Lecture 4
Differentiated Service (DiffServ)

Comparisons of IntServ and DiffServ (1)

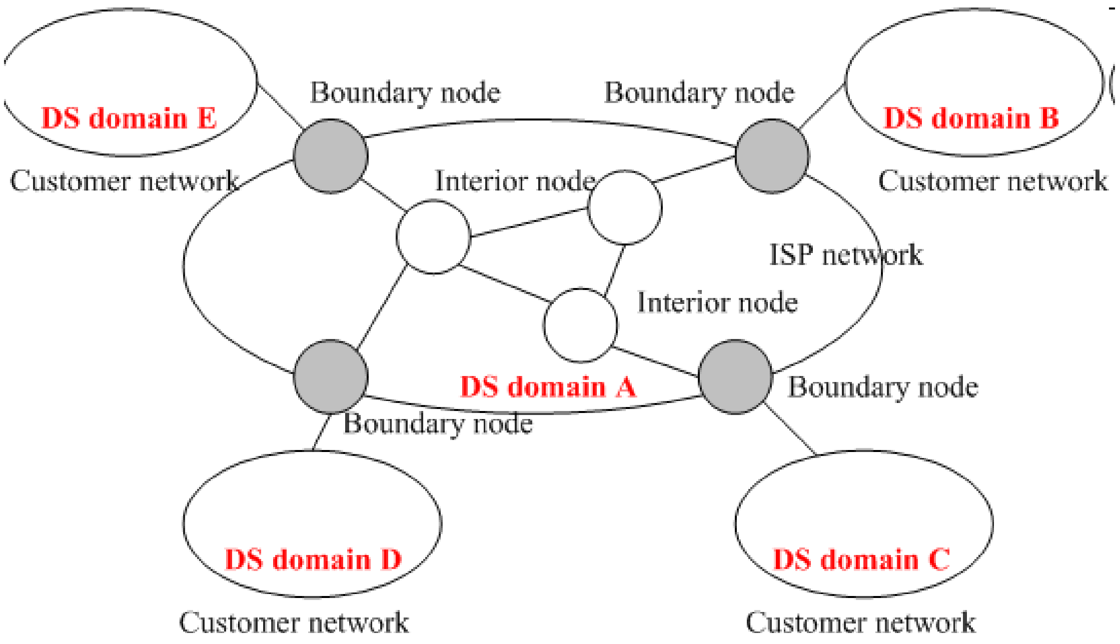
- Resource allocation to aggregated traffic (class) rather than individual flows
  - In DiffServ, performance assurance to individual flows in a forwarding class is provided through prioritization and provisioning rather than per-flow reservation
  - In IntServ, resources allocated to individual flows
- Traffic policing on the edge and class-based forwarding in the core
  - In DiffServ, only boundary nodes at the edge of the network classify traffic and mark packets; the interior nodes use the forwarding classes encoded in the packet header to determine the treatment accordingly
  - In IntServ, all nodes perform packet classification and scheduling
- Define forwarding behaviors, not services
  - DiffServ defines forwarding treatments (i.e., forwarding classes), not end-to-end services
  - IntServ defines services. The treatment of packets is not part of the standard

Comparisons of IntServ and DiffServ (2)

- Guarantee by provisioning rather than reservation
  - In DiffServ, resource assurance is through provisioning and prioritization (achieve different levels of services)
  - In IntServ, it's through per-flow resource reservation (achieve absolute guarantee)
- Emphasis on service level agreements rather than dynamic signaling
  - Purpose of DiffServ is to ensure the long-term SLA between customers and service providers
  - IntServ provides dynamic resource reservation, instead
- Focus on a single domain vs. end to end
  - The deployment of DiffServ in the Internet can be incremental
  - The IntServ model is inherently end-to-end

