Part II: Internet Protocols

Routing in the Internet

- Introduction, Autonomous Systems

- GGP, RIP, OSPF, EGP, BGP

Routing in the Internet (1)

- Routing Algorithms
  - Bellman-Ford
  - Dijkstra
- Routing Protocols
  - Distance Vector
  - Link State Routing Protocol
- Creating a hierarchy for routing in large networks
  - Interior Gateway Protocols (RIP, HELLO, OSPF, IGRP)
  - Exterior Gateway Protocols (EGP, BGP, CIDR, Policy Routing)
  - Multicasting (IGMP)

Routing in the Internet (2)

- References :
  - 1) Douglas Comer, "Internetworking with TCP/IP"
    Prentice Hall
    Chapters 14, 15, 16 and 17
  - 2) Christian Huitema, "Routing in the Internet"
    Prentice Hall
    Chapters 4, 5, 6, 7, 8, 9 and 10

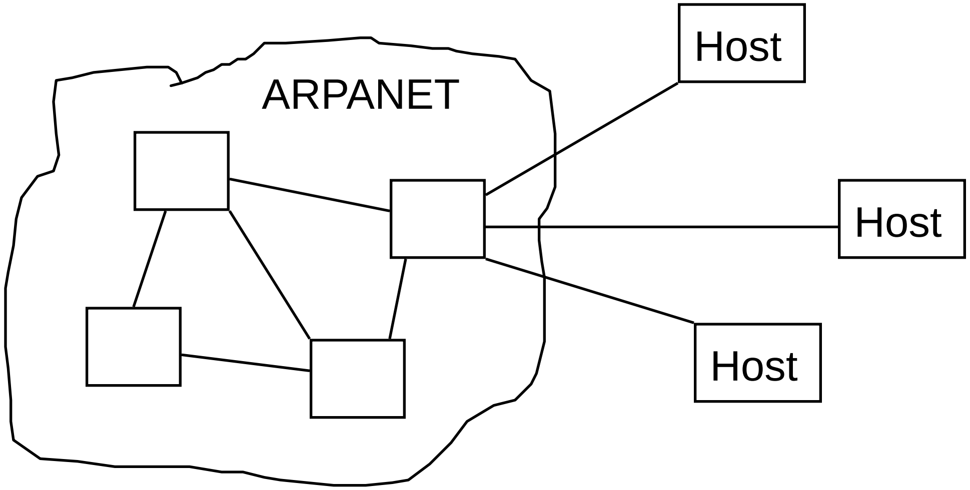
Bellman-Ford Algorithm

- The Bellman-Ford algorithm iterates on the number of hops in the path from source to destination
- At the k^{th} iteration, only paths of at most k hops have been explored, and we get the shortest path from source to destinations constrained by the k-hop limitation

Dijkstra's Algorithm

- Finds the shortest paths from a given source node to all other nodes, by developing the paths in order of increasing path length
- Proceeds in stages: by the k^{th} stage, the shortest paths to the k nodes closest to the source node have been determined

