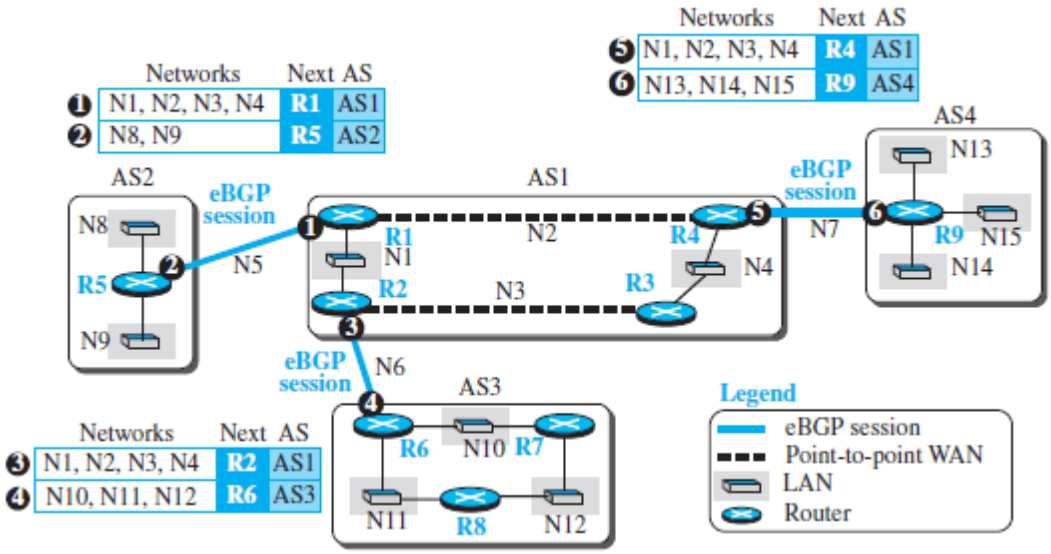


Fig: EBGP operation .[Source : Data Communications and Networking by Behrouz A. Forouzan].

The circled number defines the sending router in each case.

For example, message number 1 is sent by router R1 and tells router R5 that N1, N2, N3,and N4 can be reached through router R1 (R1 gets this information from the correspondingintradomain forwarding table).

Router R5 can now add these pieces ofinformation at the end of its forwarding table. When R5 receives any packet destinedfor these four networks, it can use its forwarding table and find that the next router is R1.

Messages

BGP four types of messages for communication between the BGP speakers acrossthe ASs and inside an AS:

Four messages are

open, update, keepalive, and notification .

All BGP packets share the same common header.

Open Message. To create a neighborhood relationship, a router running BGP opens a TCP connection with a neighbor and sends an open message.

Update Message.

The update message is usedby a router to withdraw destinations that have been advertised previously, toannounce a route to a new destination, or both.

Note that BGP can withdraw severaldestinations that were advertised before, but it can only advertise one new destinationin a single update message.

Keepalive Message. The BGP peers that are running exchange keepalive messagesregularly (before their hold time expires) to tell each other that they are alive.

Notification.A notification message is sent by a router whenever an error conditionis detected or a router wants to close the session.

Performance

BGP performance can be compared with RIP. BGP speakers exchange a lot of messagesto create forwarding tables, but BGP is free from loops and count-to-infinity.

Multicasting

In multicasting, there is one source and a group of destinations. The relationship isone to many.

In this type of communication, the source address is a unicast address,but the destination address is a group address, a group of one or more destinationnetworks in which there is at least one member of the group that is interested inreceiving the multicast datagram.

Multicasting starts with a single packet from the source that is duplicated by the routers. The destination address in each packet is the same for all duplicates.

Note thatonly a single copy of the packet travels between any two routers.

Multicast Applications

Multicasting has many applications.

Access to Distributed Databases.

Most of the large databases today are distributed.That is, the information is stored in more than one location, usually at the time ofproduction.

The user who needs to access the database does not know the location of the information. A user's request is multicast to all the database locations, andthe location that has the information responds.

Information Dissemination.

Businesses often need to send information to their customers.If the nature of the information is the same for each customer, it can be multicast.In this way a business can send one message that can reach many customers.

Teleconferencing. Teleconferencing involves multicasting. The individuals attendinga teleconference all need to receive the same information at the same time.

Distance Learning.One growing area in the use of multicasting is distance learning.

Lessons taught by one professor can be received by a specific group of students.

MULTICASTING BASICS

In multicast communication, the sender is only one, but the receiver is many, sometimes thousands or millions spread all over the world. It should be clear that wecannot include the addresses of all recipients in the packet.

The destination address of a packet, as described in the Internet Protocol (IP) should be only one. For this reason,we need multicast addresses. A multicast address defines a group of recipients, not asingle one.

A multicast address is an identifier for a group. If a new group is formed with some active members, an authority can assign an unused multicastaddress to this group to uniquely define it.

Multicast Forwarding

Important issue in multicasting is the decision a router needs to make to forward a multicast packet.

Forwarding in unicast and multicast communication isdifferent in two aspects:

1. In unicast communication, the destination address of the packet defines one single destination. The packet needs to be sent only out of one of the interfaces, the interface which is the branch in the shortest-path tree reaching the destination with the minimum cost.

   In multicast communication, the destination of the packet defines one group, but that group may have more than one member in the internet.

   To reach all of the destinations, the router may have to send the packet out of more than one interface.

In unicasting, the destination network N1 cannot be in more than one part of the internet; in multicasting, the group G1 is in more than one part of the internet.

Forwarding decisions in unicast communication depend only on the destination address of the packet.

Forwarding **decisions in multicast** communication dependon both the destination and the source address of the packet.

In other words, in unicasting,forwarding is based on where the packet should go; in multicasting, forwardingis based on where the packet has comefrom.

Figure (below) shows the concept.

