Chair for Network Architectures and Services – Prof. Carle
Department for Computer Science
TU München

Master Course Computer Networks IN2097

Prof. Dr.-Ing. Georg Carle
Christian Grothoff, Ph.D.
Dr. Nils Kammenhuber

Chair for Network Architectures and Services
Institut für Informatik
Technische Universität München
http://www.net.in.tum.de

Chapter 6 outline – Quality-of-Service Support

6.1 Link virtualization: ATM
6.2 Providing multiple classes of service
6.3 Providing Quality-of-Service (QoS) guarantees
6.4 Signalling for QoS

Virtualization of networks

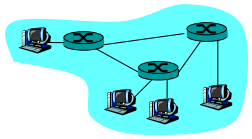
- Virtualization of resources: powerful abstraction in systems engineering:
- Computing examples: virtual memory, virtual devices
  - Virtual machines: e.g., java
  - IBM VM os from 1960's/70's
- Layering of abstractions: don't sweat the details of the lower layer, only deal with lower layers abstractly
- Virtualization of networks: hot topic also today, e.g. applicability for cloud computing

The Internet: virtualizing networks

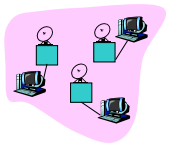
- 1974: multiple unconnected nets
  - ARPAnet
  - data-over-cable networks
  - packet satellite network (Aloha protocol)
  - packet radio network

… differing in:

- addressing conventions
- packet formats
- error recovery
- routing



ARPAnet



satellite net

"A Protocol for Packet Network Intercommunication",
V. Cerf, R. Kahn, IEEE Transactions on Communications,
May, 1974, pp. 637-648.

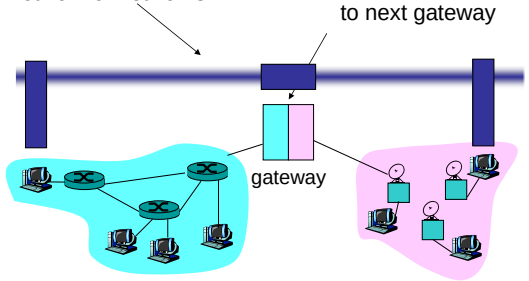
The Internet: virtualizing networks

Internetwork layer (IP):

- addressing: internetwork appears as single, uniform entity, despite underlying local network heterogeneity
- network of networks

Gateway:

- "embed internetwork packets in local packet format or extract them"
- route (at internetwork level) to next gateway



Cerf & Kahn's Internetwork Architecture

- What is virtualized?
- virtualization results in two layers of addressing: internetwork and local network
- new layer (IP) makes everything homogeneous at internetwork layer
- underlying local network technology
  - cable
  - satellite
  - 56K telephone modem
  - today: ATM, MPLS
- … "invisible" at internetwork layer.

ATM and MPLS

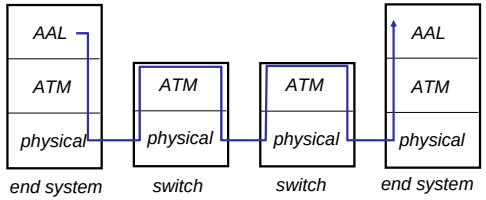
- ATM, MPLS separate networks in their own right
  - different
    - service models,
    - addressing,
    - routing

  from Internet
- viewed by Internet as logical link connecting IP routers
  - just like dialup link is really part of separate network (telephone network)
- ATM, MPLS: of technical interest in their own right

Asynchronous Transfer Mode: ATM

- ATM was 1990's/00 standard for high-speed (155Mbps to 622 Mbps and higher) *Broadband Integrated Service Digital Network* architecture
- Goal: *integrated, end-end transport to carry voice, video, data*
  - meeting timing/QoS requirements of voice, video (versus Internet best-effort model)
  - "next generation" telephony: technical roots in telephone world
  - packet-switching (fixed length packets, called "cells") using virtual circuits

ATM architecture

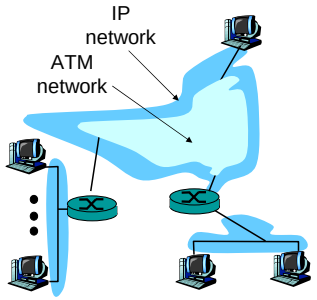


- AAL – ATM Adaptation Layer: only at edge of ATM network
  - data segmentation/reassembly
  - error detection and (optionally) error recovery
  - roughly analogous to Internet transport layer
- ATM layer: "network" layer
  - cell switching, routing
- physical layer

ATM: network or link layer?