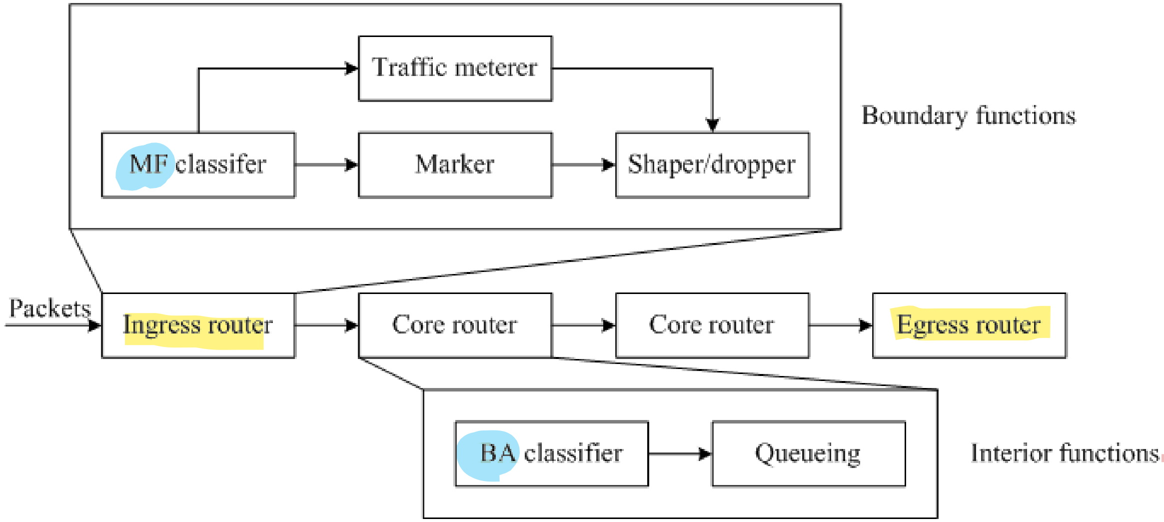
Conceptual Operations

- When traffic entering a network,
  – classified first;
  – possibly conditioned;
  – assigned a DSCP (class);
  – forwarded



Basic Approach of DiffServ Framework

- With Differentiated services (DS) traffic is divided into a small number of groups called forwarding classes
- The forwarding class is encoded in the IP packet header
- Each forwarding class represent a predefined forwarding treatment in terms of drop priority and bandwidth allocation
- Boundary node (edge node) and interior node (core node) are with different responsibilities
  - Boundary router (egress or ingress router) performs packet classification and traffic conditioning
  - Interior router (core router) performs class-based packet forwarding



Per-Hop Behaviors (PHBs)

- PHB is a description of the externally observable forwarding behavior at a DS node
- Each PHB is represented by a 6-bit value, called a Differentiated Services codepoint (DSCP)
- All packets with the same codepoint are referred to as a behavior aggregate (BA), and they receive the same forwarding treatment
- Examples of PHBs
  - Guarantee a minimal bandwidth allocation of x% of a link to a BA (absolute term)
  - Guarantee a minimal bandwidth allocation of x% of a link to a BA, with proportional fair sharing of any excess link capacity (relative term)

Services (1)

- Services are visible to customers, whereas PHBs are hidden inside the network elements
- In DiffServ, services are defined in the form of a service level agreement (SLA) between a customer and its service provider
- One of the important elements in an SLA is the traffic conditioning agreement (TCA)
- TCA details the service parameters for traffic profile and policing actions
  - Traffic profiles, such as token bucket parameters for each of the classes
  - Performance metrics, such as throughput, delay, and drop priorities
  - Actions for non-conformant packets
  - Additional marking and shaping services provided by the service provider

Services (2)

- In additional to TCA, SLA may also contain other service characteristics and business-related agreements such as availability, security, monitoring, accounting, pricing and billing
- SLA may be static or dynamic
- Services can be defined in either quantitative or qualitative terms
  - Quantitative: specify the parameters in absolute terms, e.g., 5-sec max delay
  - Qualitative: use relative terms, e.g., lower delay

Differentiated Services Field (1)

