Computer Networks : Protocols and Practice

Part 9
Quality of Service

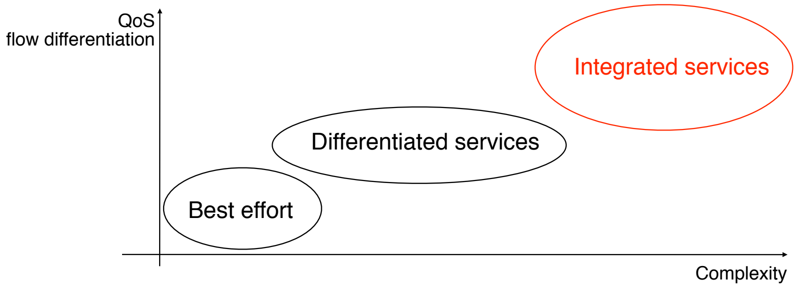
Olivier Bonaventure
http://inl.info.ucl.ac.be/

Traffic control and QoS in IP Networks
Outline

- Applications
- Packet-level traffic control mechanisms
- Standardised Services
  - Integrated Services
    - Architecture
    - RSVP : Resource reSerVation Protocol
    - The Guaranteed and Controlled Load services
  - Differentiated Services

Services

- What kind of services can we build with all these traffic control mechanisms ?



Integrated services
The "hard" approach for QoS

- Basic hypothesis
  - Some specific applications require QoS
    - delay guarantees
    - bandwidth guarantees
  - QoS will be provided to layer 4 flows
    - Each layer 4 (TCP or UDP) flow will inform the network about its QoS requirements
    - The network will accept or reject the flow based on its requirements and the current state of the network
  - QoS should support unicast and multicast
  - QoS flows should be allowed to coexist with best-effort flows in the same network
  - The existing routing protocols are left unchanged

The Intserv architecture was proposed in

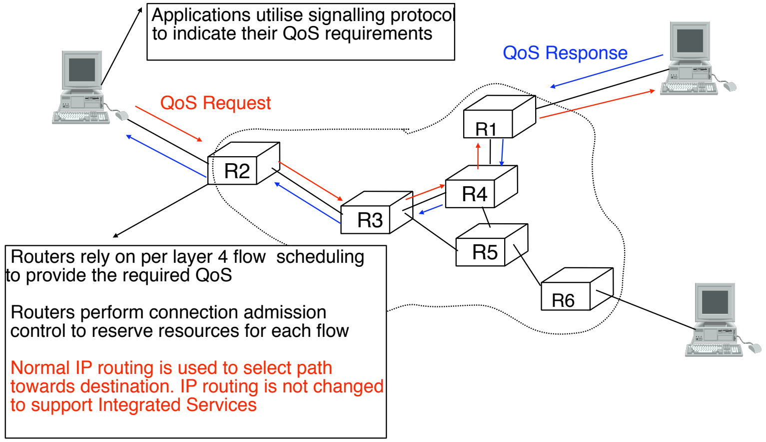
R.Braden, D.Clark, and S.Shenker. Integrated services in the Internet architecture : an overview. Internet RFC 1633, July 1994.

See also

P.White and J.Crowcroft. Integrated services in the Internet : the next stage in Internet : state of the art. *Proceedings of the IEEE*, 85(12): 1934--1946, December 1997.

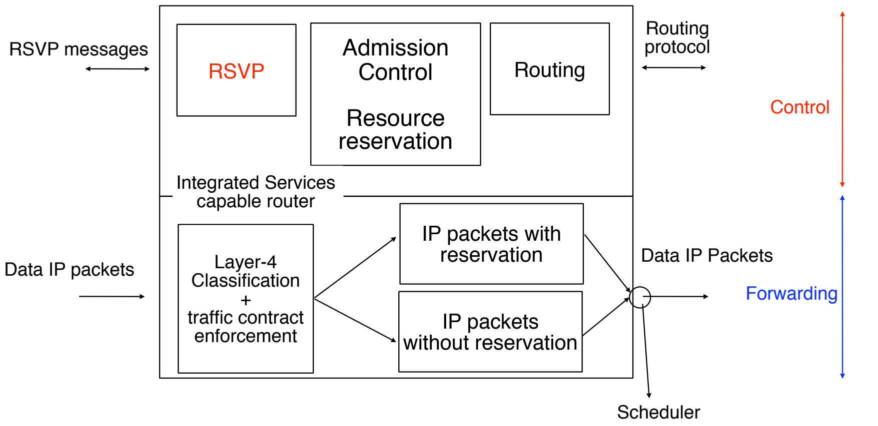
Integrated services network

- Provision of integrated services in a network



An Integrated services router

- Model of an Integrated Services router



RSVP

- RSVP : Resource Reservation Protocol
- Objectives
  - Support the establishment of unidirectional flows in IP networks
    - different types of flows
      - initially layer 4 flows
      - today mainly for MPLS LSPs
  - Suitable for IP unicast
  - Suitable for IP multicast

R.Braden, Ed., L.Zhang, S.Berson, S.Herzog, and S.Jamin. RFC 2205: Resource ReSerVation Protocol (RSVP) --- version 1 functional specification, September 1997.
F.Baker, J.Krawczyk, and A.Sastry. RFC 2206: RSVP management information base using SMIv2, September 1997. Status: PROPOSED STANDARD.

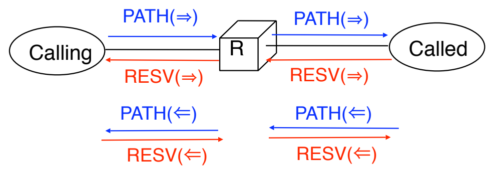
A.Mankin, Ed., F.Baker, B.Braden, S.Bradner, M.O`Dell, A.Romanow, A.Weinrib, and L.Zhang. RFC 2208: Resource ReSerVation Protocol (RSVP) --- version 1 applicability statement some guidelines on deployment, September 1997. Status: INFORMATIONAL.
J.Wroclawski. RFC 2210: The use of RSVP with IETF integrated services, September 1997. Status: PROPOSED STANDARD.

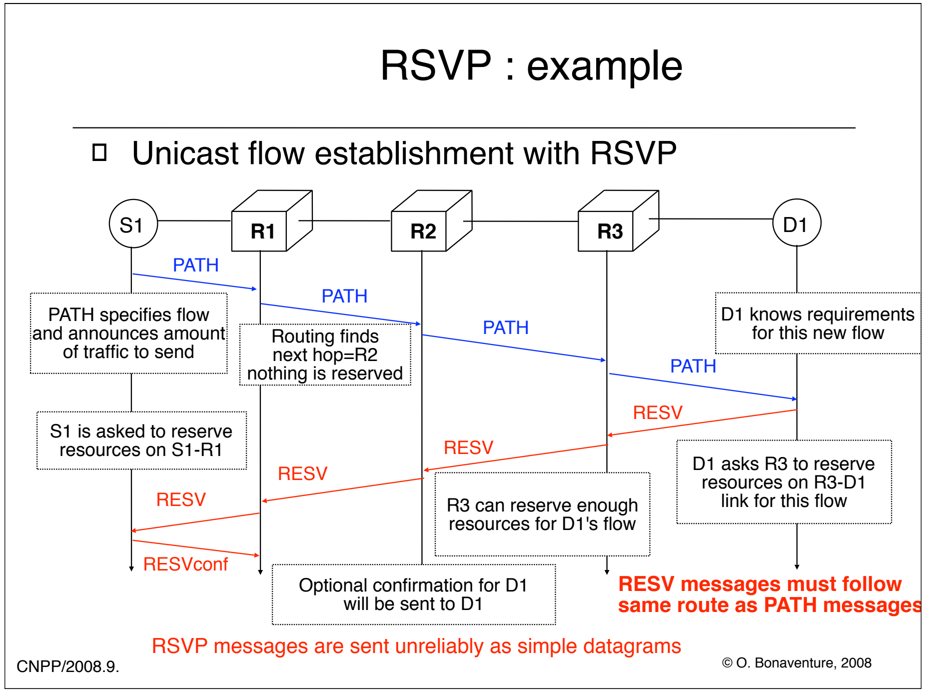
RSVP (2)

- Principles of operation
- Two important RSVP messages
  - PATH
    - used by sender to inform routers and receivers of the new flow and its required resources
      - no resources are reserved due to reception of PATH
  - RESV
    - used by receiver to actually reserve resources for the flow specified in the PATH message
      - resources are reserved for the IP packets sent by the sender towards the receiver along the path taken by the PATH message
- RSVP messages are sent inside IP packets

RSVP (3)

- RSVP flows
  - RSVP was originally defined to associate reservations to unidirectional layer 4 flows
    - RSVP flow defined as the set of packets with same
      - IP destination address
      - IP Protocol Number (e.g. TCP or UDP)
      - [Transport-level] Destination Port Number
    - A typical bi-directional voice conversation requires
      - One RSVP flow from calling to called user
      - One RSVP flow from called user to calling





It should be noted that RSVP messages always follow exactly the same route as the regular IP packets that will be send by the source and destinations.