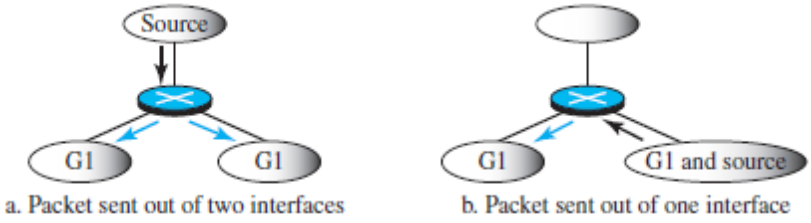


Fig: Forwarding depends on destination and source.[Source : Data Communications and Networking by Behrouz A. Forouzan].

In part a of the figure, the source is in a sectionof the internet where there is no group member. In part b, the source is in asection where there is a group member.
In part a, the router needs to send out thepacket from two interfaces; in part b, the router should send the packet only fromone interface to avoid sending a second copy of the packet from the interface it hasarrived at.

Two Approaches to Multicasting

Source-Based Tree Approach

In the **source-based tree** approach to multicasting, each router needs to create a separatetree for each source-group combination.

If there are *m* groupsand *n* sources in the internet, a router needs to create ($m \cdot n$) routing trees. In each tree,the corresponding source is the root, the members of the group are the leaves, and therouter itself is somewhere on the tree.

Group-Shared Tree Approach

In the **group-shared tree** approach, a router act like a source for each group. The designated router, which is called the core router or the rendezvouspoint router, acts as the representative for the group.

Any source that has a packet tosend to a member of that group sends it to the core center (unicast communication) and the core center is responsible for multicasting.

The core center creates one single routingtree with itself as the root and any routers with active members in the group as theleaves.

In this approach, there are *m* core routers (one for each group) and each corerouter has a routing tree, for the total of *m* trees. Therefore the number of routingtrees is reduced from (*m* * *n*) in the source-based tree approach to *m* in this approach.

INTRADOMAIN MULTICAST PROTOCOLS

Multicast Distance Vector (DVMRP)

The **Distance Vector Multicast Routing Protocol** (**DVMRP**) is the extension of theRouting Information Protocol (RIP) which is used in unicast routing. It uses the sourcebasedtree approach to multicasting.

Multicast tree in three steps:

1. The router uses an algorithm called *reverse path forwarding* (RPF) to simulate creating part of the optimal source-based tree between the source and itself.

2. The router uses an algorithm called *reverse path broadcasting* (RPB) to create a broadcast (spanning) tree whose root is the router itself and whose leaves are all networks in the internet.

3. The router uses an algorithm called *reverse path multicasting* (RPM) to create a multicast tree by cutting some branches of the tree that end in networks with no member in the group.

Reverse Path Forwarding (RPF)

The first algorithm, **reverse path forwarding (RPF),** forces the router to forward amulticast packet from one specific interface: the one which has come through the shortestpath from the source to the router.

The router does not know the shortest path from the source to itself, but it can find which is thenext router in the shortest path from itself to the source (reverse path).

The router simplyconsults its unicast forwarding table, pretending that it wants to send a packet to thesource; the forwarding table gives the next router and the interface the message that thepacket should be sent out in this reverse direction.

The router uses this information toaccept a multicast packet only if it arrives from this interface. This is needed to preventlooping. In multicasting, a packet may arrive at the same router that has forwarded it.

Ifthe router does not drop all arrived packets except the one, multiple copies of the packetwill be circulating in the internet.

Reverse Path Broadcasting (RPB)

The RPF algorithm helps a router to forward only one copy received from a source anddrop the rest.

When we think about broadcasting in the second step, we need toremember that destinations are all the networks (LANs) in the internet. To be efficient,we need to prevent each network from receiving more than one copy of the packet.

If anetwork is connected to more than one router, it may receive a copy of the packet fromeach router. RPF cannot help here, because a network does not have the intelligence toapply the RPF algorithm; we need to allow only one of the routers attached to a networkto pass the packet to the network.

One way to do so is to designate only one router as theparent of a network related to a specific source. When a router that is not the parent ofthe attached network receives a multicast packet, it simply drops the packet.

There are several ways that the parent of the network related to a network can be selected; one wayis to select the router that has the shortest path to the source (using the unicast forwardingtable, again in the reverse direction).

In other words, after this we have ashortest-path tree with the source as the root and all networks (LANs) as the leaves.

Every packet started from the source reaches all LANs in the internet travelling theshortest path. Figure shows how RPB can avoid duplicate reception in a networkby assigning a designated parent router, R1, for network N.

