Internet QoS: IntServ and DiffServ

Dr. Ying-Dar Lin
High Speed Network Lab.
Department of Computer Information Science
National Chiao Tung University
May 15, 1999

Content

- IETF Working Groups on QoS
- Per-Flow IntServ
- Per-Class DiffServ
- End-to-End Service Architecture I : DiffServ Customer and ISP
- End-to-End Service Architecture II : IntServ Customer and DiffServ ISP
- QoS Traffic Control for IntServ and DiffServ

IETF Working Groups on QoS

- Integrated Services (IntServ) Working Group

  per-flow state and processing
- RSVP Working Group

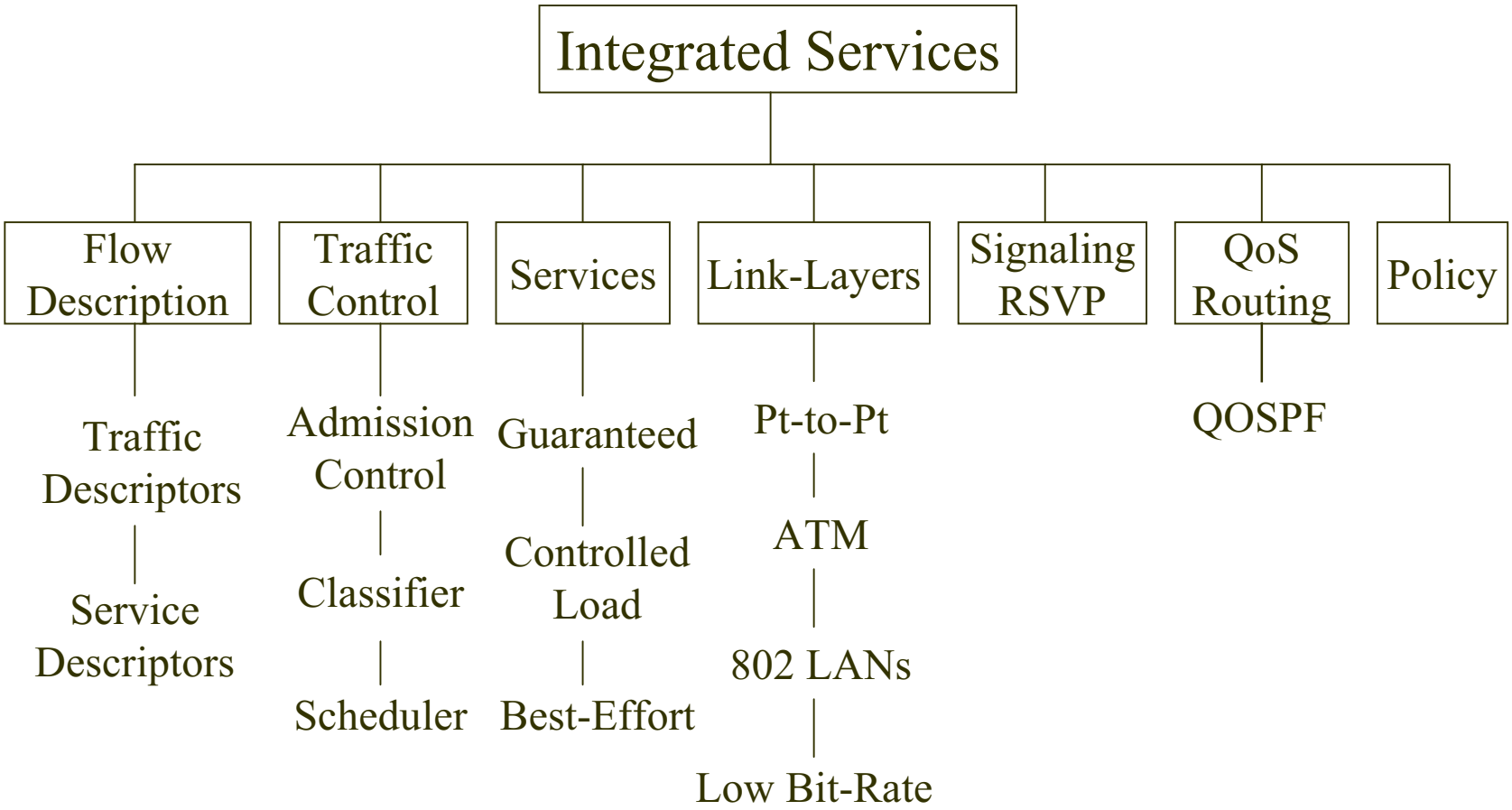
  signaling for resource reservation
- Differentiated Services (DiffServ) Working Group

  per-class state and processing

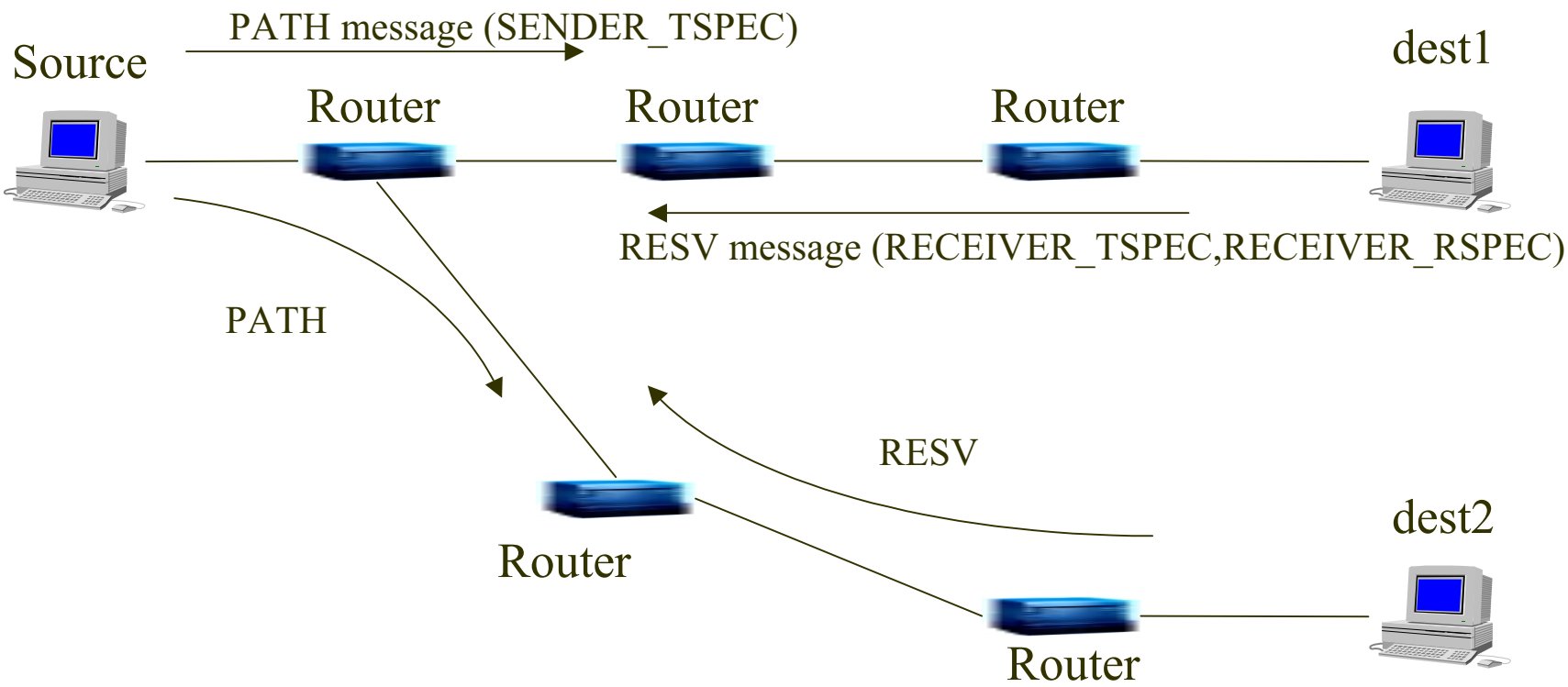
IntServ vs DiffServ

	IntServ	DiffServ	Best-Effort
Granularity of QoS	Per-flow	Per-class	None (fair to all)
Services	Guaranteed (quantitative) Controlled Load (qualitative)	Expedited Forwarding (quantitative) Assured Forwarding (qualitative)	"best-effort"
Resource Allocation	Dynamic	Static or Dynamic	None
Signaling	RSVP (for host/router)	RSVP (for host) RSVP/LDAP/COPS (for bandwidth broker) None for core router	None
Classification	Multi-field at host/router	Multi-field at the edge, DS-field at the core	None
Control	At host/router	Marking at the edge, queue mngt at the core	FIFO only
Complexity	High	Medium	Low

IntServ Elements



RSVP Signaling
(Resource Reservation Protocol)

