INF5070 – Media Servers and Distribution Systems:

Protocols with QoS Support

26/9 - 2005

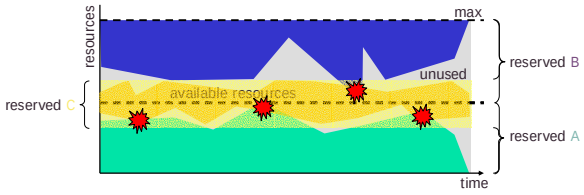
Overview

- Quality-of-Service
- Per-packet QoS
  - IP
- Per-flow QoS
  - Resource reservation
  - Tenet, ST-II, RSVP
- QoS Aggregates
  - DiffServ, MPLS
  - Network Calculus

Quality-of-Service

Quality–of–Service (QoS)

- Different semantics or classes of QoS:
  - determines reliability of offered service
  - utilization of resources



Quality–of–Service (QoS)

- Best effort QoS:
  - system tries its best to give a good performance
  - no QoS calculation (could be called *no effort QoS*)
  - ☺ simple – do nothing
  - ☹ QoS may be violated → unreliable service
- Deterministic guaranteed QoS:
  - hard bounds
  - QoS calculation based on upper bounds (worst case)
  - *premium* better name!!??
  - ☺ QoS is satisfied even in the worst case → high reliability
  - ☹ over-reservation of resources → poor utilization and unnecessary service rejects
  - ☹ QoS values may be less than calculated hard upper bound

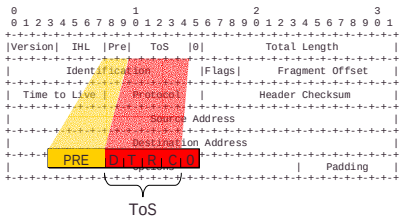
Quality–of–Service (QoS)

- Statistical guaranteed QoS:
  - QoS values are statistical expressions (served with some probability)
  - QoS calculation based on average (or some other statistic or stochastic value)
  - ☺ resource capabilities can be statistically multiplexed → more granted requests
  - ☹ QoS may be temporarily violated → service not always 100 % reliable
- Predictive QoS:
  - weak bounds
  - QoS calculation based previous behavior of imposed workload

Per-packet QoS

Internet Protocol version 4 (IPv4)

[RFC1349]



ToS

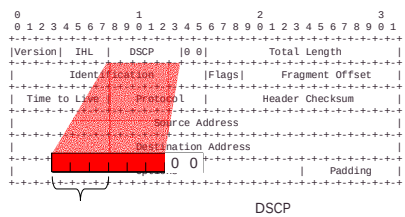
- Type of Service
  - D – minimize delay
  - T – maximize throughput
  - R – maximize reliability
  - C – minimize cost

PRE

- Precedence Field
  - Priority of the packet

Internet Protocol version 4 (IPv4)

[RFC2474]

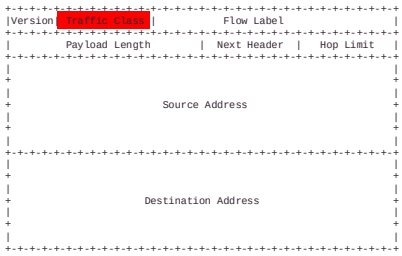


Class selector codepoints of the form xxx000

DSCP

- Differentiated Services Codepoint
  - xxxxx0 reserved for standardization
  - xxxx11 reserved for local use
  - xxxx01 open for local use, may be standardized later

Internet Protocol version 6 (IPv6)