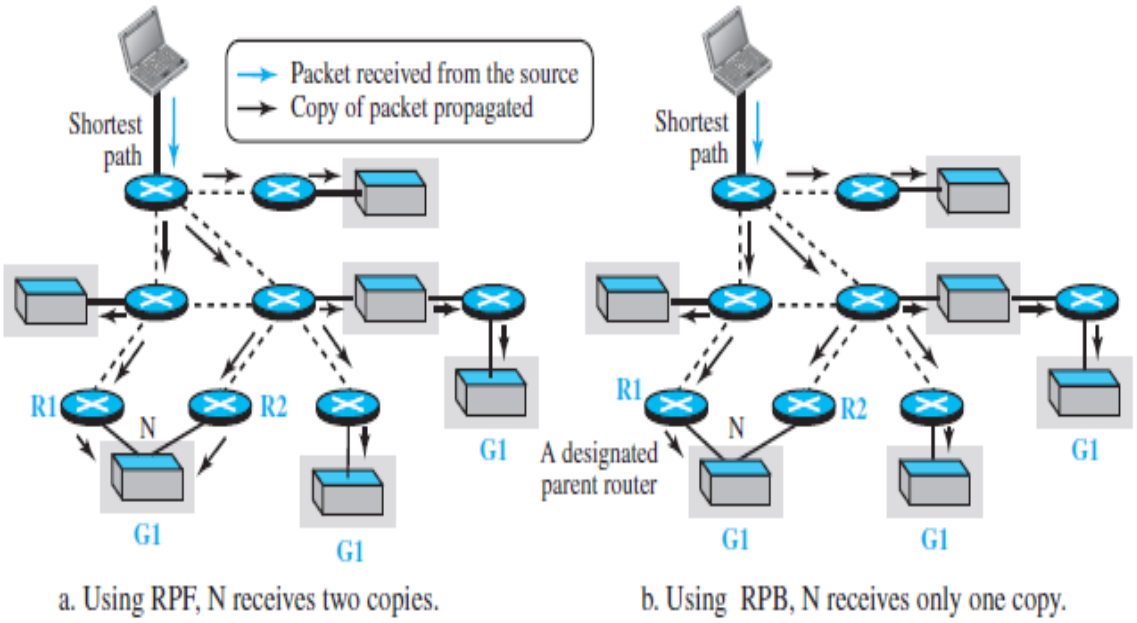


Fig: Reverse path broadcasting.[Source : Data Communications and Networking by Behrouz A. Forouzan].

Reverse Path Multicasting (RPM)

To increase efficiency, the multicast packet must reach only those networks that have active members for that particular group. This is called **reverse path multicasting (RPM).**

To change the broadcast shortest-path tree to a multicast shortest-path tree, each router needs to prune (make inactive) the interfaces that do not reach a network with active members corresponding to a particular source-group combination.

This step can be done bottom-up, from the leaves to the root. At the leaf level, the routersconnected to the network collect the membership information using the IGMP protocol.

The parent router of the network can then disseminate thisinformation upward using the reverse shortest-path tree from the router to the source, thesame way as the distance vector messages are passed from one neighbor to another.

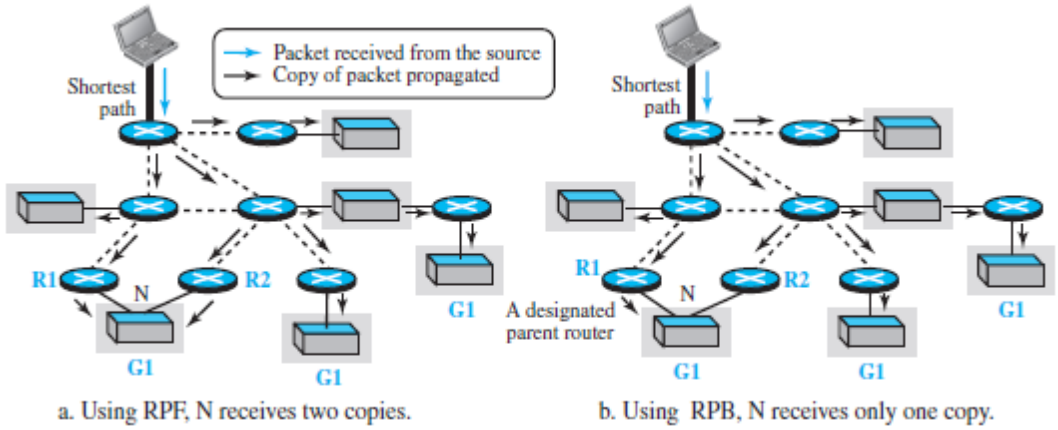


Fig:RPB vs RPF .[Source : Data Communications and Networking by Behrouz A. Forouzan].

Multicast Link State (MOSPF)

Multicast Open Shortest Path First (**MOSPF**) is the extension of the Open ShortestPath First (OSPF) protocol, which is used in unicast routing. It uses the sourcebasedtree approach to multicasting.

In multicasting, each router needs to have a database, as with thecase of unicast distance-vector routing, to show which interface has an active memberin a particular group.

A router follow these steps to forward amulticast packet received from source S and to be sent to destination G (a group ofrecipients):

The router uses the Dijkstra algorithm to create a shortest-path tree with S as the root and all destinations in the internet as the leaves. Note that this shortest-path tree is different from the one the router normally uses for unicast forwarding, in which the root of the tree is the router itself.

Here, the root of the tree is thesource of the packet defined in the source address of the packetThe router finds itself in the shortest-path tree created in the first step. In otherwords, the router creates a shortest-path subtree with itself as the root of the subtree.

The shortest-path subtree is actually a broadcast subtree with the router as the root and all networks as the leaves.

The IGMP protocol is used to find the information at the leaf level. The router can now forward the received packet out of only those interfaces that correspond to the branches of the multicast tree.

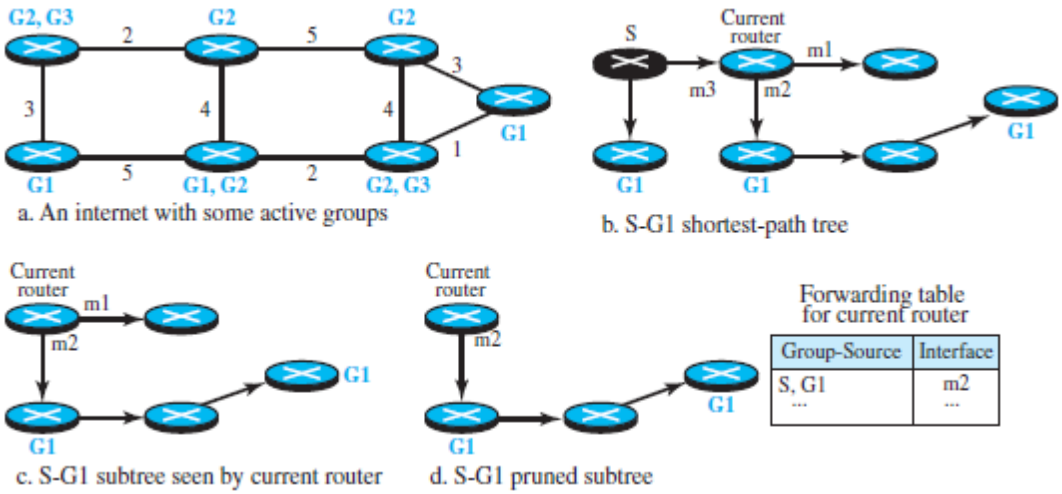


Fig: Tree formation in MOSPF.[Source : Data Communications and Networking by Behrouz A. Forouzan].

