Internetwork Operation
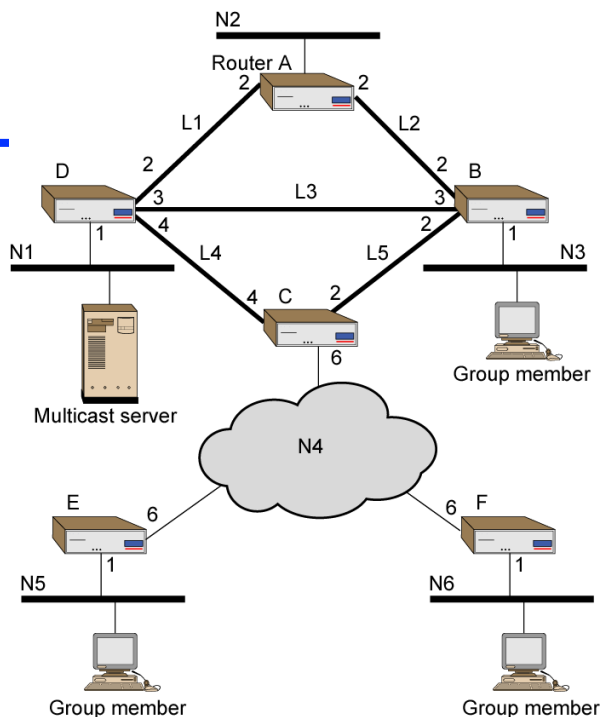
Multicasting

- Addresses that refer to group of hosts on one or more networks
- Uses
  - Multimedia "broadcast"
  - Teleconferencing
  - Database
  - Distributed computing
  - Real time workgroups

Example Config

Suppose N1 want to multicast to N3,N5 and N6

N2

Router A

2   2

L1   L2

D   2   2   B

3   L3   3

N1   4   2   N3

L4   L5

4   C   2

6

Multicast server

N4

Group member

E   6   6   F

1   1

N5   N6

Group member   Group member

Broadcast and Multiple Unicast

- Broadcast a copy of packet to each network
  —Requires 13 copies of packet
- Multiple Unicast
 —Send packet only to networks that have hosts in group
  —11 packets

2

Table 19.1 Traffic Generated by Various Multicasting Strategies

	(a) Broadcast					(b) Multiple Unicast				(c) Multicast
	S → N2	S → N3	S → N5	S → N6	Total	S → N3	S → N5	S → N6	Total	
N1	1	1	1	1	4	1	1	1	3	1
N2										
N3		1			1	1			1	1
N4			1	1	2		1	1	2	2
N5			1		1		1		1	1
N6				1	1			1	1	1
L1	1				1					
L2										
L3		1			1	1			1	1
L4			1	1	2		1	1	2	1
L5										
Total	2	3	4	4	13	3	4	4	11	8

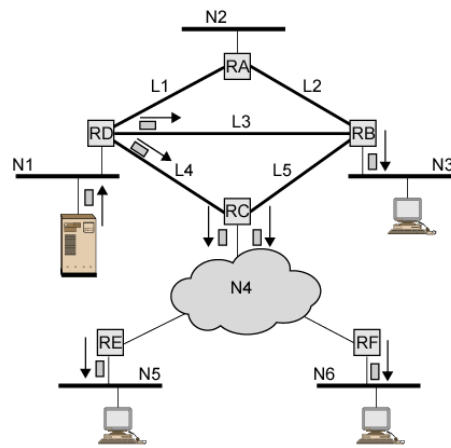# True Multicast

- Determine least cost path to each network that has host in group
  - Gives spanning tree configuration containing networks with group members
- Transmit single packet along spanning tree
- Routers replicate packets at branch points of spanning tree
- 8 packets required

3

Multicast Example
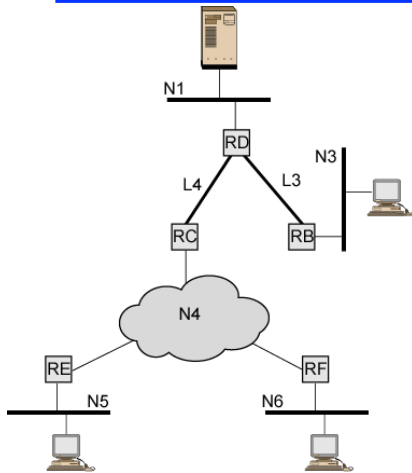
(a) Spanning tree from source to multicast group