- DiffServ uses 6 bits in the IP packet header to encode the forwarding treatment
- Current IP packet header includes an 8-bit field called the IP TOS field: 3-bit precedence (represents the priorities for the traffic), a 3-bit type of service (TOS), and 2 unused bits
- DiffServ redefines the existing TOS field to indicate the forwarding treatment
  - The first 6 bits of the DS field are used as a DSCP to encode the PHB for a packet at each DS node
  - The remaining 2 bits are reserved for future use
  - The DSCP should be treated as an index, and the mapping of DSCPs to PHBs must be configurable

Differentiated Services Field (2)

- The codepoint space is divided into 3 pools
  - A pool of 32 recommended codepoints to be standardized
  - A pool of 16 codepoints to be reserved for experimental or local use
  - A pool of 16 codepoints to be available for experimental and local use but may be subject to standardization if pool 1 is ever exhausted

IP TOS field

Precedence	D	T	R	0	0

Bit	Description
0-2	Precedence
3	0=Normal delay 1=Low delay
4	0=Normal throughput 1=High throughput
5	0=Normal reliability 1=High reliability
6-7	Reserved for future use

DS field

DSCP	CU

POOL	Codepoint space	Assignment policy
1	xxxxx0	Standard action
2	xxxx11	Experimental and local use
3	xxxx01	Experimental and local use but may be subject to standards action

Differentiated Services Field (3)

- Default PHB codepoint
  - Backward compatible with the current best effort forwarding treatment
  - Packets belonging to the default forwarding class may be sent into a network without any policing and the network will try to deliver as many of these packets as possible and as soon as possible
  - Other forwarding classes have higher priorities to network resources than the default forwarding class
  - Some minimal bandwidth is reserved for the default forwarding class to avoid starvation
  - The assigned codepint is 000000

Differentiated Services Field (4)

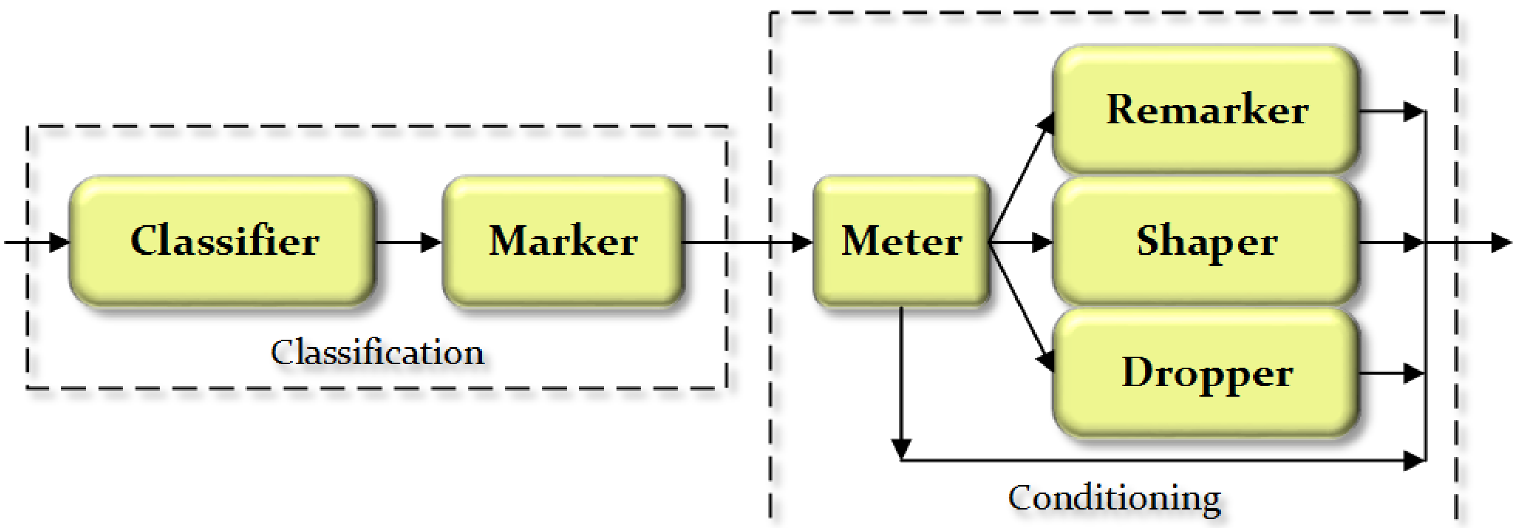
- Class selector codepoints
  - Maintain partially backward compatible with known current use of the IP precedence field
  - We name this set of PHBs class selector PHBs
  - Assigned codepoints: xxx000
  - The eight class selector PHBs must yield at least two different forwarding classes
  - The PHB mapped by a codepoint with the larger numerical value must receive better or equal forwarding treatment than the one with a lower numerical value

