# ROUTING, PART 1

**Internet Protocols**

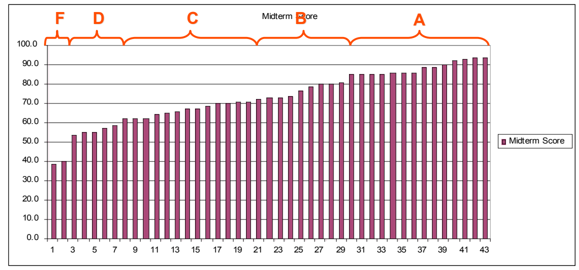
*CSC / ECE 573*

*Fall, 2005*

*N. C. State University*

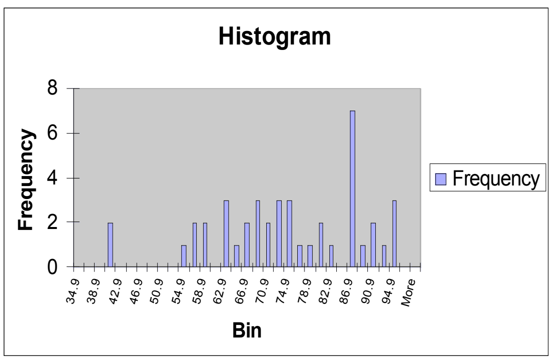
## Exam Results

| Average | Max | Min | + *Curve* |
|---|---|---|---|
| 73.5 | 94 | 39 | 7 |






## Distribution of Scores






## Today's Lecture

I. Interior Gateway Protocols

II. RIPv2

III. Improving Convergence

IV. OSPF

V. Use of Hierarchy

VI. Message and link types



# ROUTING INFORMATION PROTOCOL (RIP)

## RIP (v1:RFC1058, v2: RFC2453)

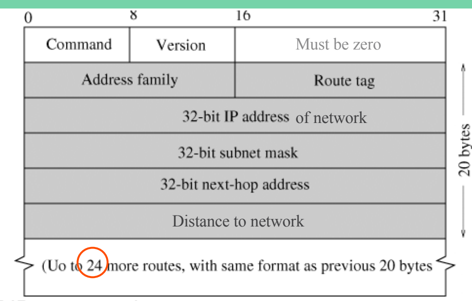
- Distance-Vector, interior gateway protocol
  - with split horizon and poison reverse
- Messages are transported over UDP
  - unreliable delivery
- Initialization
  - send request to all neighbors asking them for their complete routing table
- Neighbor discovery: none!
  - just broadcast updates and hope they are being received

## RIP Optimization Metric

- Metric = minimum hop count
  - maximum allowable value = 15 (16="infinity")
  - consequences?
- Administrators may set hopcount of a slow link to more than 1
  - consequences?
- No route will be installed unless it is strictly lower cost (smaller hop-count) than the route it replaces
  - prevents oscillation between equal cost routes



## RIPv2 Message Format




- RIP commands
  - Request for routes
  - Response (i.e., distance vectors)



## RIPv2 Message Contents

- Route Tag: origin of a route
  - must be transmitted when route is propagated
  - (used by exterior gateway routing protocols)
- Next-hop address: used to eliminate extra forwarding hops at the edge of a RIP domain
  - prevents loops and improves convergence time
- Subnet Mask supports variable-length subnet-based routing



## Policy Controls

- Policy controls not supported by RIPv2
  - instead, administrators manually configure route "filters"
  - i.e., which routing destinations are allowed to be installed in your routing table, which routes will be propagated to other routers



## Other

- Security (v2)
  - optional 16 byte cleartext password in request / response commands
  - trivial to break ☹



## RIP Timers

- Periodic: approximately (randomly) every 30 seconds…
  - send complete routing table to your neighbors
- Expiration: if a route has not been updated or renewed for 180 seconds…
  - set its metric to infinity
  - send this update to neighbors

## RIP Timers (cont'd)

- Garbage collection: after another 120 seconds…
  - actually delete route from the routing table
  - ensures route invalidation is propagated throughout network before route stops being used
  - another name: "Hold-down" period
  - disadvantages?