- Traffic class
  - Interpret like IPv4's DS field

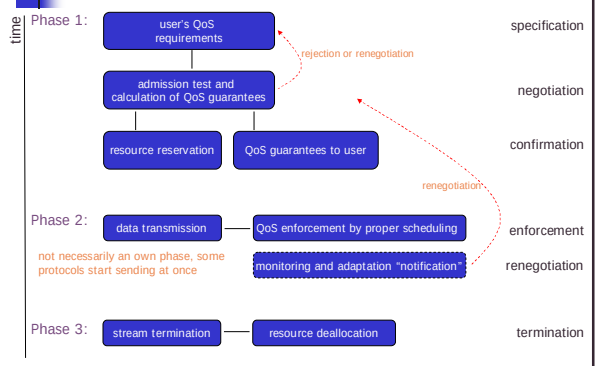
Per-flow QoS

Resource Reservation

Resource Reservation

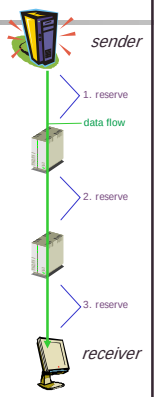
- Reservations is fundamental for reliable enforcement of QoS guarantees
  - per-resource data structure (information about all usage)
  - QoS calculations and resource scheduling may be done based on the resource usage pattern
  - reservation protocols
    - negotiate desired QoS by transferring information about resource requirements and resource usage between the end-systems and the intermediate systems participating in the data transfer
  - reservation operation
    - calculate necessary amount of resources based on the QoS specifications
    - reserve resources according to the calculation (or reject request)
  - resource scheduling
    - enforce resource usage with respect to resource administration decisions

Resource Management Phases



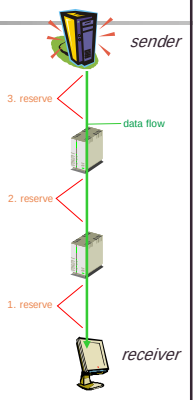
Reservation Directions

- Sender oriented:
  - sender (initiates reservation)
    - must know target addresses (participants)
    - in-scalable
    - good security



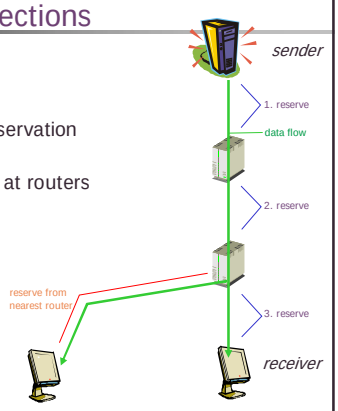
Reservation Directions

- Receiver oriented:
  - receiver (initiates reservation)
    - needs advertisement before reservation
    - must know "flow" addresses
  - sender
    - need not to know receivers
    - more scalable
    - in-secure



Reservation Directions

- Combination?
  - start sender oriented reservation
  - additional receivers join at routers (receiver based)



Per-flow QoS

Protocols

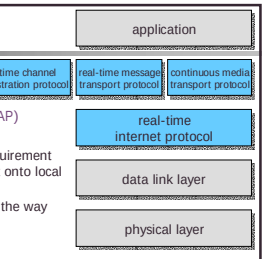
Tenet

- Late 80's/early 90's, Tenet group at Berkeley
- Aims for network support for real-time continuous media applications
- Real-time communication model
- Real-time channels: end-to-end connection with performance guarantees and traffic restrictions
  - associated with a set of nodes and links (a route) through which real-time packets pass
  - resource reservation in route nodes
  - admission control

Tenet

- Traffic specification
  - expressing peak and average load on the network
  - indication of the burstiness of the load
  - parameters
    - minimum packet inter-arrival time
    - average packet inter-arrival time
    - averaging interval
    - maximum packet size
- Supported QoS parameters (by which users describe their requirements)
  - upper bound on end-to-end message delay
  - delay violation probability bound
  - buffer overflow probability bound
  - delay jitter bound (optional)
  - a throughput guarantee is obtained from the traffic specification

Tenet



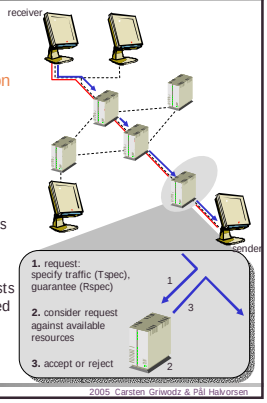
- Protocol suite:
- Real-time Channel Administration Protocol (RCAP)
  - performs channel setup
  - uses the traffic description and performance requirement to find a route and maps the global requirement onto local resources
  - performs admission control and reservations on the way
- Real-time Message Transport Protocol (RMTP)
  - intended for message based real-time transport
- Continuous Media Transport Protocol (CMTP)
  - offers a stream based interface and a time-driven mechanism for audio and video – may demand data from application
- Real-time Internet Protocol (RTIP)
  - replaces IP
  - schedules packets according to resource reservations made by RCAP

Integrated Services (IntServ)

