Route Discovery Protocols

Raj Jain

Raj Jain is now at
Washington University in Saint Louis
Jain@cse.wustl.edu
http://www.cse.wustl.edu/~jain/

- Building Routing Tables
- Routing Information Protocol Version 1 (RIP V1)
- RIP V2
- OSPF
- BGP and IDRP
- Ref: M. Naugle, "Network Protocols," McGraw Hill, 1998.

Autonomous Systems

- An internet connected by homogeneous routers under the administrative control of a single entity

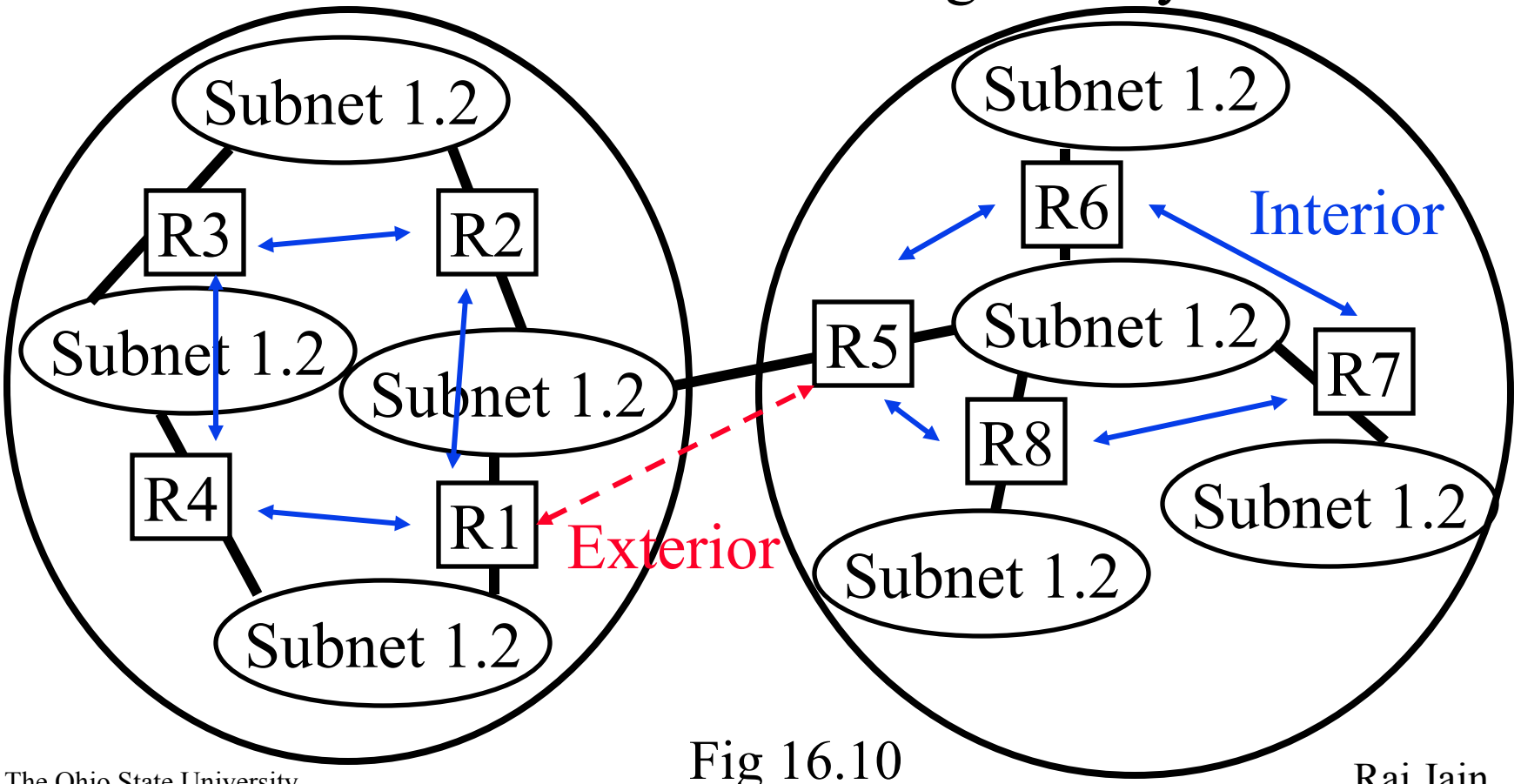


Fig 16.10

Routing Protocols

- Interior Router Protocol (IRP): Used for passing routing information among routers internal to an autonomous system. Also known as IGP.
  - Examples: RIP, OSPF
- Exterior Router Protocol (ERP): Used for passing routing information among routers between autonomous systems. Also known as EGP.
  - Examples: EGP, BGP, IDRP
    Note: EGP is a class as well as an instance in that class.

Routing Information Protocol

- RIP uses distance vector $\Rightarrow$ A vector of distances to all nodes is sent to neighbors
- Each router computes new distances:
  - Replace entries with new lower hop counts
  - Insert new entries
  - Replace entries that have the same next hop but higher cost
  - Each entry is aged.
    Remove entries that have aged out
- Send out updates every 30 seconds.

Distance-Vector Example

Fig 9.9 Stallings

RIP V1

- RFC 1058 adopted in 1988
- Implemented in Berkeley UNIX as "routed" (pronounced route d)
- Both hosts and routers can implement RIP
- Hosts use passive mode ⇒ Do not send out updates
- Runs on UDP
- RIP packets do not leave local network

RIPv1 Packet Format

Command	Version	Reserved
Family of Net 1		Reserved
Net 1 Address		
Set to 0		
Set to 0		
Distance of Network 1		
Family of Net 2		Reserved
Net 2 Address		
Set to 0		
Set to 0		
Distance of Network 2		

