Lecture 3
Integrated Service (IntServ)

Motivation of IntServ (1)

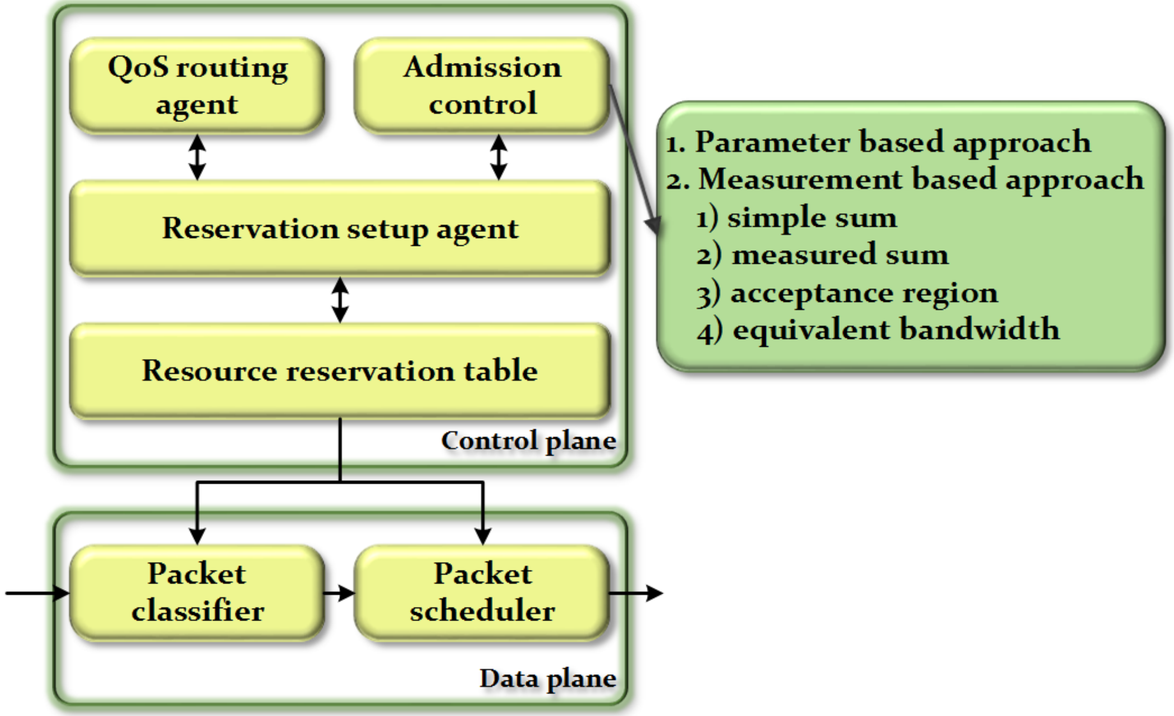
- IP network does not support QoS. Characteristics of IP network include:
  - Best effort service only
  - Fair access to all
  - FIFO queueing discipline
  - Window-based congestion control
- The goal of IntServ is to preserve the datagram model of IP-based networks and support resource reservation for real-time application
- The basic approach of IntServ is per-flow resource reservation

Assumptions of IntServ

- An implicit assumption of resource reservation is that the demand for bandwidth still exceeds supply
- The IntServ architecture assumes that the main quality of service about which the network makes commitments is the per-packet delay (in fact, the worst-case delay bound)

Key Components of IntServ

- The reference model defines two phases logically
  - Control phase: set up resource reservation
  - Data phase: forward data packets based on the reservation state
- InServ provides 3 traffic classes
  - Guaranteed service
  - Controlled load service
  - Best-effort