- Framework by IETF to provide individualized QoS guarantees to individual application sessions
- Goals:
  - efficient Internet support for applications which require service guarantees
  - fulfill demands of multipoint, real-time applications (like video conferences)
  - do not introduce new data transfer protocols
- In the Internet, it is based on IP (v4 or v6) and RSVP (described later)
- Two key features
  - reserved resources – the routers need to know what resources are available (both free and reserved)
  - call setup (admission call) – reserve resources on the whole path from source to destination

Integrated Services (IntServ)

- Admission call:
  - traffic characterization and specification
    - one must specify the traffic one will transmit on the network (Tspec)
    - one must specify the requested QoS (Rspec – reservation specification)
  - signaling for setup
    - send the Tspec and Rspec to all routers
  - per-element admission test
    - each router checks whether the requests specified in the R/Tspecs can be fulfilled
    - if YES, accept; reject otherwise

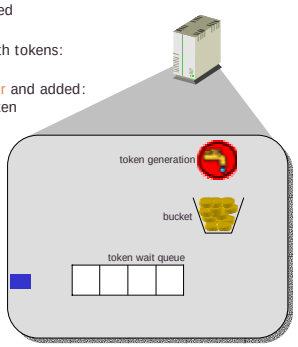


Integrated Services (IntServ)

- IntServ introduces two new services enhancing the Internet's traditional best effort:
  - guaranteed service
    - guaranteed bounds on delay and bandwidth
    - for applications with real-time requirements
  - controlled-load service
    - "a QoS closely to the QoS the same flow would receive from an unloaded network element" [RFC 2212], i.e., similar to best-effort in networks with limited load
    - no quantified guarantees, but packets should arrive with "a very high percentage"
    - for applications that can adapt to moderate losses, e.g., real-time multimedia applications

Integrated Services (IntServ)

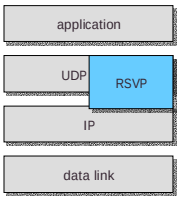
- Both service classes use token bucket to police a packet flow:
  - packets need a token to be forwarded
  - each router has a b-sized bucket with tokens: if bucket is empty, one must wait
  - new tokens are generated at a rate r and added: if bucket is full (little traffic), the token is deleted
  - the token generation rate r serves to limit the long term average rate
  - the bucket size b serves to limit the maximum burst size



Resource Reservation Protocol (RSVP)

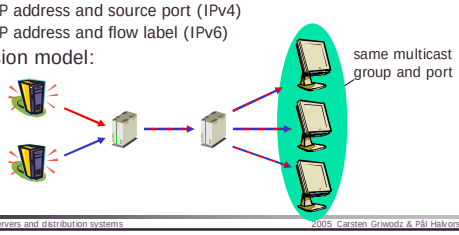
[RFC2205]

- A protocol to signal reservations of resources in the Internet
  - contains protocol elements for control
  - no support for data transfers
    - reservation signals only
  - simplex protocol
    - makes reservations for unidirectional flows
  - receiver-oriented
    - the receiver initiates and maintains resource reservations
  - maintains a "soft" state
    - graceful changes to dynamic memberships and automatic adaptation to route changes (timeouts)



Resource Reservation Protocol (RSVP)

- Sessions
  - a data flow with particular destination and transport protocol
  - defined by (destination address, protocol ID)
    - IP address
    - IP protocol ID
  - may carry multiple data flows
- Data flows are distinguished by
  - source IP address and source port (IPv4)
  - source IP address and flow label (IPv6)
- Transmission model:



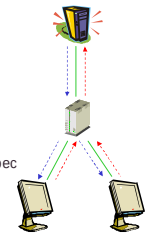
Resource Reservation Protocol (RSVP)

- Two fundamental messages
  - PATH:
    - sender sends a PATH message downstream following the data path
    - sent using same source and destination addresses
    - includes:
      - hop-addresses
      - sender template (describes data packet format)
      - sender Tspec (traffic characteristics generated by sender)
      - sender Adspec (advertisement information)
      - ...
  - RESV:
    - receiver sends a RESV message upstream using the path described in the PATH message
    - sent to previous hop only
    - includes:
      - flowspec: reservation requests, desired QoS (e.g., RFC 1363)
      - filterspec: reservation style
      - reverse data paths for the flow
      - ...

