

TCP/IP 通訊協定及應用

Spring 2002
 中央大學 吳曉光博士
<http://wmlab.csie.ncu.edu.tw/course/tcp>



Chapter 10: Dynamic Routing Protocols



Introduction

- When do we use dynamic routing?
 - If any of the following three conditions is false, dynamic routing is used
 - The network is small
 - A single connection point to other networks
 - No redundant routes
- Dynamic Routing
 - It occurs when routers talk to adjacent routers, informing each other of what networks each router is currently connected to
 - What's the routing daemon
 - The process is running protocol, communicating with its neighbor routers
 - If the daemon finds multiple routes to a destination, the daemon chooses the best one to insert into the kernel's table
 - If the daemon finds that a link has gone down, it can delete routes or alternate routes that bypass the problem



Unix Routing Daemons

- Routing Daemons in Unix System:
 - Unix systems often run the routing daemon named `routed`
 - An alternative program is `gated` that supports both IGPs and EGPs

Daemon	Interior Gateway Protocol		Exterior Gateway Protocol	
	HELLO	RIP	OSPF	BGP
<code> routed </code>		V1		
<code> gated, Version 2 </code>	*	V1		V1
<code> gated, Version 3 </code>	*	V1, V2	V2	V2, V3

Figure 10.1 Routing protocols supported by `routed` and `gated` .

- RIP (Routing Information Protocol) message format

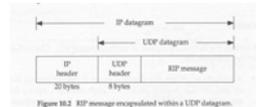


Figure 10.2 RIP message encapsulated within a UDP datagram.



RIP: Routing Information Protocol

- Format of a RIP message

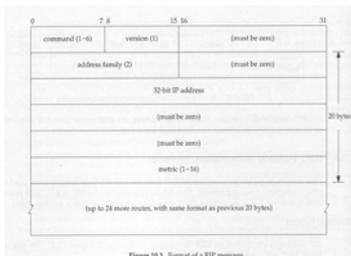


Figure 10.3 Format of a RIP message.



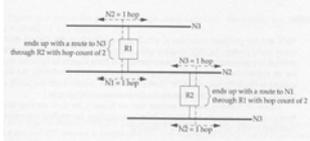
RIP: Routing Information Protocol

- Normal Operation
 - The well-known port number for RIP is UDP port 520
 - Initialization: The daemon sends a request packet out each interface, asking for the other router's complete routing table
 - Request received: If we have a route to the specified address, set the metric to our value, else set the metric to 16
 - Response received: The response is validated and may update the routing table
 - Regular routing updates: Every 30 seconds, routing table is sent to every neighbor router
 - Triggered updates: Only those entries that have changed must be transmitted
 - A route has not been updated for 3-min, metric is set to infinity(16)



RIP: Routing Information Protocol

- What's metrics
 - The metrics used by RIP are hop counts. The hop count for all directly connected interfaces is 1



- The metric to R1 for N3 is 2, and the metric to R2 for R1 is 1
- The router selects the path with the smallest hop count and ignores the other paths
- maximum number of hops between hosts is 15



RIP: Routing Information Protocol

- Drawbacks of RIP
 - RIP has no knowledge of subnet addressing
 - It takes a long time to stabilize after the failure of a router or a link
 - A maximum of 15 for the metric limits sometime is not enough
- Example 1
 - ripquery tries to send one of the undocumented requests (name "poll", a command of 5)
 - If no response is received in 5 seconds, the standard RIP request is issued (command of 1)

```
sun % ripquery -m meth
504 bytes from meth (140.252.1.183):
  140.252.1.0, metric 1
  140.252.13.0, metric 2
244 bytes from meth (140.252.1.183):
  first message contains 504 bytes
  lots of other lines deleted
  the top Ethernet in Figure 10.5
  the bottom Ethernet in Figure 10.5
  second message with remaining 244 bytes
  lots of other lines deleted.
```



RIP: Routing Information Protocol

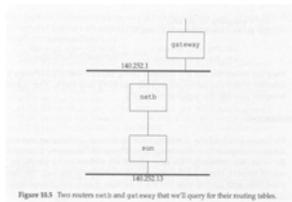


Figure 10.3. Two routers meth and gateway that we'll query for their routing tables.

```
sun % toping -m600 -i k10
1 0.0 sun.2879 > meth.route: rip-poll 24
2 5.214702 (5.0147) sun.2879 > meth.route: rip-req 24
3 5.340437 (0.5457) meth.route > sun.2879: rip-resp 25;
4 5.710231 (0.1496) meth.route > sun.2879: rip-resp 12;
```

Figure 10.4. Toping output while running ripquery program.



RIP: Routing Information Protocol

- Here the metric in 140.252.1.0 stays at 1, since that Ethernet is directly connected to both gateway and meth
- Our subnet 140.252.13.0 has the expected metric of 2
- Example 2
 - We'll run the Solaris 2.x program snoop, which is similar to tcpdump, on the host solaris
 - Figure 10.8 shows the packets captured during a 60-second period

```
sun % ripquery -m gateway
504 bytes from gateway (140.252.1.41):
  140.252.1.0, metric 1
  140.252.13.0, metric 2
lots of other lines deleted
the top Ethernet in Figure 10.3
the bottom Ethernet in Figure 10.3.
```



RIP: Routing Information Protocol

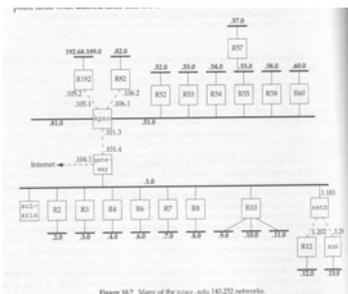


Figure 10.7. Many of the routers onto 140.252 networks.