Integrated Service Reference Model

How to Setup Resource Reservation in IntServ

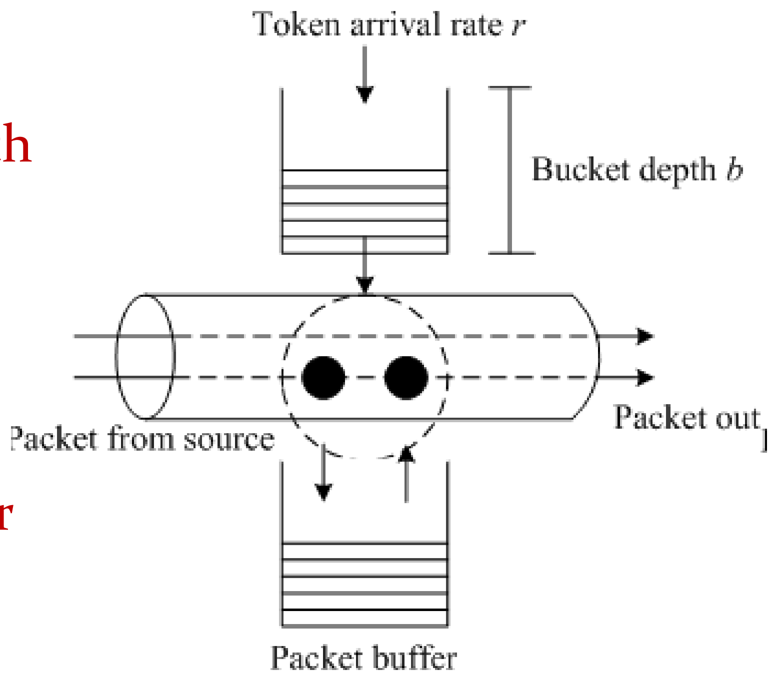
- Step 1: Flow specification
  - An application characterizes its traffic flow and specifies the QoS requirements
- Step 2: Route selection
  - The reservation setup request is sent to the network (routers)
  - Router interacts with routing module to determine the next hop to which the reservation request should be forwarded
  - The router also coordinate with the admission control to decide whether there are sufficient resources to meet the requested resources
  - Once the reservation setup is successful, the information for the reserved flow is installed into the resource reservation table
  - The information in the resource reservation is then used to configure the flow identification module and the packet scheduling module in the data plane

Flow Specification (1)

- Essentially a service contract
- Common parameters
  - Peak rate
    - The highest rate at which a source can generate traffic
  - Average rate
    - The average transmission rate over a time interval
  - Burst size
    - The maximum amount of data that can be injected into the network at the peak rate
- In Integrated Services, the admitted flow rate is implemented by means of token bucket

Flow Specification (2)

- A token bucket is with two parameters
  - Token arrival rate *r* and bucket depth *b*
  - Tokens drop into the bucket at a constant rate *r*
  - Tokens are consumed by the incoming packets
  - When a packet arrives, the regulator will send this packet only if the bucket has enough tokens



  - When a packet leaves the token bucket, the regulator removes the amount of tokens equal to the size of the outgoing packet
  - The depth *b* is the limit on the maximum amount of tokens that can be accumulated

Challenges of Route Selection

- Currently protocols typically use a simple metric, such as delay, hop count, or administrative weight to compute the shortest paths to all destination networks → lacks of necessary information about the available resources to make an intelligent decision
- This issue is further complicated by the fact that applications may have multiple requirements → NP complete problem
- Many interesting schemes have been proposed to resolve this problem, but none have been implemented in commercial products
- Therefore, IntServ <u>decouples routing from the reservation process</u>, and assumes the routing module in a router will supply the next hop