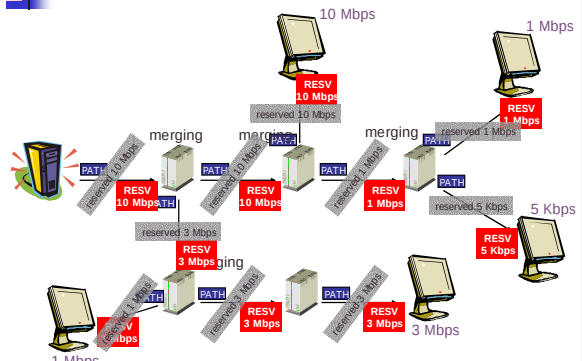
} flow descriptor

Resource Reservation Protocol (RSVP)

- Creating and maintaining a reservation state
  - the *SOURCE*
    - multicasts data flows
    - sends PATH messages with traffic characteristics (Tspec) describing flows
  - the *RECEIVER*
    - joins multicast group
    - receives the PATH message
    - determines own QoS requirements based the PATH Tspec
    - sends a RESV message with request and filters
  - the *ROUTERS*
    - reserve according to incoming flowspecs downstream
    - merge and forward the RESV messages to next node using largest flowspec
  - the reservations are maintained using "soft" states
    - the reservation has an associated timer – a timeout removes the reservation
    - periodically refreshed by PATH and RESV messages

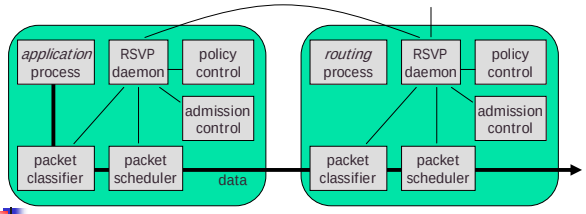


Resource Reservation Protocol (RSVP)



Resource Reservation Protocol (RSVP)

- RSVP in hosts and routers:
  1. RESV with flowspec and filterspec to RSVP daemon
  2. policy control to check privileges etc.
  3. admission control using flowspec
  4. forward RESV message
  5. control of local modules: classifier and scheduler



Resource Reservation Protocol (RSVP)

- Reservation styles
  - a reservation request includes a set of options called the reservation *style*
  - shared vs. distinct reservations
    - concerns treatment of reservations of different senders
    - shared – single reservation for all senders (e.g., video conference audio)
    - distinct – one reservation per sender (e.g., video conference video)
  - explicit vs. wildcard
    - concerns selection of senders
    - explicit – specify senders (e.g., teleteaching)
    - wildcard – automatically select all senders (e.g., video conference)

Resource Reservation Protocol (RSVP)

	distinct reservation	shared reservation
explicit sender selection		
wildcard sender selection		

Resource Reservation Protocol (RSVP)

- The RSVP standard [RFC 2205] allows to reserve link bandwidth – it does **NOT...**:
  - ...define how the network should provide the reserved bandwidth to the data flows – the routers must implement these mechanisms themselves
  - ...specify how to do resource provisioning – which must likely be done using a proper scheduling mechanism
  - ...determine the route – it is not a routing protocol, but relies on others
  - ...determine which data to drop in case of overflow, i.e., the most important data may be lost
  - ...perform an admission test, but it assumes that the routers perform admission control
- **THUS; RSVP can only be used as a small piece in the QoS guarantee puzzle**

Kurose, J. F., Ross, K. W.: "*Computer Networking: A Top-Down Approach Featuring the Internet*", 2nd edition, Addison Wesley, 2002

Resource Reservation Protocol (RSVP)

- Criticism
  - Complexity of protocol elements
    - Number of states on routers proportional to number of sessions
    - Keeping PATH and RESV states in each router
    - Merge processing
    - Reservation styles for multicast
  - Implementation-specific overhead
    - Two sending styles: protocol 46 in IP or encapsulation in UDP
    - Implementation usually in user space demons

QoS Aggregates

Protocols

Differentiated Services (DiffServ)

- IntServ and RSVP provide a framework for per-flow QoS, but they ...
  - ... give complex routers
    - much information to handle
  - ... have scalability problems
    - set up and maintain per-flow state information
    - periodically PATH and RESV messages overhead
  - ... specify only a predefined set of services
    - new applications may require other flexible services

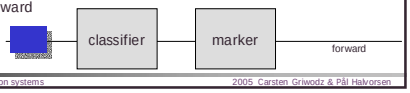
⇨ DiffServ [RFC 2475] tries to be both scalable and flexible