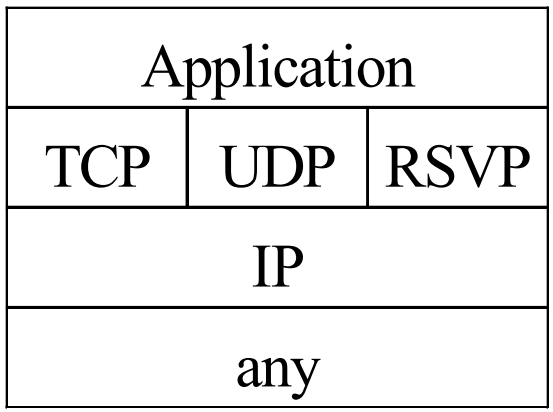


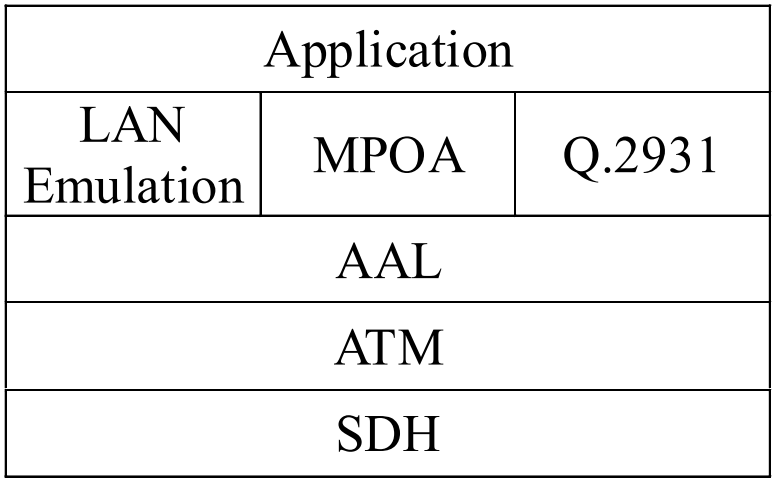
Characteristics of RSVP

- Source characterization using leaky bucket
  - Soft-state resource reservation from reverse direction
  - Separate routing and admission control (hop by hop)
- RSVP issues:
  - Per-flow state overhead and poor scalability
  - Not suitable for short-lived sessions
  - QoS routing support preferred

RSVP vs ATM Q.2931 (1/2)

- **Model:**





RSVP vs ATM Q.2931 (2/2)

- **Features**

RSVP	ATM Q.2931
Connectionless	Connection-oriented
Flow	Call
Soft state reservation	Hard state reservation
Receiver-initiated reservation	Sender-initiated reservation
Receiver-heterogeneity allowed (heterogeneous multicast possible)	Receiver heterogeneity not allowed (homogeneous multicast only)
No QoS routing so far (can work with RIP/OSPF, emerging QOSPF)	QoS PNNI/Phase 1 routing
Separate routing and admission control	Concurrent routing and admission control

- **Services**

RSVP	ATM Q.2931
Best-effort Guaranteed Controlled Load	**Constant Bit Rate Variable Bit Rate Unspecified Bit Rate Available Bit Rate**

Flow Description

Filterspec	Flowspec	
	Tspec	Rspec
•Identify which flow a packet belongs to (for classifier)	• Leaky Bucket para： p: peak rate r: token rate b: bucket size m: minimum policed packet size M: maximum packet size (for scheduler and policer)	• QoS para： R: bandwidth s : slack term (for admission control and scheduler)

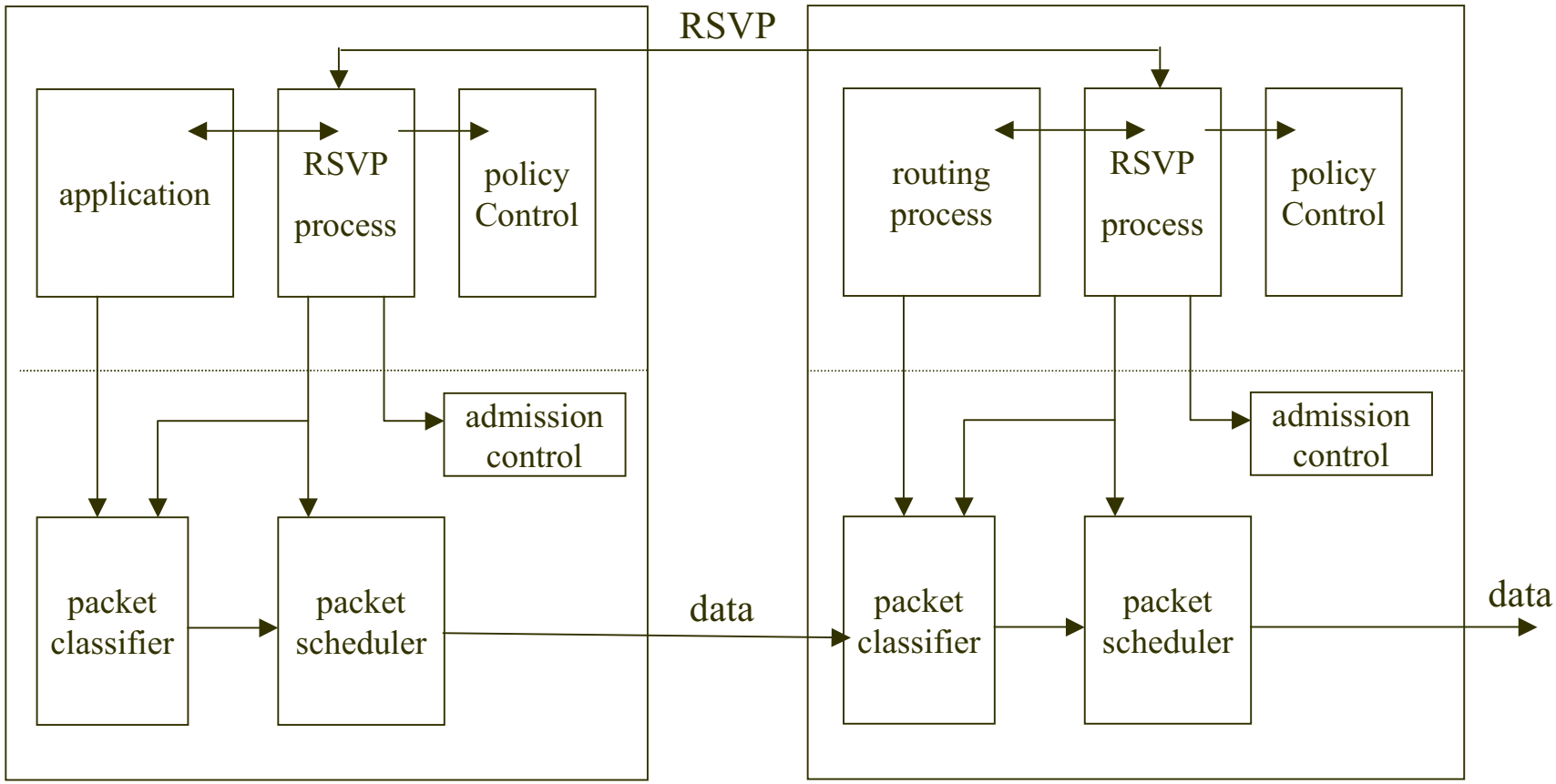
IntServ Services

	Guaranteed	Controlled Load	Best Effort
Parameters	Tspec Rspec	Tspec No Rspec	None
QoS	Delay bound No loss	Best-effort over uncongested network	None
Traffic control	Admission control Policing Scheduling	Admission control Policing Scheduling	FIFO scheduling
RFC	RFC 2212	RFC 2211	Many others

RSVP Traffic Control

- Admission control (Rspec, RSVP node state)
- Policing (Tspec)
- Classification (Tspec)
- Scheduling (Tspec, Rspec)
- QoS routing (Tspec, RSVP node state) (optional)
- Policy control (Tspec, Rspec, policy database) (optional)