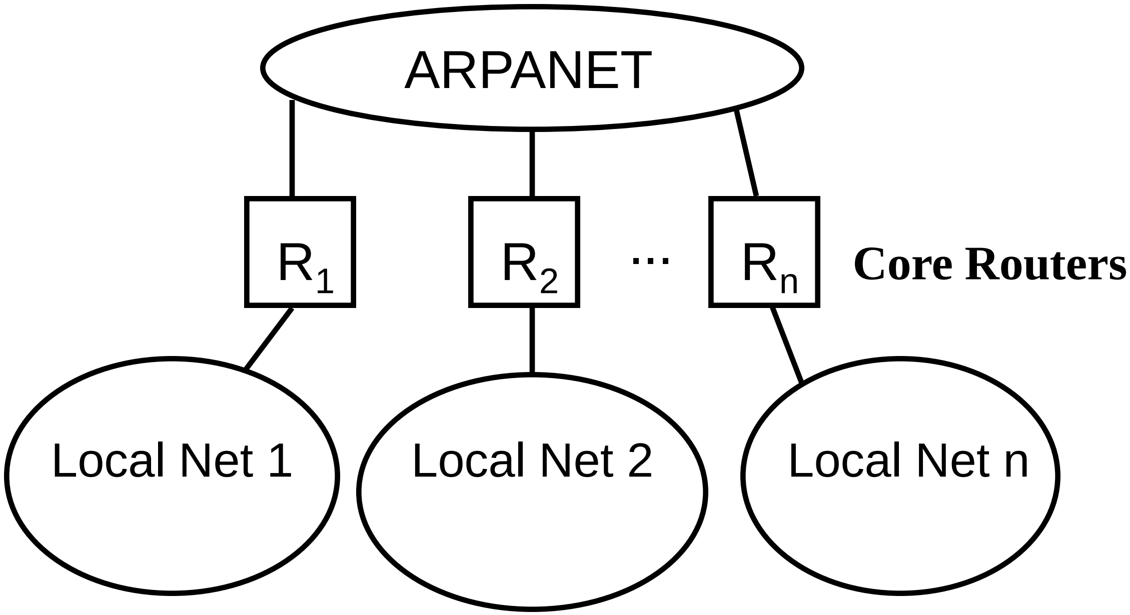
A Brief Look at
How The Internet Grew (1)



- First there was the ARPANET
  - Routers (then called “Gateways”) had complete information about all possible destinations - They are referred to as core routers
  - used the Gateway-to-Gateway Protocol (GGP), a distance vector routing protocol

A Brief Look at
How The Internet Grew (2)

- Then LANs were connected to the ARPANET

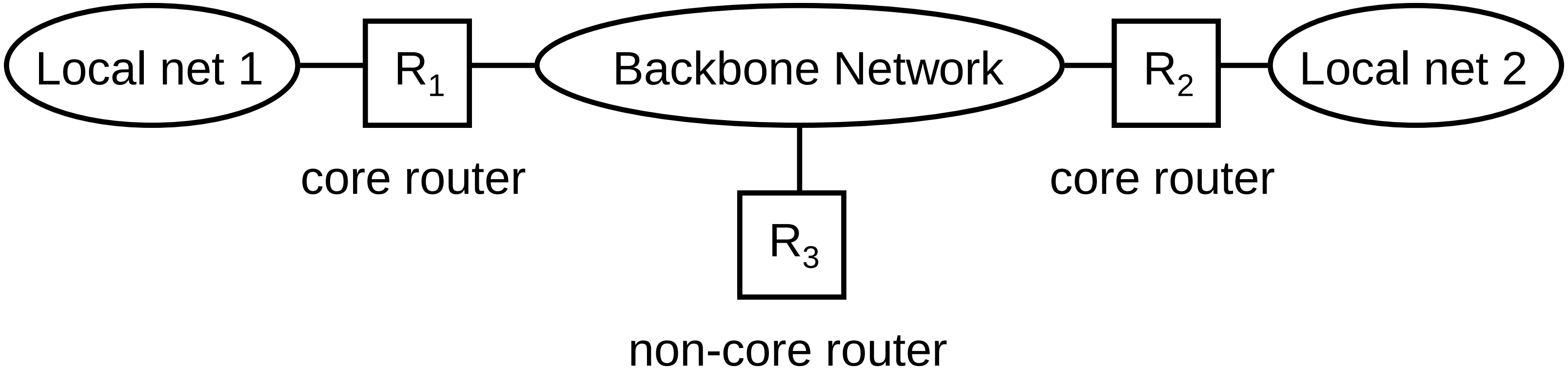


A Brief Look at
How The Internet Grew (3)

- Problems caused by this configuration:
  - routing overhead increased with the number of connected routers
    - size of routes increased with the # of connected networks
    - frequency of routing exchanges increased
    - higher likelihood that something went wrong somewhere requiring updates
  - number of different types of routers increased
  - slower deployment of new versions of routing algorithms

A Brief Look at
How The Internet Grew (4)