Differentiated Services (DiffServ)

- ISP favor DiffServ
- Basic idea
  - multicast is not necessary
  - make the core network simple due to many users
  - implement more complex control operations at the edge
  - aggregation of flows – reservations for a group of flows, not per flow
  - ⇨thus, avoid scalability problems on routers with many flows
  - do not specify services or service classes
  - instead, provide the functional components on which services can be built
  - ⇨thus, support flexible services

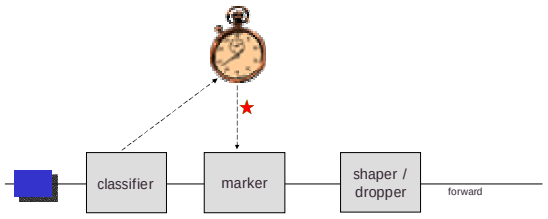
Differentiated Services (DiffServ)

- Two set of functional elements:
  - edge functions: packet classification and traffic conditioning
  - core function: packet forwarding
- At the edge routers, the packets are tagged with a DS-mark (differentiated service mark)
  - uses the type of service field (IPv4) or the traffic class field (IPv6)
  - different service classes (DS-marks) receive different service
  - subsequent routers treat the packet according to the DS-mark
  - classification:
    - incoming packet is classified (and steered to the appropriate marker function) using the header fields
    - the DS-mark is set by marker
    - once marked, forward



Differentiated Services (DiffServ)

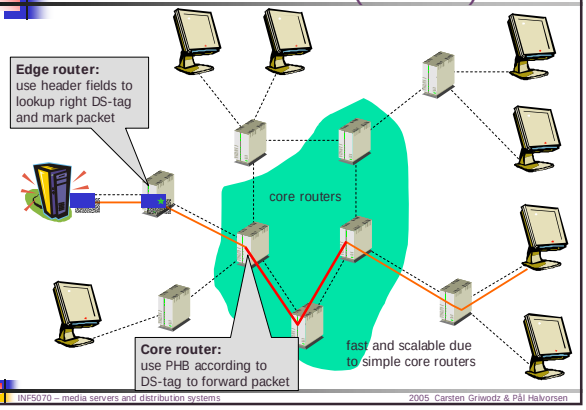
- Note, however, that there is no "rules" for classification – it is up to the network provider
- A metric function may be used to limit the packet rate:
  - the traffic profile may define rate and maximum bursts
  - if packets arrive too fast, the metric function assigns another marker function telling the router to delay or drop the packet



Differentiated Services (DiffServ)

- In the core routers, a DS-marked packet is forwarded according to a per-hop behavior (PHB) associated with the DS-tag
  - the PHB determines how the router resources are used and shared among the competing service classes
  - the PHB should be based on the DS-tag only
  - traffic aggregation
    - packets with same DS-tag are treated equally
    - regardless of source or destination
  - a PHB can result in different service classes receiving different performance
  - performance differences must be observable and measurable to be able to monitor the system performance
  - no specific mechanism for achieving these behaviors are specified

Differentiated Services (DiffServ)



Differentiated Services (DiffServ)

- Currently, two PHBs are under active discussion
  - expedited forwarding [RFC 3246]
    - specifies a *minimum* departure rate of a class, i.e., a guaranteed bandwidth
    - the guarantee is independent of other classes, i.e., enough resources must be available regardless of competing traffic
  - assured forwarding [RFC 2597]
    - divide traffic into four classes
    - each class is guaranteed a minimum amount of resources
    - each class are further partitioned into one of three "drop" categories (if congestion occur, the router drops packets based on "drop" value)

Multiprotocol Label Switching (MPLS)

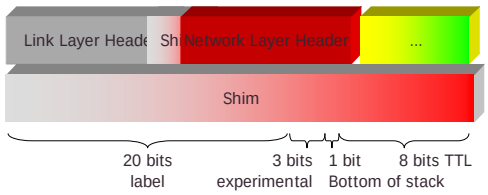
- Multiprotocol Label Switching
  - Separate path determination from hop-by-hop forwarding
  - Forwarding is based on labels
  - Path is determined by choosing labels
- Distribution of labels
  - On application-demand
    - LDP – label distribution protocol
  - By traffic engineering decision
    - RSVP-TE – traffic engineering extensions to RSVP

Multiprotocol Label Switching (MPLS)