RSVP-capable Host/Router



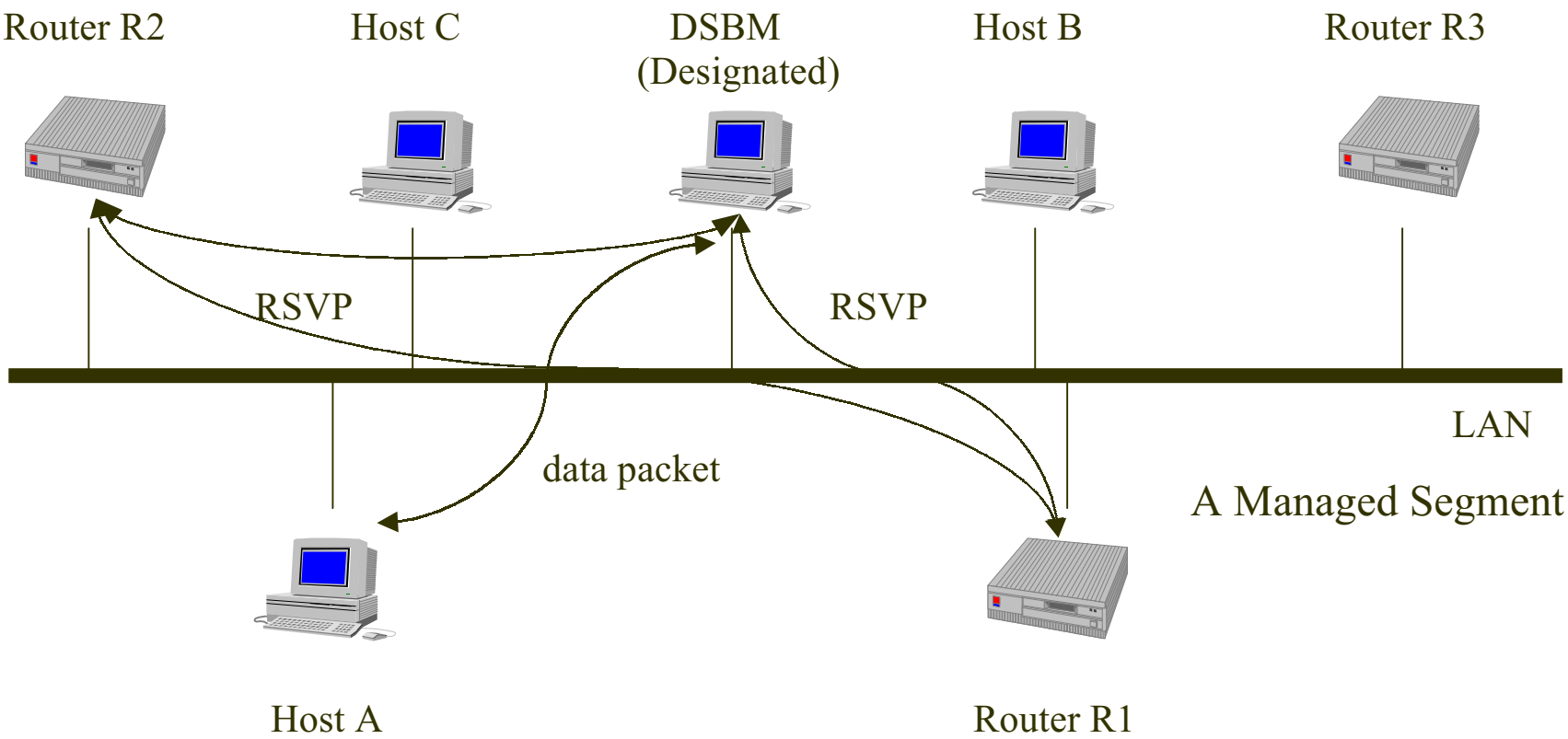
RSVP over Various Data Links

- RSVP over pt-to-pt (no extra problem)
- RSVP over ATM (parameter mapping, receiver heterogeneity, and VC management)
- RSVP over IEEE 802 LANs (SBM takes care of admission control and scheduling) (IEEE 802.1P VLAN Tag may be used)
- RSVP over low-bit rate links (Multi-class Multi-link PPP can fragment and queue packets)

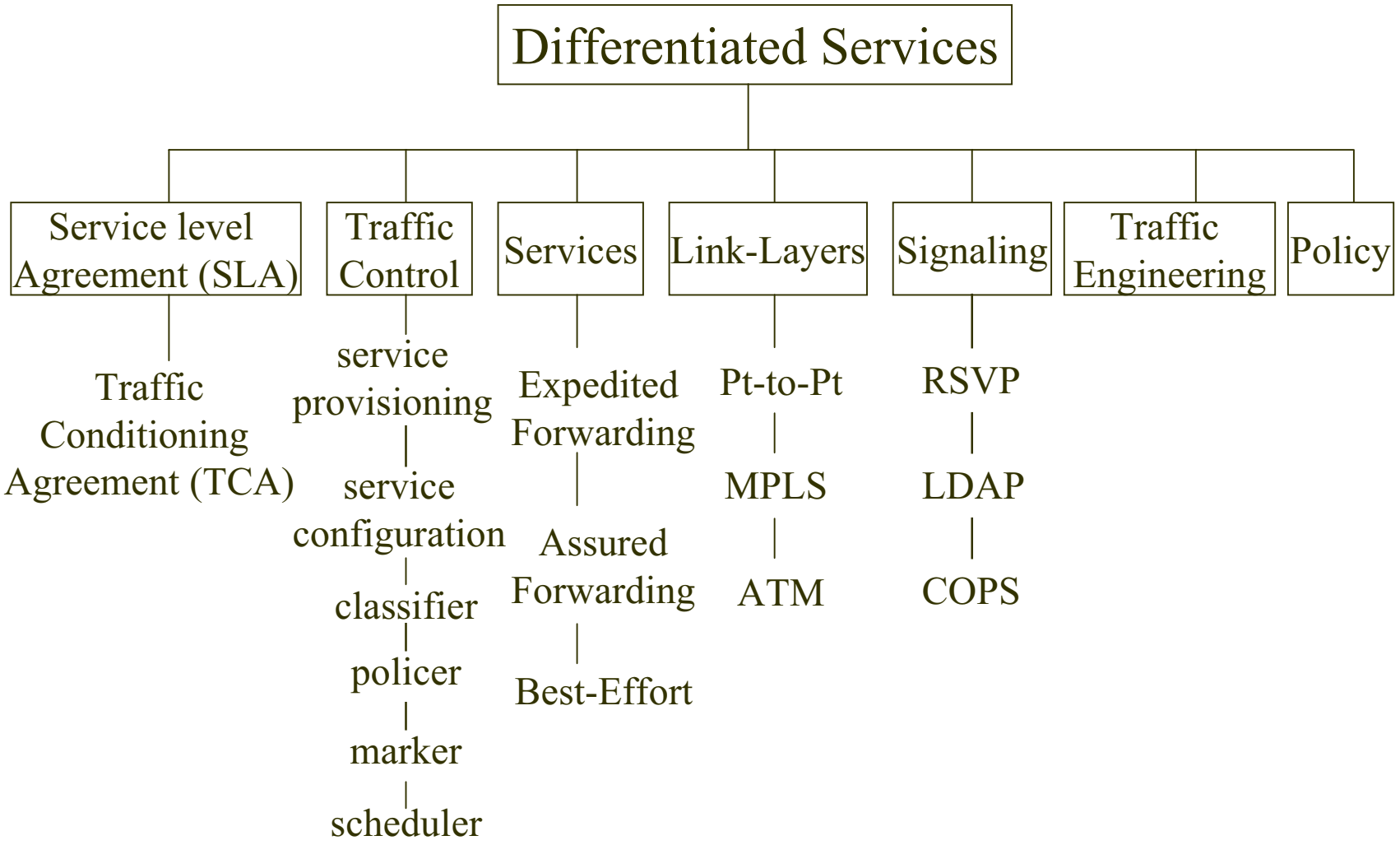
RSVP over ATM

- RSVP “rides on” ATM’s QoS
- RSVP RESV messages establish ATM switched path
  - ingress router establishes ATM SVC (upon RESV)
  - or “concatenated” SVCs to form a path
- Mapping from RSVP parameters to ATM Q.2931 parameters exists
- No heterogeneous multicast allowed
- VC management problems (soft/hard-state)