Differentiated Services Field (5)

- Current codepoint allocation (been standardized by IETF)
  - Assured forwarding (AF) (RFC 2597 and RFC 3260): assurance of delivery under conditions
  - Expedited forwarding (EF) (RFC 3246): dedicated to low-loss, low-latency traffic

Traffic Classification and Conditioning (1)

- Two responsibilities of boundary routers are classification and conditioning
- Classification module contains a classifier and a marker
- Classifier divides incoming packets streams into multiple groups based on some predefined rules



Traffic Classification and Conditioning (2)

- There are two basic types of classifiers: behavior aggregate (BA) or multifield (MF)
- BA is the simplest DS classifier and it selects packets based solely on the DSCP values
- MF uses a combination of one or more fields of the five-tuple (source address, destination address, source port, destination port, and protocol ID) in header for classification

Traffic Classification and Conditioning (3)

- Traffic conditioner is to enforce the TCA between customer and service provider
- Traffic conditioner consists of 4 basic elements: meter, marker, shaper, and dropper
- Meter
  - A meter measures the traffic flow from a customer against its traffic profile
  - In-profile packets are allowed to enter the network, while out-of-profile packets are conditioned based on the TCS
  - Most meters are implemented as token buckets
- Marker
  - Markers set the DS field of a packet to a particular DSCP, and add the marked packets to the forwarding class
  - Markers may act on unmarked packets or remark marked packets

Traffic Classification and Conditioning (4)

- Marking could be done by the applications itself, the first-hop router of the LAN, or the boundary router of a service provider
- Remarking is necessary at the boundary of two administrative domains using different DSCPs
- When a packet remarked with DSCP receives worse forwarding treatment → PHB demotion (typically)
- When receiving better forwarding treatment → PHB promotion
- Typically boundary routers demote out-of-profile packets to a DSCP with worse forwarding treatment

- Shaper
  - Shaper delays the non-conformant packets in order to bring the stream into compliance with the agree-on traffic profile
  - A marker marks non-conformant packets and lets them into the network, whereas a shaper prevents these packets from entering the network till the stream conforms to the traffic profile
  - A much stronger form of policing than marking

Traffic Classification and Conditioning (5)

- Dropper
  - Dropping is another action may be applied to out-of-profile packets
  - For a shaper, it buffers packets temporarily, while for a dropper, it just drops out-of-profile packets
  - Dropping is much easier to implement
- Location of traffic classifier and conditioner
  - Usually situated with DS ingress and egress nodes where traffic goes across domains
  - Within a source domain
    - Traffic source and intermediate nodes within the source domain may perform classification and marking before packets leave the domain by the host or the 1st hop router. This is referred to as premarking
    - Premarking allows the source domain to classify packets based on local policies (e.g., CEO's PC, or mission-critical servers)