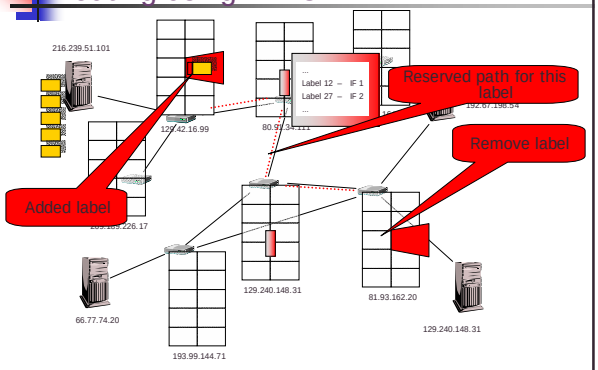
- MPLS works above **multiple** link layer **protocols**
- Carrying the **label**
  - Over ATM
    - Virtual path identifier or Virtual channel identifier
    - Maybe shim
  - Frame Relay
    - data link connection identifier (DLCI)
    - Maybe shim
  - Ethernet, TokenRing, ...
    - Shim
- Shim?

Multiprotocol Label Switching (MPLS)

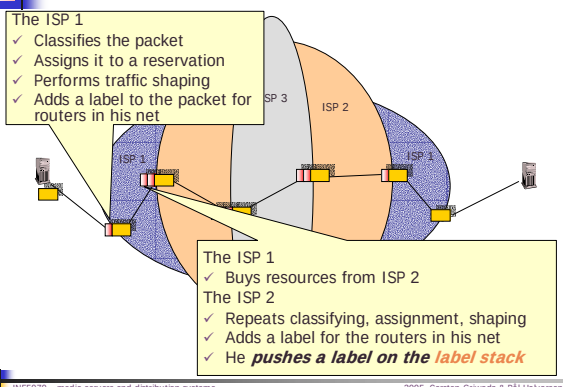
- Shim: the label itself



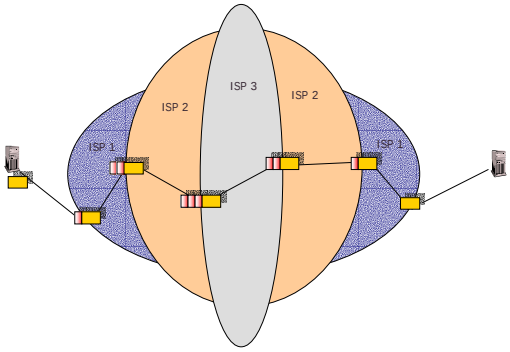
Routing using MPLS



MPLS Label Stack



MPLS Label Stack



Generalized Multi-Protocal Label Switching

- Classes of label switched routers
  - Packet-switch capable interfaces
    - Interfaces that recognize packet/cell boundaries
      - Forwarding based on the shim
      - e.g. ATM VPI/VCI
  - Time-division multiplex capable interfaces
    - Interfaces that forward data based on a time slot
      - e.g. SONET/SDH cross-connect
  - Lambda-switch capable interfaces
    - Interfaces that forward data based on the wavelength on which data is received
      - e.g. optical cross-connects that operates on wavelength
  - Fiber-switch capable interfaces
    - Interfaces that forward data based on physical link it arrives on
      - e.g. optical cross-connects that operates on fibers

RSVP-TE

[RFC3209]

- Traffic Engineering extensions for RSVP
- Goal
  - Use RSVP as a signaling protocol
  - Establish an explicitly route path by setting up MPLS labels
    - a "label-switched path"
  - Keep soft-state semantics of RSVP
    - Automatic routing away from failures, congestion and bottlenecks
- Extensions
  - Reserve for labels, not for address tuples
  - EXPLICIT_ROUTE object
    - Allows the creation of LSP tunnels
    - Object includes IP addresses or AS numbers for which a tunnel is valid

RSVP-TE

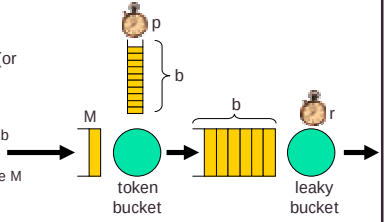
- Improvements
  - Fuzzy timer management
    - Timers below 10ms need not be sorted
    - Improvement: processing reduced by 4-11%
  - Dedicated memory management
    - Use free lists
    - Improvement: processing reduced by 16-18%
  - Refresh reduction
    - Summary refresh messages
    - Distribute refresh messages uniformly over the refresh interval
    - Improvement: processing reduced by 69%, memory use increased by 11%