Up to 25 entries

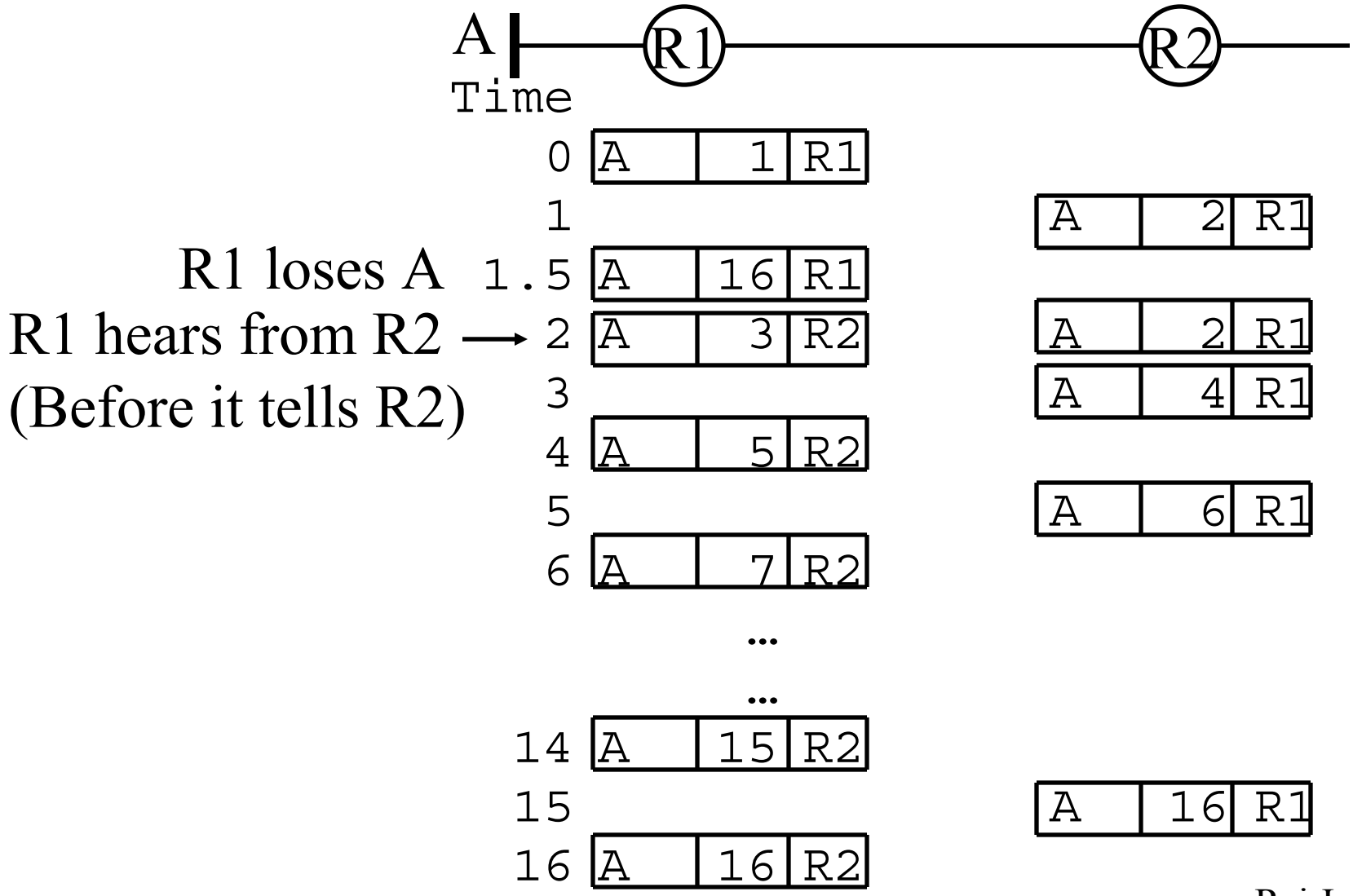
RIP V1 Packet Fields

- RIP Commands:
  - 1 = Request for partial or full routing table info
  - 2 = Response containing routing table
  - 3 = Turn on trace mode (obsolete)
  - 4 = Turn off trace mode (obsolete)
  - 5 = Sun Microsystems Internal Use
- Family of Net: Shows protocol that owns the packet
  - 2= IP
  - XNS and other protocols can also use RIP
- Addresses can be 14-byte long. IP uses only 4.
- Distance = Integers from 1 to 16.
  16 ⇒ Unreachable

Shortcomings of RIP

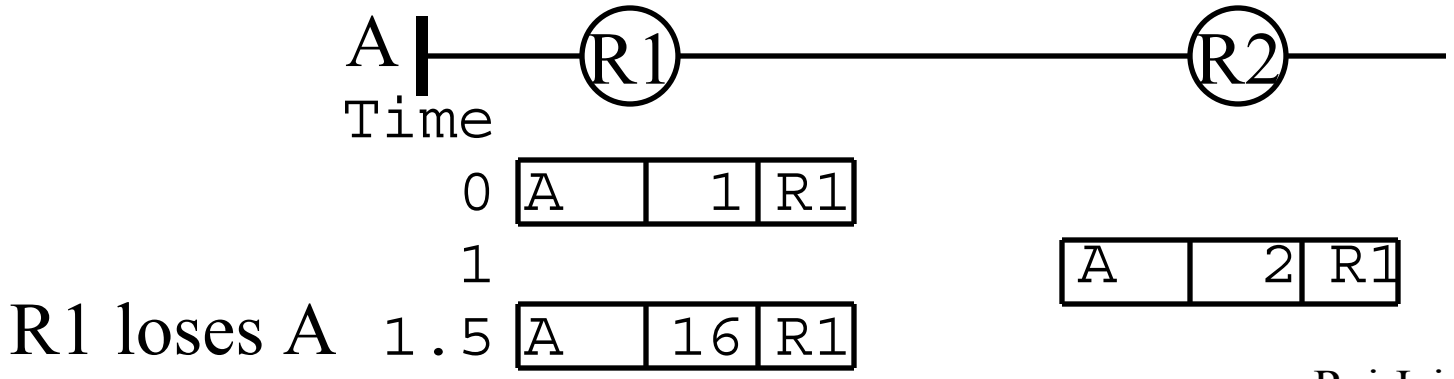
- Maximum network diameter = 15 hops
- Cost is measured in hops
  Only shortest routes. May not be the fastest route.
- Entire tables are broadcast every 30 seconds.
  Bandwidth intensive.
- Uses UDP with 576-byte datagrams.
  Need multiple datagrams.
  300-entry table needs 12 datagrams.
- An error in one routing table is propagated to all routers
- Slow convergence

Counting to Infinity Problem



Improving Convergence

- **Split Horizon**: Remember the port from which a route was learnt. Do not send the route to that port.
- **Hold-down Timer**: If a network is unreachable, ignore all updates for that network for, say, 60 s.
- **Poison Reverse and Triggered Updates**: Once a network is unreachable, it broadcasts it *immediately* to other routers and keeps the entry for some time.