(b) Packets generated for multicast transmission

Requirements for Multicasting (1)

- Router may have to forward more than one copy of packet
- Convention needed to identify multicast addresses
  — IPv4 - Class D - start 1110
  — IPv6 - 8 bit prefix, all 1, 4 bit flags field, 4 bit scope field, 112 bit group identifier
- Nodes must translate between IP multicast addresses and list of networks containing group members
- Router must translate between IP multicast address and network multicast address

Requirements for Multicasting (2)

- Mechanism required for hosts to join and leave multicast group
- Routers must exchange info
  - Which networks include members of given group
  - Sufficient info to work out shortest path to each network
  - Routing algorithm to work out shortest path
  - Routers must determine routing paths based on source and destination addresses

Spanning Tree from Router C to Multicast Group
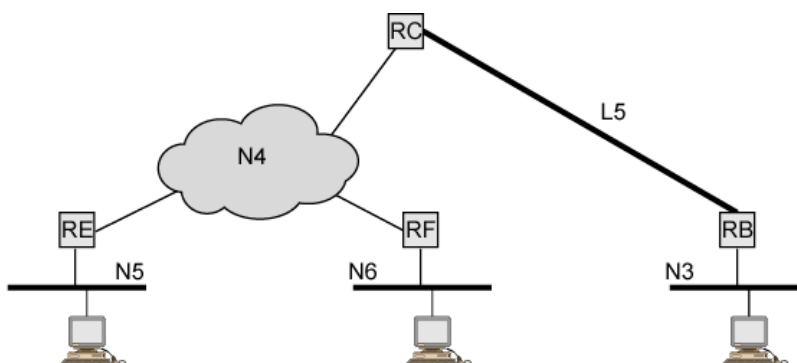
Internet Group Management Protocol (IGMP)

- RFC 3376
- Host and router exchange of multicast group info
- Use broadcast LAN to transfer info among multiple hosts and routers

Principle Operations

- Hosts send messages to routers to subscribe to and unsubscribe from multicast group
  —Group defined by multicast address
- Routers check which multicast groups are of interest to which hosts
- IGMP currently version 3
- IGMPv1
  —Hosts could join group
  —Routers used timer to unsubscribe members

Operation of IGMPv1 & v2

- Receivers have to subscribe to groups
- Sources do not have to subscribe to groups
- Any host can send traffic to any multicast group
- Problems:
  - Spamming of multicast groups
  - Even if application level filters drop unwanted packets, they consume valuable resources
  - Establishment of distribution trees is problematic
  - Location of sources is not known
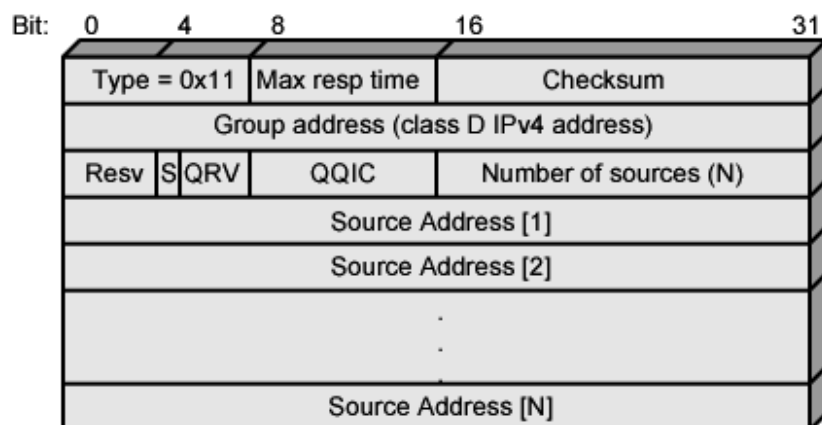 - Finding globally unique multicast addresses difficult

IGMP v3

- Allows hosts to specify list from which they want to receive traffic
  - Traffic from other hosts blocked at routers
- Allows hosts to block packets from sources that send unwanted traffic

IGMP Message Formats
Membership Query

(a) Membership query message

Membership Query

- Sent by multicast router
- General query
  —Which groups have members on attached network
- Group-specific query
  —Does group have members on an attached network
- Group-and-source specific query
  —Do attached device want packets sent to specified multicast address
  —From any of specified list of sources

Membership Query Fields (1)

- Type
- Max Response Time
  - Max time before sending report in units of 1/10 second
- Checksum
  - Same algorithm as IPv4
- Group Address
  - Zero for general query message
  - Multicast group address for group-specific or group-and-source
- S Flag
  - 1 indicates that receiving routers should suppress normal timer updates done on hearing query