More About Reservation Setup

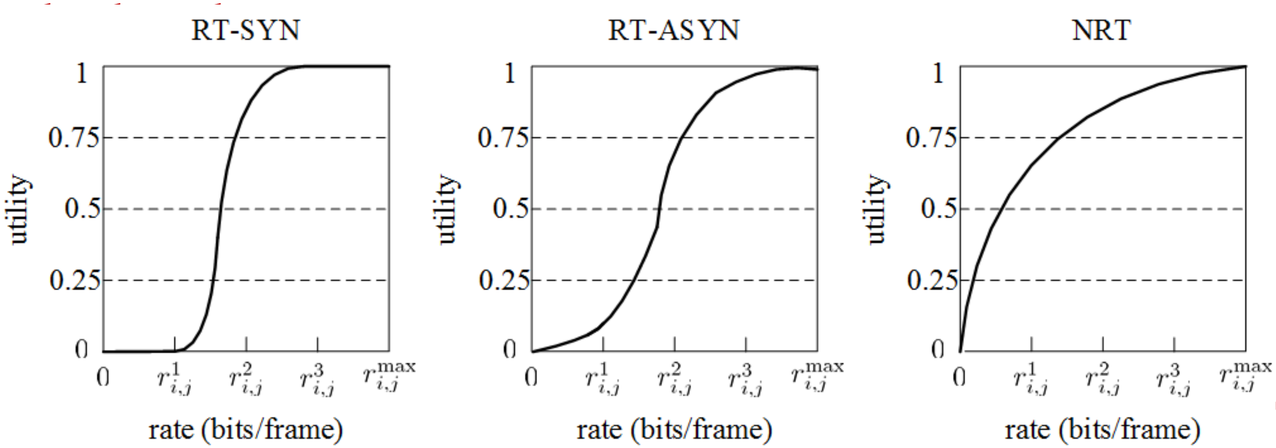
- We need a protocol to carry the information about traffic characterization and resource requirements so that each router along the path can determine to accept or deny the request
- We also need a reservation setup protocol that goes hop by hop along the path to install the reservation state in the routers
- The reservation setup protocol must deal with changes in the network topology
- In IntServ, RSVP (ReSerVation Protocol) is the adopted reservation setup protocol
- Features of RSVP
  - A receiver-initiated approach and is designed to work with IP multicast
  - Allows different reservation styles
  - Soft-state approach to deal with resource changes

Admission Control (1)

- Two functions of admission control
  - Determine if a new reservation can be setup based on the admission control policy
  - Monitor and measure the available resources
- Two basic approaches to admission control
  - Parameters based
    - A set of parameters is used to precisely characterize traffic flows
    - Difficult to give accurate, tight traffic models
  - Measurement based
    - The network measures the actual traffic load and uses that for admission control

Admission Control (2)

- Most common used methods
  - Simple sum: the sum of requested bandwidth for all current flows and the new flow does not exceed the link capacity. This is the most conservative approach
  - Measured sum: the measured sum approach uses the measured load of existing flows rather than the bandwidth requested by them
  - Acceptance region: maximize the reward of utilization increases against packet losses. Given the statistical models of traffic sources, the acceptance region for a type of traffic can be



Admission Control (3)

- Equivalent bandwidth: the equivalent bandwidth for a set of flows is defined as the bandwidth $C(p)$ such that the stationary bandwidth requirement of the set of flows exceeds this value with a probability of at most p

- How to measure the traffic load of existing flows?
  - Exponential averaging over consecutive measurement:

$$\text{New estimation} = (1 - w) \times \text{old estimation} + w \times \text{new measurement}$$

  - Time window: the average arrival rate is measured over a sampling interval. At the end of a measurement period the highest average is used as the estimated rate. Suppose there are n sample intervals in a measurement period T and C_i is the average rate measured over sampling interval i; then

$$\text{Estimated rate} = \max\left[c_1, c_2, .., c_n\right]$$

Flow Identification

- In packet processing a router must examine every incoming packet and decide if the packet belongs to one of the received RSVP flows
- The task is performed at packet classifier (inside a router actually)
- Each packet is identified by five fields (five-tuple) in the packet header
  - Source IP address
  - Destination IP address
  - Protocol ID
  - Source port
  - Destination port

Packet Scheduling

- Packet scheduling is responsible for enforcing resource allocation
- Common scheduling methods
  - Fair queueing: WFQ
  - Deadline-based: EDF
  - Rate-based: virtual clock (timestamp)

Service Models Defined in IntServ (1)