SBM (Subnet Bandwidth Manager): RSVP over IEEE 802 LANs



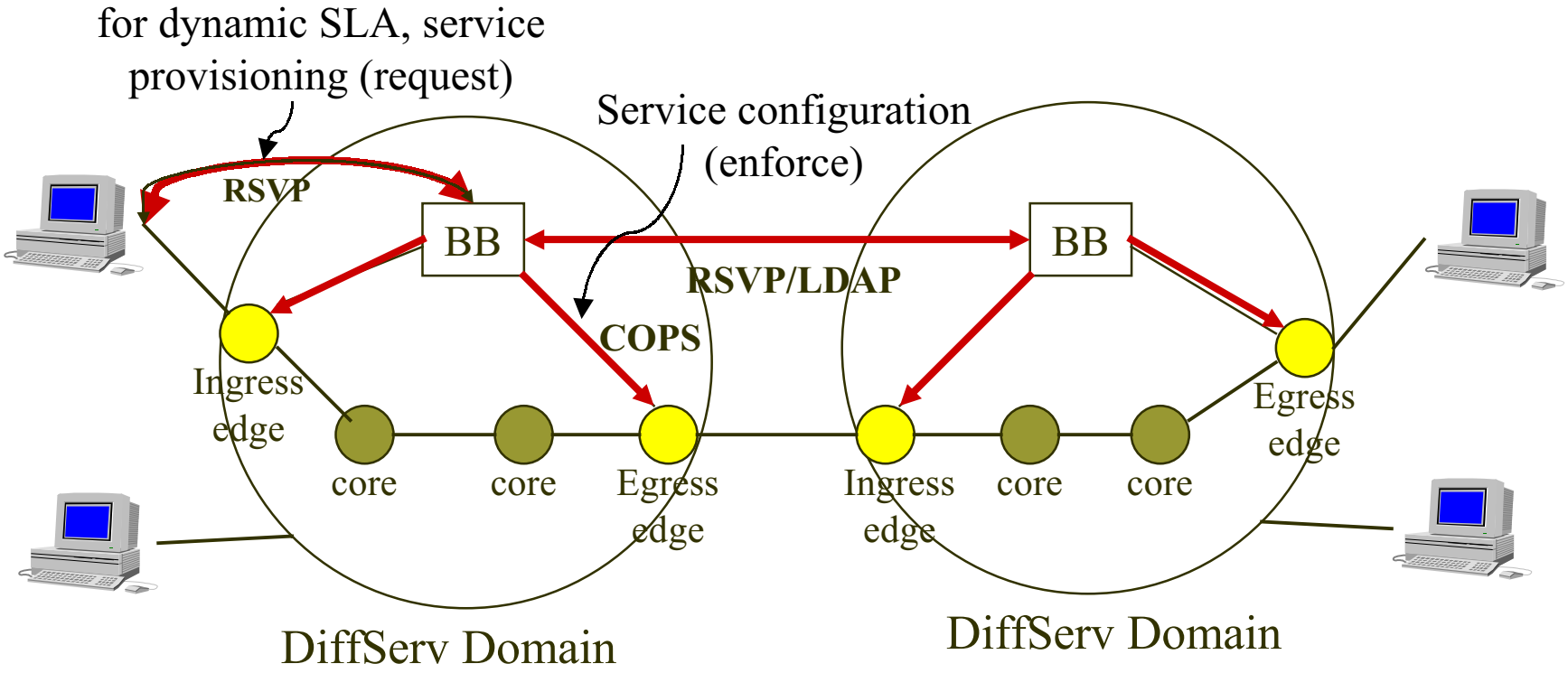
DiffServ Elements



DiffServ Overview

- No per-flow state and processing
- Provide a limited number of service classes
- Keep core routers simple
- Move control-plane mechanisms to bandwidth brokers
- Push complexity to the edges
- Exercise different user-plane mechanisms at edge and core
- Allow static and dynamic service provisioning

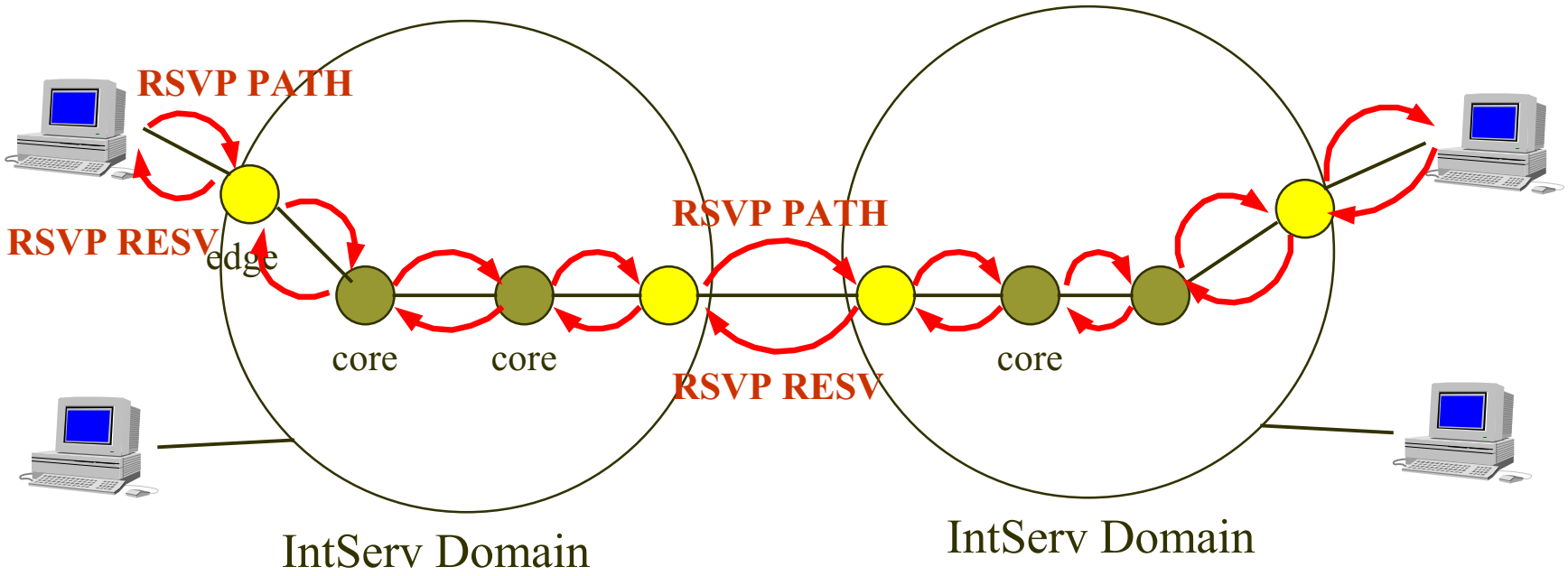
Generic DiffServ Operation Model



control-plane path
user-plane path
BB : bandwidth broker
SLA : service level agreement
MF : Multi-Field
BA : Behavior Aggregate

LDAP : Lightweight Directory Access Protocol
Ingress : MF classifier, policer, marker, scheduler
Core : BA classifier, scheduler
Egress : BA classifier, policer, shaper, remarker, scheduler
BB : provisioning, configuration

Generic IntServ Operation Model



: control-plane path

: user-plane path

edge, core : admission control, MF classifier, policer, scheduler

DiffServ Services

DSCP (DiffServ Code Point) and PHB (Per-Hop Behavior)

PHB Group	AF (Assured Forwarding)					EF (Expedited Forwarding)	Best-Effort
Features	Olymic service(an example) 4 delay priority classes, each with 3 drop precedence subclasses (quantitative service)					Premium/Virtual Leased Line Service (quantitative service)	none
Recommended DSCP in DS-field		AF1	AF2	AF3	AF4	101110	000000
	Low	010000	011000	100000	101000		
	Middle	010010	011010	100010	101010		
	High	010100	011100	100100	101100		
Traffic Control	Static SLA Policing, classification, marking RIO/WRED scheduling					Dynamic SLA Policing, classification, marking Priority/WFQ scheduling	FIFO scheduling
Non-conforming Traffic	Re-mark as Best-Effort					Drop	Forward

