Data Communications and Networking

Textbook
William Stallings, Data and Computer Communications, 6e

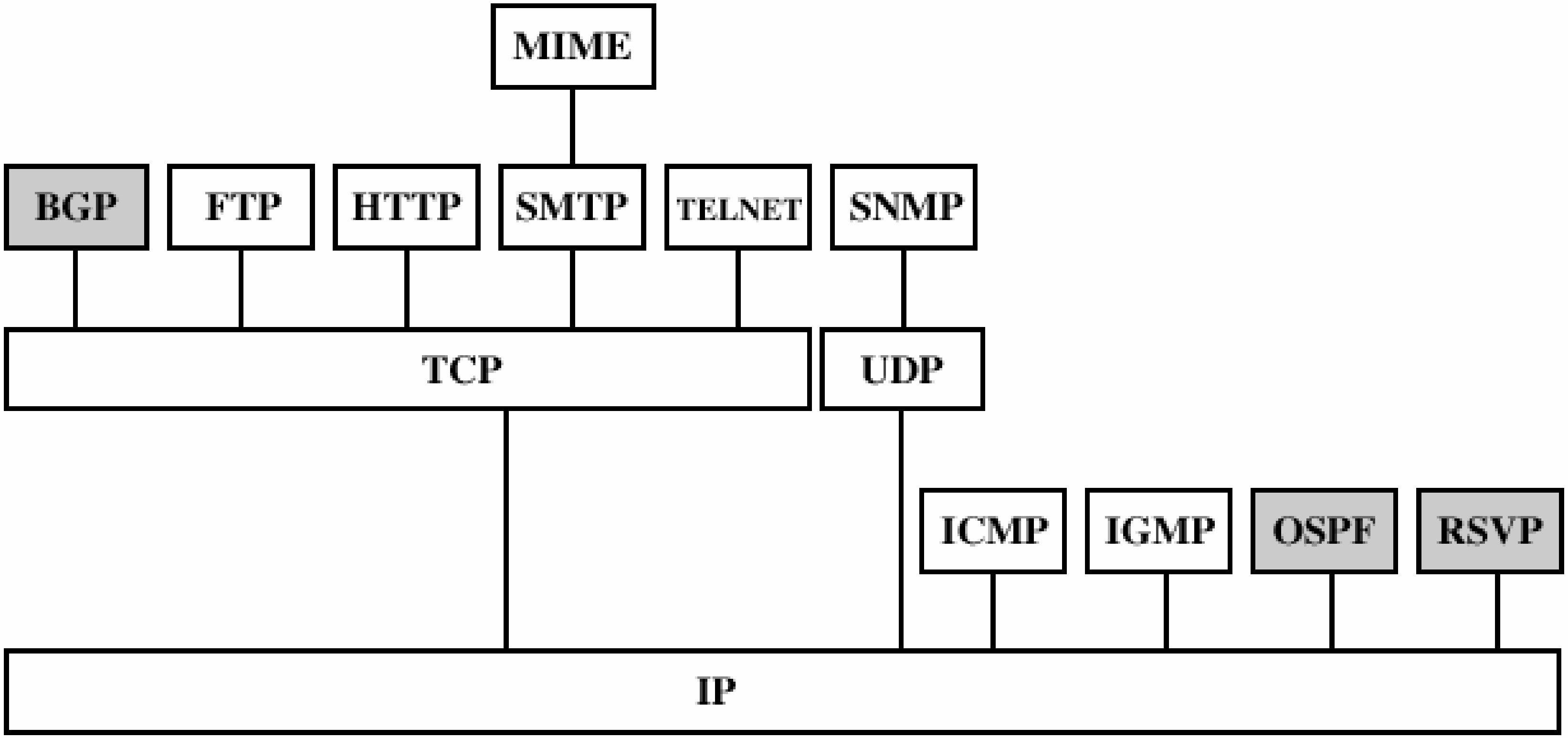
Chapter 16 Internetwork Operation

Yang Xianchun
Department of Computer Science and Technology
Nanjing University

Key Points

- Routing Protocols
  - IRP and its' instance - OSPF
  - ERP and its' instance - BGP
- ISA - Integrated Services Architecture
  - Approach, Components, and Services
- RSVP - Resource Reservation Protocol
  - Goals,Characteristics,Operation,and Mechanisms
- Differentiated Services
  - Services, Configuration and Operation

Internetworking Protocols in Context



16.1 Routing Protocols

- Autonomous Systems
  - Interior Router/Gateway Protocol (IRP/IGP)
  - Exterior Router/Gateway Protocol (ERP/EGP)
- Border Gateway Protocol (BGP)
  - Second generation EGP
- Open Shortest Path First Protocol (OSPF)

Distinguish Two Concepts

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

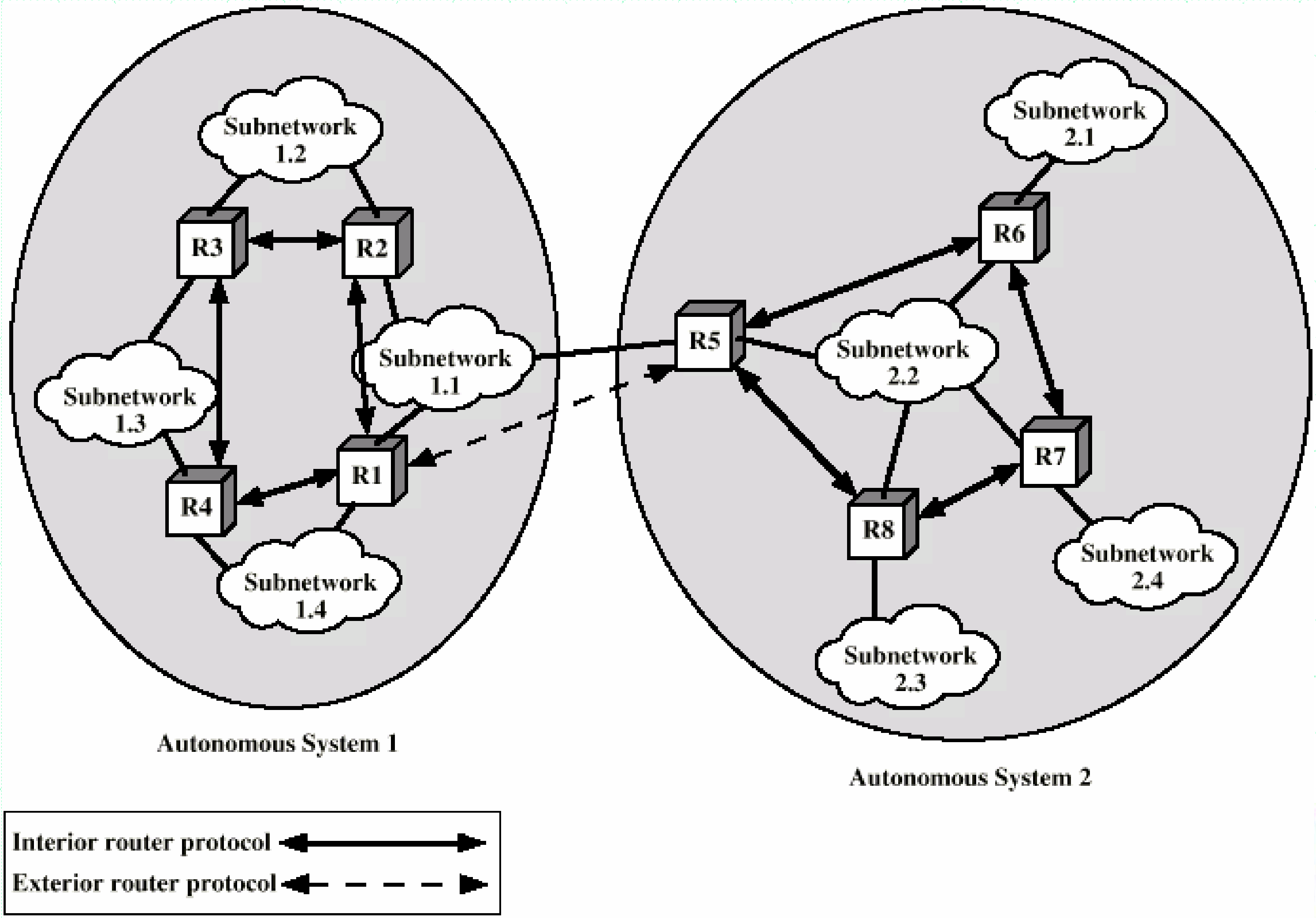
16.1.1 Autonomous Systems (AS)