RSVP messages are directly encapsulated inside IP packets, usually containing the IP router alert option.

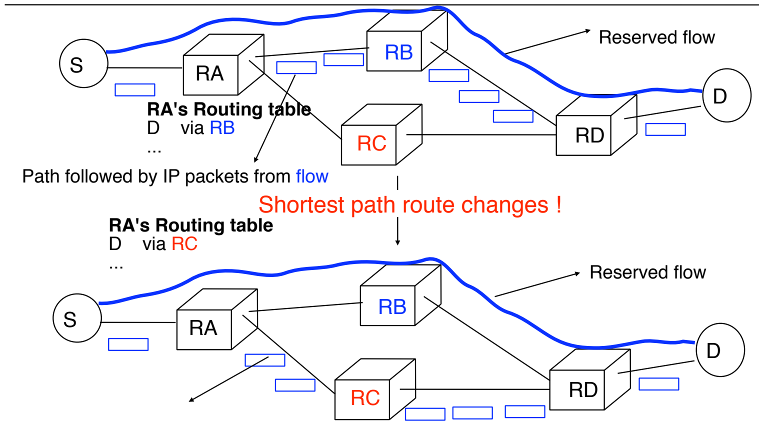
State maintenance

- RSVP routers maintain some per-flow state
- Possible solutions for state maintenance

- Hard-state
  - traditional solution used in circuit-switched network
  - state is created at flow establishment and removed at flow tear down
  - if intermediate router crashes, state+reservation are lost

- Soft-state (solution chosen by RSVP)
  - a timer is associated with each per-flow state
    - state is removed upon expiration of timer
  - hosts periodically retransmit PATH and RESV message
    - timer is resent upon reception of PATH/RESV message
  - if intermediate router crashes, state automatically reset
  - if route changes, new reservations/states are established on new path and all reservations/states expires

The problem of route changes



- Either IP packets should continue to follow the guaranteed path or the guaranteed path should change to reflect the new path for the IP packets

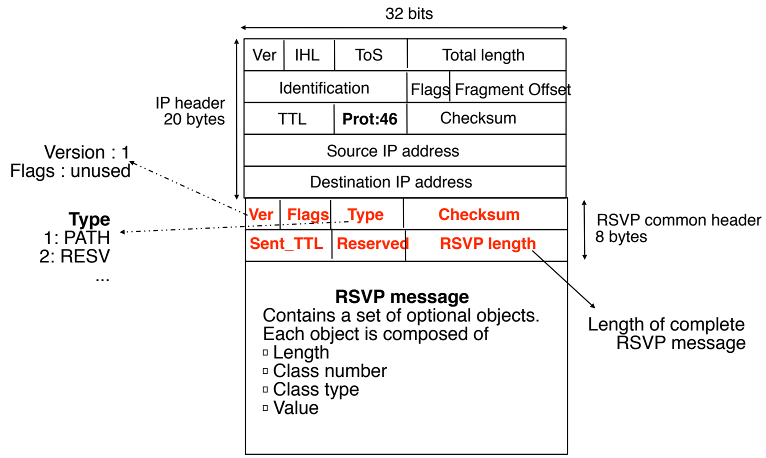
Dealing with route changes with RSVP

- Principle
  - A flow established with RSVP lasts for some time
  - To remain alive, a flow must be refreshed regularly
  - RSVP only considers unidirectional flows
- Endsystems behaviour
  - regularly send RSVP messages to maintain flow
- Routers behaviour
  - maintain some state information for each flow
  - state is removed and flow released if no signalling messages have been received during some time

RSVP : detailed example

- Issues to consider
  - How to encapsulate RSVP messages in IP packets ?
  - How to ensure that the reservation messages follow the same route as the PATH messages
    - cannot simply rely on IP for this
      - remember : we do not change anything to routing protocols
  - What kind of information is stored in the state of the intermediate routers ?

RSVP : packet format

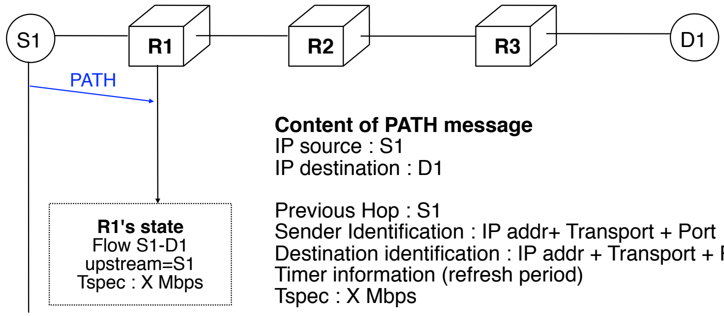


Sent_TTL

When a RSVP message is created, the sender places in Sent_TTL the same value as the TTL of the IP packet containing the RSVP message

RSVP : detailed example (1)

- What happens inside the routers ?
  - S1 announces a flow towards D1

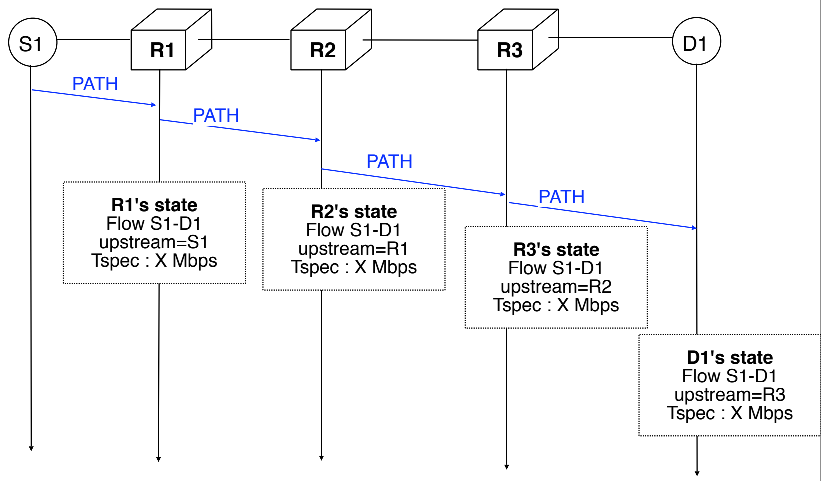


Content of PATH message
IP source : S1
IP destination : D1

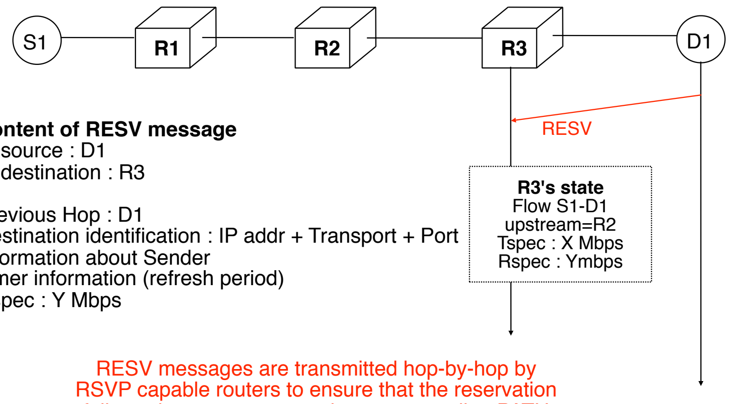
Previous Hop : S1
Sender Identification : IP addr+ Transport + Port
Destination identification : IP addr + Transport + Port
Timer information (refresh period)
Tspec : X Mbps