Membership Query Fields (2)
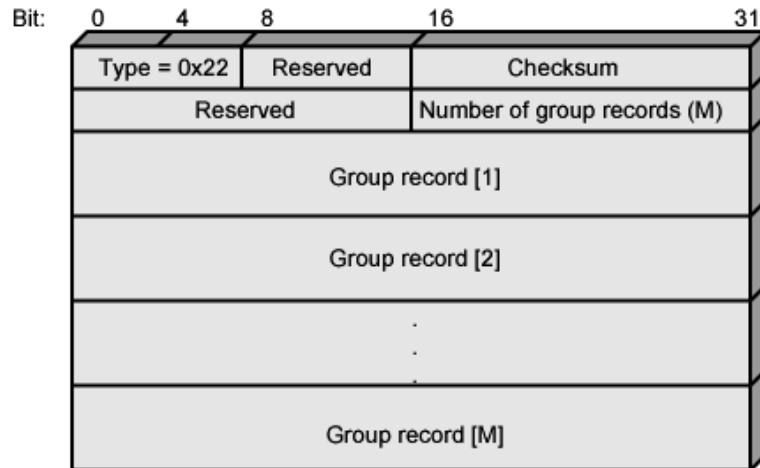
- QRV (querier's robustness variable)
  - RV value used by sender of query
  - Routers adopt value from most recently received query
  - Unless RV was zero, when default or statically configured value used
  - RV dictates number of retransmissions to assure report not missed
- QQIC (querier's querier interval code)
  - QI value used by querier
  - Timer for sending multiple queries
  - Routers not current querier adopt most recently received QI
  - Unless QI was zero, when default QI value used
- Number of Sources
- Source addresses
  - One 32 bit unicast address for each source

IGMP Message Formats
Membership Report

Bit:	0	4	8	16	31

```
+---------------+----------+---------------------------+
| Type = 0x22   | Reserved |        Checksum           |
+---------------+----------+---------------------------+
|        Reserved          | Number of group records (M)|
+--------------------------+---------------------------+
|                  Group record [1]                    |
+------------------------------------------------------+
|                  Group record [2]                    |
+------------------------------------------------------+
|                        .                             |
|                        .                             |
|                        .                             |
+------------------------------------------------------+
|                  Group record [M]                    |
+------------------------------------------------------+
```
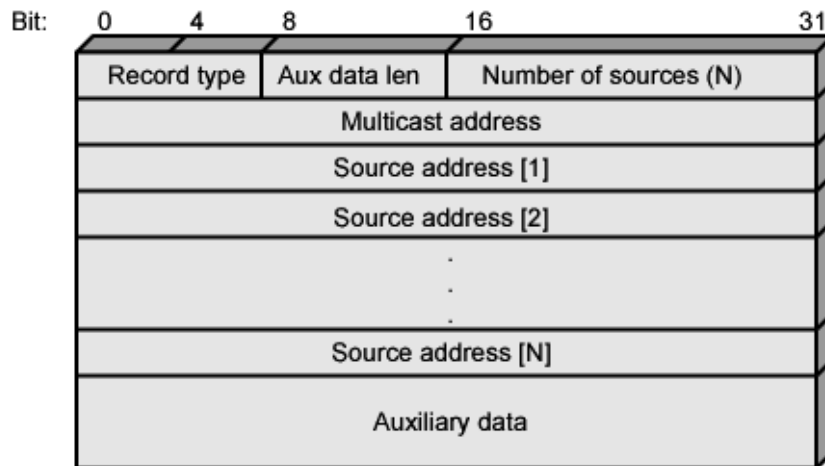
(b) Membership report message

Membership Reports

- Type
- Checksum
- Number of Group Records
- Group Records
  — One 32-bit unicast address per source

10

IGMP Message Formats
Group Record

Bit: 0 4	8	16	31
Record type	Aux data len	Number of sources (N)	
Multicast address			
Source address [1]			
Source address [2]			
. . .			
Source address [N]			
Auxiliary data			

(c) Group record

Group Record

- Record Type
  — See later
- Aux Data Length
  — In 32-bit words
- Number of Sources
- Multicast Address
- Source Addresses
  — One 32-bit unicast address per source
- Auxiliary Data
  — Currently, no auxiliary data values defined