RIP V2

- Backward compatible with RIP V1. Many new features
  - **Authentication**: If the address family is 0xFFFF, the first route entry is the password.
    (V1 ignores this family type)
  - **Subnet mask**: Added to addresses
  - **Route Tag**:
    Allows routes learnt externally (e.g., BGP)
  - **Next Hop**: Next hop for each route entry.
    Useful for multiple routers on a LAN
  - **Multicast**: RIPV2 uses IP multicast (address=224.0.0.9, 01-00-5E-00-00-09). RIPv1 uses broadcast.

RIPv2 Packet Format

Command	Version	Unused
0xFFFF		Authentication Type
Password		
Password		
Password		
Password		
Family of Net 1		Route Tag
Net 1 Address		
Subnet Mask		
Next Hop		
Distance of Network 1		

Up to 24 entries

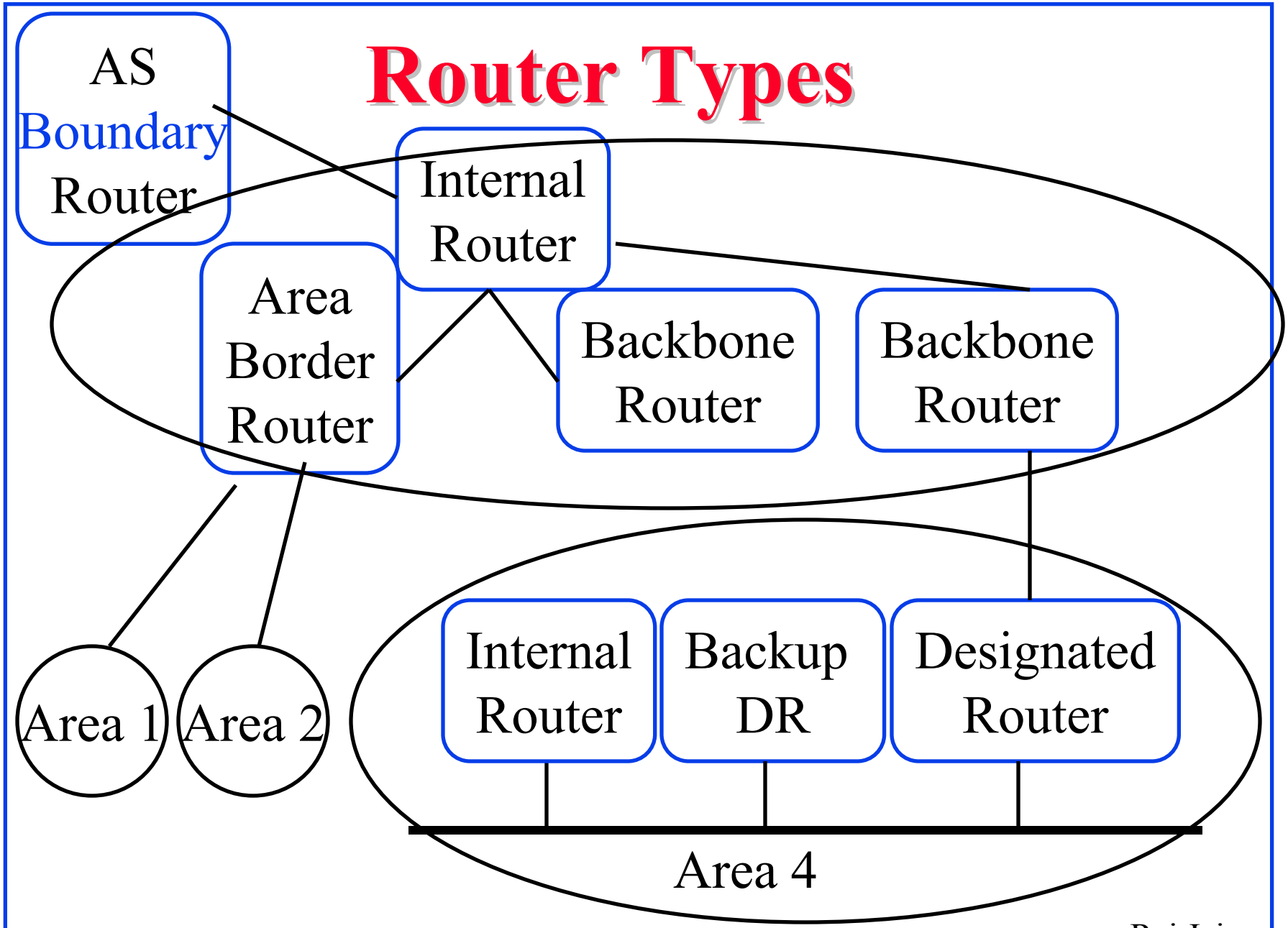
Static vs Dynamic Routing

- Static entries are put manually in the routing table. Also known as default route.
- Static entries override dynamic (learnt) entries.
- Static entry may or may not be included in the dynamic updates.
- Static entries not suitable for large highly dynamic networks.
- Static entries do not automatically change when the link goes down
- Static entries used in hub-and-spoke topologies. All branch routers are programmed to send all external packets to central office.

Open Shortest Path First (OSPF)

- Uses true metrics (not just hop count)
- Uses subnet masks
- Allows load balancing across equal-cost paths
- Supports type of service (ToS)
- Allows external routes (routes learnt from other autonomous systems)
- Authenticates route exchanges
- Quick convergence
- Direct support for multicast
- Link state routing ⇒ Each router broadcasts its connectivity with neighbors to entire network

Router Types



Router Types (Cont)