Behaviour of R1
Content of PATH message inserted in R1's
RSVP message updated and sent downstream
PHOP replaced by R1

RSVP : detailed example (2)



RSVP : detailed example (3)

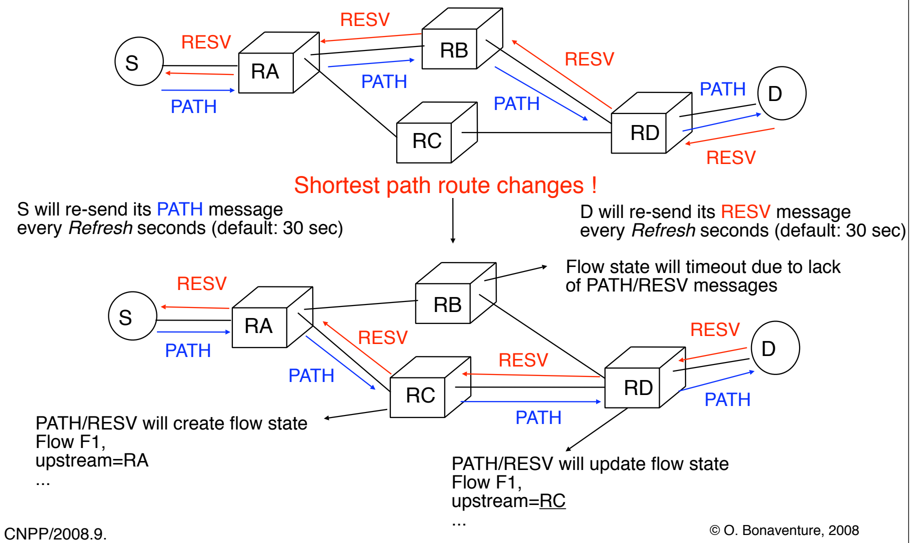


Content of RESV message
IP source : D1
IP destination : R3

Previous Hop : D1
Destination identification : IP addr + Transport + Port
Information about Sender
Timer information (refresh period)
Rspec : Y Mbps

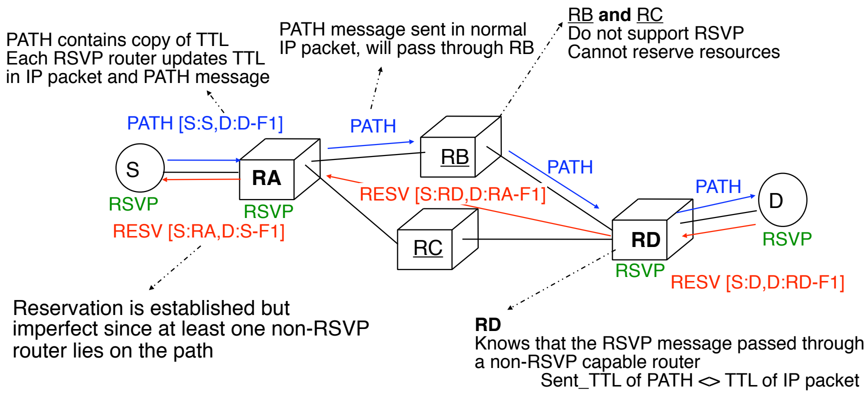
RESV messages are transmitted hop-by-hop by
RSVP capable routers to ensure that the reservation
follows the same route as the corresponding PATH

RSVP and route changes



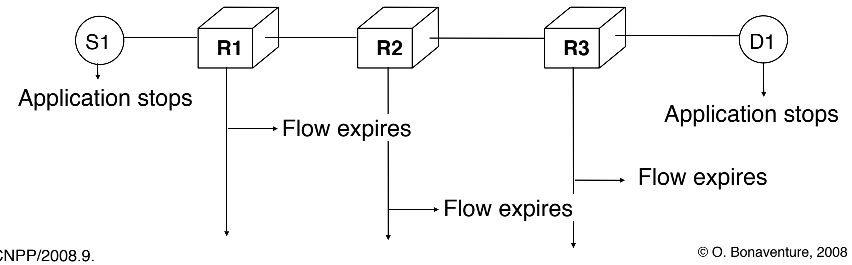
Smooth deployment of RSVP

- Problem
  - Can we utilize RSVP in heterogeneous network ?



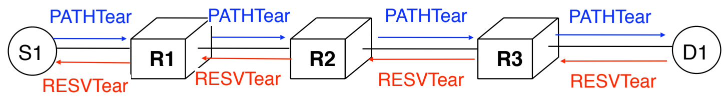
RSVP flow release

- How can we release RSVP flows ?
  - Sender and receiver stopping transmitting PATH and RESV messages
    - flow state will timeout inside routers
    - not the best solution since routers are expected to maintain flow state about five times longer than refresh period without receiving PATH/RESV messages
      - default refresh period is 30 seconds



RSVP flow release (2)

- A better solution
  - Define two new specific signalling messages
    - PATHtear
      - used by sender to announce to network end of a flow
    - RESVtear
      - used by receiver to announce to network end of a reservation



    - state timeout remains necessary since all RSVP messages (including PATHTear and RESVTear) are delivered as simple datagrams subject to packet loss

RSVP : comments

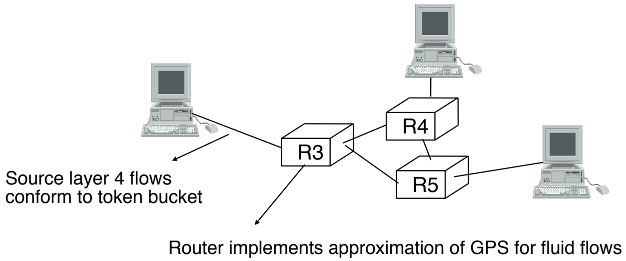
- The initial RSVP goal
  - define a signalling protocol to support large scale multicast reservations for layer-4 flows
- Best design decision
  - extensibility of the protocol and its mechanisms
- What RSVP is becoming today
  - a generic signalling protocol that can be used to establish any flow (layer 4, MPLS, SDH, ...)
  - support for multicast and unicast
  - soft-state can be completed by acknowledgements to improve reliability

Traffic control and QoS in IP Networks
Outline

- Applications
- Packet-level traffic control mechanisms
- Standardized Services
  - Integrated Services
    - Architecture
    - RSVP : Resource reSerVation Protocol
    - The Guaranteed and Controlled Load services
  - Differentiated Services

Guaranteed Service

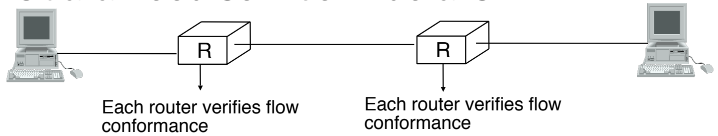
- Idea
  - provide a service as close as possible to a physical link between source and destination(s)
    - Ideal service would be fluid flow
    - the closest service we can provide in reality is an approximation to fluid flow with a firm (mathematically provable) guarantee on end-to-end delay and no losses



S.Shenker, C.Partridge, and R.Guerin. Specification of Guaranteed quality of service. Internet RFC 2212, September 1997.