Protocol Independent Multicast (PIM)

Protocol Independent Multicast (PIM) is the name given to a common protocol thatneeds a unicast routing protocol for its operation, but the unicast protocol can be eithera distance-vector protocol or a link-state protocol.

PIM uses theforwarding table of a unicast routing protocol to find the next router in a path to thedestination, but it does not matter how the forwarding table is created.

Feature of PIM:

It can work in two different modes: dense and sparse.

The term dense means that the number of active members of a group in the internet is large;the probability that a router has a member in a group is high.

For example, in a popular teleconference that has a lot of members.

The term sparse, means that only a few routers in the internet have active members in the group; the probability that a router has a member of the group is low.

For example, in a technical teleconference where a number of members are spreadsomewhere in the internet. When the protocol is working in the dense mode, it isreferred to as PIM-DM; when it is working in the sparse mode, it is referred to as PIMSM.

Protocol Independent Multicast-Dense Mode (PIM-DM)

When the number of routers with attached members is large relative to the number ofrouters in the internet, PIM works in the dense mode and is called **PIM-DM**.

In thismode, the protocol uses a source-based tree approach.

PIM-DM uses only two strategies described in DVMRP: RPF and RPM.

The two steps used in PIM-DM .

1. A router that has received a multicast packet from the source S destined for the group G first uses the RPF strategy to avoid receiving a duplicate of the packet. It consults the forwarding table of the unicast protocol to find the next router if it wants to send a message to the source S (in the reverse direction).

   If the packet has not arrived from the next router in the reverse direction, it drops the packet and sends a prune (remove things which are not needed) message in that direction to prevent receiving future packets related to (S, G).

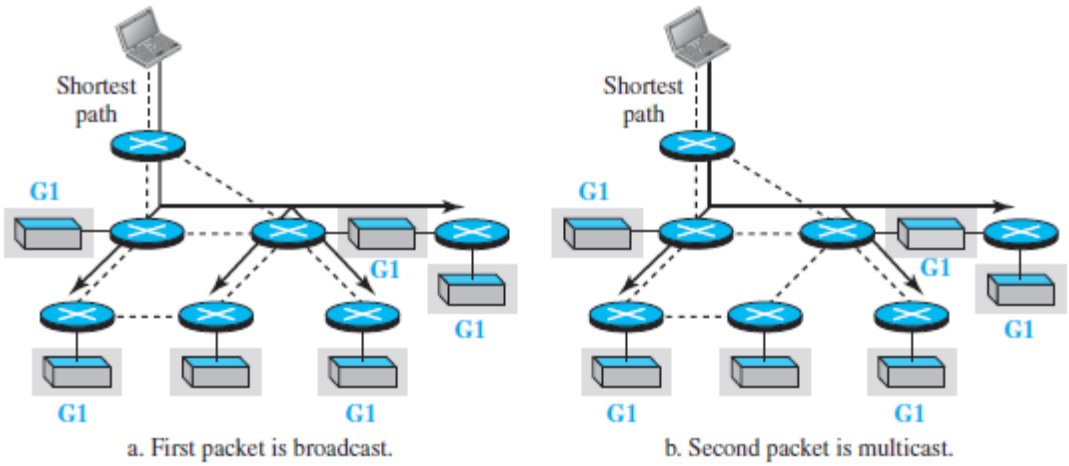


Fig: Idea behind PIM- DM.[Source : Data Communications and Networking by Behrouz A. Forouzan].

2. If the packet in the first step has arrived from the next router in the reverse direction, the receiving router forwards the packet from all its interfaces except the one from which the packet has arrived .

Note that this is broadcasting instead of a multicasting if the packet is the first packet from the source S to group G.

Each router downstream that receives an unwanted packet sends a prune message to the router upstream, and eventually the broadcasting is changed to multicasting.

Figure (above) PIM-DM. The first packet is **broadcast** to all networks, which have or do not have members. After a prune message arrives from a router with no member, the second packet is only **multicast**.

Protocol Independent Multicast-Sparse Mode (PIM-SM)

When the number of routers with attached members is small relative to the number ofrouters in the internet, PIM works in the sparse mode and is called **PIM-SM.**

In thisenvironment, PIM-SM uses a group-shared tree approach to multicasting.

The corerouter in PIM-SM is called the rendezvous point (RP). Multicast communication isachieved in two steps.

Any router that has a multicast packet to send to a group of destinationsfirst encapsulates the multicast packet in a unicast packet (tunneling) andsends it to the RP. The RP then decapsulates the unicast packet and sends the multicastpacket to its destination.

PIM-SM uses a complex algorithm to select one router among all routers in the internet as the RP for a specific group. This means that if we have m active groups, weneed m RPs, although a router may serve more than one group.

After the RP for eachgroup is selected, each router creates a database and stores the group identifier and theIP address of the RP for tunneling multicast packets to it.

PIM-SM uses a spanning multicast tree rooted at the RP with leaves pointing to designated routers connected to each network with an active member. A very interestingpoint in PIM-SM is the formation of the multicast tree for a group.

To create a multicast tree rooted at the RP, PIM-SM uses join and prune messages.

Figure (below) shows the operation of join and prune messages in PIM-SM.

First, threenetworks join group G1 and form a multicast tree. Later, one of the networks leaves thegroup and the tree is pruned.

The join message is used to add possible new branches to the tree; the **prune message**is used to cut branches that are not needed.

When a designated router finds outthat a network has a new member in the corresponding group (via IGMP), it sends ajoin message in a unicast packet destined for the RP.