## RIP Timers (cont'd)

- *Triggered* updates: changes to route metric are propagated immediately
  - instead of being required to wait until next broadcast interval
- Propagates throughout network immediately
  - small delay before propagating, to avoid generating excessive network traffic



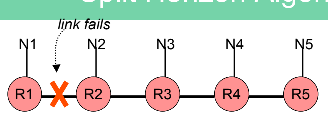
## Reliability / Consistency in RIPv2

- Reliability
  - based solely on periodic (frequent) complete retransmission
  - since routers do not have complete network topology, cannot easily detect inconsistencies or loops



## IMPROVING CONVERGENCE

## Split Horizon Algorithm




| Distance to network N1 → | R2 | R3 | R4 | R5 | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Initially |
| | 3 | 2 | 3 | 4 | After 1 exchanges |
| | . | . | . | . | … |
| | . | . | . | . | … |
| | 7 | 6 | 7 | 6 | After 5 exchanges |

- *Split horizon*: a router should not advertise to a neighbor a route for which that neighbor is the next hop
- *Example*: R3 does not advertise N1 to R2



## Split Horizon with Poison Reverse

- SH with *poison reverse*: router can advertise a route with distance ∞
  - i.e., "poisons" the route to N1 through R1
  - eliminates timer-based (slow) timeout of route
- Assessment
  - prevents routing loops involving only two routers
  - *drawback*: larger routing update messages (must advertise all networks that can't be reached)

## Split Horizon (cont'd)

- Does this completely eliminate loops?



## OPEN SHORTEST PATH FIRST PROTOCOL (OSPF)

## OSPF (v2:RFC2328)

- “Recommended” interior (intra-domain) routing protocol for TCP/IP
  - link state routing using Dijkstra's algorithm
- Goals
  1. converges faster than RIP
  2. exchange less information than RIP
  3. scale to larger networks
- Runs directly over IP (not over UDP or TCP)



## OSPF (v2:RFC2328) (cont’d)

- Possible optimization metrics
  - hop-count
  - delay
  - throughput, etc.
- Load-balancing possible
  - when several equal-cost routes exist, can send traffic along each of them
  - just make sure all the packets for one TCP connection follow the same path
    - why?
  - (not widely used)



## The LS Routing Algorithm

Each router does the following

1. Discovers the (immediately adjacent) neighbors
2. Builds an LSA containing distance to each of its neighbors
3. Broadcasts the LSA to all routers in the network (using flooding)
4. Stores the most recent LSA from every router in the network



## The LS Routing Algorithm (cont’d)

5. Creates a “map” of the network topology from these LSAs
6. Computes routes (forwarding table) from its map of the network topology

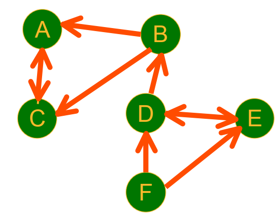
## Flooding LSA to All Other Routers

- How get LSAs to all routers if the forwarding tables are not (yet) correct?
  - solution: don't use forwarding tables, use flooding
- Inefficient method: keep broadcasting until sure everyone has the information
- Efficient method: broadcast only one time
  - have to remember what you've already broadcast
  - every node sends information once to neighbors
    - but not out the interface on which the information arrived




## "Flooding" Example

- Flooding from F to rest of network







## Link State Advertisement (LSA) Header

1. Link type (later)
2. Link ID
3. Advertising router
4. Sequence number
5. Checksum
6. Age




## LSA Sequence Number

- Originating router
  - starts LSA **`Sequence Number`** at smallest possible value
  - increments **`Sequence Number`** of successive LSAs
- Non-originating routers discard earlier LSA in favor of latest LSA from the originating router




## LSA Age Field

- Number of seconds since LSA was "originated"
- Originating router
  - sets LSA **`Age`** to 0 initially, floods LSA
  - refreshes LSA when **`Age`** > refresh interval
    - default interval = 30 minutes




## LSA Age Field (cont'd)

- Non-originating routers
  - when **`Age`** > "max age" (i.e., originating router is down)
    - assumes LSA is no longer valid
    - removes LSA from database
    - notifies other routers to remove LSA
  - max age = 60 minutes
- Can withdraw link state information immediately by re-flooding the LSA with maximum age (60 minutes)

## LSA Acknowledgment

- LSAs must be ACKed
  - receiving router tells (immediate) neighbor it received the LSA
- LSAs are not flooded / ACKed immediately upon receipt; why not?
  1. acknowledge multiple LSAs in a single packet
  2. helps to avoid "overflooding" of slightly newer LSAs
- Periodically, scan LSA database for LSAs that need to be flooded / ACKed




## OSPF Reliability

- Reliability accomplished by…
  - flooding, with neighbor acknowledgments
  - reoriginate LSAs at 30 minute intervals
  - all LSAs are checksummed



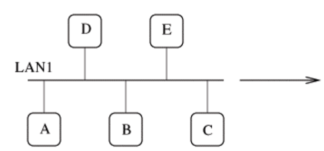

## Other

- Subnetting
  - works with variable-length subnets and CIDR
- Security
  - simple cleartext password, or
  - MD5 message digest, based on shared secret key




## LAN Abstraction

- *N* routers connected to a broadcast LAN
- Model as *N**(*N*-1)/2 point-to-point connections?? Expensive!
  - size of link state database
  - number of LSAs exchanged