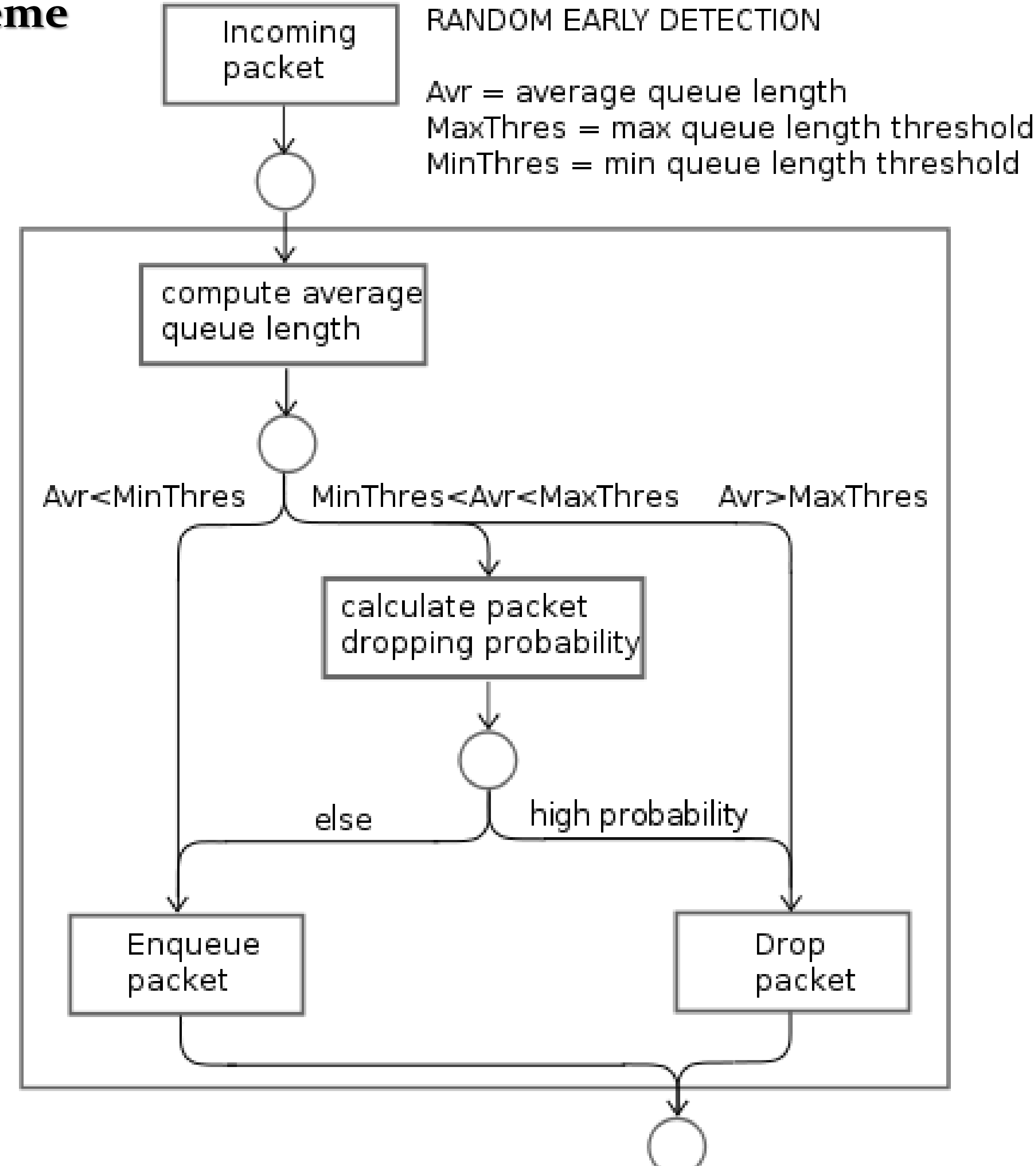
Traffic Classification and Conditioning (6)

- At the boundary of a DS domain
  - The ingress node of the downstream domain performs all necessary classification and conditioning
  - If two domains use different codepoints for a PHB, any premarked packets must be remarked at the boundary, either the egress node of the upstream domain or the ingress node of the downstream domain may perform the mapping
- In interior DS nodes
  - At some heavily congested hot spots, additional traffic policing may be applied or the traffic may be shaped at some points to ensure arrival patterns

Assured Forwarding (1)