IGMP Operation - Joining

- Host using IGMP wants to make itself known as group member to other hosts and routers on LAN
- IGMPv3 can signal group membership with filtering capabilities with respect to sources
  - EXCLUDE mode – all group members except those listed
  - INCLUDE mode – Only from group members listed
- To join group, host sends IGMP membership report message
  - Address field multicast address of group
  - Sent in IP datagram with Group Address field of IGMP message and Destination Address encapsulating IP header same
  - Current members of group will receive learn of new member
  - Routers listen to all IP multicast addresses to hear all reports

IGMP Operation – Keeping Lists Valid

- Routers periodically issue IGMP general query message
  - In datagram with all-hosts multicast address
  - Hosts that wish to remain in groups must read datagrams with this all-hosts address
  - Hosts respond with report message for each group to which it claims membership
- Router does not need to know every host in a group
  - Needs to know at least one group member still active
  - Each host in group sets timer with random delay
  - Host that hears another claim membership cancels own report
  - If timer expires, host sends report
  - Only one member of each group reports to router

IGMP Operation - Leaving

- Host leaves group, by sending leave group message to all-routers static multicast address
- Send membership report message with EXCLUDE option and null list of source addresses
- Router determine if there are any remaining group members using group-specific query message

Routing Protocols

- Routing Information
  —About topology and delays in the internet
- Routing Algorithm
  —Used to make routing decisions based on information

Autonomous Systems (AS)

- Group of routers
- Exchange information
- Common routing protocol
- Set of routers and networks managed by signle organization
- A connected network
  —There is at least one route between any pair of nodes

Interior Router Protocol (IRP)
Exterior Routing Protocol (ERP)

- Passes routing information between routers within AS
- May be more than one AS in internet
- Routing algorithms and tables may differ between different AS
- Routers need some info about networks outside their AS
- Used exterior router protocol (ERP)
- IRP needs detailed model
- ERP supports summary information on reachability

Application of IRP and ERP

Autonomous System 1

Autonomous System 2

| Interior router protocol | ◄────────► |
| Exterior router protocol | ◄─ ─ ─ ─ ► |
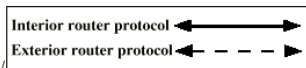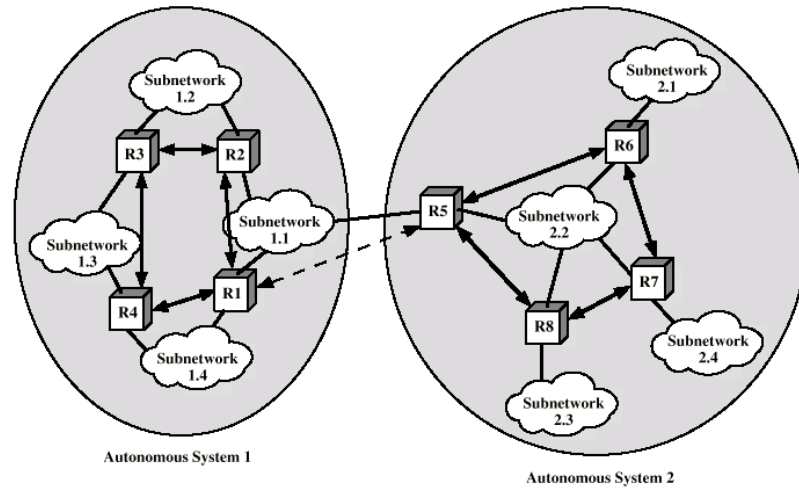
Approaches to Routing – Distance-vector

- Each node (router or host) exchange information with neighboring nodes
  - Neighbors are both directly connected to same network
- First generation routing algorithm for ARPANET
- Node maintains vector of link costs for each directly attached network and distance and next-hop vectors for each destination
- Used by Routing Information Protocol (RIP)
- Requires transmission of lots of information by each router
  - Distance vector to all neighbors
  - Contains estimated path cost to all networks in configuration
  - Changes take long time to propagate

15

Approaches to Routing – Link-state

- Designed to overcome drawbacks of distance-vector
- When router initialized, it determines link cost on each interface
- Advertises set of link costs to all other routers in topology
  - Not just neighboring routers
- From then on, monitor link costs
  - If significant change, router advertises new set of link costs
- Each router can construct topology of entire configuration
  - Can calculate shortest path to each destination network
- Router constructs routing table, listing first hop to each destination
- Router does not use distributed routing algorithm
  - Use any routing algorithm to determine shortest paths
  - In practice, Dijkstra's algorithm
- Open shortest path first (OSPF) protocol uses link-state routing.
- Also second generation routing algorithm for ARPANET