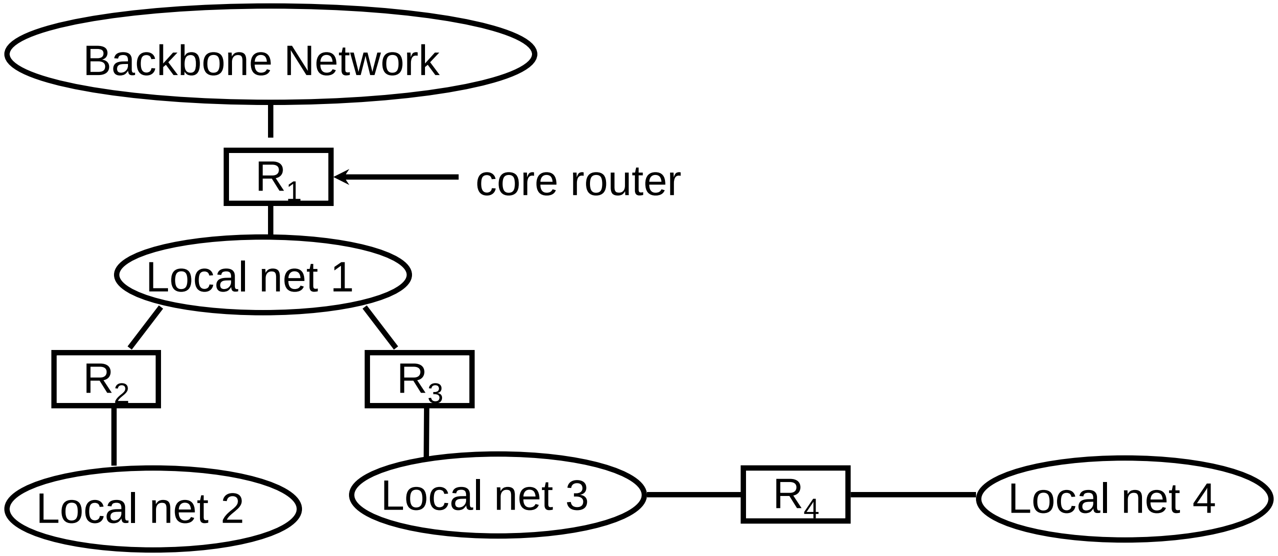
- The extra hop problem for non-core routers.



- Non-core routers connected to the backbone must learn routes from core routers to have optimal routing.

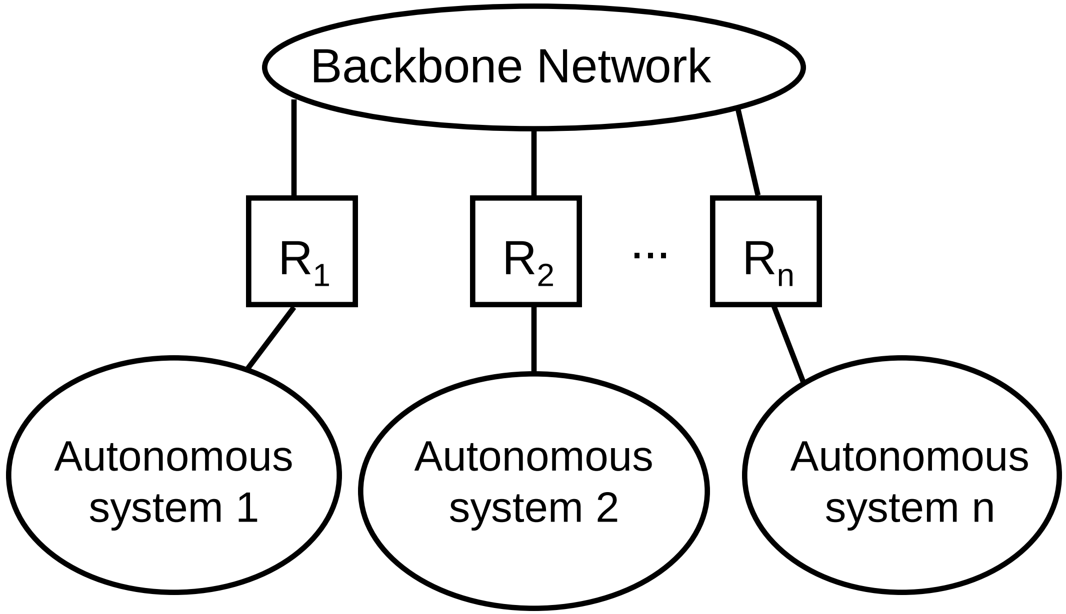
A Brief Look at
How The Internet Grew (5)

- Another Problem : Hidden Local Networks



- An example of multiple networks and routers with a single backbone connection. A mechanism is needed to pass reachability information about additional local networks to the core system.