- Group of routers
  - Exchange information
  - Common routing protocol
- Set of routers and networks
  - Managed by single organization
- A connected network (graph)
  - There is at least one route between any pair of nodes

Interior Router Protocol (IRP)

- Passes routing information between routers within an AS
  - No need to implement IRP outside of the AS
  - Allows IRPs to be custom-tailored
- May be more than one AS in internet
  - linked together through a wide area network
  - Routing algorithms and tables may differ between different ASs
  - Routers need some info about networks outside their ASs
  - Use ERP to pass routing info between routers in different ASs

Application of IRP and ERP



Different Flavor between IRP and ERP

- IRP needs to build up a rather detailed inter-connection model of routers within an AS
  - Calculate least-cost path from a given router to any network within AS
- ERP supports to exchange summary reach-ability info between separately administered ASs
- ERP is simpler and uses less detailed info than IRP

16.1.2 Border Gateway Protocol (BGP)

- BGP-4 (RFC 1771)
  - For use with TCP/IP internets
  - Preferred ERP(EGP) of the Internet
- Messages sent over TCP connections

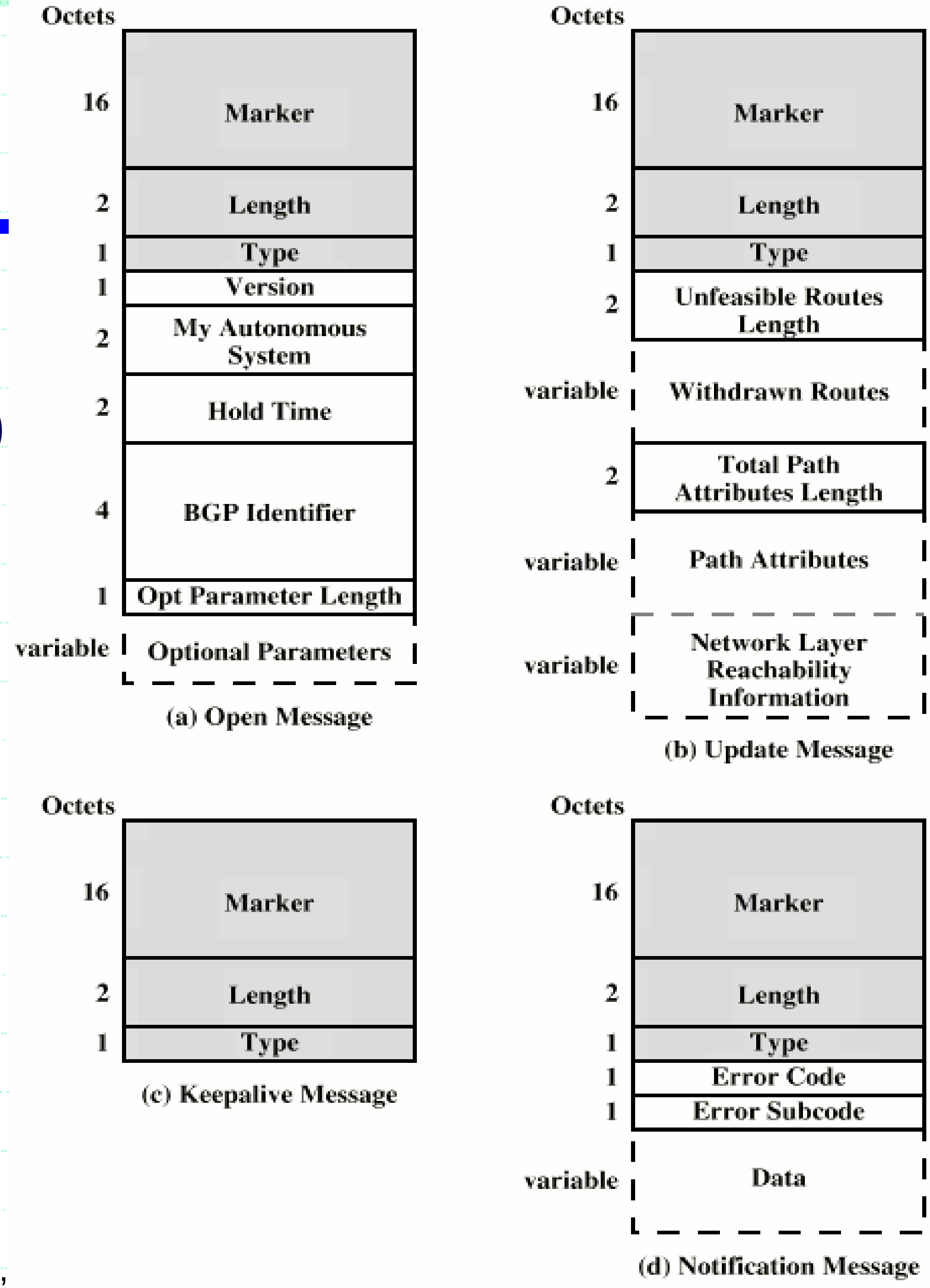
Open	Used to open a neighbor relationship with another router.
Update	Used to (1) transmit information about a single route and/or (2) list multiple routes to be withdrawn.
Keepalive	Used to (1) acknowledge an Open message and (2) periodically confirm the neighbor relationship.
Notification	Send when an error condition is detected.