Guaranteed service (2)

- Guaranteed service in details



  - Guarantees are provided for conforming packets
    - conformance defined by token bucket
    - guarantee on maximum delay guarantee
      - upper-bound on end-to-end network delay
    - no guarantee on average delay
    - no guarantee on delay jitter
    - loss guarantee
      - packet loss due to queue overflow cannot occur
      - implies reservation of buffers at intermediate routers
      - packet loss due to random bit errors cannot be avoided
  - non-conforming packets treated as best-effort

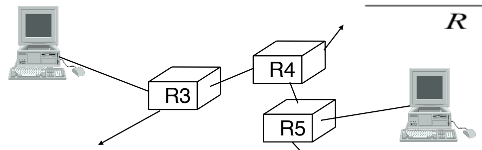
Guaranteed Service (2)

- Tspec, specified by sender in PATH messages
  - M = maximum packet size (bytes)
    - packets larger than M are considered as best-effort, IP fragmentation is not supported for guaranteed service flows
  - Token Buckets
    - Peak rate token bucket
      - bucket depth = M
      - p = peak rate of flow (bytes/sec)
      - during T seconds, flow will not send more than M+pxT bytes
    - Average rate token bucket
      - b = bucket depth (bytes)
      - r = token bucket rate (bytes/sec)
      - during T seconds, flow will not send more than b+rxT bytes
    - During T sec, traffic ≤ M+min[pxT, rxT+b-M]
    - m = minimum policed unit (bytes)
    - packets smaller than m considered of size m by token bucket

GS : delay bound

- What is the delay bound provided by GS ?
  - Fluid flow with GPS
    - transmission and processing delay
    - queuing delay ≤ b/r
  - Packet flows
    - transmission and processing delay
    - Queuing delay depends on scheduler
    - what about heterogeneous networks ?

$$\frac{B+n\times P_{max}}{R}+\sum_{i=1}^{n}\frac{P_{max}}{C_i}$$



$$\frac{B+n\times P_{max}}{R}+\sum_{i=1}^{n}\frac{K_i\times P_{max}}{C_i}$$

$$\frac{B+n\times P_{max}}{R}+\sum_{i=1}^{n}\frac{P_{max}}{C_i}$$

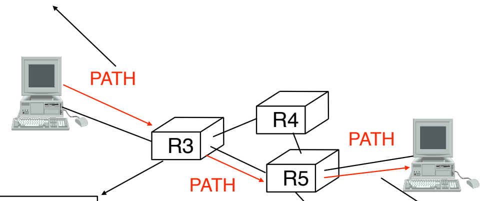
GS : delay bound (2)

- How can we support different schedulers ?
  - Compute delay bound of each scheduler as a delta compared to the GPS bound
  - Queuing delay bound = $\frac{b}{r}+\frac{C}{r}+D$
    - C : rate dependent error term
    - D : rate independent error term
  - Each scheduler advertises its C and D error terms in ADSPEC of PATH messages
    - ADSPEC is modified by each intermediate router
    - ADSPEC contains

$$C_{tot}=\sum_i C_i \quad D_{tot}=\sum_i D_i$$

GS : delay bound (3)

PATH message contains
- TSPEC
- ADSPEC (hop count, Path bandwidth, fixed delay, MTU, Ctot, Dtot)



R3 relies on WFQ
ADSPEC updated

$$C_{tot} = P_{max}$$

$$D_{tot} = \frac{P_{max}}{C}$$

R5 relies on SCFQ
ADSPEC updated

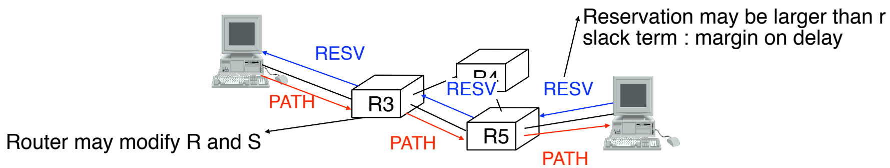
$$C_{tot} = P_{max} + P_{max}$$

$$D_{tot} = \frac{P_{max}}{C} + \frac{K \times P_{max}}{C}$$

Destination knows
- MTU for whole path
- Ctot
- Dtot
- Fixed Delay
- Min bandwidth of links in path

GS : delay bound (4)

- RESV message
  - destination specifies its reservation in RSPEC
    - Token bucket parameters (r, b, p, M, m)
    - R : rate (bytes per second)
      - R ≥ r
      - can be used to request a smaller queuing delay
    - S : Slack term (microseconds)
      - can be used by the destination when the announced delay is smaller than what is needed
      - intermediate routers could then benefit from part of this slack term to reduce the reservation attached to this flows



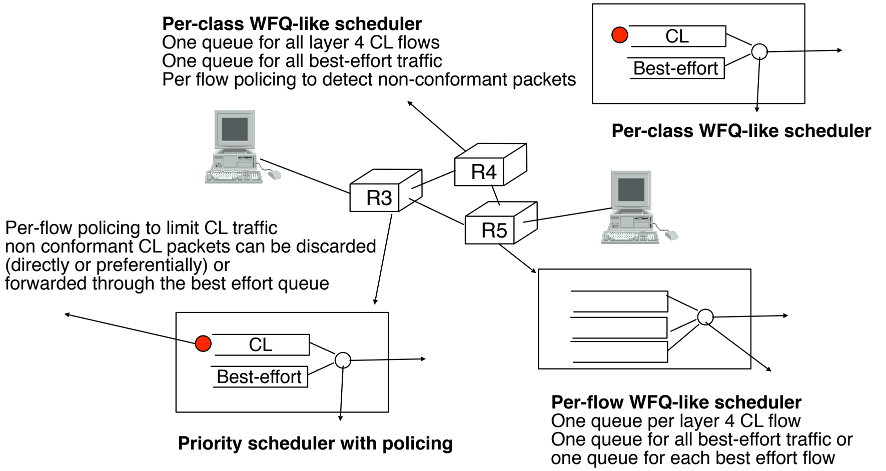
Controlled Load

- Idea
  - provide a service with the same performance as best-effort service in a lightly loaded and uncongested network
- Details
  - no deterministic guarantee on losses
    - but packet losses should be almost as rare as with GS
  - no guarantee on end-to-end delay or delay jitter
    - but on average queuing delay through intermediate routers should be low for conforming packets
- Utilisation
  - applications requiring bandwidth reservation

J.Wroclawski. Specification of the Controlled-Load network element service. Internet RFC 2211, September 1997.

Controlled Load implementation

- Several possible implementations



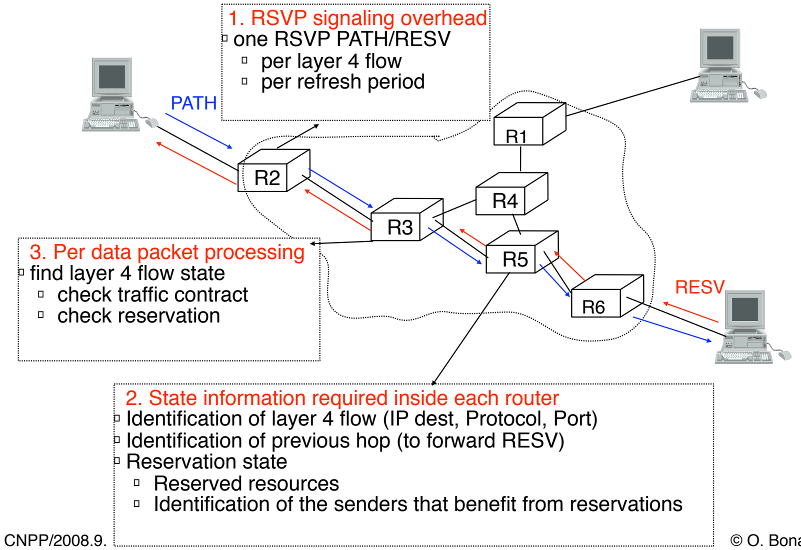
A critique of Integrated services

- Advantages
  - provides per layer 4 flow QoS guarantees
  - GS with delay/bandwidth guarantees
  - CL with bandwidth guarantees
- Drawbacks
  - Requires each intermediate router to perform some operations for each layer 4 flow
    - RSVP message processing
    - per layer 4 flow classification
      - classification can become complex for multicast !
    - per layer 4 flow policing/queuing/scheduling
      - a backbone router may see thousands of simultaneous flows !
  - Not all applications are able to express precisely their traffic and QoS requirements

For example, one average during periods of 1 second, one 2.4 Gbps link on the Sprint backbone saw 15.000 different layer four flows : http://ipmon.sprintlabs.com/packstat/viewresult.php?9:activeconnx:nyc-20.0-021121:

But on another 2.4 Gbps link more heavily loaded, the average number of layer four flows per second was ten times higher (170.000) http://ipmon.sprintlabs.com/packstat/viewresult.php?41:activeconnx:sj-22.0-021121:

Scaling issues with integrated services

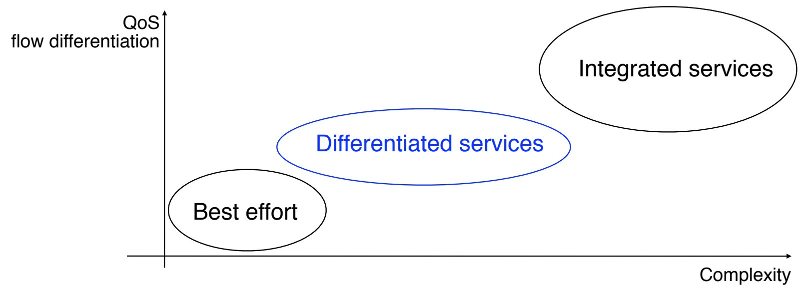


Traffic control and QoS in IP Networks
Outline

- Applications
- Packet-level traffic control mechanisms
- Standardised Services
  - Integrated Services
  - Differentiated Services
    - The architecture
    - The services
    - The role of MPLS

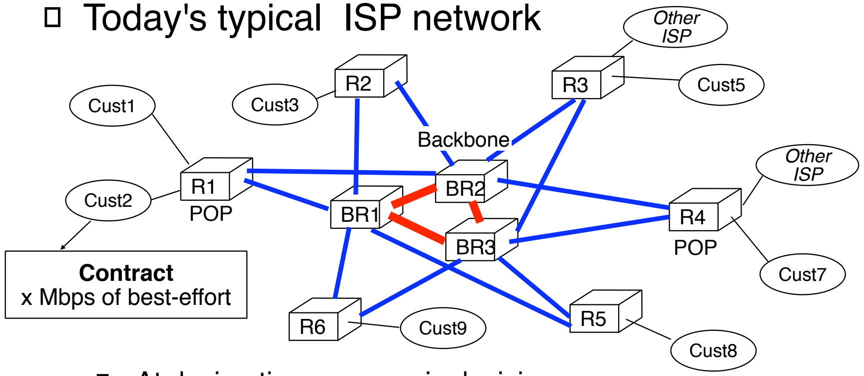
Services

- What kind of services can we build with all these traffic control mechanisms ?



Today's best effort IP networks

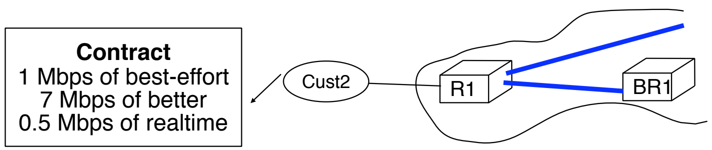
- Today's typical ISP network



- At design time, economic decision
- For each customer, we know subscribed amount of traffic, but usually not precisely destination
- Follow closely the load on links and contracts
  - upgrade links when load becomes too high

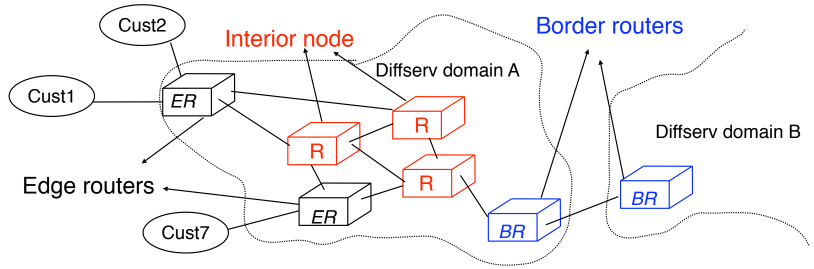
A pragmatic approach to QoS

- Idea
  - Allow customers to specify contracts for a few classes of traffic
    - Example classes
      - best effort, better than best effort, realtime, ...
    - different QoS commitments associated with each class
      - average delay
      - average losses
  - ISP configures its routers to efficiently support the requirements of these traffic classes
    - routers must remain simple



Differentiated Services : Architecture

- Principles
  - Network is divided in two parts



  - Complex mechanisms are implemented only on boundary nodes (border and edge routers)
  - Interior nodes are simple to operate at high speeds
  - No change to existing routing protocols

CNPP/2008.9.

40

The architecture is described in

S.Blake, D.Black, M.Carlson, E.Davies, Z.Wang, and W.Weiss. An architecture for differentiated services. Internet RFC 2475, December 1998.

See also

D.Grossman. New terminology and clarififcation for diffserv. RFC3260, April 2002.

The MIB is being finalised, see

F.Baker, K.Chan, and A.Smith. Management information base for the differentiated services architecture. Internet draft, draft-ietf-diffserv-mib-16.txt, work in progress, November 2001.

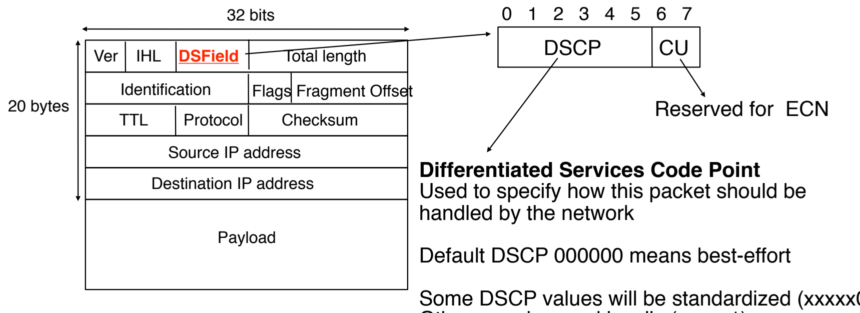
One of the original papers on DiffServ
K.Nichols, V.Jacobson, and L.Zhang. A two-bit differentiated architecture for the internet. Internet draft draft-nichols-diff-svc-arch-00.txt, work in progress, available from ftp://ftp.ee.lbl.gov/papers/dsarch.pdf, November 1997.

Differentiated Services : Architecture (2)

- Provision of different types of service
  - The type of service is indicated explicitly inside each IP packet
    - Marking can be performed by
      - customer
      - edge router
      - border router
  - Interior routers only rely on this indication to provide the different types of services
    - complexity of interior routers depends on number of different services, not on number of layer x flows

Differentiated Services : Packet Marking

- Packet marking
  - redefine the semantics of the rarely used ToS byte inside IP header



Differentiated Services Code Point
Used to specify how this packet should be handled by the network

Default DSCP 000000 means best-effort

Some DSCP values will be standardized (xxxxx0)
Others can be used locally (xxxxx1)

See

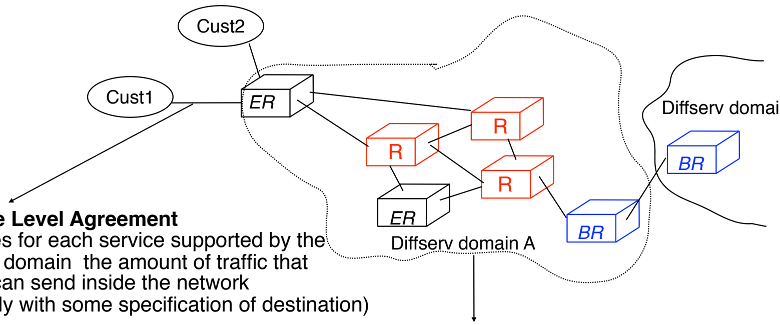
K.Nichols, S.Blake, F.Baker, and D.Black. Definition of the differentiated serivices field (DS field) in the IPv4 and IPv6 headers. Internet RFC 2474, December 1998.

For previous usage of the Diffserv field, see

P.Almquist. Type of service in the internet protocol suite. Internet RFC 1349, 1992.

F.Baker. Requirements for IP version 4 routers. Internet RFC 1812, June 1995.