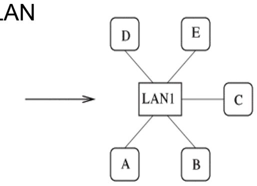






## LAN Abstraction Example

- Instead, create one LSA for the LAN
  - each router is a "neighbor" of the LAN
  - generate one LSA, send to all routers attached to the LAN
- A *designated router* is responsible for originating the LSA representing the LAN







## HIERARCHICAL OSPF NETWORKS

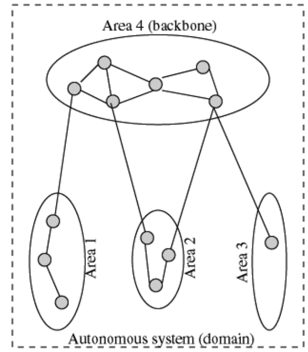
## Area Routing

- Routing domain is split into *areas*
  - area = generalization of a subnet
  - its topology and details are not visible outside the area
  - maximum area size = a few hundred routers
- Goal: reduction in LSAs and link-state database size
  - each area runs a copy of the link-state protocol
  - result: smaller link-state databases



## Area Routing (cont'd)

- Network consists of "backbone" and attached areas
- areas do not overlap
- 2-level hierarchy; each area must attach directly to backbone






## Routing Between Areas

- Routers at boundaries of areas communicate with other boundary routers
  - summarize (aggregate) routing prefixes within area, advertise fewer routes
- Exchange of summary LSA's across the backbone: uses Distance Vector approach!
  - but using flooding as method of distribution

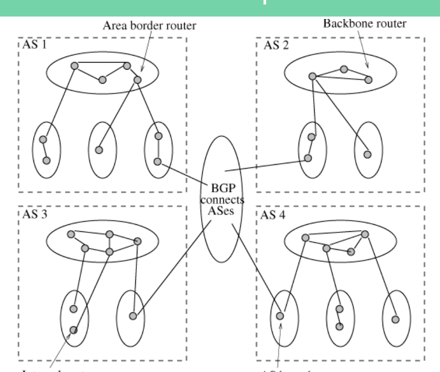


## OSPF Classes of Routers

- Internal routers: wholly within an area
- Area border routers: connect to backbone
- Backbone routers
- AS boundary routers: talk to routers in other ASes



## Example






## OSPF MESSAGE AND LINK TYPES

## Message Types

- Hello message
  - at boot time, used to discover who the neighbors are
  - also used to periodically test reachability
- Link State Update message
  - flooded to the whole network
- Database Description message
  - when a link between routers is brought up, gives `Sequence Number` of each link state entry the new neighbor has



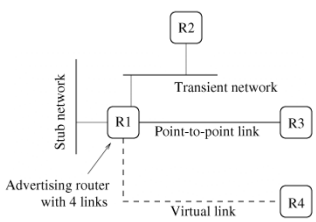
## Message Types (cont'd)

- Link State Request
  - after exchanging Database Description messages, sender requests Link State information for links for which the other router has more recent info



## Types of Links

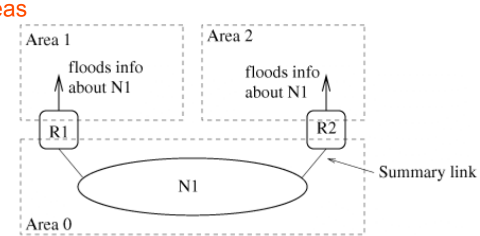
- Router Link: lists router's interfaces, and …
  - the cost
  - the type of network or router connected to
  - range of IP addresses directly accessible from interface






## Types of Links (cont')

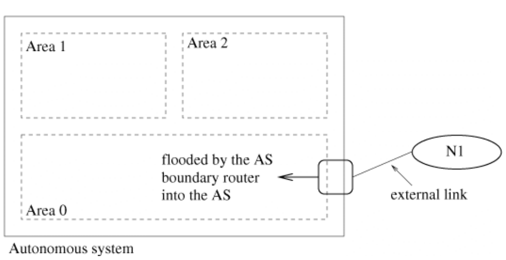
- Network link: performs the LAN abstraction
  - lists all routers connected to the network
- Summary Link: advertises IP addresses between areas






## Types of Links (cont')

- External Link: imports / exports routing information from / to other AS






## Summary

- RIP is a Distance-Vector protocol
  - simple, widely used
  - getting it to converge quickly requires care
  - doesn't scale to larger networks
- OSPF is a Link-State protocol
  - more powerful, more scalable
  - but, more complex to configure, administer
- *Note*: There are other interior routing protocols: EIGRP, IS-IS, etc.

## Next Lecture

- Routing, Part 3 (BGP4)