QoS Aggregates

Network Calculus

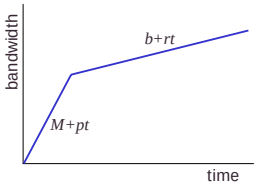
Using Network Calculus

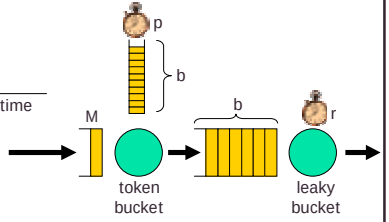
- Guaranteed Service
  - An assured level of bandwidth
  - A firm end-to-end delay bound
  - No queuing loss for data flows that conform to a TSpec
- TSpec
  - Describes how traffic arrives from the user in the worst case
  - Double token bucket (or combined token bucket/leaky bucket)
    - Token bucket rate r
    - Token bucket depth b
    - Peak rate p
    - Maximum packet size M



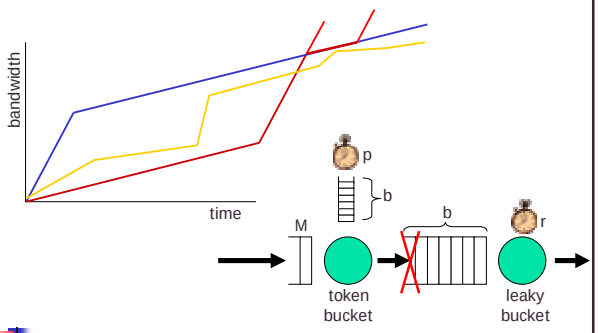
Using Network Calculus

arrival curve: $a(t) = \begin{cases} M + pt & t < \frac{b-M}{p-r} \\ b + rt & t \geq \frac{b-M}{p-r} \end{cases}$





Using Network Calculus



Using Network Calculus

- Service curve
  - The network's promise
  - Based on a "fluid model"

Service curve: $c(t) = R(t-V)^+$

Service rate: $R \geq r$

Deviations: $V = \frac{C}{R} + D \approx D$

$R \geq p \geq r \quad d_{\max} = \frac{M}{R} + D$

$p \geq R > r \quad d_{\max} = \frac{(b-M)(p-R)}{R(p-r)} + \frac{M}{R} + D$ Delays in the network

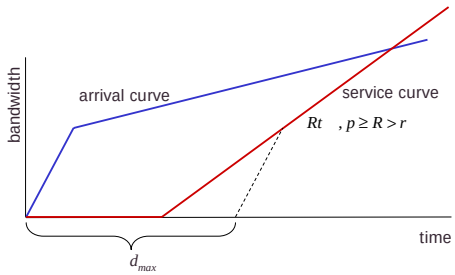
But: delay d_{max} is usually part of the user-network negotiation

Required service rate dependent on requested d_{max}

$R \geq p \geq r \quad R = \frac{M}{d_{\max} - D}$

$p \geq R > r \quad R = \frac{p\frac{b-M}{p-r} + M}{d_{\max} + \frac{b-M}{p-r} - D}$

Using Network Calculus

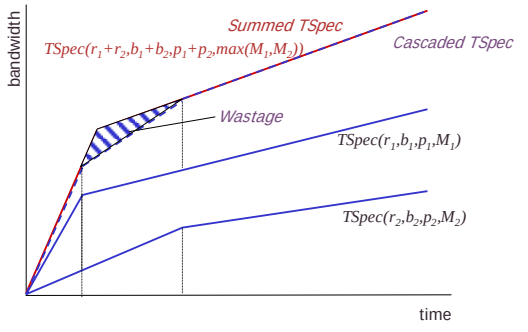


Using Network Calculus

- Using network calculus to scale
- Aggregation
  - Less state in routers
    - One state for the aggregate
  - Share buffers in routers
    - Buffer size in routers depends on the TSpec's rates
  - Use scheduling to exploit differences in d_{max}
    - Schedule flows with low delay requirements first