The packet travels through the unicastshortest-path tree to reach the RP. Any router in the path receives and forwards thepacket, but at the same time, the router adds two pieces of information to its multicastforwarding table.

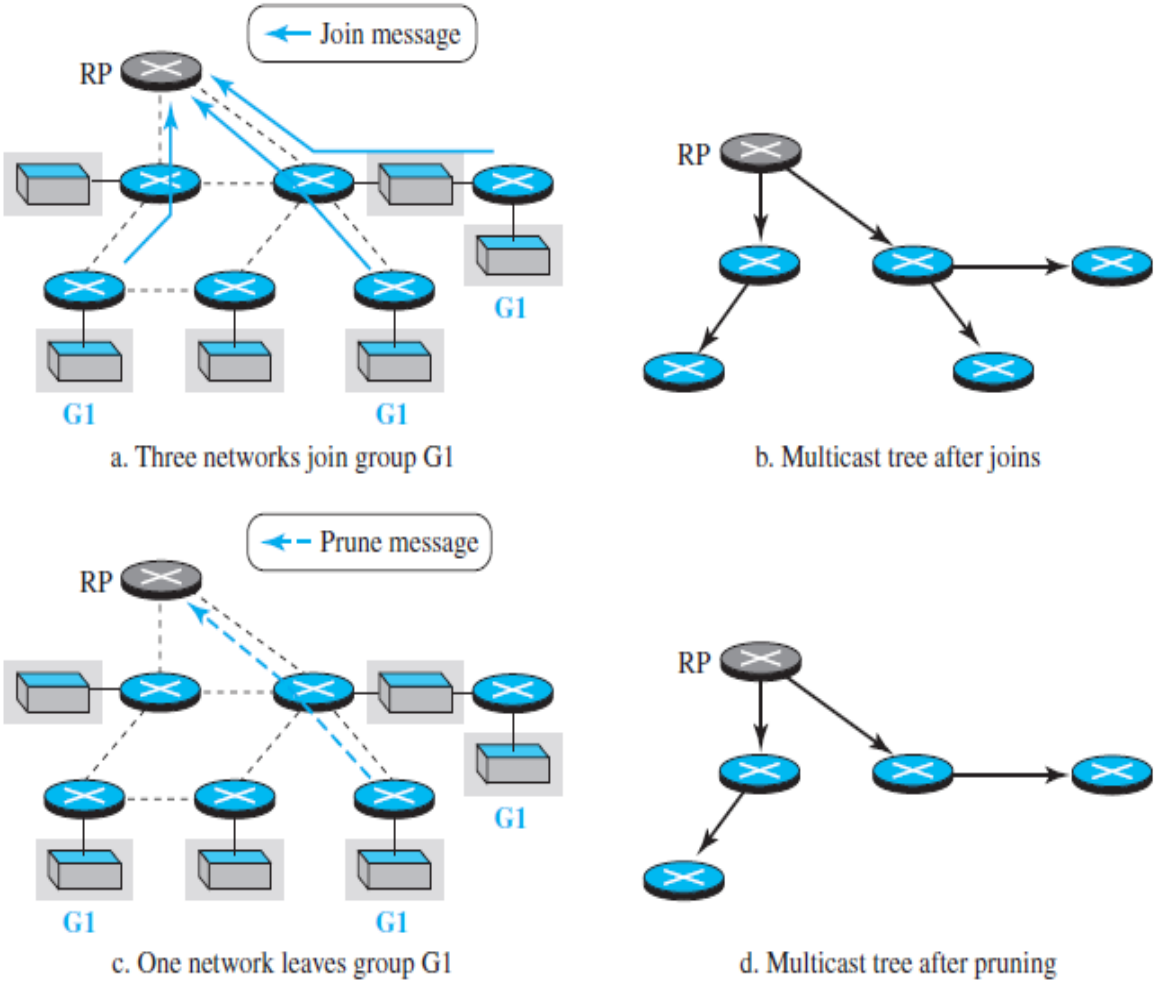


Fig: Join and Prune message format .[Source : Data Communications and Networking by Behrouz A. Forouzan].

The number of the interface through which the join message was sent to the RP is marked(if not already marked) as the only interface through which the multicast packet destinedfor the same group should be received.

In this way, the first join message sent by a designated router creates a path from the RP to one of the networks with groupmembers.

To avoid sending multicast packets to networks with no members, PIM-SM uses the prune message.

INTERDOMAIN MULTICAST PROTOCOLS

When the members of the groups are spread among different domains (ASs), we need an interdomain multicast routing protocol.

One common protocol for interdomain multicast routing is called Multicast Border Gateway Protocol (MBGP), which is the extension of BGP .

MBGP provides two paths between ASs: one for unicasting and one for multicasting.

Information about multicasting is exchanged between border routersin different ASs. MBGP is a shared-group multicast routing protocol in which onerouter in each AS is chosen as the rendezvous point (RP).

The problem with MBGP protocol is that it is difficult to inform an RP about the sources of groups in other ASs. The Multicast Source Discovery Protocol (MSDP) is anew suggested protocol that assigns a source representative router in each AS to informall RPs about the existence of sources in that AS.

Switch basics

Ethernet switches link Ethernet devices together by relaying Ethernet frames between thedevices connected to the switches. By moving Ethernet frames between the switch ports, aswitch links the traffic carried by the individual network connections into a larger Ethernetnetwork.
Ethernet switches perform their linking function by bridging Ethernet frames between Ethernet segments. To do this, they copy Ethernet frames from one switch port to another, basedon the Media Access Control (MAC) addresses in the Ethernet frames. Ethernet bridging wasinitially defined in the 802.1D IEEE Standard for Local and Metropolitan Area Networks: MediaAccess Control (MAC) Bridges.

The standardization of bridging operations in switches makes it possible to buy switchesfrom different vendors that will work together when combined in a network design. That’s theresult of lots of hard work on the part of the standards engineers to

define a set of standards thatvendors could agree upon and implement in their switch designs.