Exterior Router Protocols – Not Distance-vector

- Link-state and distance-vector not effective for exterior router protocol
- Distance-vector assumes routers share common distance metric
- ASs may have different priorities
  - May have restrictions that prohibit use of certain other AS
  - Distance-vector gives no information about ASs visited on route

Exterior Router Protocols – Not Link-state

- Different ASs may use different metrics and have different restrictions
    —Impossible to perform a consistent routing algorithm.
- Flooding of link state information to all routers unmanageable
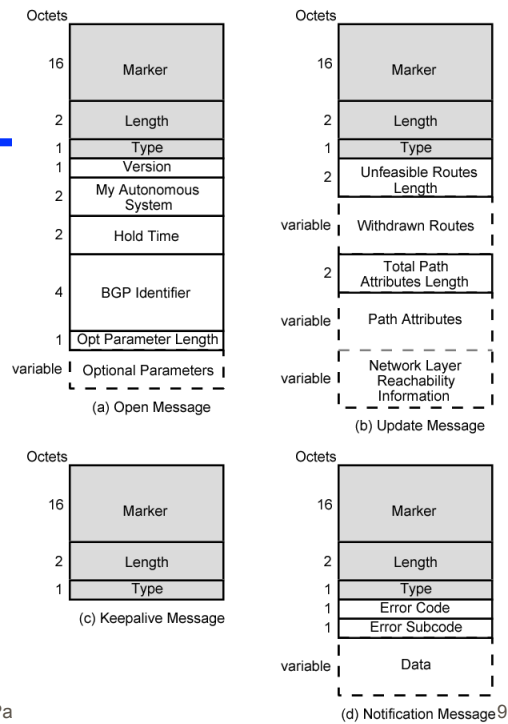
Exterior Router Protocols – Path-vector

- Dispense with routing metrics
- Provide information about which networks can be reached by a given router and ASs crossed to get there
    — Does not include distance or cost estimate
- Each block of information lists all ASs visited on this route
    — Enables router to perform policy routing
    — E.g. avoid path to avoid transiting particular AS
    — E.g. link speed, capacity, tendency to become congested, and overall quality of operation, security
    — E.g. minimizing number of transit ASs

Border Gateway Protocol (BGP)

- For use with TCP/IP internets
- Preferred EGP of the Internet
- Messages sent over TCP connections
  — Open
  — Update
  — Keep alive
  — Notification
- Procedures
  — Neighbor acquisition
  — Neighbor reachability
  — Network reachability

BGP Messages

Octets	
16	Marker
2	Length
1	Type
1	Version
2	My Autonomous System
2	Hold Time
4	BGP Identifier
1	Opt Parameter Length
variable	Optional Parameters

(a) Open Message

Octets	
16	Marker
2	Length
1	Type
2	Unfeasible Routes Length
variable	Withdrawn Routes
2	Total Path Attributes Length
variable	Path Attributes
variable	Network Layer Reachability Information

(b) Update Message

Octets	
16	Marker
2	Length
1	Type

(c) Keepalive Message

Octets	
16	Marker
2	Length
1	Type
1	Error Code
1	Error Subcode
variable	Data

(d) Notification Message

18

BGP Procedure

- Open TCP connection
- Send Open message
  - Includes proposed hold time
- Receiver selects minimum of its hold time and that sent
  - Max time between Keep alive and/or update messages

Message Types

- Keep Alive
  - To tell other routers that this router is still here
- Update
  - Info about single routes through internet
  - List of routes being withdrawn
  - Includes path info
    - Origin (IGP or EGP)
    - AS_Path (list of AS traversed)
    - Next_hop (IP address of boarder router)
    - Multi_Exit_Disc (Info about routers internal to AS)
    - Local_pref (Inform other routers within AS)
    - Atomic_Aggregate, Aggregator (Uses address tree structure to reduce amount of info needed)

Uses of AS_Path and Next_Hop

- AS_Path
  - —Enables routing policy
    - Avoid a particular AS
    - Security
    - Performance
    - Quality
    - Number of AS crossed
- Next_Hop
  - —Only a few routers implement BGP
    - Responsible for informing outside routers of routes to other networks in AS

Notification Message

- Message header error
  - — Authentication and syntax
- Open message error
  - — Syntax and option not recognized
  - — Unacceptable hold time
- Update message error
  - — Syntax and validity errors
- Hold time expired
  - — Connection is closed
- Finite state machine error
- Cease
  - — Used to close a connection when there is no error