- Guarantee service
  - Provided guarantee bandwidth and strict bounds on end-to-end queueing delay (max) for conforming flows
  - The end-to-end behavior of a path that supports guaranteed service can be viewed as a virtual circuit (more flexible than real circuit) with guaranteed bandwidth
  - TSpec (traffic specification)
    - Bucket rate *r* (bytes/second): the rate at which tokens arrive at the token bucket
    - Peak rate *p* (bytes/second): the max rate at which packets can transmit
    - Bucket depth *b* (bytes): the size of the token bucket
    - Min policed unit *m* (bytes): any packet with a size smaller than *m* will be counted as *m* bytes
    - Max packet size *M* (bytes): the max packet size that can be accepted

Service Models Defined in IntServ (2)

- RSpec (service specification)
  - Service rate R (bytes/second): the bandwidth requirement
  - Slack term S (microseconds): the extra amount of delay that a node may add that still meets the end-to-end delay requirement
- Delay calculation
  - Best-case end-to-end queueing delay: b/R ($p \rightarrow \infty$, $R \geq r$)
  - Worst-case end-to-end queueing delay: ??? ($p > R \geq r$)
  - **Check the definition of b, p, R and r**

Service Models Defined in IntServ (3)

- Controlled-load service
  - Emulate a lightly loaded network for applications that request this service
  - More efficient than guarantee service by allowing statistical multiplexing
  - Well fit with adaptive applications that require some degree of performance assurance but not absolute bounds
  - A service model between the best effort and guaranteed service through appropriate admission control and traffic isolation mechanisms (better-than-best-effort service)
  - End-to-end behavior
    - Very high probability that transmitted packets are successfully delivered to the destination
    - End-to-end transit delay measured for a very high percentage of packets not greatly exceeding the **min** end-to-end transit delay

Service Models Defined in IntServ (4)

- The acceptance of a request for the controlled load service implies that the network has sufficient resources to accommodate the traffic without causing congestion
- TSpec is allowed to specify this desired traffic parameters
- New request can be accepted only when all nodes over the path have sufficient resources available to accommodate the new flow
- Admission control is left as a local matter
- Traffic policing
  - When non-conformant packets arrive, the network must continue to provide the contracted service guarantees to conforming flows
  - The network should prevent the excessive controlled load traffic from unfairly impacting the best-effort traffic
  - The network must attempt to deliver the excessive traffic when it can do so without violating the previous two requirement

RSVP (1)

- ReSerVation Protocol [RFC 2205]
- RSVP is used
  - by hosts to communicate service requirement to the network, and
  - by routers in the network to establish a reservation state along a path
- RSVP is not
  - a routing protocol
  - providing network services
  - the only IP reservation protocol

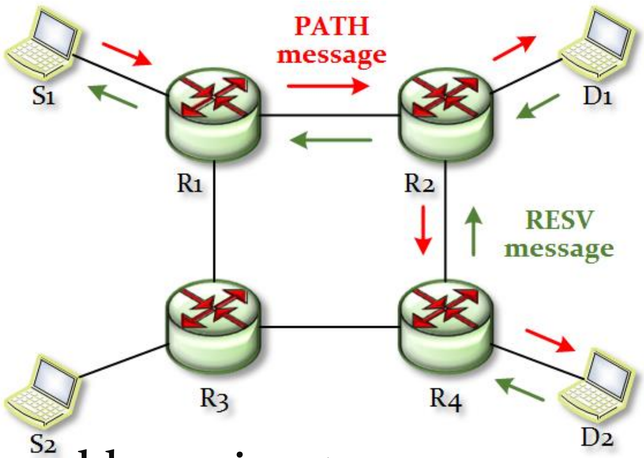
RSVP (2)

- Two key concepts: flow and reservation
  - Flow
    - Traffic streams from a sender to one or more receivers
    - Flowspec describes both the characteristics of the traffic stream sent by the source and the QoS the flow requires
    - Workable for unicast and multicast flows

RSVP (3)

- Reservation
  - Simplex reservation: RSVP makes a reservation in only one direction
  - Receiver-initiated: to accommodate heterogeneous and dynamic receivers in a multicast group
  - Routing independent: RSVP process consults the forwarding table (constructed by the routing protocol) and sends the RSVP messages accordingly
  - Policy independent: RSVP provides a general mechanism for creating and maintaining a reservation state over a multicast tree or a unicast path; the control parameters carried in the RSVP message are opaque from the RSVP perspective
  - Soft state: to support dynamic membership and network topology changes by setting a timer for each reservation state
  - Reservation style: to specify how reservation for the same multicast group should be aggregated at the immediate routers