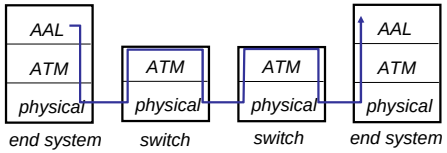
Vision: end-to-end transport: "ATM from desktop to desktop"
- ATM *is* a network technology

Reality: used to connect IP backbone routers
- "IP over ATM"
- ATM as switched link layer, connecting IP routers



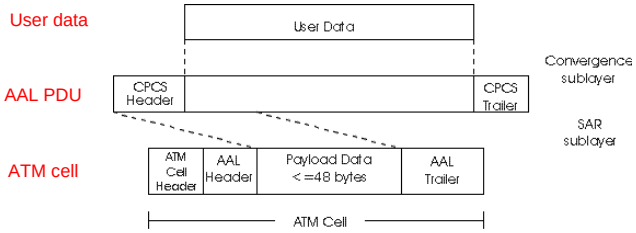
ATM Adaptation Layer (AAL)

- ATM **Adaptation Layer** (AAL): "adapts" upper layers (IP or native ATM applications) to ATM layer below
- AAL present in data plane **only in ATM end systems**, not in switches (however, signalling in switches needs AAL)
- AAL layer segment (header/trailer fields, data) fragmented across multiple ATM cells
  - analogy: TCP segment in many IP packets



ATM Adaptation Layer (AAL) [more]

Different versions of AAL layers, depending on ATM service class:
- AAL1: for CBR (Constant Bit Rate) services, e.g. circuit emulation
- AAL2: for VBR (Variable Bit Rate) services, e.g., MPEG video
- AAL5: for data (eg, IP datagrams)



(CPCS: Common Part convergence Sublayer)

ATM Layer

ATM Service: transport cells across ATM network

- analogous to IP network layer
- very different services than IP network layer

Network Architecture	Service Model	Guarantees ? Bandwidth	Loss	Order	Timing	Congestion feedback
Internet	best effort	none	no	no	no	no (inferred via loss)
ATM	CBR	constant rate	yes	yes	yes	no congestion
ATM	VBR	guaranteed rate	yes	yes	yes	no congestion
ATM	ABR	guaranteed minimum	no	yes	no	yes
ATM	UBR	none	no	yes	no	no

(ABR: Arbritrary Bit Rate
UBR: Unspecified Bit Rate)

ATM Layer: Virtual Circuits

- "source-to-destination path behaves much like telephone circuit"
  - performance-wise
  - network actions along source-to-destination path
- VC transport: cells carried on virtual circuit (VC) from source to destination
  - call setup, teardown for each call *before* data can flow
  - each packet carries VC identifier (not destination ID)
  - *every* switch on source-destination path maintain "state" for each passing connection
  - link, switch resources (bandwidth, buffers) may be *allocated* to VC: to get circuit-like performance
- Permanent VCs (PVCs)
  - long lasting connections
  - typically: "permanent" route between to IP routers
  - configuration by network management
- Switched VCs (SVC):
  - dynamically set up on per-call basis (signalling)

ATM VCs

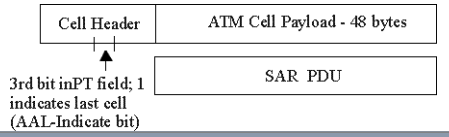
- Advantages of ATM VC approach:
  - QoS performance guarantee for connection mapped to VC (bandwidth, delay, delay jitter)
- Drawbacks of ATM VC approach:
  - Inefficient support of datagram traffic
  - one PVC between each source/destination pair) does not scale (N*2 connections needed)
  - SVC introduces call setup latency, processing overhead for short lived connections

ATM Layer: ATM cell

- 5-byte ATM cell header
- 48-byte payload
  - Why?: small payload ⇨ short cell-creation/transmission delay for digitized voice
  - halfway between 32 and 64 (compromise!)
- Benefit of cells over variable-length packets: avoiding that some packets must wait while a packet of maximum size is transmitted. (ATM is still attractive for slow links (e.g. access technologies such as DSL.)