- Basic idea behind AF came from the RED with In and Out (RIO) scheme
- What's RIO?
  - A service profile specifies the expected capacity for the user
  - Boundary nodes monitor the traffic flows and tag the packets as being in or out of their profiles
  - During congestion the packets tagged as out will be dropped first
  - The service provider should provide their networks to meet the expected capacities for all in-profile packets and allow out-of-profile packets only when excessive bandwidth is available

RIO Scheme


Incoming packet
RANDOM EARLY DETECTION
Avr = average queue length
MaxThres = max queue length threshold
MinThres = min queue length threshold
compute average queue length
Avr<MinThres
MinThres<Avr<MaxThres
Avr>MaxThres
calculate packet dropping probability
else
high probability
Enqueue packet
Drop packet

Assured Forwarding (2)

- AF standard extended the basic in or out marking in RIO into a structure of four forwarding classes, and within each forwarding class, three drop precedences
- Each forwarding class is allocated a minimum amount of buffers and bandwidth
- When backlogged packets from an AF forwarding class exceed a specified threshold, packets with the highest priority are dropped first and then packets with the lower drop priority
- Note that comparing two drop priorities in two AF classes may not be meaningful

Assured Forwarding (3)

- An example of AF implementation
  - Assume link capacity be 100 Mbps
  - Minimum bandwidth requirements of 4 forwarding classes are 2, 4, 8, and 16 Mbps
  - Proportional sharing the excessive bandwidth
  - Weight: (1:2:4:8)
  - Excessive bandwidth shares are 4 Mbps, 8 Mbps, 16 Mbps, and 32 Mbps

AF PHB codepoints

	Class 1	Class 2	Class 3	Class 4
Lower drop precedence	001010	010010	011010	100010
Medium drop precedence	001100	010100	011100	100100
High drop precedence	001110	010110	011110	100110

Expedited Forwarding

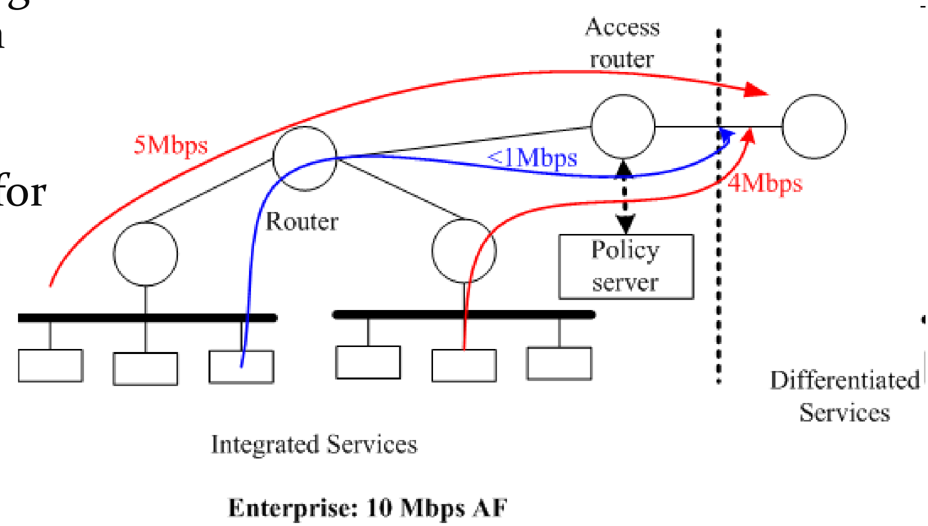
- EF is to create low loss, low delay, and assured bandwidth services
- EF is defined as a forwarding treatment for a traffic aggregate that the departure rate of the aggregate's packet from any DS node must equal or exceed a configurable rate
- EF traffic should receive this rate independent of the intensity of any other traffic attempting to transit the node
- EF traffic can preempt other traffic within a certain limit
- Assigned codepoint is 101110

End-to-End Resource Management (1)