- **Internal Router (IR)**: All interfaces belong to the same area
- **Area Border Router (ABR):** Interfaces to multiple areas
- **Backbone Router (BR):** Interfaces to the backbone
- **Autonomous System Boundary Router (ASBR):** Exchanges routing info with other autonomous systems
- **Designated Router (DR):** Generates link-state info about the subnet
- **Backup Designated Router (BDR):** Becomes DR if DR fails.

OSPF Packet Header

- All OSPF packets have the same header. Body varies.

Version	Type	Packet Length
Router ID		
Area ID		
Checksum		Authentication Type
Authentication		
LSA Specific 1=Hello, 2= DB Description, 3=LS Request, 4= LS Update, 5=LS Ack		

OSPF Message Types

- Type 1 - Router Link-State Advertisements (LSAs):
  Neighbor's address and cost
  Flooded within the area by all routers.
- Type 2 - Network LSAs:
  Addresses of all routers on the LAN and cost
  Flooded within the area by Designated Router
- Type 3 - Summary LSAs: Flooded into area by ABR.
  Describes reachable networks in other areas.
- Type 4 - AS Boundary Router Summary LSAs:
  Describes cost from the router to ASBR.
  Flooded into the area by ABR.

Message Types (Cont)

- Type 5 - AS External LSAs:
  Flooded to all areas by ASBR.
  Describes external network reachable via the ASBR.
- Type 6 - Multicast Group Membership LSAs:
- Type 7 - Multicast OSPF
- All LSAs contain 32-bit sequence numbers.
  Used to detect duplicate and old LSAs.
- All database entries have an expiration timer (age field)

Metrics (Cost)

- RFC 1253: Metric = 10^8/Speed

Bit Rate	Metric
9.6 kbps	10,416
19.2 kbps	5208
56 kbps	1785
64 kbps	1562
T1 (1.544 Mbps)	65
E1 (2.048 Mbps)	48
Ethernet/802.3 (10 Mbps)	10
100 Mbps or more	1