Cell header

Cell format



ATM cell header

- **VPI/VCI:**
  - ID-space of virtual connections structured into Virtual Path Identifier (VPI) und Virtual Channel Identifier (VCI)
  - may *change* from link to link through network
- **PT:** Payload type
  - e.g. Resource Management (RM) cell versus data cell
- **CLP:** Cell Loss Priority bit
  - CLP = 1 implies low priority cell, can be discarded if congestion
- **HEC:** Header Error Checksum
  - cyclic redundancy check

Datagram or VC network: why?

Internet

- data exchange among computers
  - "elastic" service, no strict timing requirements
- "smart" end systems (computers)
  - can adapt, perform control, error recovery
  - simple inside network, complexity at "edge"
- many link types
  - different characteristics
  - uniform service difficult

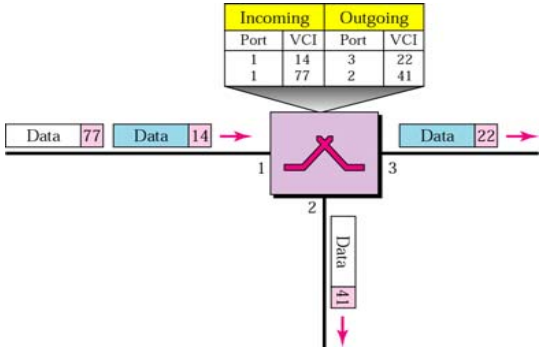
ATM

- evolved from telephony
- human conversation:
  - strict timing, reliability requirements
  - need for guaranteed service
- "dumb" end systems
  - telephones
  - complexity inside network

Virtual Circuits and Label Swapping

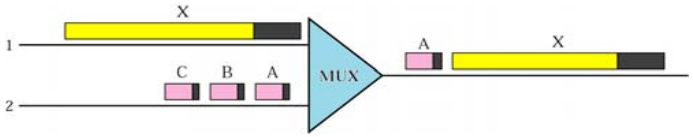
- Virtual Circuit Switching
- Multiplexing of Variable vs. Fixed Size Packets
- ATM Cell
- Virtual Path Identifiers and Virtual Channel Identifiers
- ATM Virtual Connections

Virtual Circuit Switching

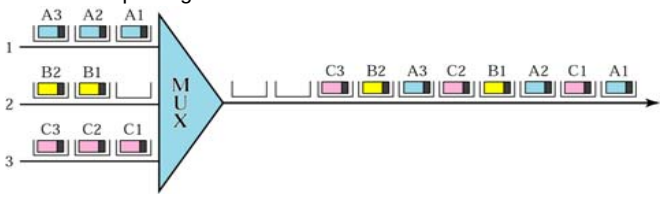


Multiplexing of Variable vs. Fixed Size Packets

- Multiplexing of variable size packets

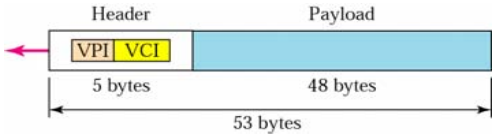


- ATM Multiplexing

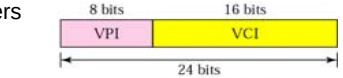


ATM Identifiers

- ATM Cell



- Virtual Path Identifiers and Virtual Channel Identifiers

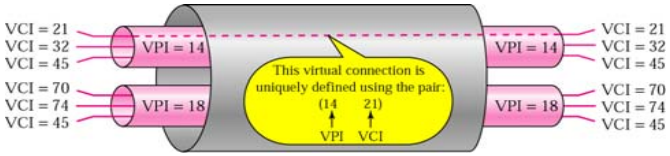


a. VPI and VCI in a UNI

b. VPI and VCI in an NNI

(UNI: User-to-Network-Interface
NNI: Network-to-Network-Interface)

ATM Virtual Connections



ATM Physical Layer (more)

Two pieces (sublayers) of physical layer:

- Transmission Convergence Sublayer (TCS): adapts ATM layer above to Physical Medium Dependent (PMD) sublayer below
- Physical Medium Dependent: depends on physical medium being used

TCS Functions:

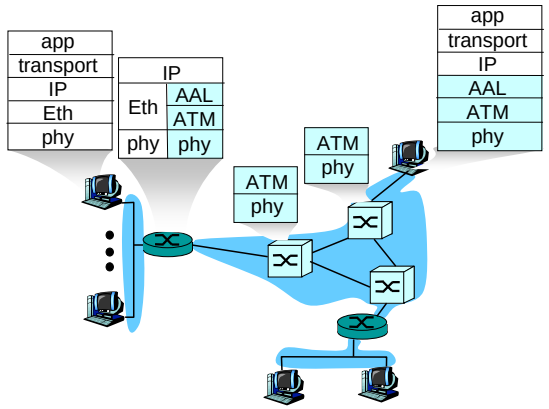
- Header **checksum** generation: 8 bits CRC
- Cell **delineation**
- With "unstructured" PMD sublayer, transmission of **idle cells** when no data cells to send

ATM Physical Layer

Physical Medium Dependent (PMD) sublayer

- SONET/SDH: transmission frame structure (like a container carrying bits);
  - bit synchronization;
  - bandwidth partitions (TDM);
  - several speeds:
    OC3 = 155.52 Mbps;
    OC12 = 622.08 Mbps;
    OC48 = 2.45 Gbps,
    OC192 = 9.6 Gbps
- TI/T3: transmission frame structure (old telephone hierarchy): 1.5 Mbps / 45 Mbps
- unstructured: just cells (busy/idle)

IP-Over-ATM



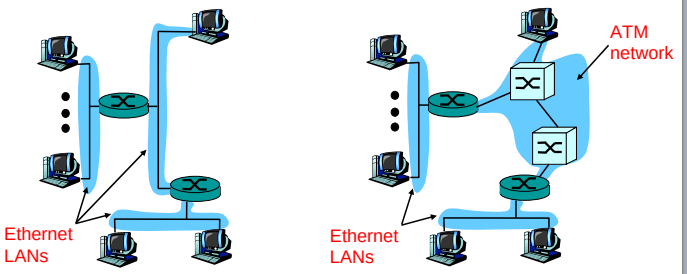
IP-Over-ATM

Classic IP only

- 3 “networks” (e.g., LAN segments)
- MAC (802.3) and IP addresses

IP over ATM

- replace “network” (e.g., LAN segment) with ATM network
- ATM addresses, IP addresses



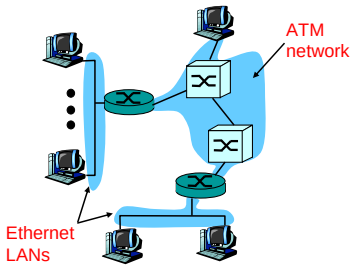
Datagram Journey in IP-over-ATM Network

- at Source Host:
  - IP layer maps between IP, ATM destination address (using ARP)
  - passes datagram to AAL5
  - AAL5 encapsulates data, segments cells, passes to ATM layer
- ATM network: moves cell along VC to destination
- at Destination Host:
  - AAL5 reassembles cells into original datagram
  - if CRC OK, datagram is passed to IP

IP-Over-ATM

Issues:

- IP datagrams into ATM AAL5 PDUs
- from IP addresses to ATM addresses
  - just like IP addresses to 802.3 MAC addresses!



Chapter 6 outline – Quality-of-Service Support

6.1 Link virtualization: ATM

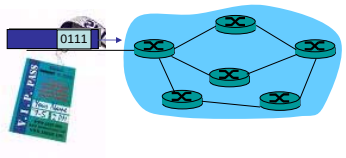
6.2 Providing multiple classes of service

6.3 Providing QoS guarantees

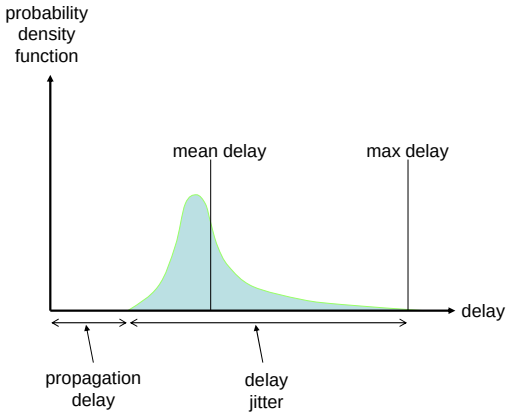
6.4 Signalling for QoS

Providing Multiple Classes of Service

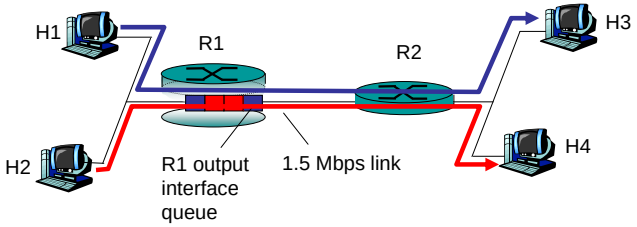
- Traditional Internet approach: making the best of best effort service
  - one-size fits all service model
- Alternative approach: multiple classes of service
  - partition traffic into classes
  - network treats different classes of traffic differently (analogy: VIP service vs regular service)
- granularity:
  differential service among multiple classes, not among individual connections
- history:
  ToS bits in IP header