BGP Procedures

- Neighbor acquisition
  - Acquisition: **Open**
  - Response: **Keep alive**
  - Error: Notification
- Neighbor reachability
  - To tell other routers that this router is still here
  - Periodically exchange info each other: **Keep alive**
  - Error: Notification
- Network reachability
  - Maintain a route database
  - Change or withdraw routing info: **Update**
  - Error: Notification

BGP Messages

- ⌘Header
  - Marker (16 octets)
  - Length (2 octets)
  - Type (1 octets)



Open and Keepalive Messages

- Open TCP connection to neighbor router
- Send **Open message**
  - Identifies the AS to which the sender belongs
  - Provides IP address of the router
  - Includes proposed hold time
- Target router return a **Keepalive message** in response to accept the request
- Receiver selects minimum of its hold time and that in received Open message
  - Max time that elapse between Keep alive and/or update messages received successively by sender
  - Each router often issues these messages to each of its peers enough to prevent Hold Timer from expiring

Update Message

- May contain one or two types of info
  - Info about single routes through internet
    - Network Layer Reachability Information - NLRI
    - Total Path Attributes Length
    - Path Attributes
  - List of (one or more) routes being withdrawn
    - Previously advertised by this router
    - Identified by IP address of destination network
- Path Attributes field includes path info
  - Origin (IRP or ERP)
  - 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

Add a word “不”
in Line 5
on Page 453

BGP Routing Information Exchange

- ⌘ Consider R1 in AS1, in Fig. 16.2
- ⌘ Within AS, router builds topology picture using IRP
- ⌘ Router issues **Update message** to other routers (e.g. R5) outside AS using BGP
  - **AS-Path**: the identity of AS1
  - **Next-Hop**: the IP address of R1
  - **NLRI**: a list of all of networks in AS1

This message informs R5 that all of networks listed in NLRI are reachable via R1 and that the only AS traversed is AS1

- ⌘ These routers exchange info with other routers in other AS
- ⌘ Routers must then decide best routes
- ⌘ Internal neighbor may exchange BGP info
  - No ID of common AS added to **AS-Path** by sending router
- ⌘ Multiple entry points into an AS available to a border router in another AS, use **Multi-Exit_Disc** to choose

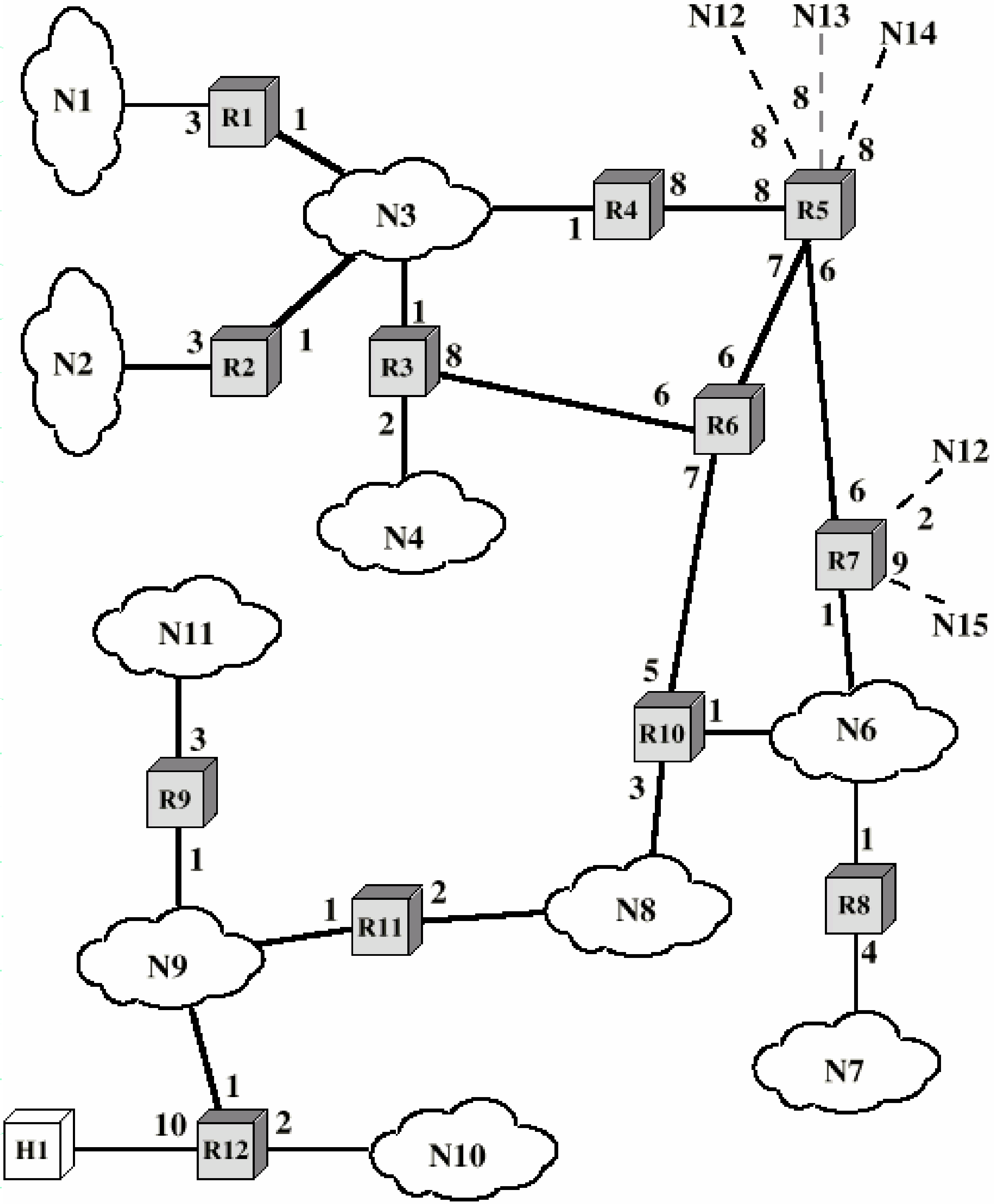
16.1.3 Open Shortest Path First (1)

- OSPF (RFC 2328)
- IRP (IGP) of Internet
- Replaced Routing Information Protocol (RIP)
- Uses Link State Routing Algorithm
  - What is the routing algorithm used in BGP-4 ?
  - 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
- Route computed on least cost based on user cost metric