RSVP (4)



- Two major messages of RSVP
  - Path message
    - Generated by the sender
    - Travel in the routing direction, determined by unicast or multicast routing protocols
    - Purposes
      - Distributing source info to receivers
      - Passing on path's characteristics
      - Providing info for receivers to reach senders
    - Path spec
      - Flow ID (sender Template): sender IP address and optionally the UDP/TCP sender port
      - Flowspec (i.e., sender TSpec, not modified by the intermediate nodes)

RSVP (5)

    - Adspec: optional element in the PATH messages that is used to carry the OPWA (One Pass with Advertising). Adspec is passed to the local traffic control at each node, which returns an updated Adspec and then is forwarded in PATH messages sent downstream
    - PHop (Previous Hop)
- Reservation message
  - Generated by the receiver
  - Travel in the reverse direction of Path message
  - RESV spec
    - Flowspec: specifies the desired QoS and parameters to be used in packet scheduling
    - Reservation style
    - NHop
  - Merging reservation requests whenever possible

RSVP (6)

- RSVP operation
  - If Path state and Resv state exist,
    - Path messages and Resv messages are sent by sender and receiver, respectively, periodically to refresh existing states
  - Otherwise,
    - Sender sends a Path message first to set up the Path state in the routers, and then receivers send Resv requests in the reverse direction
  - Admission control and policy control
  - RSVP is an one pass reservation model
  - An enhancement is one pass with advertising (OPWA)

RSVP (7)

- Packet filtering
  - A separate function from resource reservation
  - A function which selects those packets that can use the reserved resources
- Reservation style: how multiple requests are merged and which resource requests are forwarded to the upstream node
  - Fixed filter (FF)
    - Explicit sender selection + distinction reservation
    - E.g., FF($S_1\{Q_1\}$, $S_2\{Q_2\}$,.., $S_n\{Q_n\}$), where Q_1, Q_2, .., Q_n are corresponding flowspecs, and S_1, S_2, .., S_n are senders

RSVP (8)

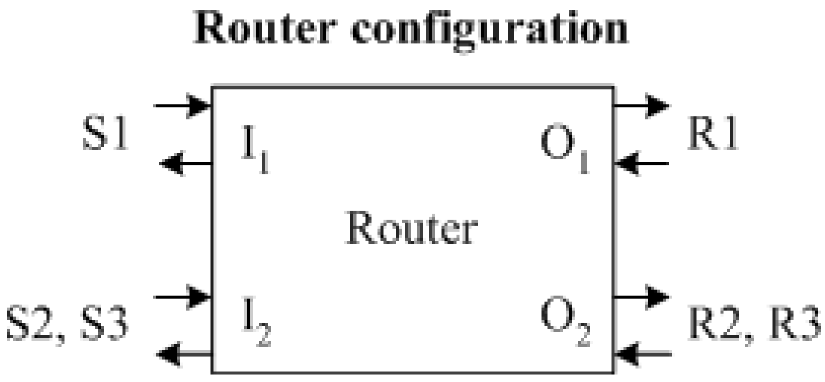
- Wildcard filter (WF)
  - Wildcard sender selection + shared reservation
  - All receivers share a single reservation whose size is the largest of the resource requests
  - All upstream senders can use the reservation
  - E.g., WF(*, {Q}), where Q is the flowspec, * represents the wild-card sender selection
- Shared explicit (SE)
  - Explicit sender selection + shared reservation
  - E.g., SE((S_1, S_2, S_4{Q}), where Q is the flowspec, and S_1, S_2, S_3, S_4 are senders
- Shared reservation (WF and SE) are designed for multicast applications
  - Typically only one or two people can speak at the same time in audio conferencing. An WF or SE reservation request for twice the bandwidth for one sender should be sufficient in such cases