Delay Distributions

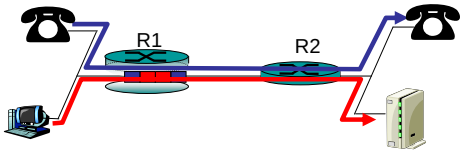


Multiple classes of service: scenario



Scenario 1: mixed FTP and audio

- Example: 1Mbps IP phone, FTP or NFS share 1.5 Mbps link.
  - bursts of FTP or NFS can congest router, cause audio loss
  - want to give priority to audio over FTP

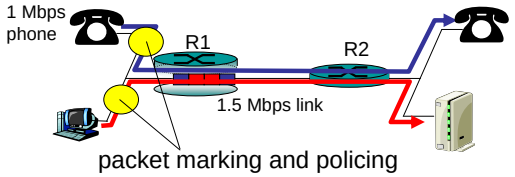


Principle 1

packet marking needed for router to distinguish between different classes; and new router policy to treat packets accordingly

Principles for QOS Guarantees (more)

- what if applications misbehave (audio sends higher than declared rate)
  - policing: force source adherence to bandwidth allocations
- marking and policing at network edge:
  - similar to ATM UNI (User Network Interface)

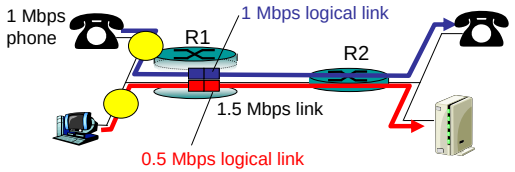


Principle 2

provide protection (*isolation*) for one class from others

Principles for QOS Guarantees (more)

- Allocating *fixed* (non-sharable) bandwidth to flow: *inefficient* use of bandwidth if flows doesn't use its allocation

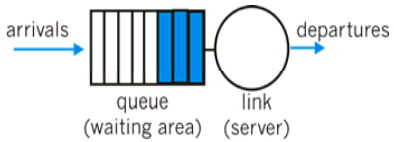


Principle 3

While providing **isolation**, it is desirable to use resources as efficiently as possible

Scheduling And Policing Mechanisms

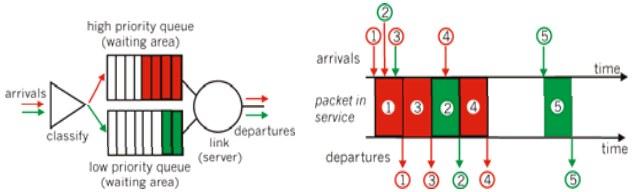
- scheduling: choose next packet to send on link
- FIFO (first in first out) scheduling: send in order of arrival to queue
  - ⇨real-world example?
  - discard policy: if packet arrives to full queue: who to discard?
    - Tail drop: drop arriving packet
    - priority: drop/remove on priority basis
    - random: drop/remove randomly



Scheduling Policies: more

Priority scheduling: transmit highest priority queued packet

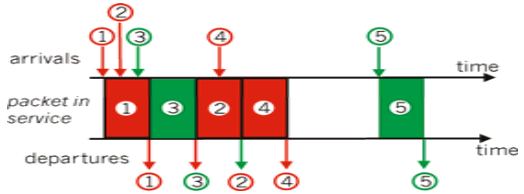
- multiple *classes*, with different priorities
  - class may depend on marking or other header info, e.g. IP source/dest, port numbers, etc..