BGP Routing Information Exchange

- Within AS, router builds topology picture using IGP
- Router issues Update message to other routers outside AS using BGP
- These routers exchange info with other routers in other AS
- Routers must then decide best routes

Open Shortest Path First (1)

- OSPF
- IGP of Internet
- Replaced Routing Information Protocol (RIP)
- Uses Link State Routing Algorithm
  - Each router keeps list of state of local links to network
  - Transmits update state info
  - Little traffic as messages are small and not sent often
  - RFC 2328
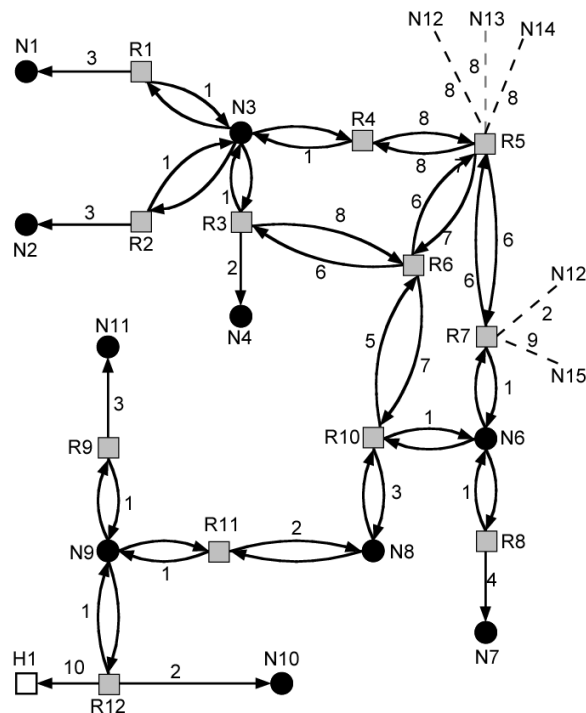- Route computed on least cost based on user cost metric

21

Open Shortest Path First (2)

- Topology stored as directed graph
- Vertices or nodes
  - Router
  - Network
    - Transit
    - Stub
- Edges
  - Graph edge
    - Connect two router
    - Connect router to network

Sample AS
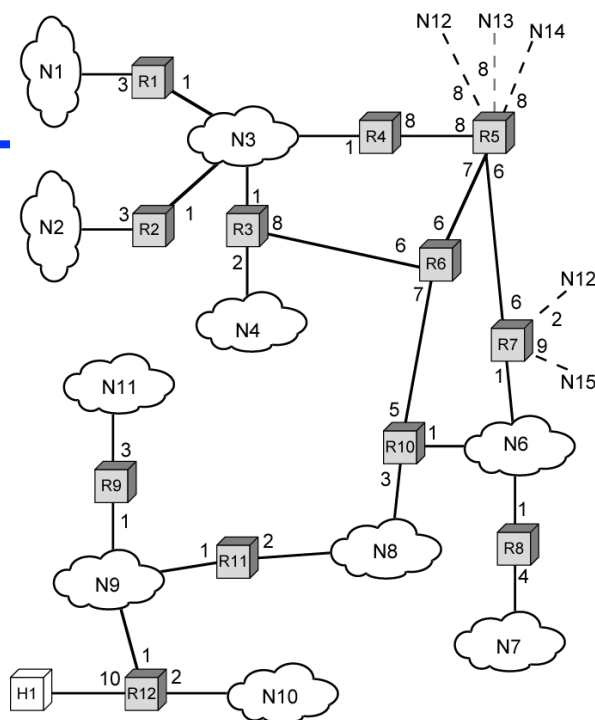
Directed Graph of AS

Operation

- Dijkstra's algorithm used to find least cost path to all other networks
- Next hop used in routing packets

23

SPF Tree for Router 6

Integrates Services Architecture

- Changes in traffic demands require variety of quality of service
- Internet phone, multimedia, multicast
- New functionality required in routers
- New means of requesting QoS
- ISA
- RFC 1633

Internet Traffic

- Elastic
    - Can cope with wide changes in delay and/or throughput
        - FTP sensitive to throughput
        - E-Mail insensitive to delay
        - Network Management sensitive to delay in times of heavy congestion
        - Web sensitive to delay
- Inelastic
    - Does not easily adapt to variations
    - e.g. real time traffic

Internet Traffic

- we will skip the rest of this chapter