The Concept of Autonomous Systems (1)



- Architecture of an internet with autonomous systems at backbone sites. Each autonomous system consists of multiple networks and routers under a single administrative authority

The Concept of Autonomous Systems (2)

- What is an autonomous system?
  - "a set of routers and networks under the same administration" e.g.
    - a single router directly connecting one local network to the internet
    - a corporate network linking several local networks through a corporate backbone
    - a set of client networks served by a single Internet provider
  - IMPORTANT (from a routing point of view) : all parts of an AS must remain connected.

The Concept of Autonomous Systems (3)

- Internal connectivity within the AS means:
  - all routers must be interconnected
  - (two local networks that belong to the same organizations but that rely on the core AS for connectivity cannot constitute a single AS)
  - all routers must exchange routing information in order to maintain the connectivity; (normally achieved by selecting a single routing protocol and running it between all routers)
- Routers inside an AS are called “interior gateway” and the protocol that they use is called an “Interior Gateway Protocol” (IGP)

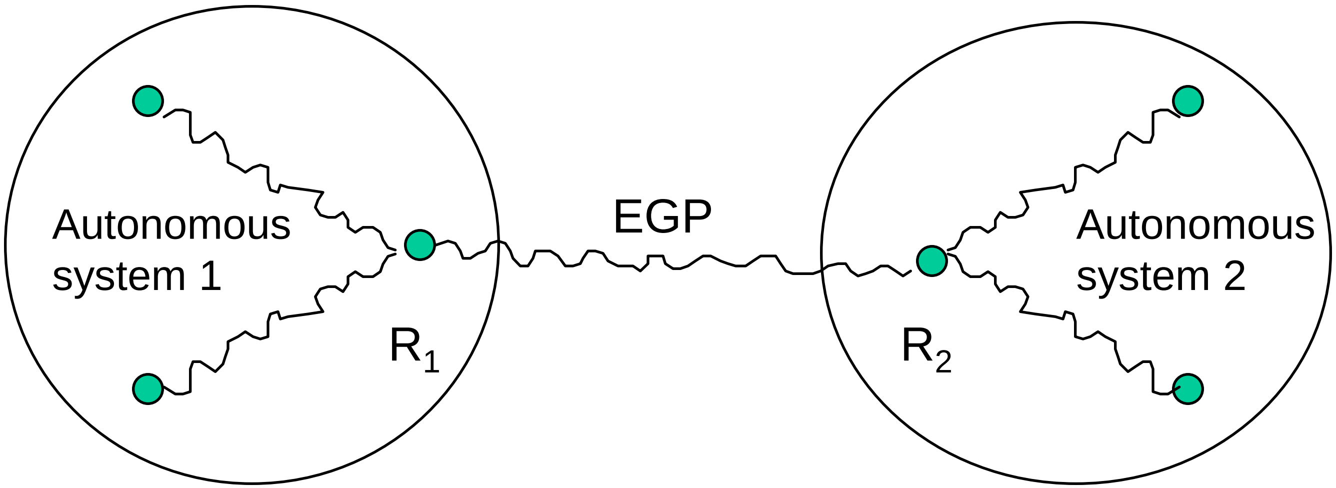
The Concept of Autonomous Systems (4)

- In 1982 the protocol of choice was GGP
- IGP's in use today are :
  - RIP
  - OSPF
  - IGRP
- Each AS is identified by a 16-bit "AS number" .
- Number is assigned by the numbering authorities.

The Concept of Autonomous Systems (5)

- Benefits of AS's:
  - lowering routing overhead
  - easing network management
  - computing new routes is easier
  - distributing new versions of software is easier
  - isolating failing elements is easier
  - Autonomous Systems use EGP (Exterior Gateway Protocol) to exchange information about reachability.