Scheduling Policies: still more

round robin scheduling:

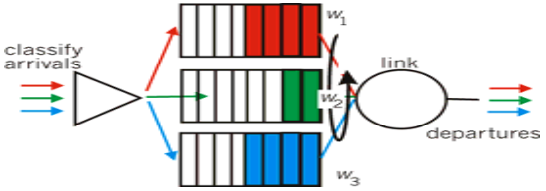
- multiple classes
- cyclically scan class queues, serving one from each class (if available)



Scheduling Policies: still more

Weighted Fair Queuing:

- generalized Round Robin
- each class gets weighted amount of service in each cycle



Policing Mechanisms

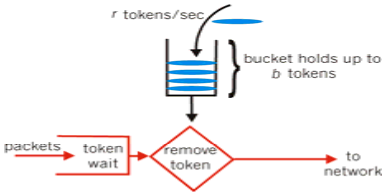
Goal: limit traffic to not exceed declared parameters

Three common-used criteria:

- *(Long term) Average Rate:* how many packets can be sent per unit time (in the long run)
  - crucial question: what is the interval length: 100 packets per sec or 6000 packets per min have same average!
- *Peak Rate:* e.g., 6000 packets per min. (ppm) avg.; 1500 ppm peak rate
- *(Max.) Burst Size:* max. number of packets sent consecutively (with no intervening idle)

Policing Mechanisms

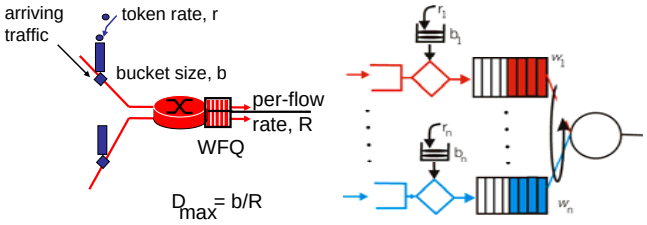
Token Bucket: limit input to specified Burst Size and Average Rate.



- bucket can hold b tokens
- tokens generated at rate *r token/sec* unless bucket full
- *over interval of length t: number of packets admitted less than or equal to (r t + b).*

Policing Mechanisms (more)

- token bucket, WFQ combined provide guaranteed upper bound on delay, i.e., *QoS guarantee*!



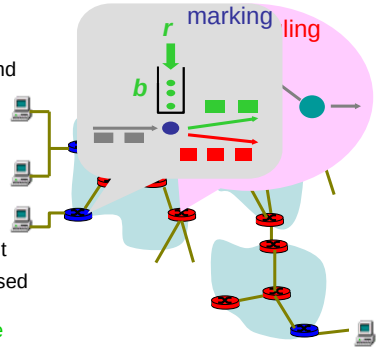
IETF Differentiated Services

- want "qualitative" service classes
  - "behaves like a wire"
  - relative service distinction: Platinum, Gold, Silver
- *scalability:* simple functions in network core, relatively complex functions at edge routers (or hosts)
  - signaling, maintaining per-flow router state difficult with large number of flows
- don't define define service classes, provide functional components to build service classes

Diffserv Architecture

Edge router:

- per-flow traffic management
- marks packets as in-profile and out-profile

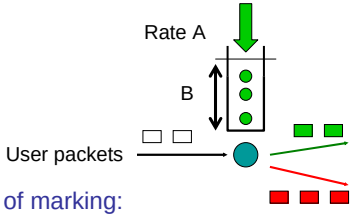


Core router:

- per class traffic management
- buffering and scheduling based on marking at edge
- preference given to in-profile packets

Edge-router Packet Marking

- profile: pre-negotiated rate A, bucket size B
- packet marking at edge based on per-flow profile



Possible usage of marking:

- class-based marking: packets of different classes marked differently
- intra-class marking: conforming portion of flow marked differently than non-conforming one

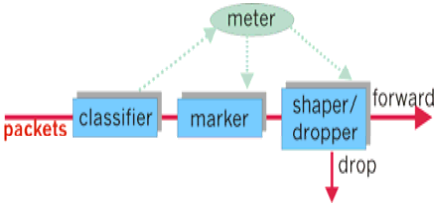
Classification and Conditioning

- Packet is marked in the Type of Service (TOS) in IPv4, and Traffic Class in IPv6
- 6 bits used for Differentiated Service Code Point (DSCP) and determine PHB that the packet will receive
- 2 bits can be used for congestion notification

Classification and Conditioning

May be desirable to limit traffic injection rate of some class:

- user declares traffic profile (e.g., rate, burst size)
- traffic metered, shaped if non-conforming



Forwarding (PHB)

- PHB result in a different observable (measurable) forwarding performance behavior
- PHB does not specify what mechanisms to use to ensure required PHB performance behavior
- Examples:
  - Class A gets x% of outgoing link bandwidth over time intervals of a specified length
  - Class A packets leave first before packets from class B

Forwarding (PHB)