Provision of Differentiated Services

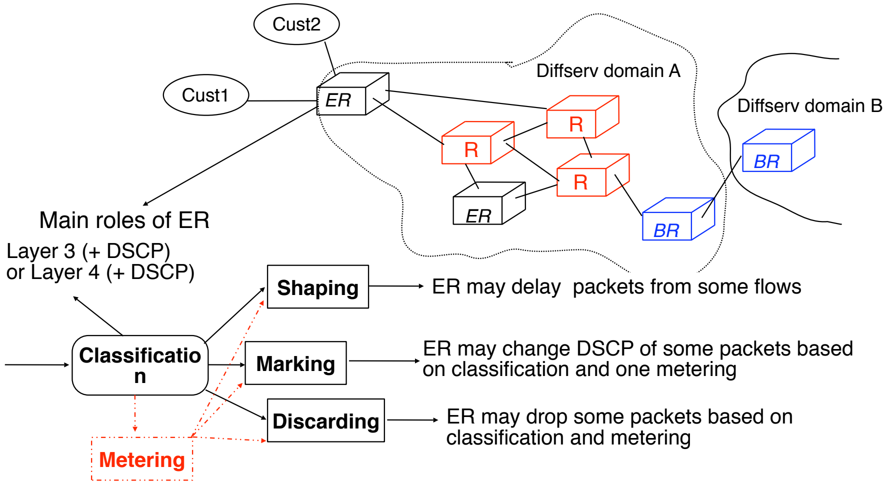


Service Level Agreement
specifies for each service supported by the diffserv domain the amount of traffic that Cust2 can send inside the network
(possibly with some specification of destination)

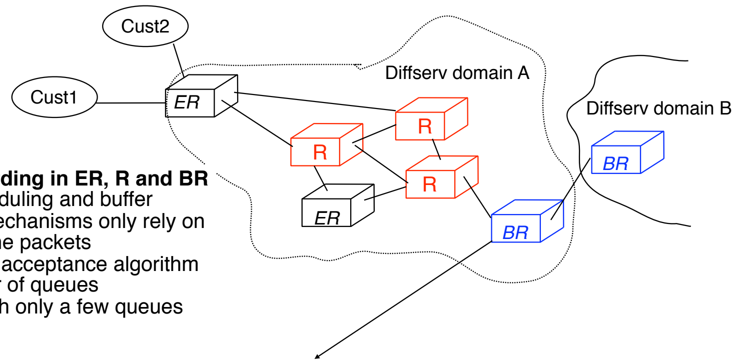
Network will commit to some performance figures
- packet loss rate over some period
- end to end delay at key points
- network reliability
- ...

Diffserv Codepoints
Network provider defines the services that he wants to provides and associates one or several DSCP to each service

Provision of Differentiated Services (2)



Provision of Differentiated Services (3)



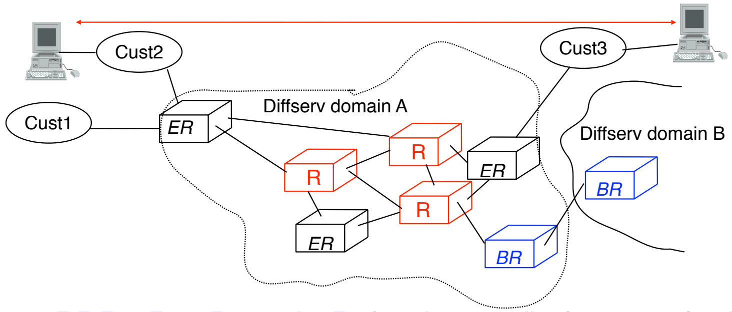
Packet forwarding in ER, R and BR
Queuing, scheduling and buffer acceptance mechanisms only rely on the DSCP of the packets
- simple buffer acceptance algorithm
- small number of queues
- scheduler with only a few queues

Roles of Border Routers
Supports same functions as ER, but usually on higher bandwidth links and with interdomain traffic
May also need to remark traffic when Diffserv domains A and B utilize different DSCPs for the same service

Traffic control and QoS in IP Networks Outline

- Applications
- Packet-level traffic control mechanisms
- Standardised Services
  - Integrated Services
  - Differentiated Services
    - The architecture
    - The services
    - The role of MPLS

The Differentiated services



PDB : Per-Domain-Behaviour or "edge-to-edge" service provided by a Diffserv network depends on :

- *Traffic conditioning functions* (shaping, policing, marking,...) located only on edge and border routers
- *Packet handling functions* (queuing, scheduling, buffer acceptance) located on each router
  - PHB : per hop behaviour in Diffserv terminology

K.Nichols and B.Carpenter. Definition of differentiated services per domain behaviors and rules for their specification. RFC 3086, April 2001.

The Per-Hop Behaviours

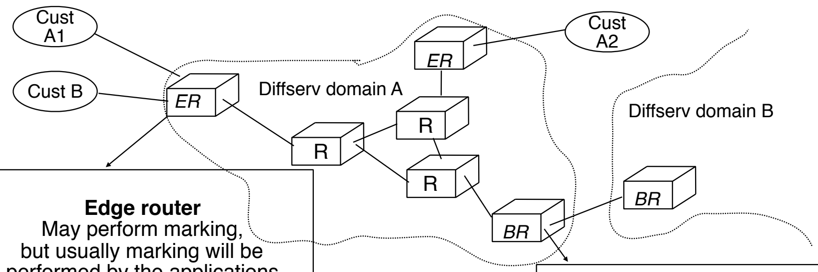
- Defined PHBs
  - Best-effort service
    - classical router behaviour
  - Class Selector
    - backward compatibility with Precedence field
  - Expedited forwarding
    - should provide across a diffserv domain a service with
      - low packet loss ratio
      - low end-to-end packet delay
      - low delay jitter
  - Assured forwarding
    - reserve different amount of resources (bandwidth, buffers) for 4 classes of traffic in each router
    - classes of traffic should be served independently

Class Selector PHB

- Objective
  - Backward compatibility with IP Precedence
  - 8 different class selectors
- Definition
  - Use of class selector should yield to at least 2 independently forwarded traffic classes
  - Router should give a higher probability of timely forwarding to CS_x packets than CS_y packets if x>y
  - Packets with CS6 and CS7 should receive a better treatment than best-effort traffic
    - several routing protocols use CS6/7 IP packets

The Class Selector PHB corresponds to packets with DSCP=xxx000 (i.e. The three high order bits are used like the former Precedence bis)

Sample CS implementations



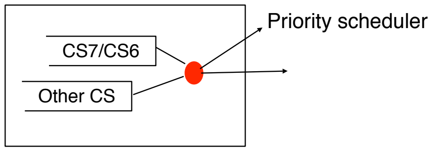
Edge router
May perform marking,
but usually marking will be
performed by the applications
If CS7/CS6 is used used by routing
protocols inside domain A, edge
routers should block or remark
incoming packets with CS7/CS6

Border router
Same as edge router + remarking

Per Hop Behavior
Rely on scheduler to timely serve packets
with higher Class Selectors

Sample CS implementations (2)

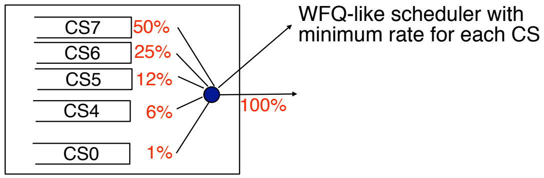
- Priority-based implementation



- Advantages
  - Simple to implement
  - CS7/CS6 are timely served
- Drawbacks
  - CS7/CS6 traffic may starve other traffic
    - assume BGP uses CS7 packets and BGP router reloads...
    - be careful in case of DoS attacks with CS7/CS6
  - Difficult to generalise to N CS codepoints
    - but only CS7/CS6 are used in practice

Sample CS implementations (3)

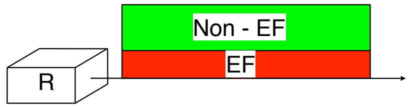
- WFQ-based implementation



- Advantages
  - Starvation is not possible anymore
  - Packets with high CS are timely served
- Drawbacks
  - more configuration is required than with PQ
  - needs some knowledge of traffic mix

EF : Expedited Forwarding

- Objective
  - Provide low delay, low loss, low jitter service
- Principle



  - On each node, the administrator should be able to configure a rate R_{EF} for EF traffic
  - The EF traffic should be served at rate $>= R_{EF}$ independently of the intensity of any other traffic through the node

The original definition

V.Jacobson, K.Nichols, and K.Poduri. An expedited forwarding PHB. Internet RFC2598, June 1999.