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

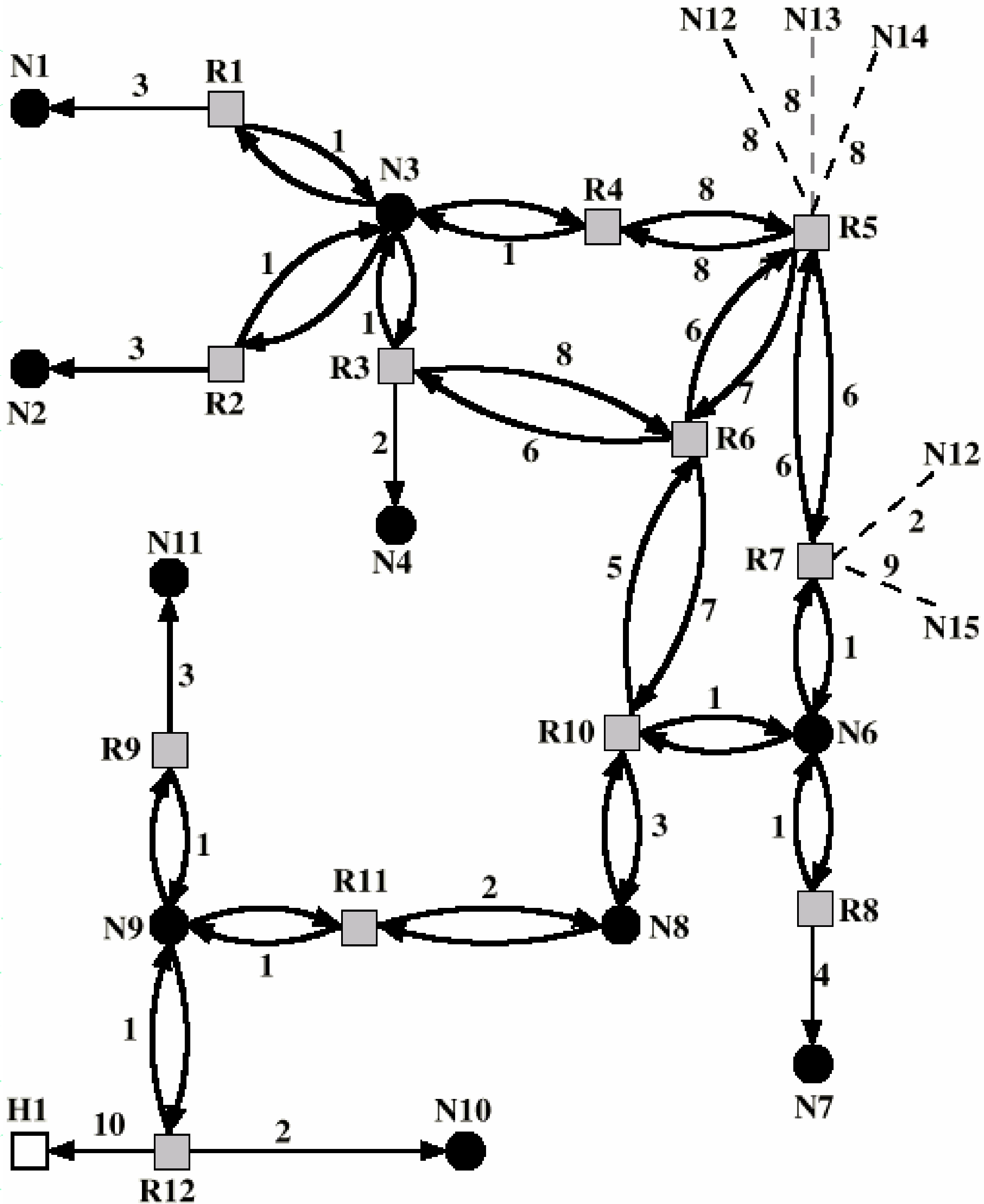
Sample AS


N1
R1
N2
R2
N3
R3
N4
R4
R5
N12
N13
N14
R6
R7
N12
N15
N11
R9
R10
N6
R8
N7
N8
R11
N9
H1
R12
N10

AS Represent as Directed Graph

- Two router are joined by a point-to-point link
  - directly connected by a pair of edges, one in each direction (e.g. R6 and R10)
- Multiple routers are attached to a network
  - all routers bidirectionally connected to network vertex (e.g. R1,R2,R3,and R4 all connected to N3)
- A single router attached to a network
  - The network will appear in graph as a stub connection (N7)
- A host directly connected to a router
  - It is depicted in the corresponding graph (e.g. H1)
- A router is connected to other AS
  - Each network in other AS is represented as a stub (e.g. N12 through N15)

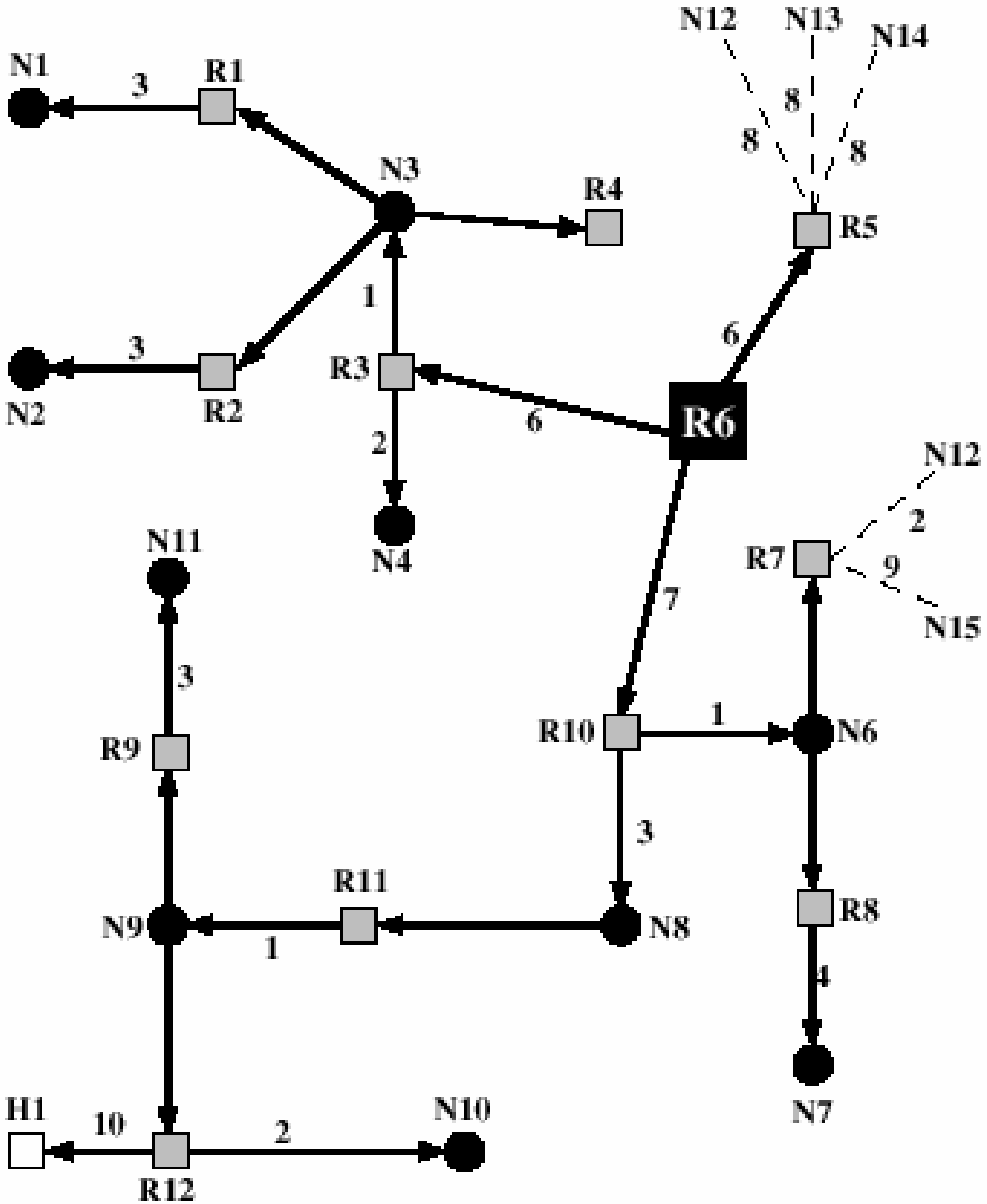
Directed Graph of AS



Operation

- Dijkstra's algorithm (Appendix 10A) used to find least cost path to all other networks
- Next hop used in routing packets

SPF Tree for R6



Routing Table for R6

Destination	Next Hop	Distance
N1	R3	10
N2	R3	10
N3	R3	7
N4	R3	8
N6	R10	8
N7	R10	12
N8	R10	10
N9	R10	11
N10	R10	13
N11	R10	14
H1	R10	21
R5	R5	6
R7	R10	8
N12	R10	10
N13	R5	14
N14	R5	14
N15	R10	17

16.2 Integrates Services Architecture

- Changes in traffic demands require variety of quality of service
- Internet phone, multimedia, multicast
- ATM schemes are costly
- New means of requesting QoS in TCP/IP
- New functionality required in routers
- ISA
  - A suite of standards that IETF is developing
- RFC 1633