Operation of Ethernet Switches

Networks exist to move data between computers. To perform that task, the network software organizes the data being moved into Ethernet frames. Frames travel over Ethernet networks, and the data field of a frame is used to carry data between computers.

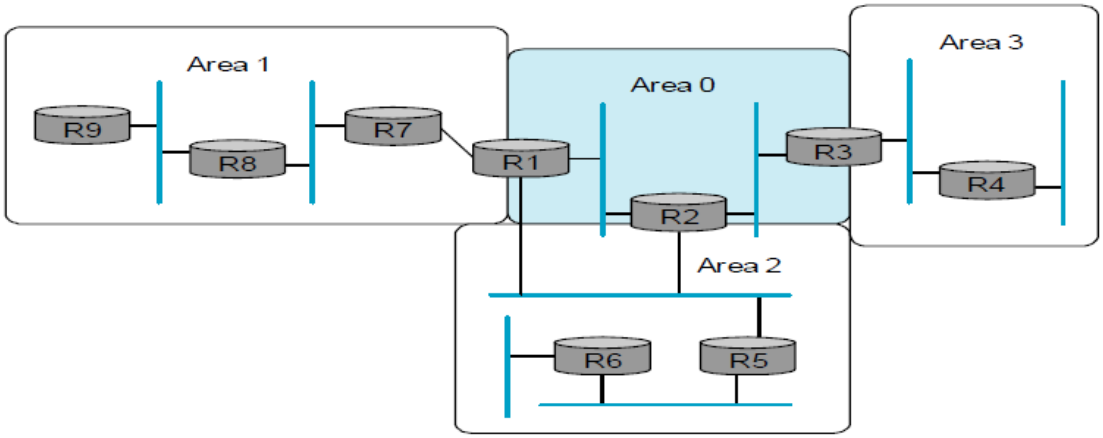
Frames are nothing more than arbitrary sequences of information whose format is defined in a standard.

The format for an Ethernet frame includes a destination address at the beginning, containing the address of the device to which the frame is being sent.

Next comes a source address, containing the address of the device sending the frame. The addresses are followed by various other fields, including the data field that carries the data being sent between computers

---

Global Internet Areas:



A router that is a member of both the backbone and a non-backbone area (R1) is calleda area router.

• Border routers "summarize" routing information and make it available to other areas -- act like proxies --reflect costs to reach networks from an area.

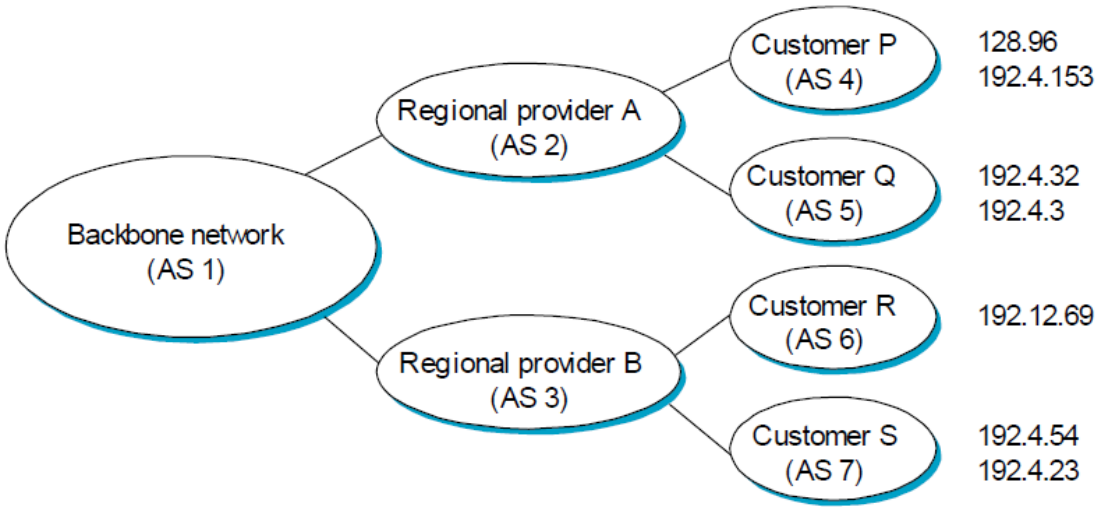
• When there are many possible routes, routers choose cost info to forward packets.

• Trade-offs -- Optimality versus scalability -- All packet have to pass through the backbone area (may not be optimal).

Border Gateway Protocol (**BGP**) is a standardized exterior gateway protocol designed toexchange routing and reachability information between autonomous

systems (AS) on theInternet. The protocol is often classified as a path vector protocol but is sometimes also classedas a distance-vector routing protocol.

• BGP supports flexibility -- paths could be chosen by a provider based on a policy.
• To configure BGP, each AS admin picks at least one node to be the “BGP” speaker -- a spokesperson node for the entire AS.
– The BGP speaker establishes a BGP session with other BGP speakers in otherASes.
• In addition, there are border gateways using which packets enter/leave ASes.
• Source advertises complete paths (unlike distance vector or link state routing) -- thus loops are prevented.