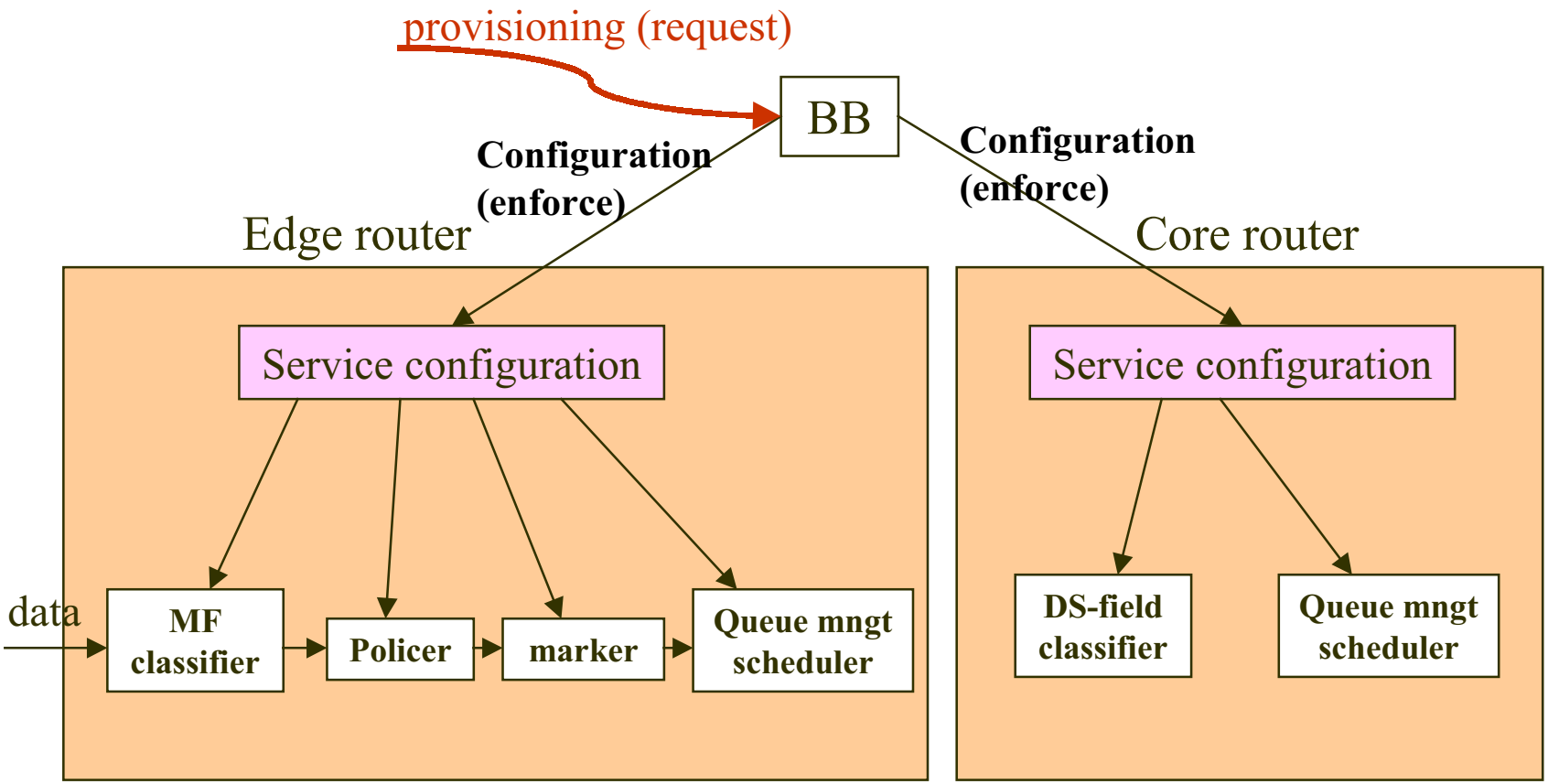
Service Level Agreement

- SLA: service contract between customer domain and DS ISP domain
- SLA may include TCA, classification rules, routing constraints, pricing and billing, authentication and encryption, etc.
- TCA (Traffic Conditioning Agreement) : subset of SLA
  - performance parameters
  - traffic profile
  - etc.
- Scope of SLA: unrestricted, one-to-one, one-to-many
  - governing transmitted traffic from ingress
  - but how about received traffic to egress?
- Static(pre-determined) and dynamic(on-demand) SLA

DiffServ Traffic Control

		Bandwidth Broker	Ingress Router	Core Router	Egress Router
Control-plane	Service Provisioning (request)	Intra-domain/inter domain (RSVP/LDAP)			
	Service Configuration (enforce)	Intra-domain (COPS)	Yes	Yes	Yes
User-plane	Classification		MF (multi-field) Classification	BA (Behavior Aggregate) (DS-field) classification	BA (DS-field) Classification
	Policing		Yes		Yes
	Marking		Yes		possibly remark
	Scheduling		Possibly drop/remark, Queue management (RIO/WFQ/priority over A-queue/P-queue)	Queue management (RIO/WFQ/priority over A-queue/P-queue)	Possibly drop/remark, Queue management (RIO/WFQ/priority over A-queue/P-queue)

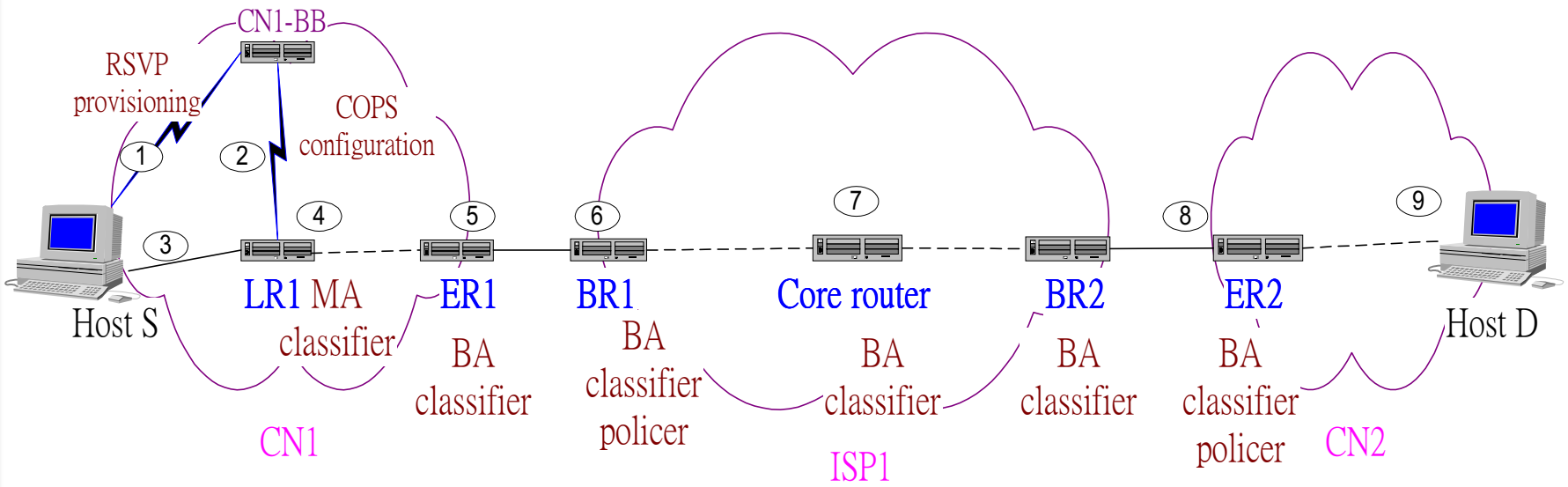
DS-capable Edge/Core Router



End-to-End Service Architecture I: DiffServ Customer/ISP

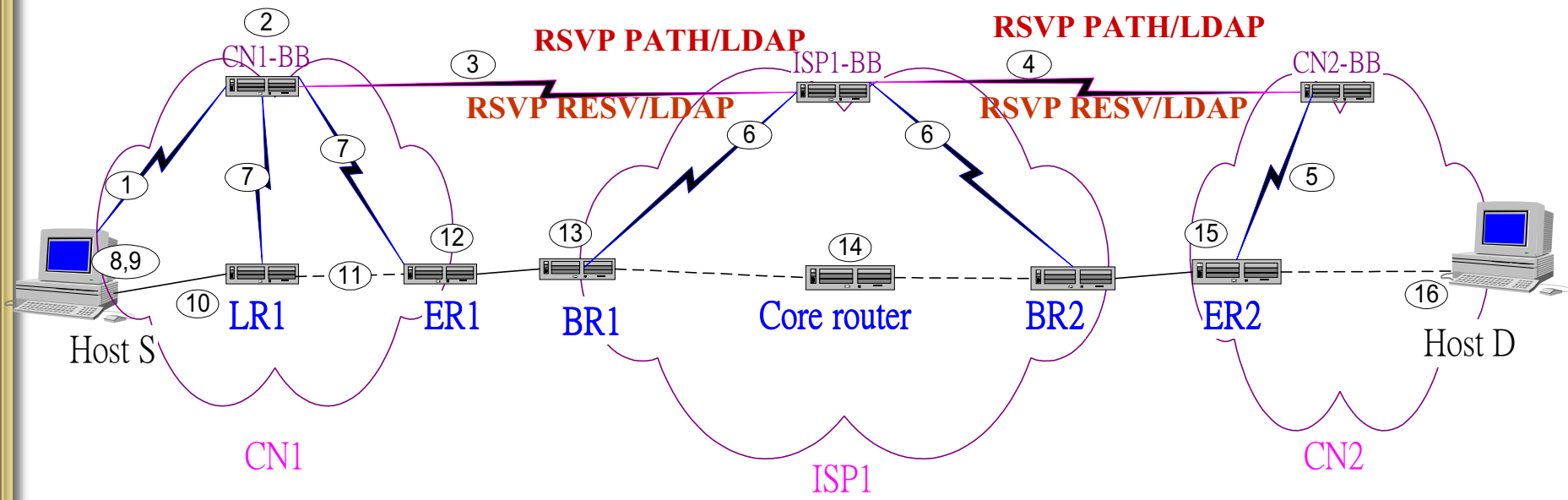
- AF with static SLA
- EF with dynamic SLA
- EF with dynamic SLA in MPLS-based DiffServ
- MPLS overview
- MPLS-based DiffServ ISP vs pure DiffServ ISP