16.2.1 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

How stringent the quality-of-service requirements are

Application	Reliability	Delay	Jitter	Bandwidth
E-mail	High	Low	Low	Low
File transfer	High	Low	Low	Medium
Web access	High	Medium	Low	Medium
Remote login	High	Medium	Medium	Low
Audio on demand	Low	Low	High	Medium
Video on demand	Low	Low	High	High
Telephony	Low	High	High	Low
Videoconferencing	Low	High	High	High

Requirements for Inelastic Traffic

- Throughput
  - Absolutely require a given minimum value
- Delay
  - As short as possible
- Jitter
  - Delay variation require a reasonable upper bound
- Packet loss
  - Real-time applications vary in amount
- Require preferential treatment for certain types of traffic
- Require elastic traffic to be supported as well

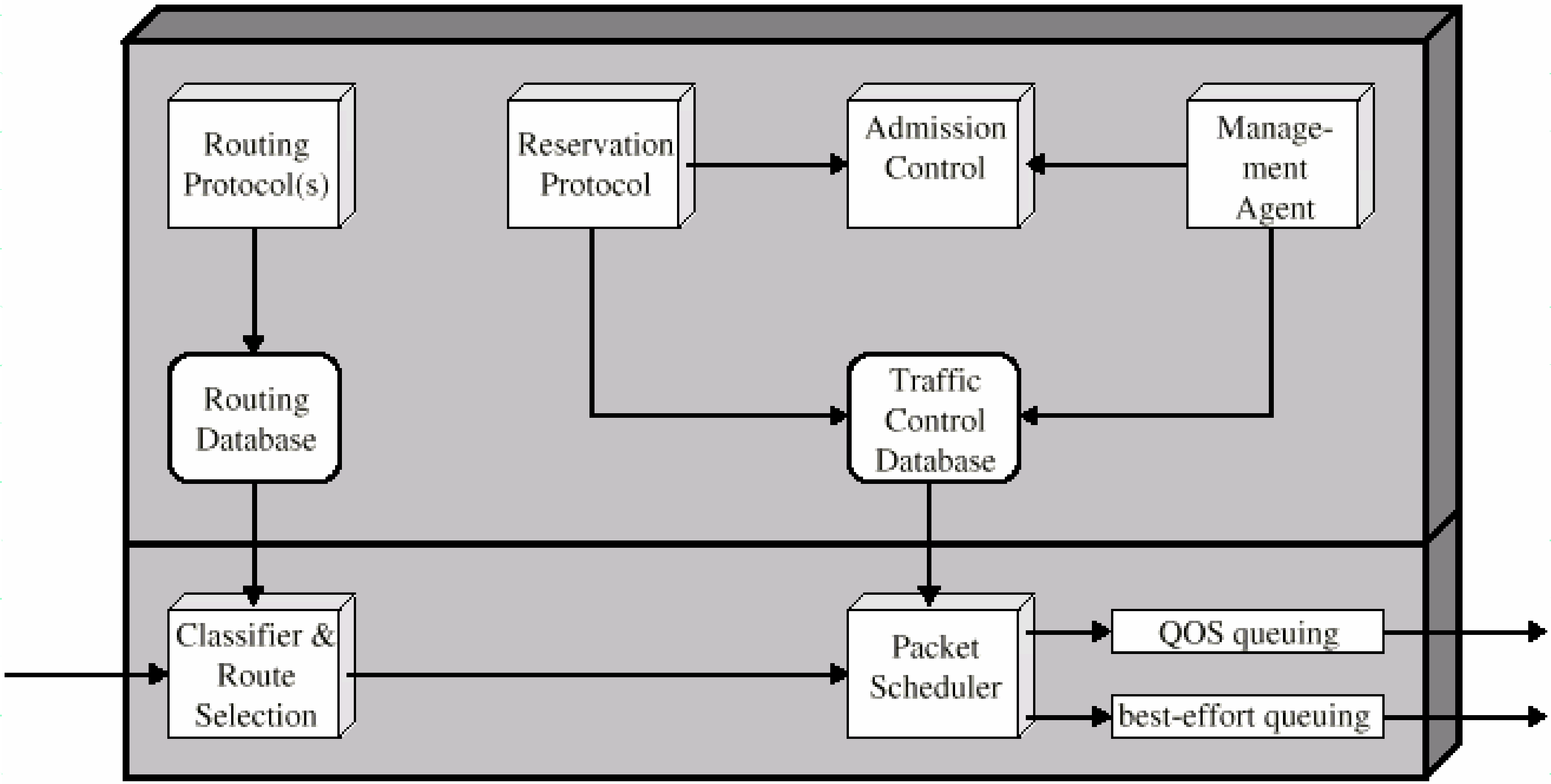
16.2.2 ISA Approach

- Congestion controlled by
  - Routing algorithms - minimum delay
  - Packet discard - the most recent
- Associate each packet with a flow
  - e.g. Consists of a transport connection or a video stream
  - A flow differ from a TCP connection
    - Unidirectional
    - Can be multicast
- Manage congestion and provide QoS
  - Admission Control - require RSVP for a new flow
  - Routing Algorithm - OSPF select routes based on QoS
  - Queuing discipline - effective
  - Discard policy - manage congestion and guarantee QoS

16.2.3 ISA Components

- Principal background functions of ISA in a router
  - Reservation Protocol
  - Admission Control
  - Management Agent
  - Routing Protocol
- Background functions support the main task of router
  - Forwarding of Packets
- Functional area that accomplish forwarding
  - Classifier and routing selection
  - Packet scheduler