The Concept of Autonomous Systems (6)



Interior Gateway Protocols

- GGP
- RIP
- HELLO
- OSPF
- IGRP

Gateway-to-Gateway Protocol

- The “old” ARPANET routing protocol
- Definition can be found in RFC823
- A distance vector routing protocol
  - only core routers participate in GGP
- GGP messages travel in IP datagrams, with Protocol Type = 3
- GGP measures distance in router hops : the number of hops along a path refers to the number of routers

GGP Message Types

- 4 types of GGP messages
  - GGP Routing Update Message (Type = 12)
  - GGP Acknowledgment Messages (Types 2 or 10)
  - GGP Echo Request or Reply Messages (Type 0 or 8)

GGP Routing Update Message

TYPE (12)	UNUSED (0)
SEQUENCE NUMBER	
UPDATE	NUM. DISTANCES
DISTANCE D_1	NUM. NETS AT D_1
FIRST NET AT DISTANCE D_1	
SECOND NET AT DISTANCE D_1	
…	
LAST NET AT DISTANCE D_1	
DISTANCE D_2	NUM. NETS AT D_2
FIRST NET AT DISTANCE D_2	
SECOND NET AT DISTANCE D_2	
…	
LAST NET AT DISTANCE D_N	

- The format of a GGP routing update message. A router sends such a message to advertise destination networks it knows how to reach. Network numbers contain either 1, 2 or 3 octets, depending on whether the network is class A, B or C

Routing Update Fields

- GGP routing update message
  - To keep size of message small, networks are grouped by distance
  - 16- bit sequence number is used to validate a GGP message
  - UPDATE : binary value that specifies whether the sender gateway is requesting an update from the receiver
  - NUM DISTANCES : specifies how many distance groups are present in this update

GGP Acknowledgement Message

0 8	16	31
TYPE (2 or 10)	UNUSED (0)	SEQUENCE

- The format of a GGP acknowledgment message.
  - Type 2 identifies the message as a positive acknowledgment;
  - Type 10 identifies the message as a negative acknowledgment
  - In Positive Acknowledgment messages : sequence specifies the sequence number of update message that receiver is acknowledging.
  - In Negative Acknowledgment, it is the sequence number that the receiver last received correctly.

GGP Echo Message

0	8	31
TYPE (0 or 8)	UNUSED (0)	

- The format of a GGP echo request or reply message. Type 8 identifies the message as an echo request, while type 0 identifies the message as an echo reply.
  - Used to allow one router to test whether another is responding

RIP
Routing Information Protocol

- A distance vector based IGP
- Similarity to GGP
- Designed at UC Berkeley
- Based on earlier work done at Xerox PARC named Xerox XNS Routing Information Protocol
- Distributed with 4 BSD Unix
- Basis for local routing, and then used in larger networks.
- First RFC standard appeared in June 1988 (RFC1058)

RIP Details

- Routers are active machines
  - advertise their routes to others
- Hosts are passive machines (or silent)
  - they listen and update their routes but do not advertise
- Advertisement are sent every 30 seconds
  - contents of their routing tables
- RIP uses hop count metric
- RIP messages are transmitted using UDP on Port 520