Hello Protocol

- Routers periodically transmit hello packet
  Multicast to "All-SPF-Routers" (224.0.0.5)
- Used to find neighours and elect DR and BDR
- Packets stay on local subnet.
  Not forwarded by routers.
- Packet contains:
  - Router's selection of DR and BDR
  - Priority of DR and BDR
  - Timers: Hello interval and dead interval (time before a router is declared down)
  - List of neighbor routers from which hellos have been received

Adjacency

- Adjacency is formed between:
  - Two routers on a point-to-point link
  - DR or BDR and routers on LANs
  - Other routers on the LAN do not form adjacency between them
- Adjacent routers should have "synchronized databases"
- Routers send to adjacent routers a summary list of LSAs using database description packets
- Routers then compares the databases and request missing information.
- Database is synchronized ⇒ Fully adjacent. Dykstra algorithm is then run to find OSPF routes.

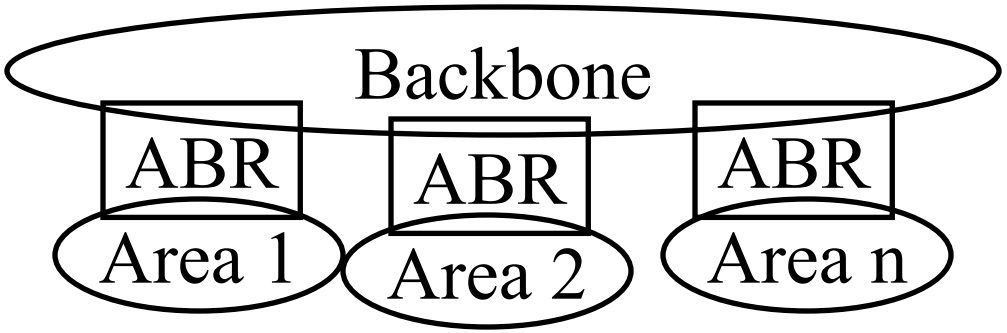
Maintaing the Database

- Databases are continually checked for synchronization by flooding LSAs
- All flooded LSAs are acked. Unacked LSAs are flooded again.
- Database information is checked. If new info, it is forwarded to other adjacencies using LSAs.
- When an entry is aged out, the info is flooded.
- Dykstra algorithm is run on every new info, to build new routing tables.

OSPF Areas

- LSAs are flooded throughout the area
- Area = domain
- Large networks are divided into areas to reduce routing traffic.
- Each area has a 32-bit area ID.
- Although areas are written using dot-decimal notation, they are locally assigned.
- The backbone area is area 0 or 0.0.0.0
  Other areas may be 0.0.0.1, 0.0.0.2, …
- Each router has a router ID. Typically formed from IP address of one of its interfaces.