The updates to clarify (notably mathematically the EF specification)

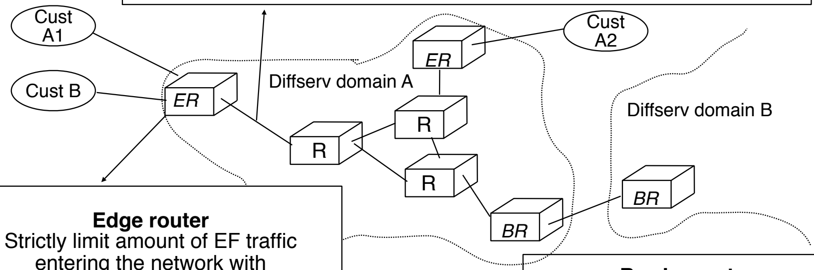
B.Davie, A.Charny, F.Baker, J.Bennet, J.-Y. Leboudec, K.Benson, A.Chiu, W.Courtney, S.Davari, V.Firoiu, C.Kalmanek, K.K. Ramakrishnan, and D.Stiliadis. An expedited forwarding PHB. RFC 3246, March 2002.

G.Armitage, A.Casati, J.Crowcroft, J.Halpern, B.Kumar, and J.Schnizlein. A delay bound alternative revision of RFC2598. RFC3248, March 2002.

A.Charny, F.Baker, J.Bennet, B.Davie, K.Benson, A.Chiu, W.Courtney, S.Davari, V.Firoiu, C.Kalmanek, K.K. Ramakrishnan, and D.Stiliadis. Supplemental information for the new definition of the EF PHB. RFC 3247 March 2002.

Sample EF implementations

On each link
Provision link so that amount of EF traffic is limited to small fraction of bandwidth



Edge router
Strictly limit amount of EF traffic entering the network with
- policing (with dropping)
  - simple for the routers, but may cause bursts inside network
- shaping (preferred)
  - complex for the routers, but reduces bursts inside network

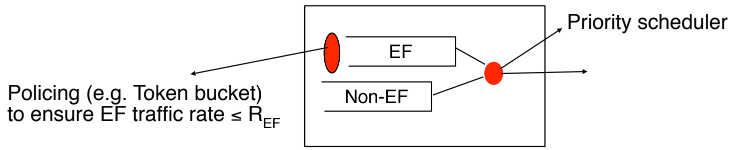
based on the customer's SLA

Border router
Same as edge router + remarking

Per Hop Behaviour
Determine R_{EF} for each link at provisioning
Configure scheduler on each node so that EF traffic is always served at rate >= R_{EF}

Sample EF implementations (2)

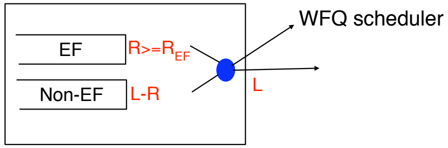
- Priority-based implementation



- Comments
  - provides smallest possible delay for EF packets
  - traffic on each link should be regularly monitored to ensure that EF rate is not larger than R_{EF}
    - if the network is well engineered, no packets will be dropped by EF policer on interior nodes
    - if EF packets are dropped inside interior nodes, this might be due to over subscription or network problems (DoS, configuration errors, ...)
      - EF drops should be monitored

Sample EF implementations (3)

- WFQ-based implementation



- Comments
  - provides larger queuing delay than PQ
    - average delay and delay jitter will decrease when R increases
  - policing is not required
    - policing (at R_{EF}) should be used when $R >> R_{EF}$
  - traffic on each link should be regularly monitored to ensure tha EF rate is not larger than R_{EF}
    - EF drops from EF queue should be monitored

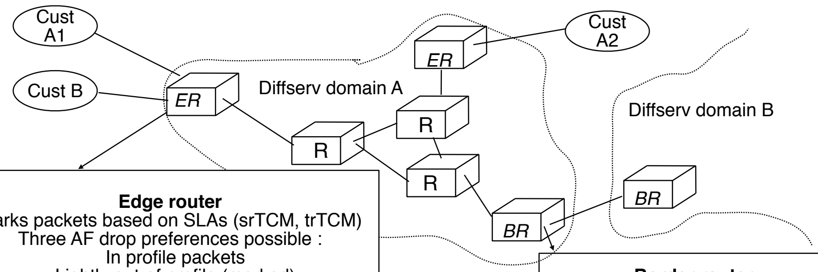
AF : Assured Forwarding

- Goal
  - provide differently provisioned services on top of a single IP network
    - business service with large over provisioning
    - premium ADSL service correctly provisioned
    - under provisioned cheap ADSL service
- Principle
  - Colour assigned to each traffic class
  - edge routers colour packets based on their class
  - network resources are divided into a few classes
    - business service has a lot of resources to avoid congestion
    - premium ADSL has fewer resources, but can benefit from unus business resources
    - cheap ADSL service utilises the leftover resources

J.Heinanen, F. .Baker, W.Weiss, and J.Wrocklawski. Assured forwarding PHB group. Internet RFC 2597, June 1999.

N.Seddihg, B.Nandy, and J.Heinanen. An assured rate per-domain behaviour for differentiated services. Internet draft, draft-ietf-diffserv-pdb-ar-00.txt, work in progress, February 2001.

Sample AF implementation



Edge router
marks packets based on SLAs (srTCM, trTCM)
Three AF drop preferences possible :
- In profile packets
- Lightly out-of profile (marked)
- Severely out-of-profile (marked or dropped)

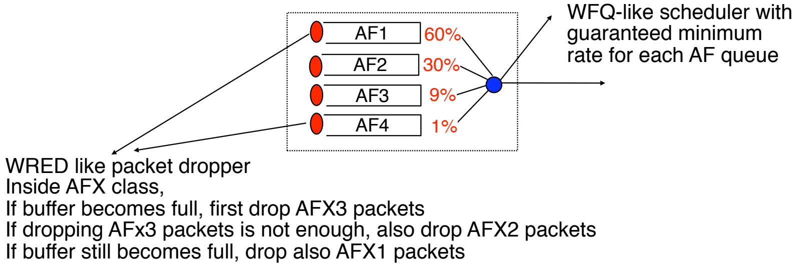
Border router
Same as edge router + remarking

Per Hop Behaviour
Four AF classes can be supported
- AF1 should rarely see congestion
- AF2 should see more congestion than AF1
- AF3 should see more congestion than AF1 and AF2

...
Inside each AF class
- first drop severely out-of-profile packets
- then drop lightly out-of-profile packets
- only drop in-profile packets if dropping out-of-profile is not enough

Sample AF implementation (2)

- Typical AF implementation



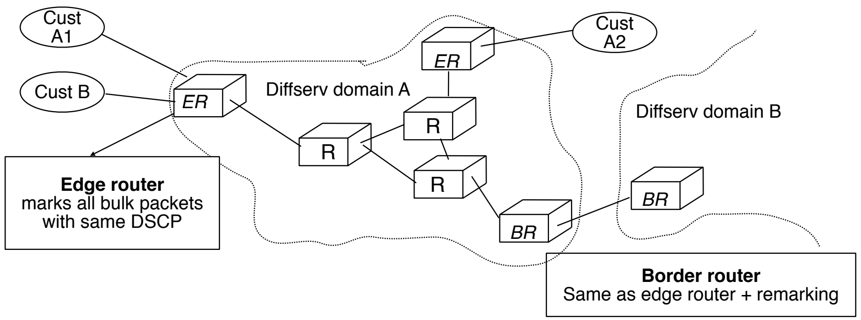
- Comments
  - a network does not need to support the 4 AF classes and the 3 drop preferences inside each class
  - weight associated to each queue depends on expected amoun AF traffic in each class

Bulk Handling PDB

- Objectives
  - Force bulk traffic to be handled during off-peak
    - Non-important applications should only consume bandwidth during off-peak hours
      - example : `netnews` for ISP or company
    - Non-important users should be discouraged to utilise network resources during peak hours
      - example : free ISP user
- Deployment
  - based on Class Selector PHB
  - based on Assured Forwarding PHB

B.Carpenter and K.Nichols. A bulk handling per-domain behavior for differentiated services. Internet draft,draft-ietf-diffserv-pdb-bh-00.txt, work in progress, October 2000.