Assured Forwarding with static SLA



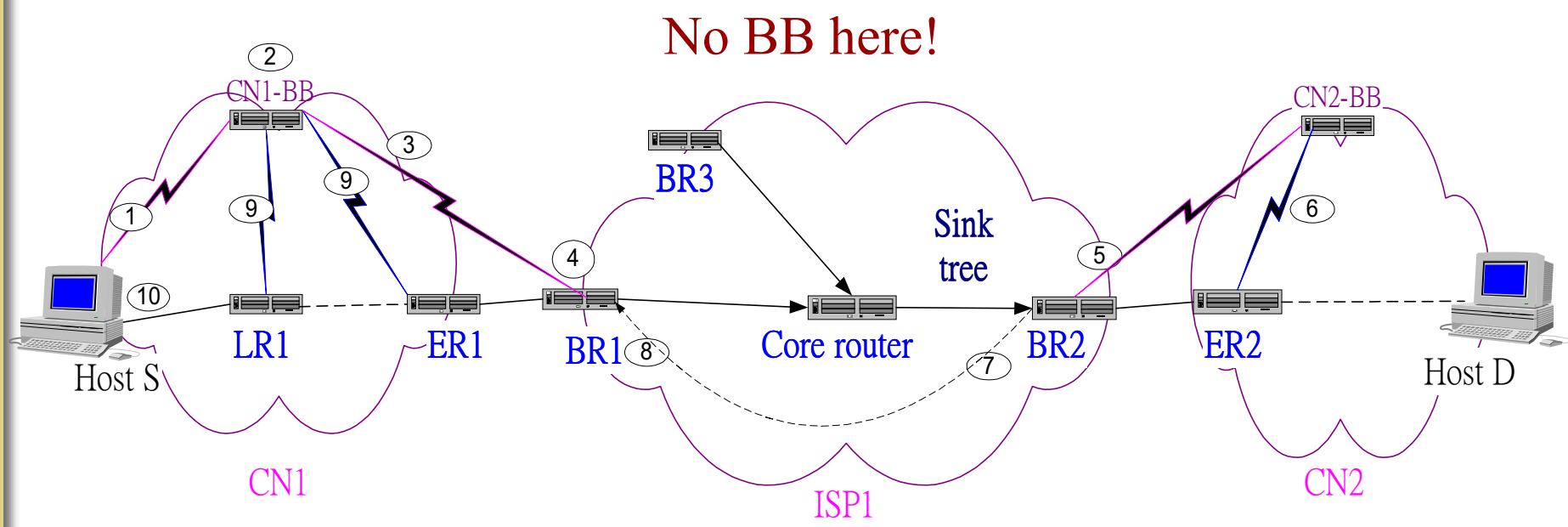
Source: IEEE Network, March/April 1999, Xiao and Ni

Expideted Forwarding with Dynamic SLA



Source: IEEE Network, March/April 1999, Xiao and Ni

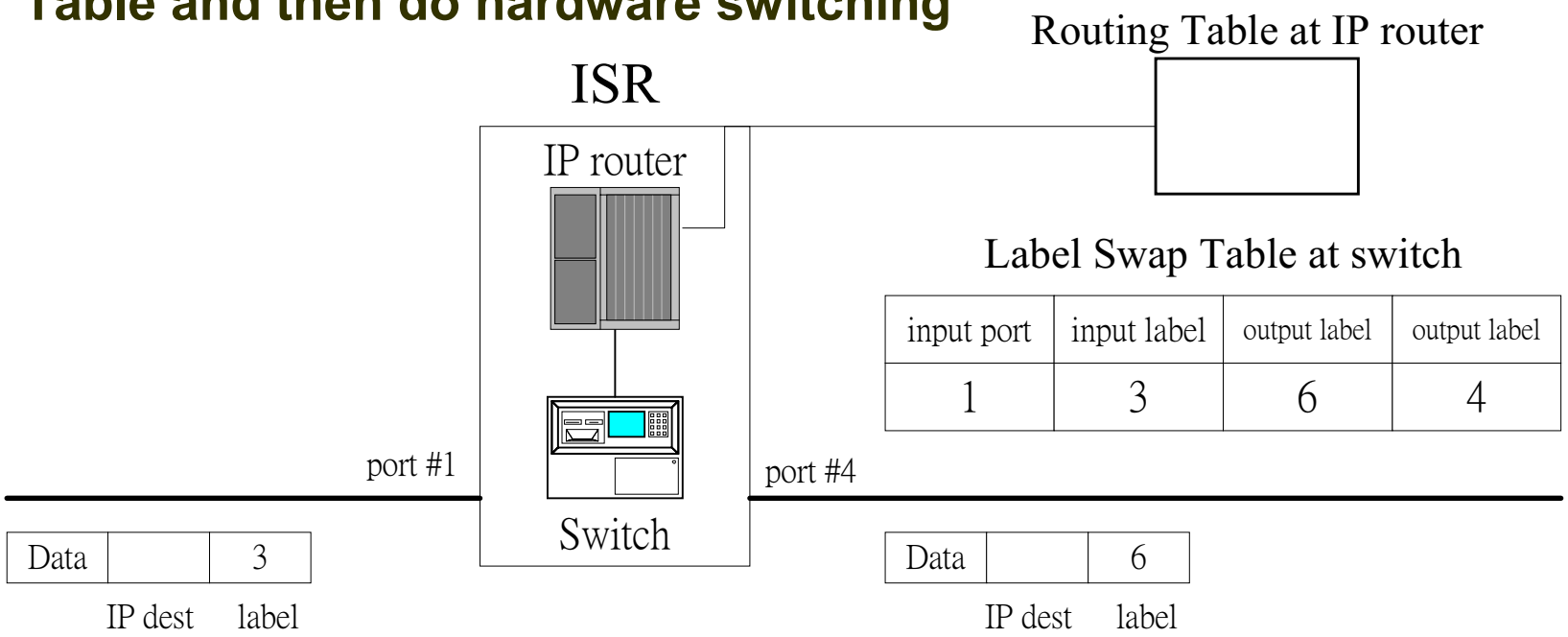
Expedited Forwarding with Dynamic SLA in MPLS-based DiffServ



Source: IEEE Network, March/April 1999, Xiao and Ni

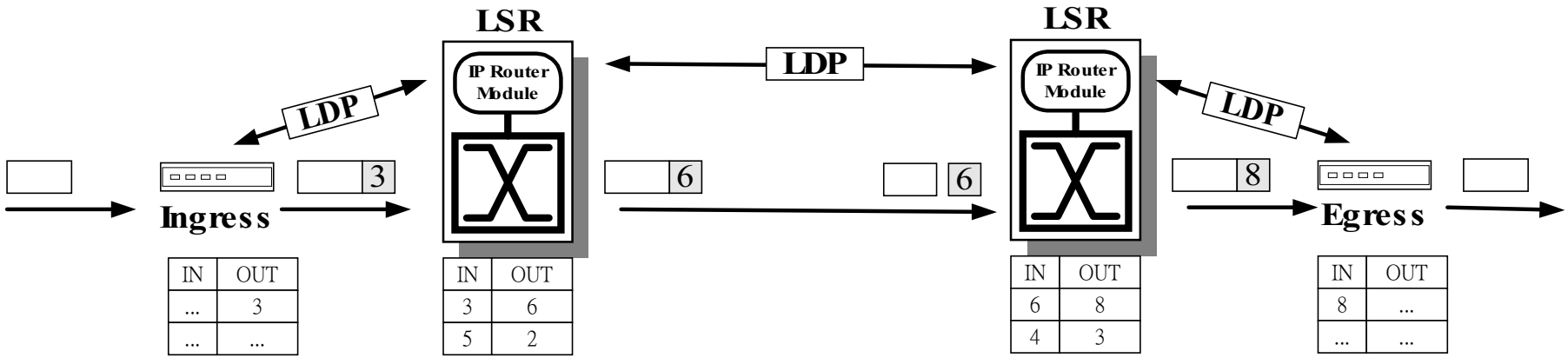
MPLS Overview

- ISR (Integrated Switch Router) includes IP router and layer-2 switch
- Packets are sent hop-by-hop by IP router in layer-3 or sent by switch in layer-2
- The data of IP routing or flow mapped into labels
- The label pasted on packets compare with Label Swap Table and then do hardware switching



Label Swapping

- Control protocol distributes route-to-label mappings
- Classify, label and forward at ingress
- Label swap over L2 path to egress
- Remove label and forward at the egress