ISA Implemented in Router



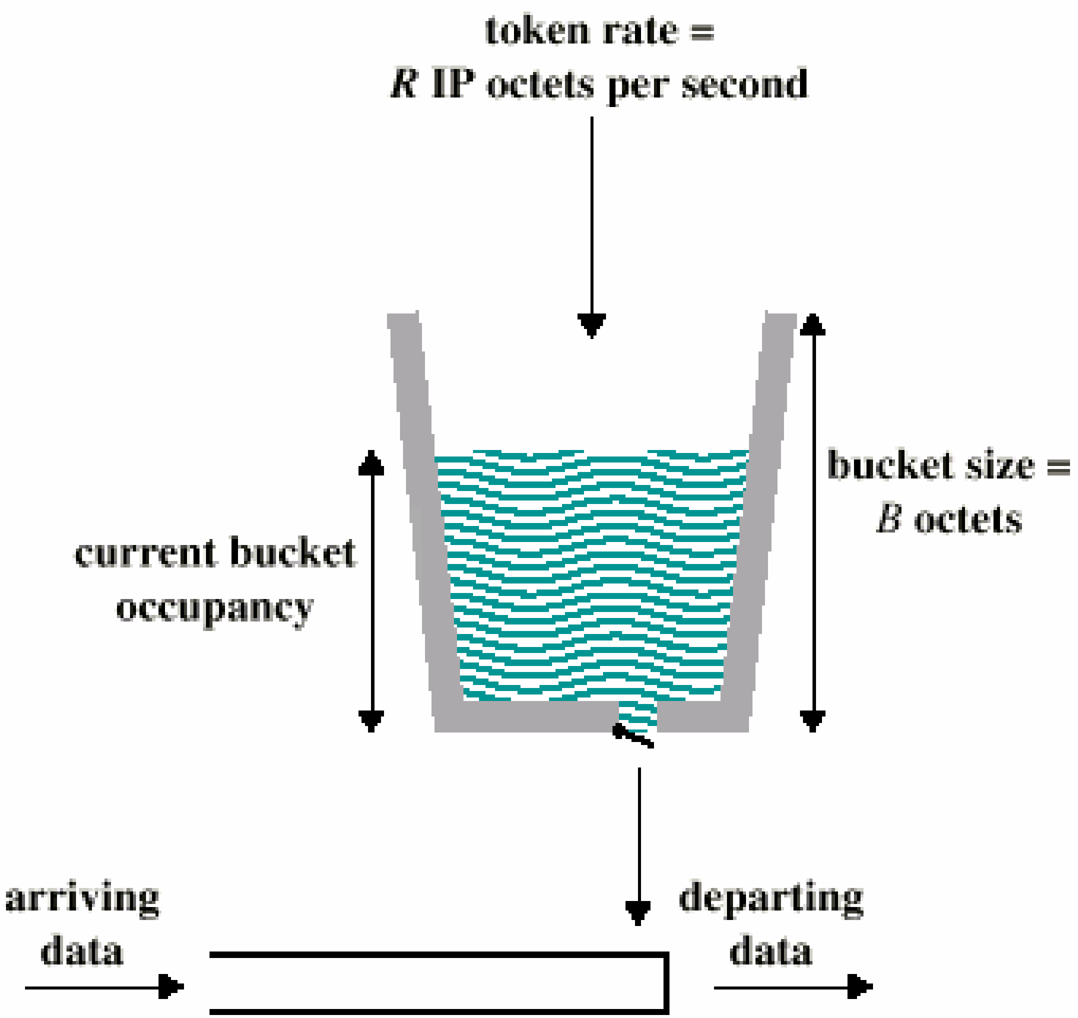
16.2.4 ISA Services

- Guaranteed
- Controlled load
- Best effort

Token Bucket Traffic Specification

- Token replenishment rate R
  - Continually sustainable data rate
- Bucket size B
  - Amount that data rate can exceed R for short period
  - During time period T amount of data sent can not exceed $RT + B$

Token Bucket Scheme



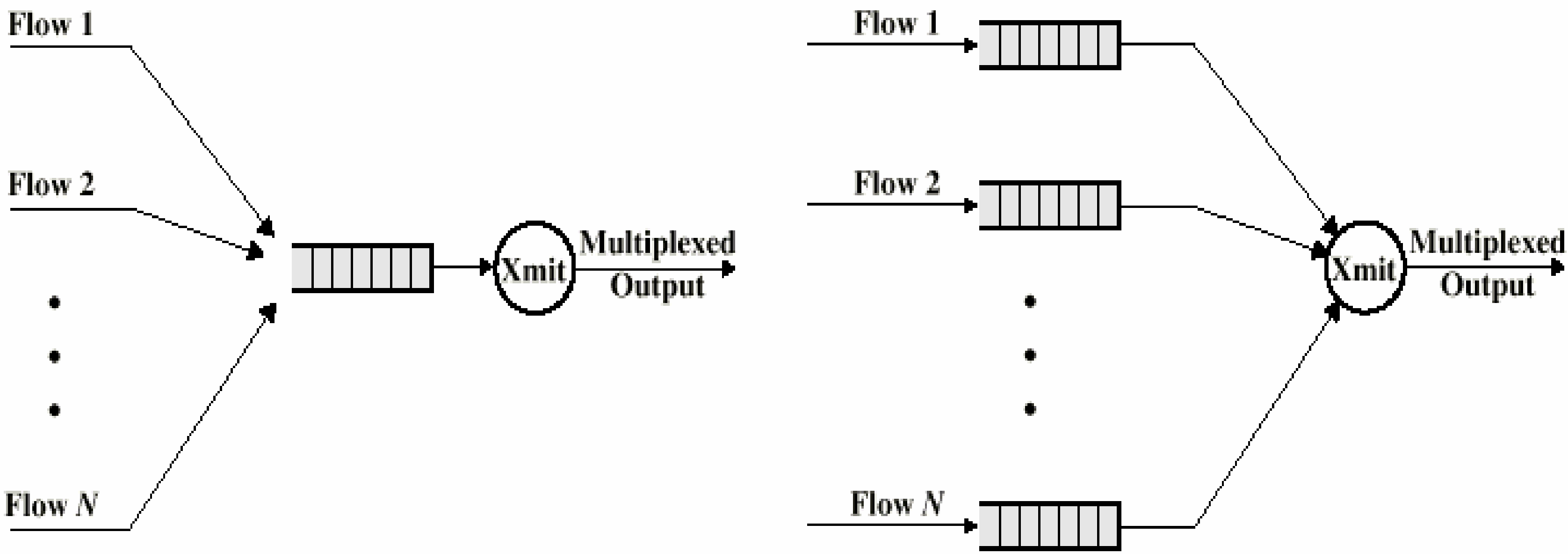
ISA Services

- Guaranteed
  - Assured data rate
  - Upper bound on queuing delay
  - No queuing loss
  - Real time playback
- Controlled load
  - Approximates behavior to best efforts on unloaded network
  - No specific upper bound on queuing delay
  - Very high delivery success

12.6.5 Queuing Discipline

- Traditionally FIFO
  - No special treatment for high priority flow packets
  - Large packet can hold up smaller packets
  - Greedy connection can crowd out less greedy connection
- Fair queuing
  - Queue maintained at each output port
  - Packet placed in queue for its flow
  - Round robin servicing
  - Skip empty queues
  - Can have weighted fair queuing

FIFO and Fair Queue



(a) FIFO Queuing

(b) Fair Queuing

16.3 Resource Reservation: RSVP

- Unicast applications can reserve resources in routers to meet QoS
- If router can not meet request, application informed
- Multicast is more demanding
- May be reduced
  - Some members of group may not require delivery from particular source over given time
    - e.g. selection of one from a number of “channels”
  - Some group members may only be able to handle a portion of the transmission

Soft State

- Set of state info in router that expires unless refreshed
- Applications must periodically renew requests during transmission
- Resource ReSerVation Protocol (RSVP)
- RFC 2205

RSVP Goals

- Ability for receivers to make reservations
- Deal gracefully with changes in multicast group membership
- Specify resource requirements such that aggregate resources reflect requirements
- Enable receivers to select one source
- Deal gracefully with changes in routes
- Control protocol overhead
- Independent of routing protocol