RIP: Routing Information Protocol

- Why R10 is advertising four networks ?

```
solaris % snoop -P -t4 udp port 520
0.000000 R0.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
4.497000 R4.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
6.200000 R2.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
11.483137 R7.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
16.197900 R8.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
18.471311 R3.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
17.821887 gateway.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
20.460000 R10.tuc.nasa.edu -> BROADCAST RIP R (4 destination)
28.479400 R6.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
34.502000 R4.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
36.333900 R2.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
41.345400 R7.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
44.102070 R8.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
46.521930 R3.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
47.018970 gateway.tuc.nasa.edu -> 140.252.1.255 RIP R (1 destination)
50.464530 R10.tuc.nasa.edu -> BROADCAST RIP R (4 destination)
```

Figure 10.8. RIP broadcasts captured at solaris over a 60-second period.

```
RIP: Address Metric
RIP: 140.252.0.0 16 (not reachable)
RIP: 140.252.9.0 1
RIP: 140.252.10.0 1
RIP: 140.252.11.0 1
```



RIP: Routing Information Protocol

- The router gateway advertises 15 routes. We can run snoop with the -v flag and see the entire contents of the RIP message

```

olaris> snoop -P -v -xz udp port 520 host gateway
RIP: Opcode = 2 (route response)
RIP: Version = 1
RIP: Address      Metric
RIP: 140.252.101.0 1
RIP: 140.252.104.0 1
RIP: 140.252.81.0  2
RIP: 140.252.81.6  2
RIP: 140.252.105.0 2
RIP: 140.252.106.0 2
RIP: 140.252.52.0  3
RIP: 140.252.53.0  3
RIP: 140.252.54.0  3
RIP: 140.252.55.0  3
RIP: 140.252.56.0  3
RIP: 140.252.60.0  3
RIP: 140.252.62.0  3
RIP: 140.48.189.0  3
RIP: 140.252.57.0  4
    
```

Figure 80.9 RIP response from gateway.

RIP Version 2

- Compare RIP and RIP-2

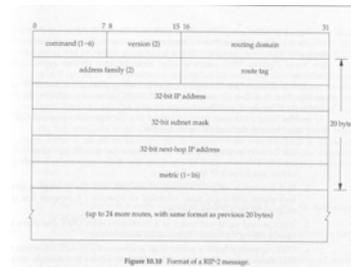


Figure 80.10 Format of a RIP-2 message.

- RIP-2 supports multicasting. This can reduce the load on hosts that not listening for RIP-2 messages

OSPF: Open Shortest Path First

- The difference of OSPF and RIP
 - OSPF is a link-state protocol, RIP is a distance-vector protocol
 - The distance-vector means each router updates from its neighbors
 - In a link-state protocol, each router actively tests the status of its link to each of its neighbors, sends this information to its other neighbors
 - A link-state protocol will always converge faster than a distance-vector protocol
- Features of OSPF that superior to RIP
 - OSPF can calculate a separate set of routes for each IP TOS
 - Each interface is assigned a dimensionless cost
 - OSPF distributes traffic equally among the routes (load balancing)
 - OSPF supports subnets
 - PPP links between routers do not need an IP address at each end

BGP: Border Gateway Protocol

- A cleartext password can be specified, similar to the RIP-2 scheme
- OSPF uses multicasting instead of broadcasting to reduce the load
- Local traffic and transit traffic
 - Local traffic in an AS either originates or terminates in that AS
 - Anything else is called transit traffic
 - A major goal of BGP is to reduce transit traffic
- Category of an AS
 - A stub AS has only a single connection to one other AS
 - A multihomed AS has connections to more than one other AS
 - A transit AS has connections to more than one other AS and is designed, to carry both local and transit traffic

BGP: Border Gateway Protocol

- BGP allows for policy-based routing
- BGP uses TCP as its transport protocol
- BGP is a distance vector protocol
- BGP detects the failure of either the link or the host by sending keepalive message to its neighbor on a regular basis
- What's CIDR?
 - CIDR is a way to prevent explosion in the size of the Internet routing tables. It is also called supernetting
 - The basic concept in CIDR is to allocate multiple IP addresses summarization into a smaller number of routing table entries
- Three features are needed to allow this summarization
 - Multiple IP addresses must share the same high-order bits

CIDR: Classless Interdomain Routing

- The routing tables and routing algorithms must be extended to base their routing decisions on a 32-bit IP address and a 32-bit mask
- The routing protocols being used must be extended to carry the 32-bit mask in addition to the 32-bit address
- CIDR will slow down the growth of the Internet routing tables
- Summary
 - Two routing protocols: IGP and EGP
 - The most popular IGP is RIP
 - OSPF being widespread use
 - The popular EGP is BGP
 - CIDR is to reduce the size of the size of the Internet routing tables
 - Other protocols: IDRP, IS-IS