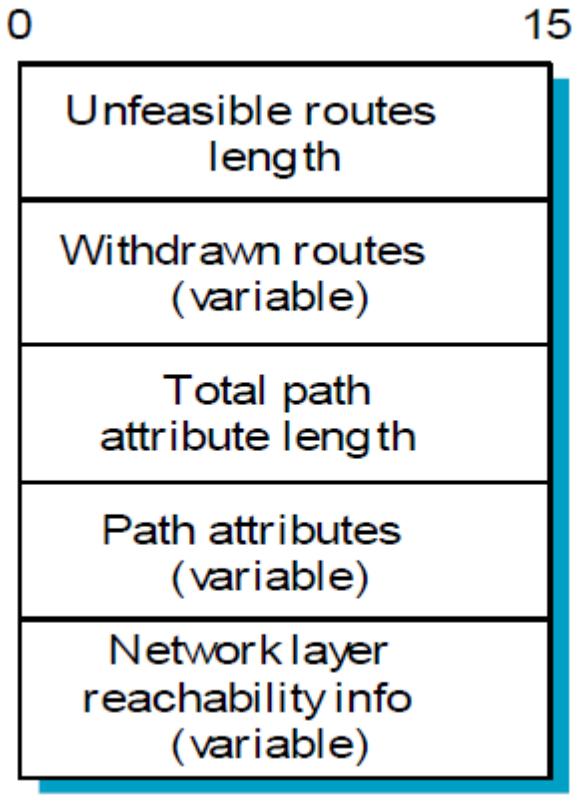


•

-AS 2 says 128.96, 192.4.15, 192.4.32, 192.4.3 can be reached via AS 2.
• AS 1 advertises that these networks can be reached via <AS1, AS2> --note full path description.
• Loops are avoided.

BGP Messages:
BGP has four types of messages
– OPEN: Establish a connection with a BGP peer
• Note: BGP connection is TCP based ! (Port no. 179).
– UPDATE -- advertise or withdraw routes to a destination
• Note --BGP speaker needs to be able to cancel previously advertised paths if nodes or links fail. This form of negative advertisements are said to advertise “withdrawn routes”.
• KEEPALIVE: Inform a peer that the sender is still alive but has no information to send.
• NOTIFICATION: Notify that errors are detected.
• 16 byte fields.
• For more detail look at book.

• Important thing --- BGP updates are of the type prefix/length
– 192.4.16/20



Routing with BGP:
• For stub AS -- border router injects a default route into the intra-domain routing protocol.
• If there are more than one border router, each injects specific routes that they have learned from outside the AS.
• IBGP or Interior BGP is used to distribute the information to all other routers in the domain (and the speaker).

IPv6 ADDRESSING

IPv4 has the small size of the address space.

An IPv6 address is 128 bits or 16 bytes(octets) long, four times the address length in IPv4.

IPv6 address, in hexadecimal format, is very long, many of the digits are zeros.

In this case, the leading zeros of a section canbe omitted. Using this form of abbreviation, 0074 can be written as 74, 000F as F, and0000 as 0.

Address Space

The address space of IPv6 contains 2128 addresses. This address space is 296 times theIPv4 address—definitely no address depletion—as shown, the size of the space is

340, 282, 366, 920, 938, 463, 374, 607, 431, 768, 211, 456.

Three Address Types

In IPv6, a destination address can be of three categories: unicast, anycast, and multicast.

Unicast Address
A unicast address defines a single interface (computer or router). The packet sent to aunicast address will be routed to the intended recipient (Receiver).

Anycast Address

An **anycast address** defines a group of computers that all share a single address. Apacket with an anycast address is delivered to only one member of the group, the mostreachable one.

An anycast communication is used, for example, when there are severalservers that can respond to an inquiry. The request is sent to the one that is most reachable.

Multicast Address

A multicast address defines a group of computers.

Differencebetween anycasting and multicasting.

In anycasting, only one copy of the packet is sentto one of the members of the group; in multicasting each member of the group receivesa copy.

Address Space Allocation
Like the address space of IPv4, the address space of IPv6 is divided into several blocksof varying size and each block is allocated for a special purpose. Most of the blocks are still unassigned and have been set aside for future use.

IPv6 Packet Format

The IPv6 packet is shown in Figure (below).

Each packet has a base header followedby the payload. The base header occupies 40 bytes, payload is upto 65,535 bytes of information.

The description of fields follows.

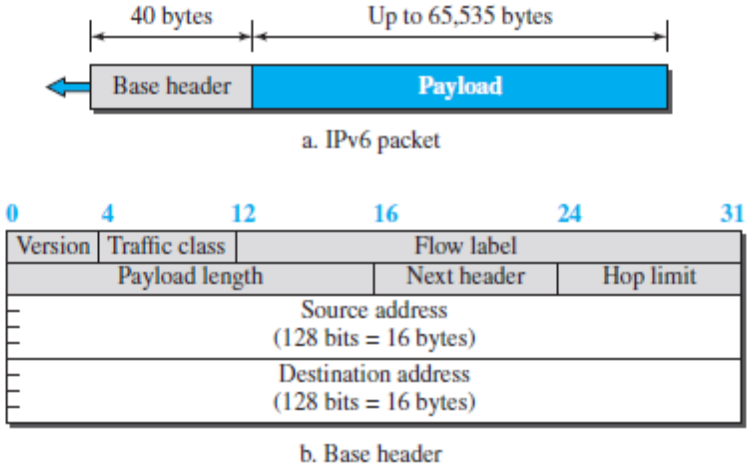


Fig: IPv6 datagram.[Source : Data Communications and Networking by Behrouz A. Forouzan].

Version. The 4-bit version field defines the version number of the IP. For IPv6, thevalue is 6.

Traffic class. The 8-bit traffic class field is used to distinguish different payloadswith different delivery requirements. It replaces the type-of-service field in IPv4.

Flow label. The flow label is a 20-bit field that is designed to provide special handlingfor a particular flow of data.

Payload length. The 2-byte payload length field defines the length of the IP datagram excluding the header.

Note that IPv4 defines two fields related to the length: header length and total length.

In IPv6, the length of the base header isfixed (40 bytes); only the length of the payload needs to be defined.

Next header. The **next header** is an 8-bit field defining the type of the first extension header (if present) or the type of the data that follows the base header in thedatagram.

Hop limit. The 8-bit hop limit field serves the same purpose as the TTL field in IPv4.

Source and destination addresses. The source address field is a 16-byte (128-bit)Internet address that identifies the original source of the datagram.

The destinationaddress field is a 16-byte (128-bit) Internet address that identifies the destination ofthe datagram.

Payload. Compared to IPv4, the payload field in IPv6 has a different format andmeaning, as shown in Figure .