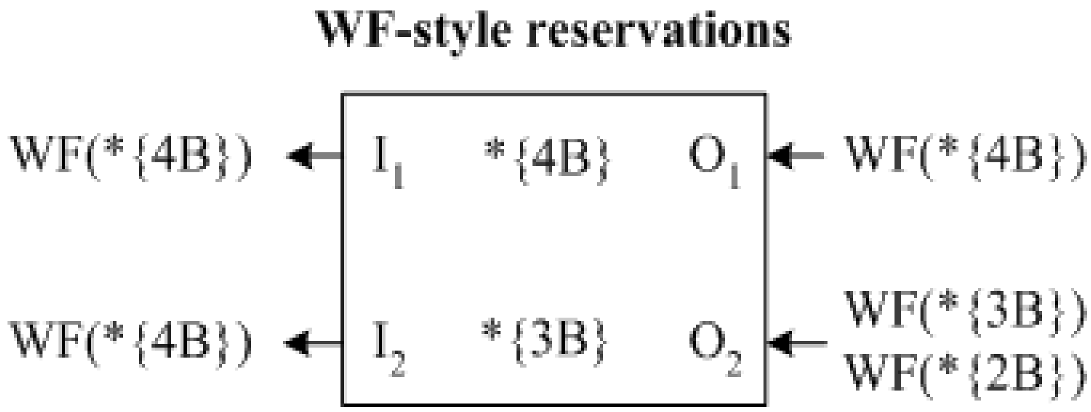
RSVP (9)

- Explicit vs. wildcard server selection
  - Explicit: the reservation made only for senders explicitly listed in the RESV message
  - Wildcard: for any sender
- Distinct vs. shared reservation
  - Distinct: each sender has its own RSVP flow
  - Shared: using the same RSVP flow for multiple senders

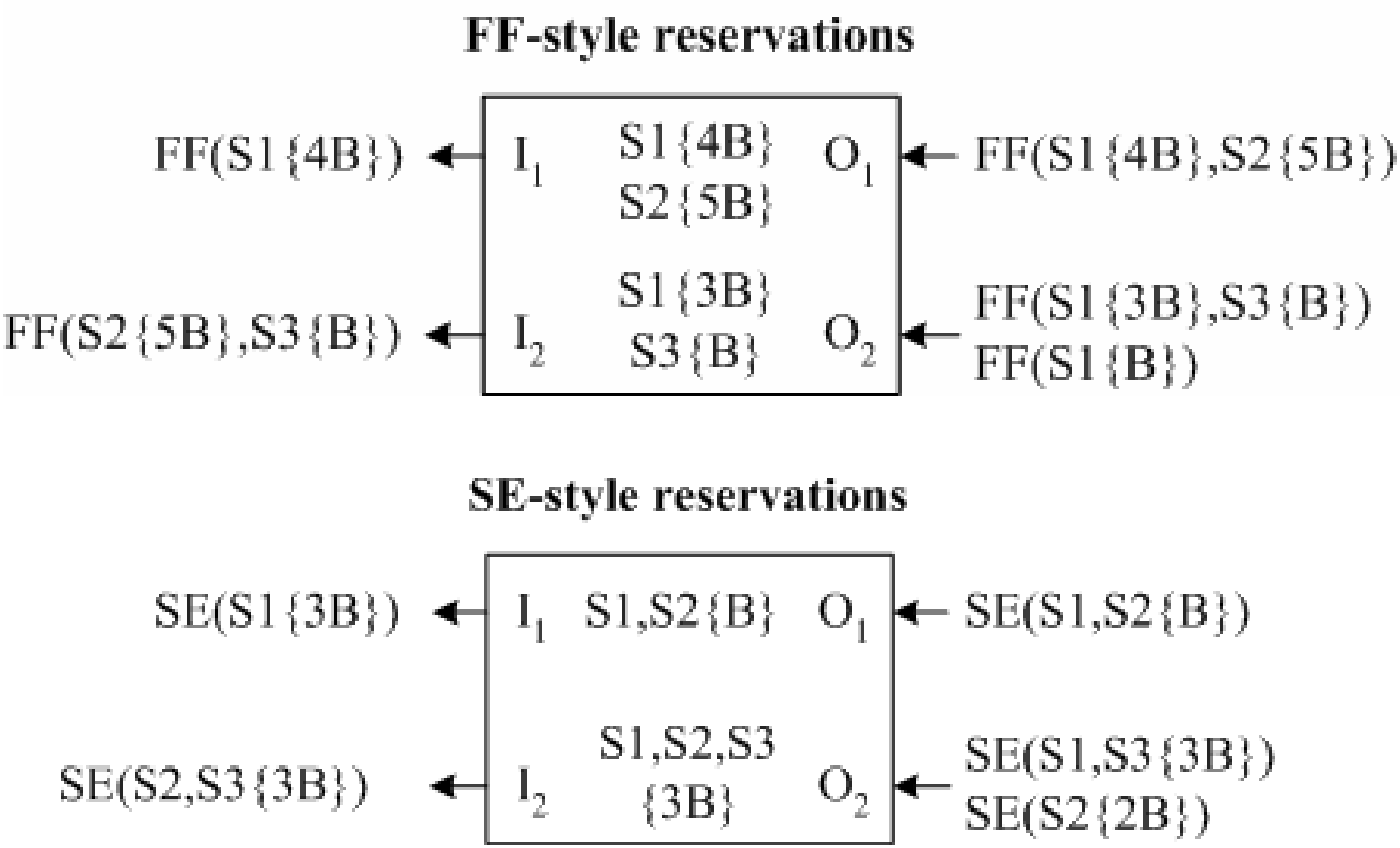
RSVP (10)