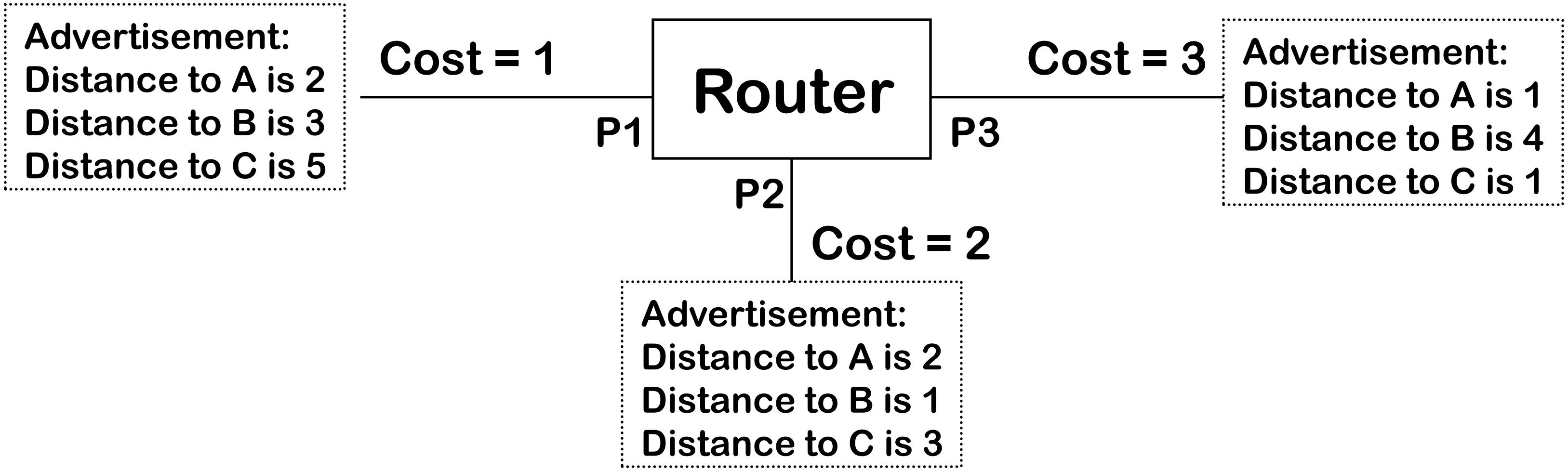
RIP Route Computation

- To each link in the network, there is a "cost" associated with using that hop (typically cost=1 $\Rightarrow$ number of hops)
- A router receives, from each of its neighbors, an advertisement showing that neighbor's distance to each destination in the network.
- For each destination in the network:
  - The router takes each received advertisement, and adds to it the cost to reach the neighbor that sent the advertisement. This is the distance to that destination through the neighbor.
  - The router selects the lowest of these as the path to that destination

Algorithm Properties

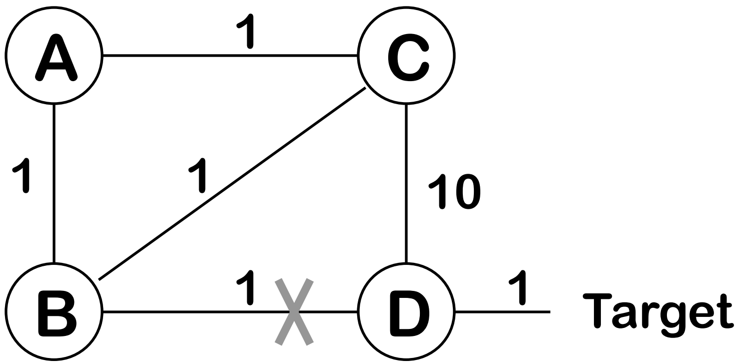
- This algorithm will converge to the shortest paths in a finite time, in the absence of topology changes.
- The starting value of the distance estimates to each destination can be any non-negative number.
- Algorithm makes no assumptions on when the updates are sent or when distances are computed.
  - Each router can work based on its own clock and send its updates asynchronously.
- If the network changes, routes converge to a new equilibrium point.

Example



Distance to Destination	Through Port P1	Port P2	Port P3
A	**3**	4	4
B	4	**3**	7
C	6	5	**4**

Preventing Instability



Routes to Target:
A: route via B, distance 3
B: route via D, distance 2
C: route via B, distance 3
D: direct, distance 1

Assume that the B ↔ D link goes down, and B notices:

							Iterations						
From	**Via**	**Dist**	**Via**	**Dist**	**Via**	**Dist**	**Via**	**Dist**	**…**	**Via**	**Dist**	**Via**	**Dist**
A	B	3	C	4	C	5	C	6		C	11	C	12
B	unrec	-	C	4	C	5	C	6		C	11	C	12
C	B	3	A	4	A	5	A	6		A	11	D	11
D	dir	1	dir	1	dir	1	dir	1		dir	1	dir	1

What if the C ↔ D link also goes down? ("counting to infinity" problem)