Backbone Area

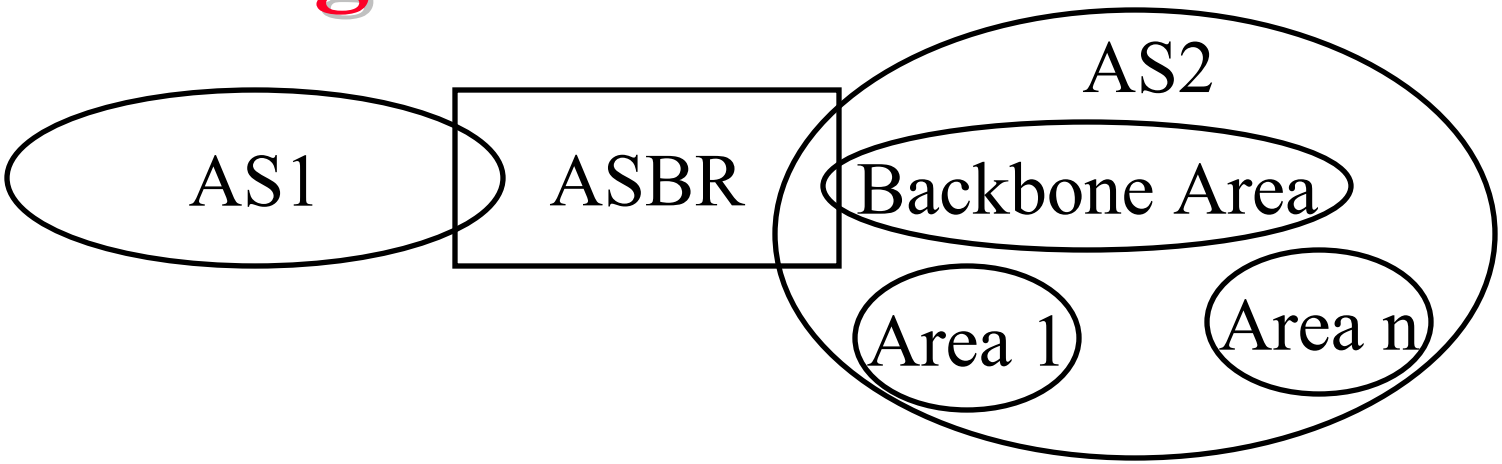


- Area border routers (ABRs) summarize the topology and transmit it to the backbone area
- Backbone routers forward it to other areas
- ABRs connect an area with the backbone area.
  ABRs contain OSPF data for two areas.
  ABRs run OSPF algorithms for the two areas.
- If there is only one area in the AS, there is no backbone area and there are no ABRs.

Inter-Area Routing

- Packets for other areas are sent to ABR
- ABR transmits the packet on the backbone
- Backbone routers send it to the destination area ABR
- Destination ABR forwards it in the destination area.

Routing Info from Other ASs



- Autonomous Systems Boundary Router (ASBR) exchanges "exterior gateway protocol (EGP)" messages with other autonomous systems
- ASBRs generate "external link advertisements." These are flooded to all areas of the AS. There is one entry for every external route.

RIPv1 vs RIPv2 vs OSPF

Feature	RIPv1	RIPv2	OSPF
RFC	RFC1058	RFC1723	RFC2178
Method	Distance-Vector	Distance-Vector	Link-state
Range of Metrics	16	16	65,535
Update Frequency	30s	30s	Changes or 30 mins
Dead interval	300s	300s	Variable setting
Authentication	No	Yes	Yes
Convergence	Slow	Fast	Fast
Subnet Support	No	Yes	Yes
Type of Service	No	No	Yes
Multipath	No	No	Yes
Max Net Diameter	15	15	65,535
Ease of Use	Yes	Yes	Complex Setup