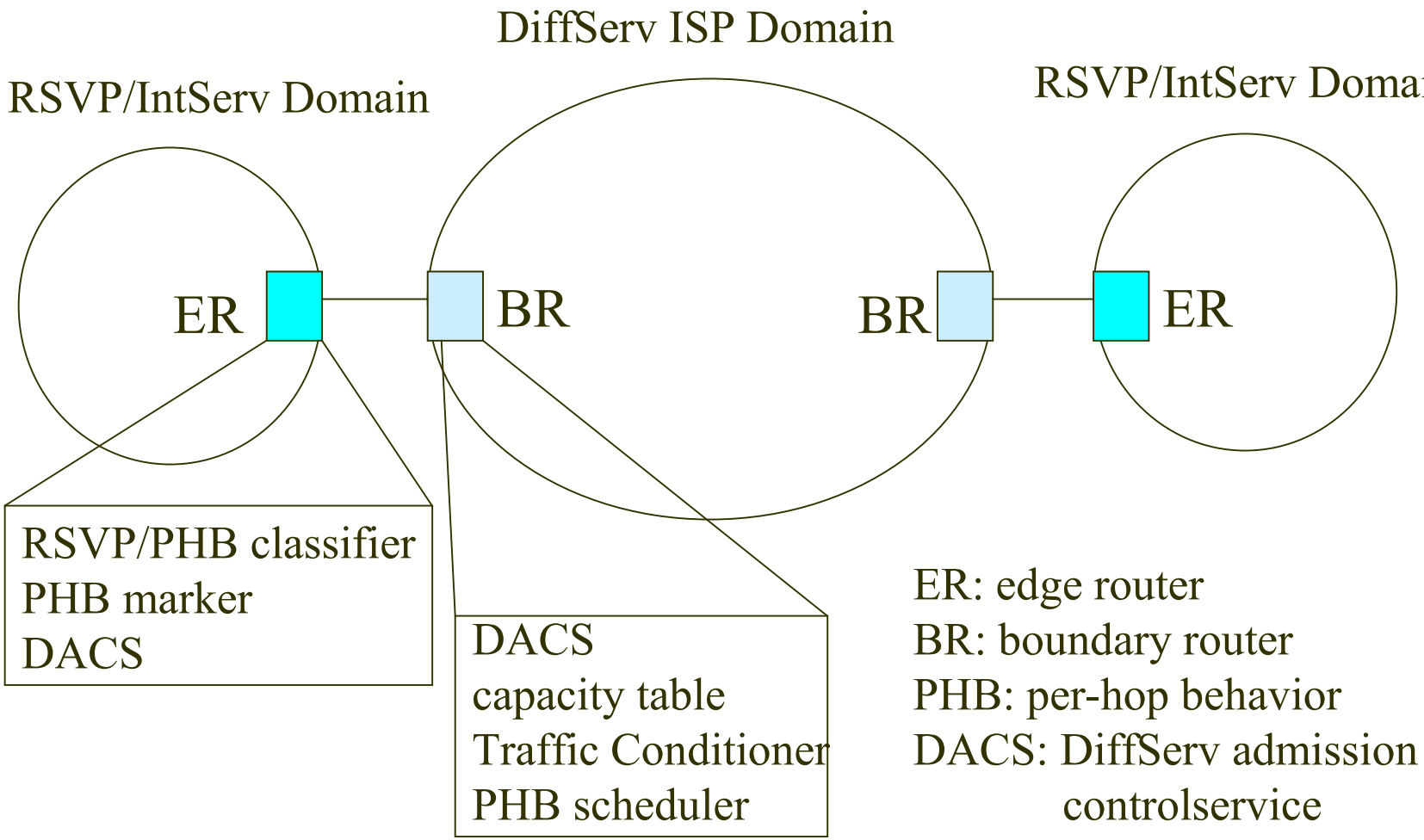


- LDP : Label Distribution Protocol
- LSR : Label Switch Router

MPLS-based DiffServ ISP vs Pure DiffServ ISP

- MPLS header:
  - label-for switching
  - COS-for class of service
- MPLS header inserted between layer 2 header and layer 3 (IP) header
- For pure DiffServ, routers check destination IP address and DS-field

End-to-End Service Architecture II:
IntServ Customer and DiffServ ISP



QoS Traffic Control

Admission Control
Policing
Classification
Scheduling

QoS Routing
Traffic Engineering
Policy Control

Admission Control for IntServ

- Main reference:
  - Tspec (Leaky Bucket parameter)
  - Rspec (B-bandwidth, S-slack term)
- Allocated bandwidth: R — between p (peak rate) and r (token rate).
- To satisfy Guaranteed Service:
  - low utilization(below 10%).
  - Controlled Load Service(no Rspec): flexible and higher utilization (50%~80%)
- Measurement-based(vs computation-based) admission control admit more flows subject to loss constraint.
- Adjusting the parameters of measurement is more important than what parameters to measure.

Admission Control for DiffServ

- Implicit: for static SLA
- Explicit: for dynamic SLA
- Capacity reconfiguration may be triggered when low-water mark is reached.

Classification

- How to distinguish the packets, according to some fields within packets

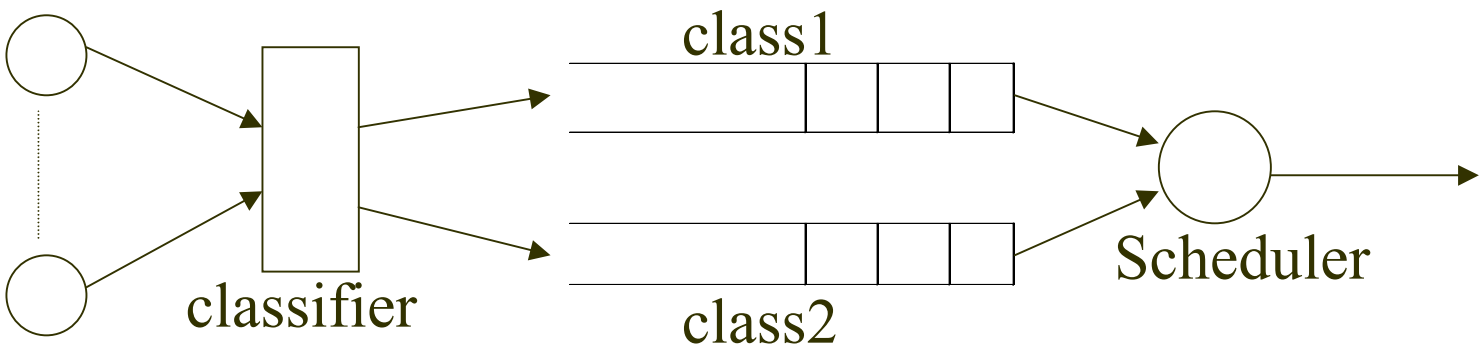
	MF (IntServ, DiffServ edge)	BA (DiffServ core)
Fields	eg. src/dest IP, proto id, src/dest port (flow id)	DS-byte
Decision	Send to per-flow or aggregated queue	Drop, mark or put at some location of a queue
Complexity	eg. 128 bits	6 bits

Scheduling for IntServ

- **Fair Queueing :**
  - **per-flow queues**
  - **bit-by-bit round robin (BR)**
  - **compute finish-time for each packet**
    (finish time: when packet would have left using BR)
  - **the one with smallest finish-time gets txed first**
- **Weighted Fair Queueing:**
  - **sum of weights = total bandwidth**
  - **each flow gets** $\frac{\text{its weight}}{\text{sum of wieghts}}$ **share**
  - **work-conserving**
  - **bounded delay**

Scheduling for IntServ (Cont.)

- **Class-based Queueing:**
  - per-class queues
  - priority & bandwidth for each class
  - scheduled based on priority & bandwidth
  - regulated flows (dropped or moved to lower class for violating traffic)
  - less state required than WFQ