Using Network Calculus

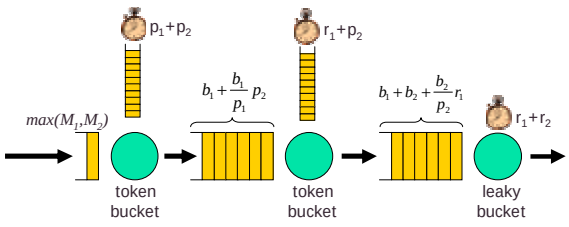
Aggregation



Using Network Calculus

Aggregation

Cascaded TSpec: n+1 token buckets



Summary

Directions of Network QoS

[Liebeherr]

- Old-style QoS is dead
  - ATM, IntServ, DiffServ, Service overlays didn't take hold
  - Causes?
    - No business case
    - Bothed standardization
    - Naïve implementations
    - No need
- Future QoS
  - Look for fundamental insights
  - Develop design principles
  - Develop analytical tools
    - Network calculus

[Crowcroft,Hand,Mortier,Roscoe,Warfield]

- Old-style QoS is dead
  - X.25 too little, too early
  - ATM too much, too late
  - IntServ too much, too early
  - DiffServ too little, too late
  - IP QoS not there
  - MPLS too isolated
- QoS through overlays can't work
- Future QoS
  - Single bit differentiation
  - Edge-based admission control
  - Micropayment

Dire

[Liebeherr

- Old-style Q
  - ATM, IntServ, DiffServ, Service o hold
  - Causes?
  - No busin
    - Bothe
    - Naïve
    - No ne
- Future QoS
  - Look for
  - Develop
  - Develop
    - Netwo

,Roscoe,Warfield]

is dead
, too early
h, too late
uch, too early
ttle, too late
ere
ated

verlays can't

erentiation
dmission control
t

- Companies do provide QoS
  - AT&T
    - MPLS
  - Equant
    - MPLS
  - Cable and Wireless
    - ATM
    - MPLS
  - TeliaSonera
    - SDH
    - WDM
    - ATM
  - Nortel
    - MPLS
    - SONET/SDH
    - WDM

Summary

- Timely access to resources is important for multimedia application to guarantee QoS – reservation might be necessary
- Many protocols have tried to introduce QoS into the Internet, but no protocol has yet won the battle...
  - often NOT only technological problems, e.g.,
    - scalability
    - flexibility
    - ...
  - but also economical and legacy reasons, e.g.,
    - IP rules – everything must use IP to be useful
    - several administrative domains (how to make ISPs agree)
    - router manufacturers will not take the high costs (in amount of resources) for per-flow reservations
    - pricing
    - ...

Some References

1. ATM Forum, http://www.atmforum.com
2. ATM Forum: "ATM Service Categories", http://www.atmforum.com/aboutatm/6.html
3. Danthine, A., Bonaventure, O.: "*From Best Effort to Enhanced QoS*", in: Spaniol, O., Danthine, A., Effelsberg, W. (Eds.): "Architecture and Protocols for High-Speed Networks", Kluwer Achademic Publishers, 1994, pp. 179-201
4. Kurose, J. F., Ross, K. W.: "*Computer Networking: A Top-Down Approach Featuring the Internet*", 2nd edition, Addison Wesley, 2002
5. Steinmetz, R., Nahrstedt, C.: "*Multimedia: Computing, Communications & Applications*", Prentice Hall, 1995
6. Tenet group: http://tenet.berkeley.edu/tenet.html

- The RFC repository maintained by the IETF Secretariat can be found at http://www.ietf.org/rfc.html
  The following RFCs might be interesting with respect to this lecture:
  - RFC 791: Internet Protocol
  - RFC 1883: Internet Protocol, Version 6 (IPv6)
  - RFC 2460: Internet Protocol, Version 6 (IPv6), Obsoletes: 1883
  - RFC 2212: Specification of Guaranteed Quality of Service
  - RFC 2205: Resource ReSerVation Protocol (RSVP)
  - RFC 1363: A Proposed Flow Specification
  - RFC 2475: An Architecture for Differentiated Services
  - RFC 3246: An Expedited Forwarding PHB (Per-Hop Behavior)
  - RFC 2597: Assured Forwarding PHB Group
  - RFC 1190: Experimental Internet Stream Protocol, Version 2 (ST-II)
  - RFC 1819: Internet Stream Protocol Version 2 (ST2): Protocol Specification - Version ST2+