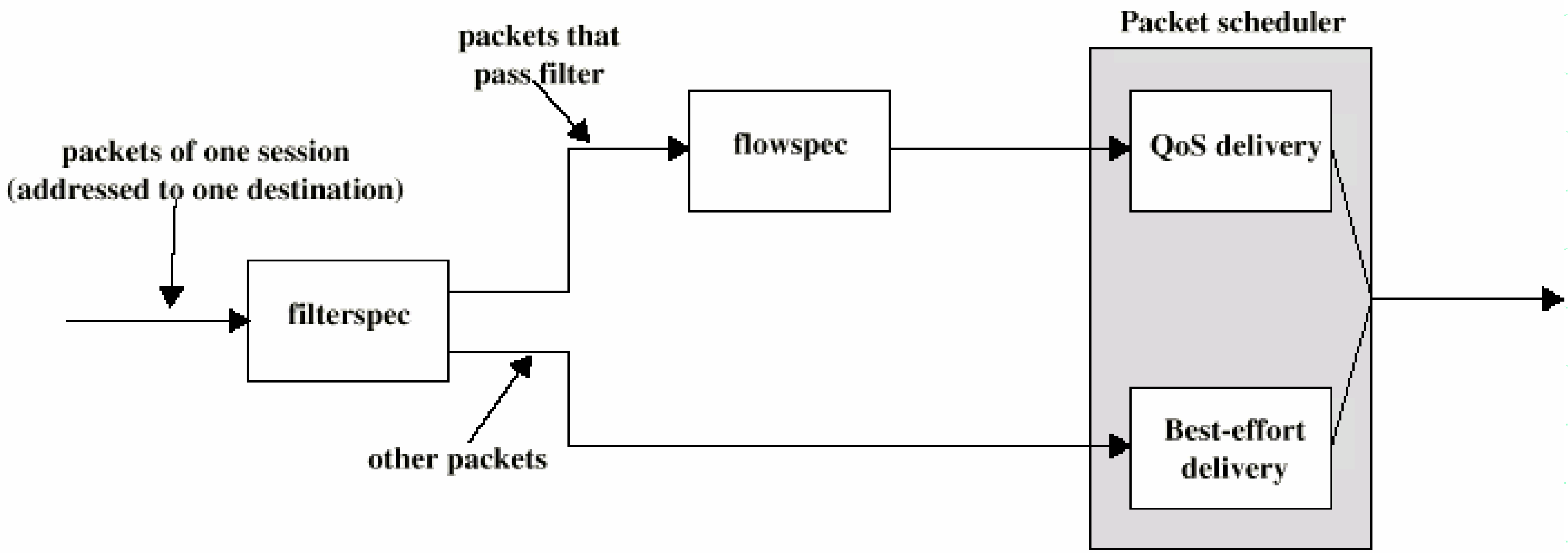
RSVP Characteristics

- ⌘Unicast and Multicast
- ⌘Simplex
- ⌘Receiver initiated reservation
- ⌘Maintain soft state in the internet
- ⌘Provide different reservation styles
- ⌘Transparent operation through non-RSVP routers
- ⌘Support for IPv4 and IPv6

Data Flow Concepts

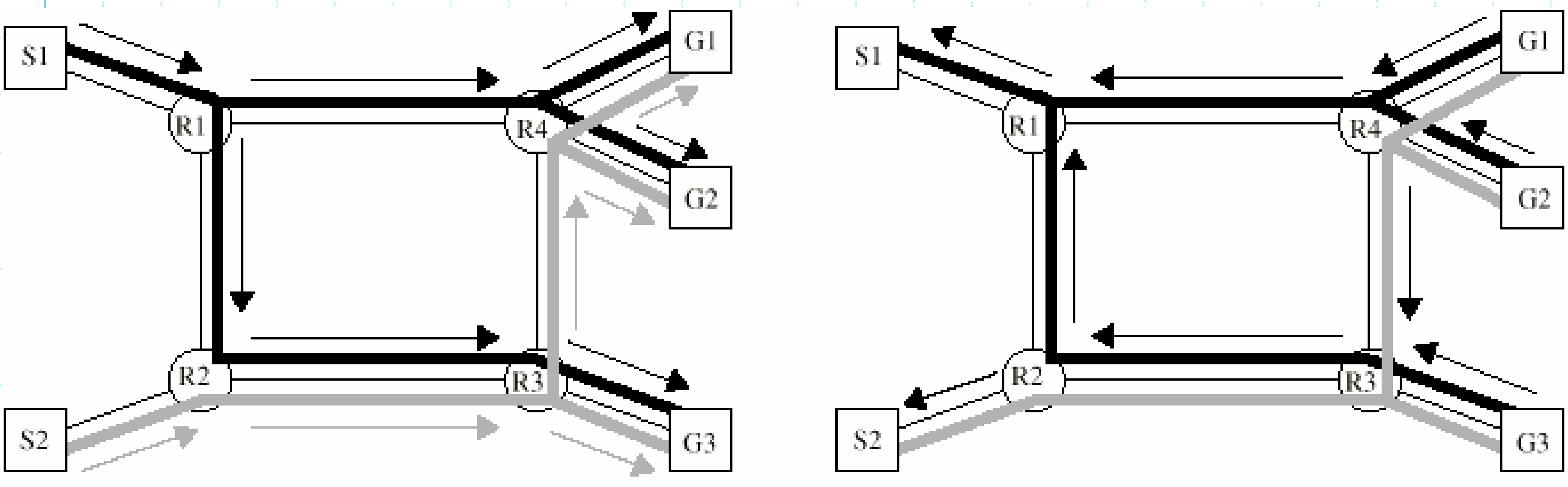
- Session
  - Data flow identified by its destination
- Flow descriptor
  - Reservation request issued by destination
  - Made up of flowspec and filterspec
  - **Flowspec** gives required QoS
    - Service class
    - Rspec
    - Tspec
  - **Filterspec** defines set of packets for which reservation is required

Treatment of Packets



⌘ Relationship among session, flowspec, and filterspec

RSVP Operation



(a) Data distrubution to a multicast group

(b) Merged Resv Messages

RSVP Message Types

- Resv
  - Originate at multicast receivers
  - Propagate upstream through distribution tree
  - Create soft states within routers
  - Reach sending host enabling it to set up traffic control for first hop
- Path
  - Provide upstream routing information

RSVP Operation From Host Perspective

- Receiver joins multicast group (IGMP)
- Potential sender issues Path message
- Receiver gets message identifying sender
- Receiver has reverse path info and may start sending Resv messages
- Resv messages propagate through internet and is delivered to sender
- Sender starts transmitting data packets
- Receiver starts receiving data packets

16.4 Differentiated Services

- Provide simple, easy to implement, low overhead tool to support range of network services
- differentiated on basis of performance
  - IP Packets labeled for differing QoS using existing IPv4 **Type of Service** or IPv6 **Traffic Class**
  - Service level agreement (SLA) established between provider and customer prior to use of DS
  - Built in aggregation
    - Good scaling to larger networks and loads
  - Implemented by queuing and forwarding based on DS octet
    - No state info on packet flows stored
- Reference – RFC 2475

16.4.1 DS Services

- Defined within DS domain
  - Contiguous portion of internet over which consistent set of DS policies are administered
  - Typically under control of one organization
  - Defined by service level agreements (SLA)

SLA Parameters

- Detailed service performance
  - Expected throughput
  - Drop probability
  - Latency
- Constraints on ingress and egress points
- Traffic profiles
  - e.g. token bucket parameters
- Disposition of traffic in excess of profile

Example Services

- Level A - low latency
- Level B - low loss
- Level C - 90% of traffic < 50ms latency
- Level D - 95% in profile traffic delivered
- Level E - allotted twice bandwidth of level F traffic
- Traffic with drop precedence X higher probability of delivery than that of Y

16.4.2 DS Octet - Code Pools

- Placed in Type of Service field of IPv4 header or Traffic Class field of IPv6 header
- Leftmost 6 bits used
- 3 pools of code points
- xxxxx0
  - assignment as standards
- xxxx11
  - experimental or local use
- xxxx01
  - experimental or local but may be allocated for standards in future

DS Octet - Precedence Field

- Xxx000
  - Compatibility with IPv4 precedence service
- Three approaches
  - Routing selection
  - Network service
  - Queuing discipline
    - Queue service
    - Congestion control