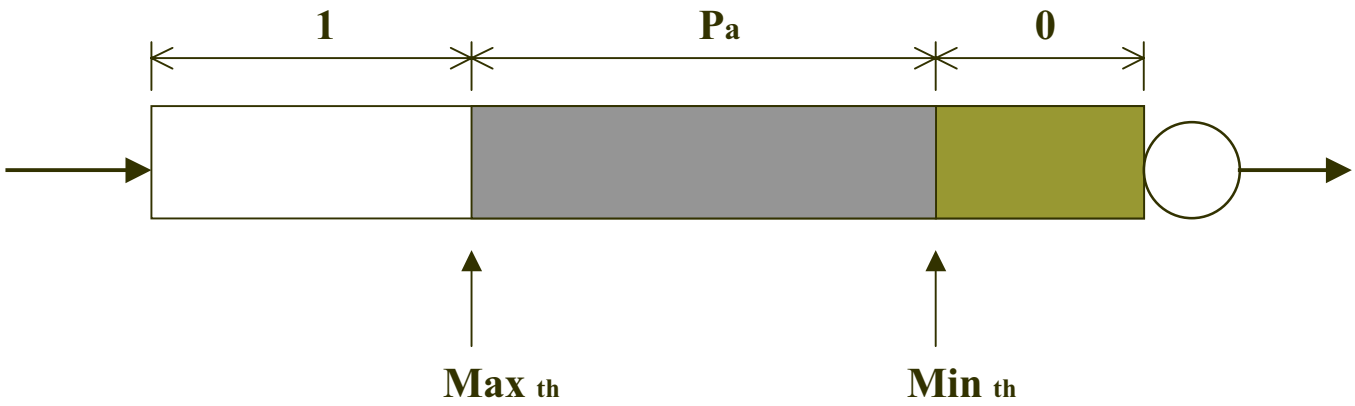


Scheduling for DiffServ

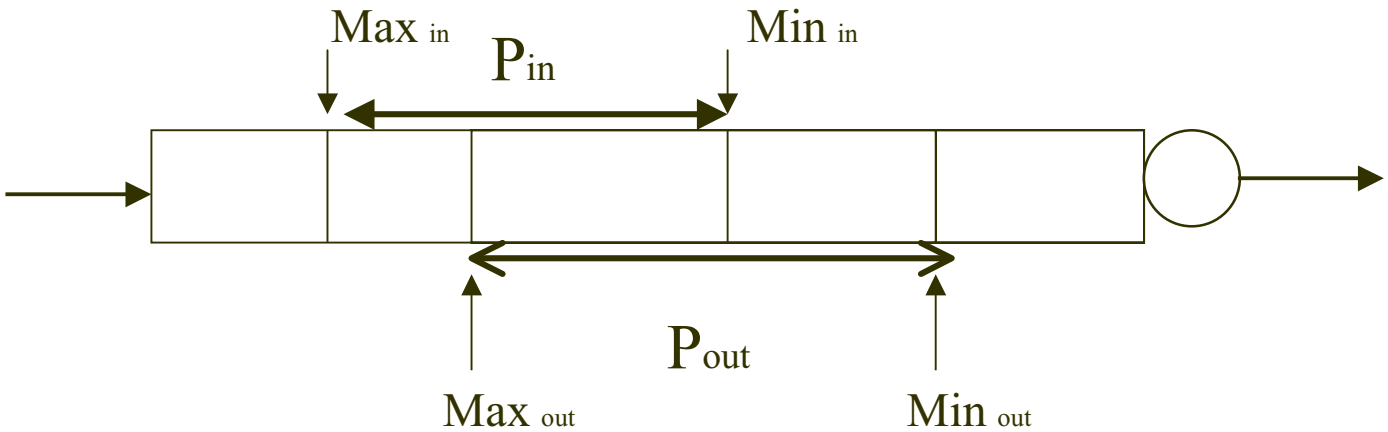
- **Mostly single queue management**
- **Common strategies:**
  - Threshold discard
  - Push out
  - Precedence
  - Weighted
- **Over a dozen of proposed schemes:**
  - RED,WRED,HOL,HLPO,RIO,PPP,etc.

RED and RIO

- **RED (Random Early Detection)**



- **RIO (RED with in and out)**

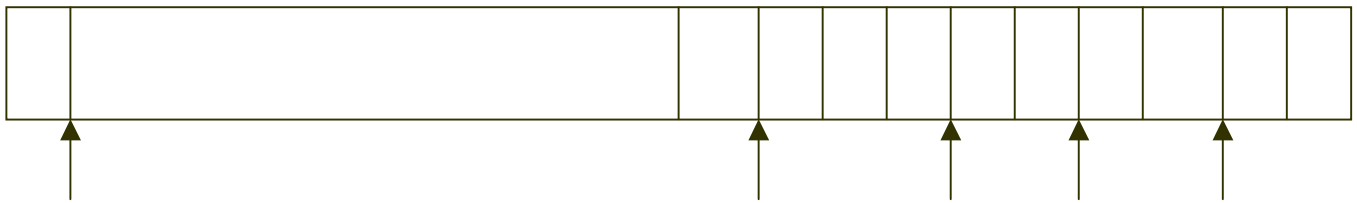


WRED (Weighted Random Early Detection) -- Cisco

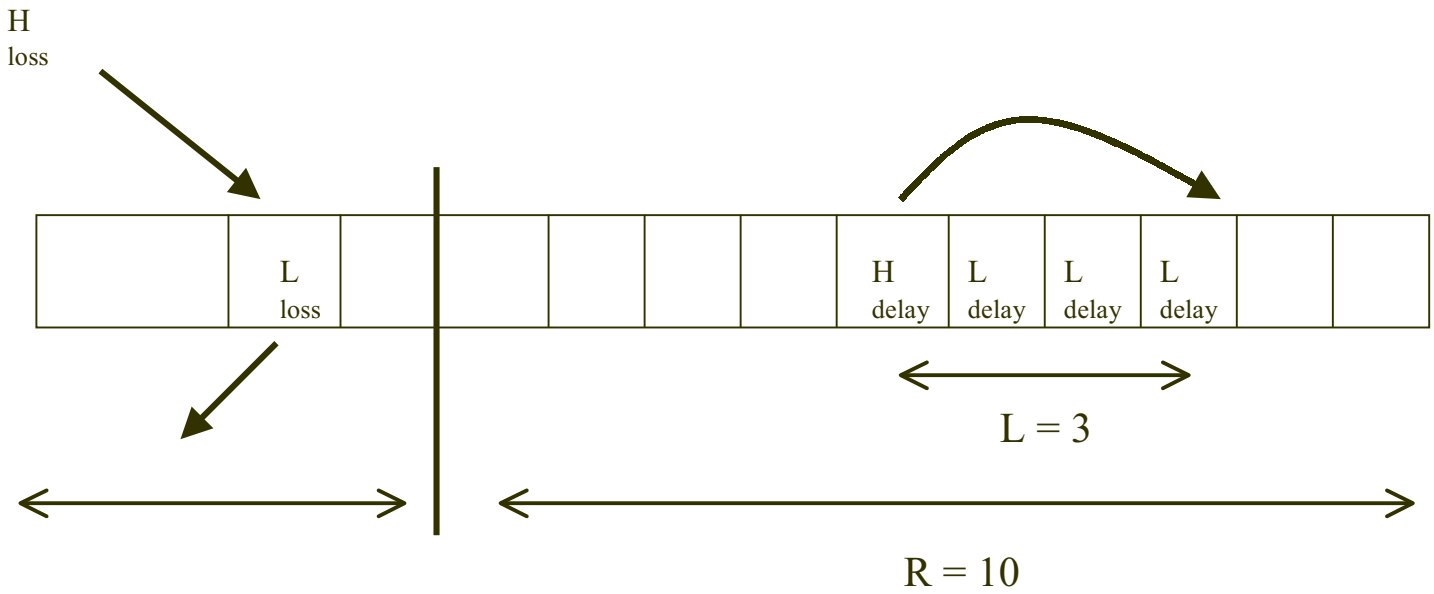
* Used in "**core routers**", rather than "**edge routers**"

* Edge routers assign IP precedence to packets

* WRED uses these precedence to determine weights and thresholds



PPP (Precedence with Partial Push out)



Current Internet Routing: RIP and OSPF

	RIP	OSPF
message exchanged	distance vector	link state
message propagation	between neighbors	flooding
state information	local	global
path computation algorithm	Bellman-Ford	Dijkstra

Problems as in Integrated Services RSVP Networks:

- not take link bandwidth into account ⇒ high blocking probability
- per-destination routing table⇒ routing table update(path change) ⇒ QoS interrupt

QoS Routing: Comparison of QOSPF

QoS Routing	QOSPF/Z	QOSPF/G
Criteria	constrained shortest path parameter:(residual_bw, delay)	constrained widest-shortest path (WSP) parameter: (residual_bw, #hops)
Triggering Path computation*	on-demand, topology change, or incoming LSA	pre-compute periodically, or threshold (number of updates)
Triggering Link-state update*	threshold	threshold or periodical
Algorithm	modified Dijkstra	modified Bellman-Ford
Routing	hop-by-hop or explicit (source) routing	hop-by-hop or explicit (source) routing
Forwarding Cache Granularity	per-pair	per-flow
Route pinned by	RESV message	RESV message

* Triggering rate can be limited by hold-down timer

Related IETF RFCs and Drafts (1/2)

1. R. Braden, Ed., “Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification,” Internet RFC 2205”, September 1997.
2. R. Braden and L. Zhang, “Resource ReSerVation Protocol (RSVP) - Version 1 Message Processing Rules,” Internet RFC 2209, September 1997.
3. J. Wroclawski, “The Use of RSVP with Integrated Services,” Internet RFC 2210, September 1997.
4. J. Wroclawski, “Specification of the Controlled-Load Network Element Service,” Internet RFC 2211, September 1997.
5. S. Shenker, C. Partridge, and R. Guerin, “Specification of Guaranteed Quality of Service,” Internet RFC 2212, September 1997.
6. S. Shenker and J. Wroclawski, “General Characterization Parameters for Integrated Service Network Elements,” Internet RFC 2215, September 1997.

Related IETF RFCs and Drafts (2/2)

7. T. Li, Y. Rekhter, “A Provider Architecture for Differentiated Services and Traffic Engineering”, Internet RFC 2430, October 1998.
8. K. Nichols, S. Blake, F. Baker, D. Black, “Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers”, Internet RFC 2474, December 1998.
9. S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, ” An Architecture for Differentiated Service”, Internet RFC 2475, December 1998.
10. P. Ford and Y. Bernet, "Integrated Services over Differentiated Services," Internet Draft, draft-ford-issll-diff-svc-00.txt, March 1998.
11. Y. Bernet, R. Yavatkar, P. Ford, F. Baker, L. Zhang, "A Framework for End-to-End QoS Combining RSVP/IntServ and Differentiated Services," Internet Draft, draft-bernet-intdiff-00.txt, March 1998.
12. Eric C. Rosen, Daniel Tappan, Tony Li, Alex Conta, “MPLS Label Stack Encoding”, Internet Draft, draft-ietf-mpls-label-encaps-04.txt, April 1999