Solutions

- Some solutions to the problem
  - Split Horizon :
    - (If node A is routing packets bound to destination X through B, it makes no sense for B to try to reach X through A; so A should not announce to B that X is a short distance from A)
    - Instead of broadcasting the same distance vector on all outgoing links, the nodes will send different versions of this message; removing destinations reached via the outgoing link on which the message is to broadcast
  - Split Horizon with Poisonous Reverse:
    - nodes include all destinations in distance vector message
    - but will set the corresponding distance to infinity if the destination is routed on the link

RIP Timers

- Route updates are sent every 30 seconds (plus a small random offset).
- Routes not updated in 180 seconds “expire” and are removed.
- Triggered updates:
  - Split horizon can prevent loops with two gateways, but not with three or more.
  - To avoid this, a gateway is required to send an immediate update when any route changes.
  - Flood of triggered updates resolve loops faster

RIP Message Format

0	8	16	24 31
COMMAND (1-5)	VERSION (1)	MUST BE ZERO	
FAMILY OF NET 1		MUST BE ZERO	
IP ADDRESS OF NET 1			
MUST BE ZERO			
MUST BE ZERO			
DISTANCE TO NET 1			
FAMILY OF NET 2		MUST BE ZERO	
IP ADDRESS OF NET 2			
MUST BE ZERO			
MUST BE ZERO			
DISTANCE TO NET 2			
....			

Message Format: Command Field

Command	Meaning
1	Request for partial or full routing information
2	Response containing network-distance pairs from sender’s routing table
3	Turn on trace mode (obsolete)
4	Turn off trace mode (obsolete)
5	Reserved for Sun Microsystems internal use

Message Format - Addresses

- Address format is not limited to TCP/IP
- RIP can be used with multiple network protocol suites
- Family of net 1 :
  - identifies the protocol family under which the network address should be interpreted
  - IP addresses are assigned value 2

RIP Metrics and Updates

- By default, RIP uses a very simple metric : hop count.
  - integers 1 to 15
  - 16 denotes infinity
- Packets are normally sent every 30 seconds.
- If a route is not refreshed within 180 seconds, distance is set to infinity and later entry is removed

RIP Routing Table

- Entry in a routing table can represent a host, a network or a subnetwork
- no address type specification in RIP packets
  - routers have to analyze the address to understand what is being passed
  - First they separate the <network part> from the <subnet + host> as a function of the address class A, B, C
- if <subnet + host> is null
  - entry represents a network
- if <subnet + host> is not null
  - entry represents either a subnet or a host.(In order to discriminate between the two, one must know the mask used within a given network)
  - if the host part is null
    - this is a subnet address
  - if the host part is not null
    - this is a host address

RIP Routing Table (cont.)

- RFC 1058 does not assume that the subnet mask is available outside of the network itself
  - subnet entries are not supposed to be propagated outside the network itself.
  - all of them are replaced by a single entry for the network itself.
- Support for host routes is optional
  - routers that do not want to maintain inflated table are allowed to drop the host entries
- 0.0.0.0 default route e.g. toward network outside the autonomous system

Routing Table Contents

- Entry in a routing table contains :
  - address of the destination
  - the metric associated to that destination
  - the address of the “next router”
  - a “recently updated” flag
  - timer

Input Processing

- Processing incoming RIP Response message
  - Examine entries one by one
    - Validation check :
    - address is a valid class A, B or C
    - network # $\neq$ 127 (loop back)
    - host part is not a “ broadcast” address
    - metric is not larger than infinity (16)
  - Incorrect entries are ignored (and should be reported as errors)

Input Processing (cont.)

- Metric for entry (not infinity) is increased by link cost
- Routing table is searched for an entry corresponding to the destination :
  - if the entry is not present, add it
  - if the entry is present but with larger metric
    - update metric and next router field
    - restart timer
  - entry is present and next router is sender of response message,
    - update metric
    - restart timer
  - all other cases
    - ignore entry received

RIP Responses

- A separate response is prepared for all connected interfaces
  - information may vary due to
    - a) split horizon processing
    - b) subnet summarization
- For triggered updates : may include only those entries that have been updated since the last transmission.
- Maximum message size : 512 bytes (up to 25 entries)
  - Multiple messages have to be sent if more than 512 bytes
- IP address is that of the interface on which the message is sent