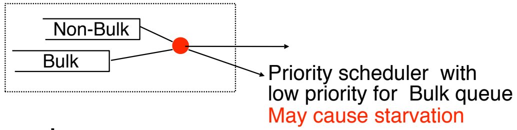
Supporting Bulk Handling



Per Hop Behaviour
Serve BH packets only if nothing needs to be served

Supporting Bulk Handling (2)

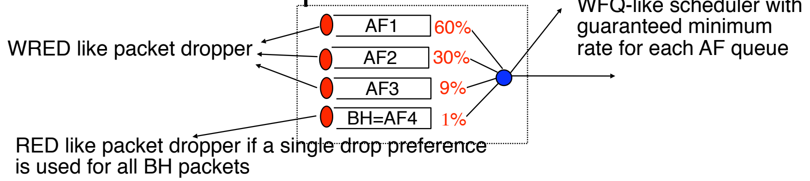
- Priority-based implementation



- Comments
  - BH packets will only be served when no other packets needs to be transmitted
    - nice from a network utilisation point of view
  - may cause starvation that could lead to problems with TCP-based applications
    - after N unsuccessful retransmissions of the same packet during starved periods, TCP will terminate the connection
    - some applications may not be ready to deal with terminated connections

Supporting Bulk Handling (3)

- AF-based implementation



- Comments
  - Assumes that AF4 class can be used
    - set rate of BH queue at lowest value supported by scheduler
      - BH packets always utilise some bandwidth, but this is usually better tha PQ with starvation for TCP if AF4 queue is large
    - utilise RED-like packet dropper for AF4 queue
  - If all four AF classes or queues are already used, alternative is
    - use AF43 for BH packets and place BH packets in AF4 queue and drops them as soon as queue fills

Back to schedulers

- Bandwidth distribution for most schedulers

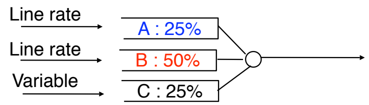
$$rate[i] = Bandwidth \times \left(\frac{(Weight[i])}{(\sum_{j=queues} Weight[j])} \right)$$

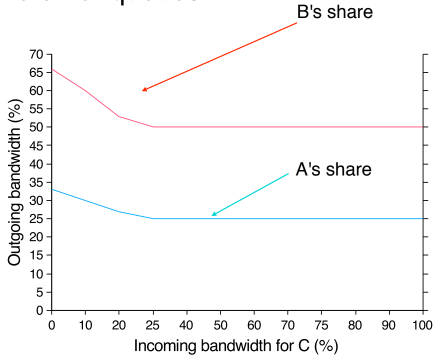
  - provided all queues are active

- What happens if a queue is not active ?
  - The bandwidth unused by this queue is distributed among the active queues in proportion to their weight
    - few schedulers allow to control how the leftover bandwidth is distributed among the active queues

Back to schedulers (2)

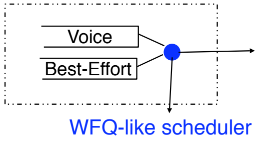
- Consequence
  - the bandwidth allocated by a scheduler to a queue is thus function of
    - the weight associated to the queue
    - the amount of traffic in the other queues





Supporting Voice with AF

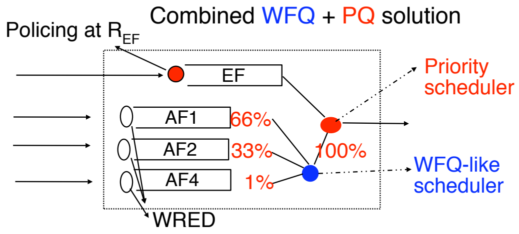
- Is it possible to use AF for Voice traffic ?
- Characteristics of Voice traffic
  - requires a low delay
  - volume of traffic is bounded by sources or policing
- Example
  - Voice traffic consumes 10% of bandwidth



- Bandwidth-based rate allocation
  - Voice : 10 % of bandwidth
  - Best-Effort : 90% of bandwidth
  - no voice packets will be lost
  - delay for voice will be high
- Delay-based rate allocation
  - Voice : 95% of bandwidth
  - Best-effort : 5% of bandwidth
  - only works if voice traffic is bounded

Combining AF and EF

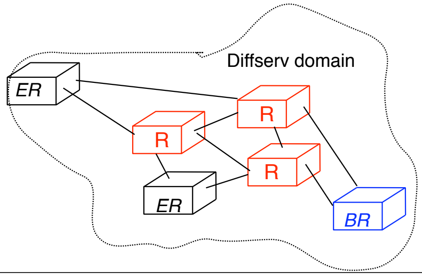
- Principle
  - Rely on two schedulers



- Comments
  - EF will obtain low delay thanks to PQ scheduler
    - policing of EF is required to ensure that AF will be not starved by AF
  - WFQ scheduler will serve AF once EF has been served
    - allocation between AF classes depends on WFQ weights

Can we change routing protocols ?

- The Diffserv assumption was that routing protocols cannot be changed to support QoS
  - ToS based routing was part of OSPFv1
    - but only one vendor implemented it
- However a small but important change was recently introduced in IS-IS and OSPF
  - Multi-topology routing



OSPF multitopology is described in

P. Psenak, S. Mirtorabi, A Roy, L. Nguyen, P. Pillay-Esnault, Multi-Topology (MT) Routing in OSPF, RFC4915, 2007

Traffic control and QoS in IP Networks Outline

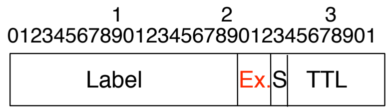
- Applications
- Packet-level traffic control mechanisms
- Standardised Services
  - Integrated Services
  - Differentiated Services
    - The architecture
    - The services
    - The role of MPLS

MPLS and IP QoS

- What could be the benefit of MPLS to support IP QoS ?
  - With differentiated or integrated services, the path followed by IP packets is independent of their QoS since those architectures did not change routing
  - When MPLS is used, IP packets with distinct QoS requirements may be placed inside distinct LSPs that follow different paths inside the network
    - MPLS allows to utilize distinct routes for packets with distinct QoS requirements

MPLS and Differentiated Services

- How can we support Differentiated services inside a MPLS network ?
- LSR must know QoS required by each packet
  - Two complementary methods
    - Indicate the QoS required by all packets of a given LSP at LSP establishment time (e.g. With special RSVP objects)
    - Encode QoS info inside MPLS label



- EXP field is part of MPLS header and contains 3 bits
- special RSVP objects can be used to map Exp field with DSCP

The MPLS support of Diffserv is discussed in :

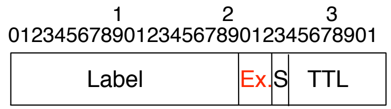
F.Le Faucheur, L.Wu, B.DavieS. Davari, P.Vaananen, R.Krishnan, P.Cheval, and J.Heinanen. Mpls support for differentiated services. Internet draft, draft-ietf-mpls-diff-ext-09.txt, work in progress, April 2001.

Other information may be found in :

S.Ganti, S.Seddigh, and B.Nandy. Mpls support of differentiated services using e-lsp. Internet draft, draft-ganti-mpls-diffserv-elsp-00.txt, work in progress, April 2001.

Diffserv support with MPLS

- The "*IP*" way, aka E-LSPs
  - a single LSP may carry packets receiving several differentiated services
  - each MPLS router relies on EXP header field to determine the service for each received packet
    - EXP field is part of MPLS header and contains 3 bits



    - useful to reduce the number of required LSPs in large networks

Diffserv support with MPLS (2)

- The "*ATM*" way, aka L-LSPs
  - one LSP carries packets receiving a single service
    - the EXP field of the MPLS header may be used to specify the drop preference for each packet (e.g. For AF)
    - MPLS router decides service for a receiving packet based on label value
  - the service used by the LSP is specified at LSP establishment time
  - useful when number of LSPs is not a constraint
  - each L-LSP may have its own explicit route
  - more L-LSPs than E-LSPs will be needed