- Integrated services over differentiated services
  - Two-tier resource allocation mode: DiffServ distributes aggregated resources in the core networks to customer networks; IntServ further allocates the resources at finer granularity to individual users or flows
  - At the boundary of the network the IntServ requests are mapped onto the underlying DiffServ network, included mappings are
    - Selecting an appropriate PHB or a set of PHBs for the requested services
    - Performing appropriate policing at the edge of the DiffServ network
    - Exporting IntServ parameters from the DiffServ network

End-to-End Resource Management (2)

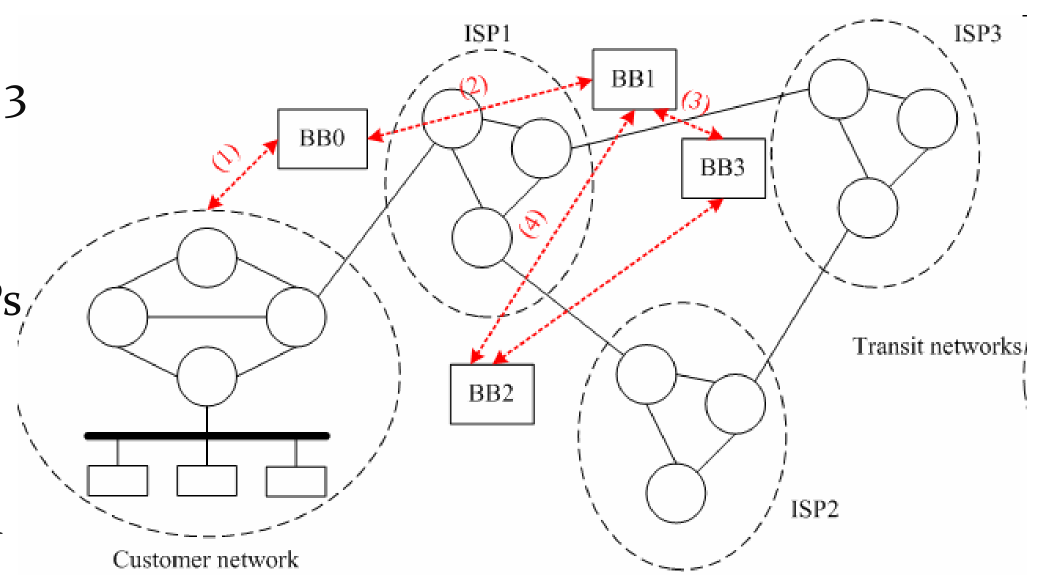
- An illustrative example of IntServ over DiffServ
  - An enterprise network with 4 routers and 2 LANs
  - The access router connects the enterprise network to its ISP and is responsible for traffic classification
  - When a sender initiates a transaction, it exchanges RSVP PATH and RESV messages with the receiver
  - The RSVP messages are ignored through the ISP that supports only DiffServ
  - When RESV messages pass through the access router, it consults with the policy server to decide about admission control
  - Suppose the enterprise has paid for 10 Mbps service and 9 Mbps has been allocated, less than 1 Mbps new reservation will be admitted



- Policy server can decide how IntServ requests should be mapped to DiffServ model

End-to-End Resource Management (3)

- Interdomain bandwidth allocation
  - Different domains are usually separately owned and administrated, and the relationship between any two different domains is governed by the SLA
  - The SLA could be dynamically and infrequently changed, and bandwidth broker (BB) approach can achieve dynamic interdomain resource negotiation
  - An illustrative example
    - The backbone consists of 3 ISPs and an enterprise network that can communicate with many receivers in any of the ISPs
    - The BB of the enterprise network performs admission control for reservation requests from its own users



End-to-End Resource Management (4)

- In essence, BBs act as resource management agents for their ISPs and perform the following 3 main functions
  - Admission control: decisions are based on the predefined provisioning algorithms for the networks
  - Policy control: handle the administrative (e.g., priority and special restrictions) and pricing policies
  - Reservation aggregation: collect multiple requests from users and make a single request for resources to improve the system scalability
- The drawback of BB approach is that BBs must know where the new traffic is going in order to make accurate bandwidth requests