- An illustrative example
  - A router with two incoming interfaces I_1 and I_2, and two outgoing interfaces O_1 and O_2
  - There are 3 senders S_1, S_2, and S_3, and 3 receivers R_1, R_2, and R_3
  - Let B be the basic unit of resources to be reserved





RSVP (11)


FF-style reservations
FF(S1{4B})
I1
S1{4B}
S2{5B}
O1
FF(S1{4B},S2{5B})
FF(S2{5B},S3{B})
I2
S1{3B}
S3{B}
O2
FF(S1{3B},S3{B})
FF(S1{B})
SE-style reservations
SE(S1{3B})
I1
S1,S2{B}
O1
SE(S1,S2{B})
SE(S2,S3{3B})
I2
S1,S2,S3
{3B}
O2
SE(S1,S3{3B})
SE(S2{2B})

Disadvantages of IntServ

- When IP flow is terminated, the resources are not immediately released
- Lack of scalability
- Refreshing signaling is bandwidth consuming

More About Adspec (1)

- This object is modified by an intermediate node, pass information to the next hop, and finally the receiver can determine the characteristics of the end-to-end path
- Adspec object has 3 components
  - Default general parameters fragment, which contains
    - Minimum path latency
    - Path bandwidth
    - Integrated Services hop count: the total umber of hops that are capable of support the Integrated Services
    - Global break bit: 0 set by the sender. If any node along the path does not support Integrated Services, the bit is set to 1 and the information is passed on to the receiver
    - Path MTU: the maximum transmission unit of the path

More About Adspec (2)

- Guaranteed service fragment, which contains
  - Ctot: the sum of rate-dependent delay (rate & packet length) over a path
  - Dtot: the sum of rate-independent delay (e.g., router pipeline delay) over a path
  - Csum: partial sum of C between shaping points
  - Dsum: partial sum of D between shaping points
- Controlled load service fragment: no extra data in the Adspec

One Pass v.s. One Pass with Advertising (OPWA)

- One pass model does not support the capability of determining the characteristics of the path or whether the path is able to support the desired QoS
- With OPWA, the sender includes an Adspec in its PATH message to collect the information about the path

Slack Term

- For guaranteed service, there is a “Slack term” field in the RSpec in an RSVP message
- Assume a flow needs 2.5 Mbps bandwidth
- Slack term is utilized to compensate the increased delay
- Slack term is the difference between the desired delay and the actual delay obtained with current bandwidth reservation (i.e., delay credit)