Type of Service (TOS) field

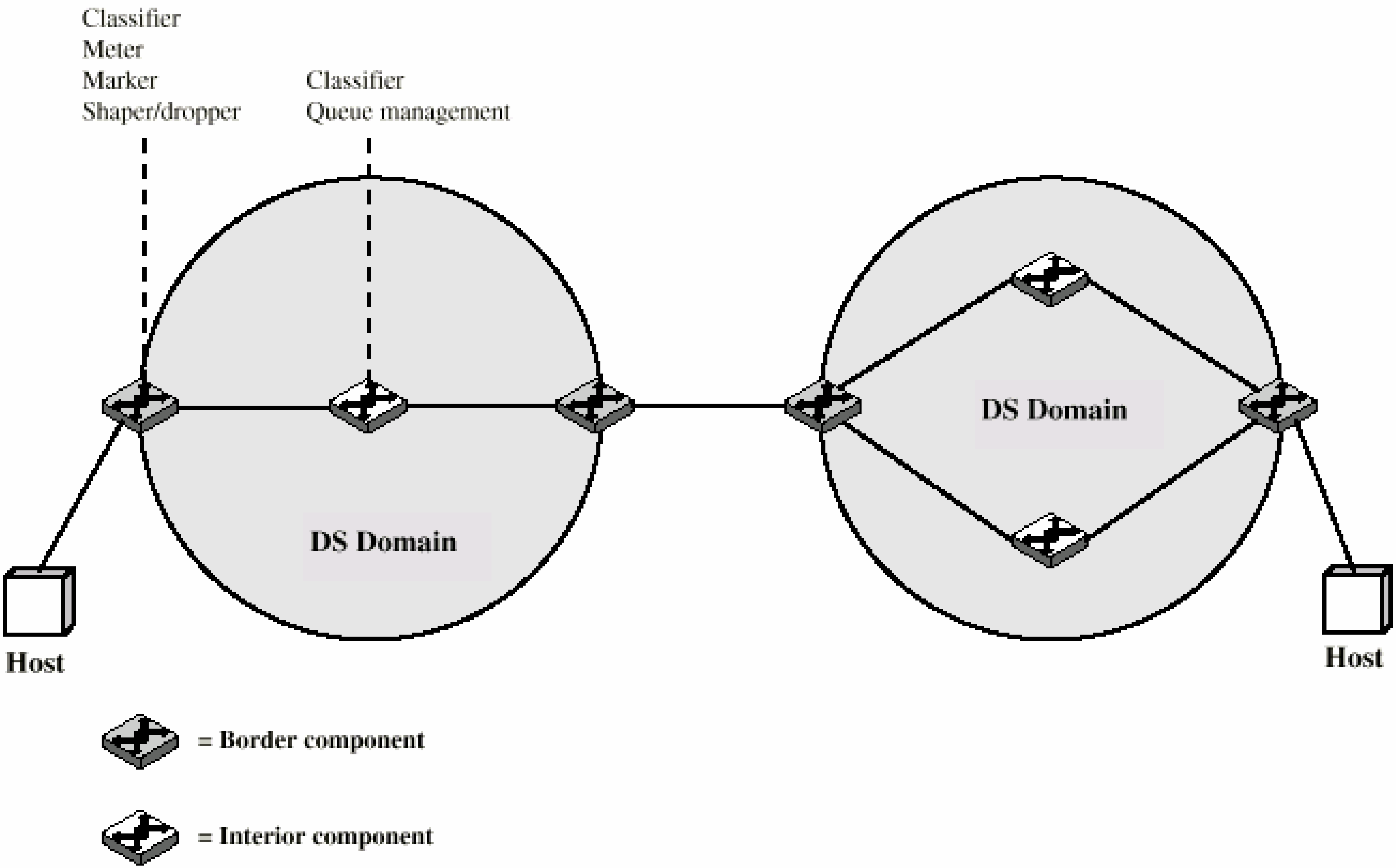
- **3-bit precedence subfield**

  provides guidance to IP entity in the source or router on selecting the next hop for this datagram

- **4-bit TOS subfield**

  provides guidance about the relative allocation of router resources for this datagran

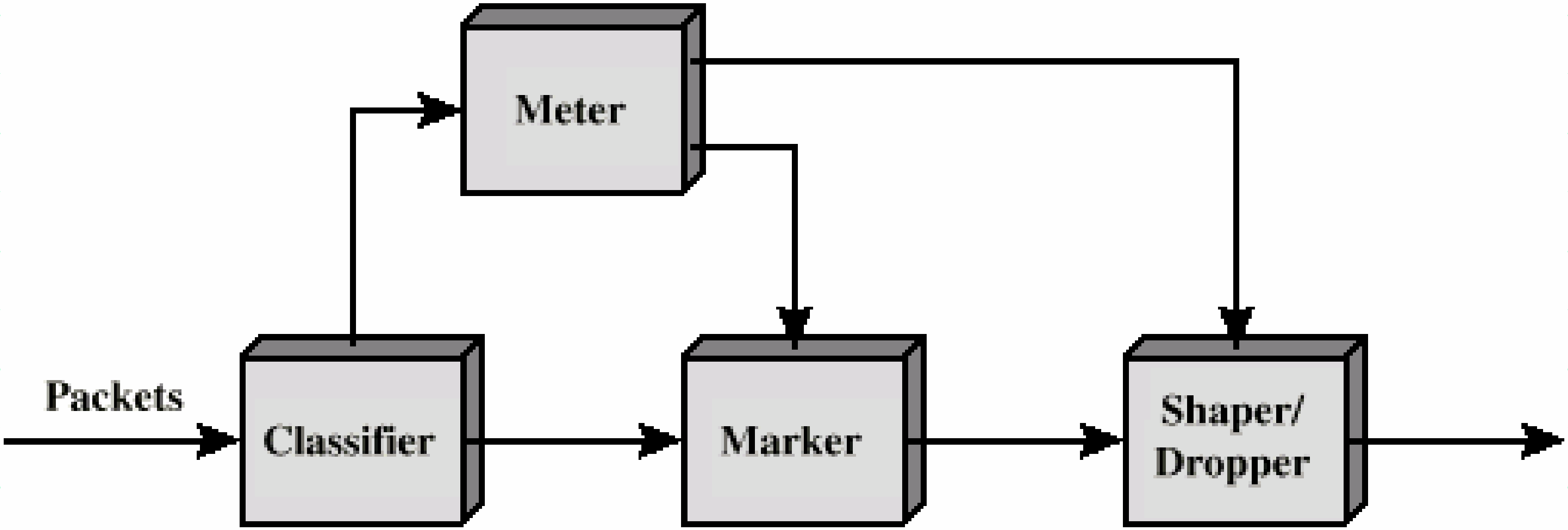
DS Domains


Classifier
Meter
Marker
Shaper/dropper
Classifier
Queue management
DS Domain
DS Domain
Host
Host
= Border component
= Interior component

16.4.3 DS Configuration and Operation

- Within domain, interpretation of DS code points is uniform
- Routers in domain are boundary nodes or interior nodes
- Traffic conditioning functions
  - Classifier
  - Meter
  - Marker
  - Shaper
  - Dropper

DS Traffic Conditioner



Required Reading

- ⌘Stallings chapter 16
- ⌘RFCs identified in text
- ⌘Comer, Internetworking with TCP/IP volume 1
- ⌘Comer, Computer networks and Internet
- ⌘谢希仁，计算机网络，**7.3**节

Problem

- ⌘16.3