Extension Header
An IPv6 packet is made of a base header and some extension headers. The length of thebase header is fixed at 40 bytes.

To give more functionality to the IP datagram,the base header can be followed by up to six **extension headers.**

Six types of extension headers have been defined.

Theseare hop-by-hop option, source routing, fragmentation, authentication, encrypted securitypayload, and destination option (see Figure below).

Hop-by-Hop Option

The hop-by-hop option is used when the source needs to pass information to all routersvisited by the datagram. For example, routers must be informed about certainmanagement, debugging, or control functions.

Destination Option
The **destination option** is used when the source needs to pass information to the destinationonly.

Intermediate routers are not permitted access to this information.

Source Routing

The source routing extension header combines the concepts of the strict source route and the loose source route options of IPv4.

Fragmentation

The concept of **fragmentation** in IPv6 is the same as that in IPv4.

In IPv6, only the original source can fragment. A source must use a**Path MTU Discovery technique** to find the smallest MTU supported by any network on the path. The source then fragments using this knowledge.

If the source does not use a Path MTU Discovery technique, it fragments the datagramto a size of 1280 bytes or smaller. This is the minimum size of MTU required foreach network connected to the Internet.

Authentication

The **authentication** extension header has a dual purpose:

It validates the messagesender and ensures the integrity of data. It is needed so the receiver can besure that a message is from the genuine sender and not others.

Encrypted Security Payload

The **encrypted security payload (ESP)** is an extension that provides confidentiality and guards against eavesdropping.

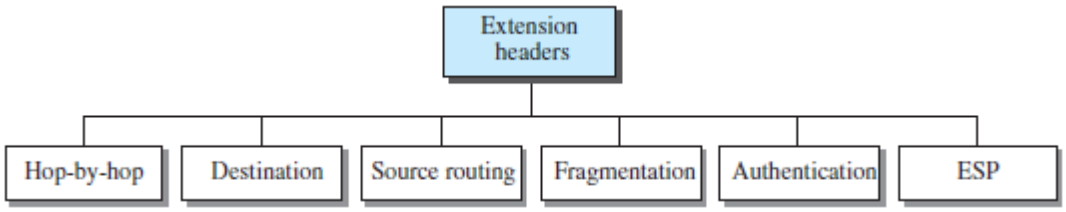


Fig: Extension header types.[Source : Data Communications and Networking by Behrouz A. Forouzan].

TRANSITION FROM IPv4 TO IPv6

Strategies

Three strategies are used for transition: dual stack, tunneling, and header translation.

One or all of these three strategies can be implemented during the transition period.

Dual Stack

It is recommended that all hosts, before migrating completely to version 6, have a **dual stack** of protocols during the transition. In other words, a station must run IPv4 and IPv6 simultaneously until all the Internet uses IPv6.

Figure shows the layout of a dual-stack configuration.

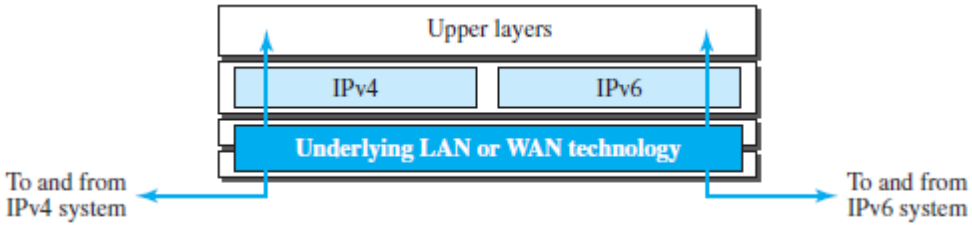


Fig:Dual stack.[Source : Data Communications and Networking by Behrouz A. Forouzan].

To determine which version to use when sending a packet to a destination, the source host queries the DNS. If the DNS returns an IPv4 address, the source host sends an IPv4 packet. If the DNS returns an IPv6 address, the source host sends an IPv6 packet.

Tunneling

Tunneling is a strategy used when two computers using IPv6 want to communicate with each other and the packet must pass through a region that uses IPv4.

To pass through this region, the packet must have an IPv4 address. So the IPv6 packet is encapsulatedin an IPv4 packet when it enters the region, and it leaves its capsule when itexits the region.

It seems as if the IPv6 packet enters a tunnel at one end and emerges atthe other end. To make it clear that the IPv4 packet is carrying an IPv6 packet as data, the protocol value is set to 41.

Tunneling is shown in Figure (below).

Header Translation

Header translation is necessary when the majority of the Internet has moved to IPv6but some systems still use IPv4.

The sender wants to use IPv6, but the receiver does notunderstand IPv6. Tunneling does not work in this situation because the packet must bein the IPv4 format to be understood by the receiver.

In this case, the header format must be totally changed through header translation. The header of the IPv6 packet is convertedto an IPv4 header.

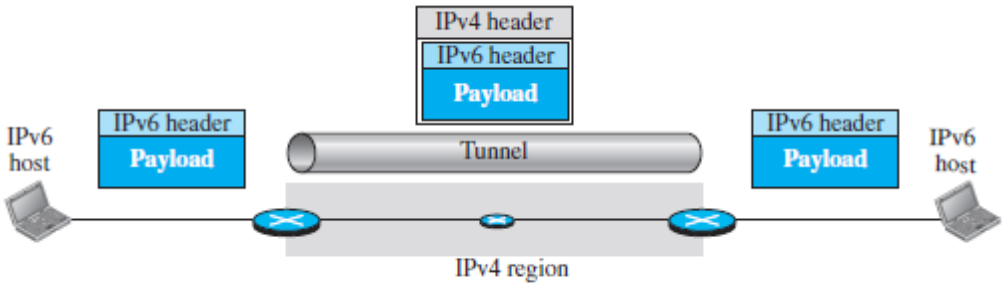


Fig: Tunneling strategy.[Source : Data Communications and Networking by Behrouz A. Forouzan].