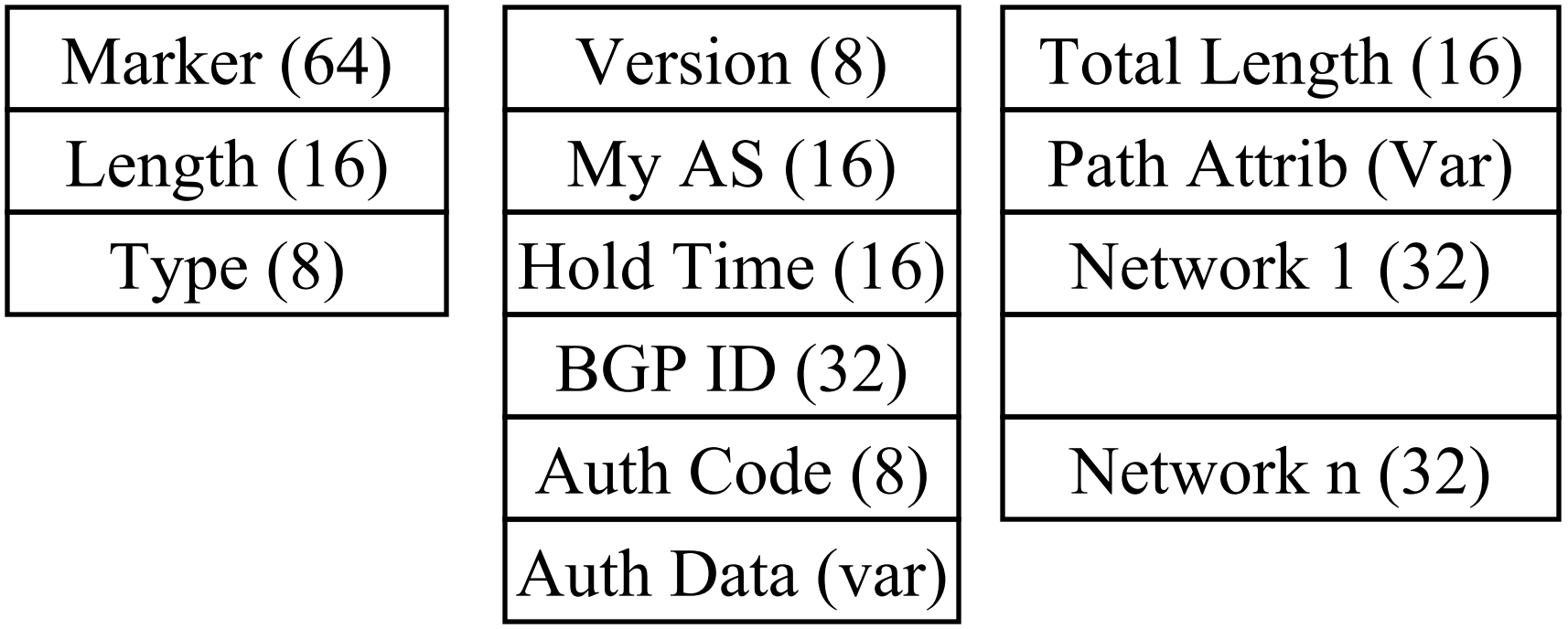
Border Gateway Protocol

- Inter-autonomous system protocol [RFC 1267]
- Used since 1989 but not extensively until recently
- Runs on TCP (segmentation, reliable transmission)
- Advertises all transit ASs on the path to a destination address
- A router may receive multiple paths to a destination
  $\Rightarrow$ Can choose the best path
- No loops and no count-to-infinity problems

BGP Operations

- BGP systems initially exchange entire routing tables. Afterwards, only updates are exchanged.
- BGP messages have the following information:
  - Origin of path information: RIP, OSPF, …
  - AS_Path: List of ASs on the path to reach the dest
  - Next_Hop: IP address of the border router to be used as the next hop to reach the dest
  - Unreachable: If a previously advertised route has become unreachable
- BGP speakers generate update messages to all peers when it selects a new route or some route becomes unreachable.

BGP Messages



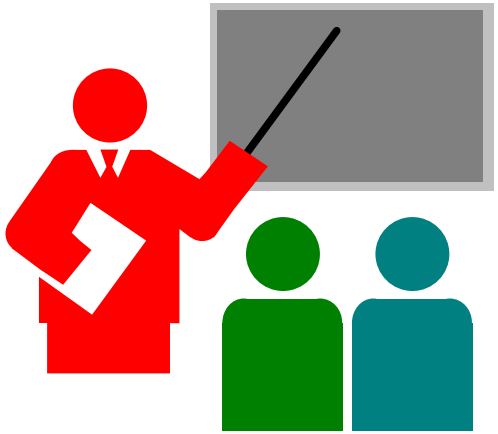
BGP Messages (Cont)

- Marker field is used for authentication or to detect a lost of synch
- Types of messages: Open, update, notification, keep-alive
- Open messages are used to establish peer relationship
  - Hold time: max time between successive keep-alive, update, or notification messages
  - BGP ID: IP address of one of the sender interfaces. Same value is used for all interfaces.
- Update messages are used to exchange routing info.
  - Path attributes = bit mask indicating optional/required, partial/full, etc.

IDRP

- Interdomain Routing Protocol (an EGP)
- Recent extension of BGP concepts
- Distributes path vectors
- Allows multiple routes to a destination
- Allows an additional hierarchy entity: Routing domain confederation ⇒ A domain can belong to several RDCs
- Each domain has a Routing Domain Identifier (RDI)
- Each RDC has a RDC identifier (RDCI)
- Uses link attributes, such as, throughput, delay, security
- IDRP has its own reliability mechanism
  ⇒ Does not need TCP

Summary



- RIP uses distance-vector routing
- RIP v2 fixes the slow convergence problem
- OSPF uses link-state routing and divides the autonomous systems into multiple areas.
  Area border router, AS boundary router, designated router
- BGP and IDRP are exterior gateway protocols