PHBs being developed:

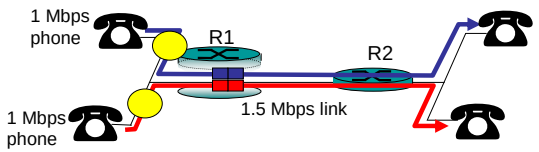
- Expedited Forwarding: packet departure rate of a class equals or exceeds specified rate
  - logical link with a minimum guaranteed rate
- Assured Forwarding: e.g. 4 classes of traffic
  - each guaranteed minimum amount of bandwidth
  - each with three drop preference partitions

Chapter 6 outline – Quality-of-Service Support

6.1 Link virtualization: ATM
6.2 Providing multiple classes of service
6.3 Providing QoS guarantees
6.4 Signalling for QoS

Principles for QOS Guarantees (more)

- *Basic fact of life:* can not support traffic demands beyond link capacity

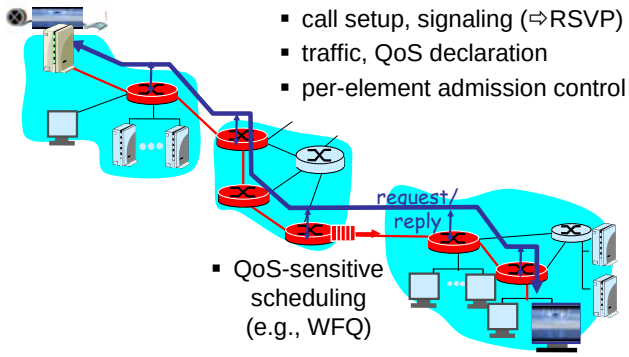


Principle 4

Call Admission: flow declares its needs, network may block call (e.g., busy signal) if it cannot meet needs

QoS Guarantee Scenario

- Resource reservation
  - call setup, signaling (⇨RSVP)
  - traffic, QoS declaration
  - per-element admission control



- QoS-sensitive scheduling (e.g., WFQ)

IETF Integrated Services

- architecture for providing QOS guarantees in IP networks for individual application sessions
- resource reservation: routers maintain state info (as for VCs) of allocated resources, QoS requests
- admit/deny new call setup requests:

Question: can newly arriving flow be admitted with performance guarantees while not violated QoS guarantees made to already admitted flows?

Call Admission

Arriving session must :

- declare its QOS requirement
  - R-spec: defines the QOS being requested
- characterize traffic it will send into network
  - T-spec: defines traffic characteristics
- signaling protocol: needed to carry R-spec and T-spec to routers (where reservation is required)
  - RSVP

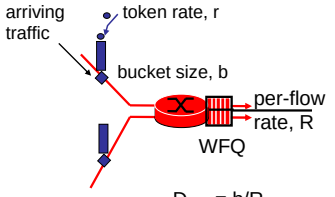
Intserv QoS: Service models [RFC 2211, RFC 2212]

Guaranteed service:

- worst case traffic arrival: leaky-bucket-policed source
- simple (mathematically provable) *bound* on delay [Parekh 1992, Cruz 1988]

Controlled load service:

- "a quality of service closely approximating the QoS that same flow would receive from an unloaded network element."



$D_{max} = b/R$

Chapter 6 outline – Quality-of-Service Support

6.1 Link virtualization: ATM
6.2 Providing multiple classes of service
6.3 Providing QoS guarantees
6.4 Signalling for QoS

Signaling in the Internet

connectionless (stateless) forwarding by IP routers + best effort service = no network signaling protocols in initial IP design

- New requirement: reserve resources along end-to-end path (end system, routers) for QoS for multimedia applications
- RSVP: Resource Reservation Protocol [RFC 2205]
  - " … allow users to communicate requirements to network in robust and efficient way." i.e., signaling !
- earlier Internet Signaling protocol: ST-II [RFC 1819]

RSVP Design Goals

1. accommodate heterogeneous receivers (different bandwidth along paths)
2. accommodate different applications with different resource requirements
3. make multicast a first class service, with adaptation to multicast group membership
4. leverage existing multicast/unicast routing, with adaptation to changes in underlying unicast, multicast routes
5. control protocol overhead to grow (at worst) linear in # receivers
6. modular design for heterogeneous underlying technologies

RSVP: does not…

- specify how resources are to be reserved
  - rather: a mechanism for communicating needs
- determine routes packets will take
  - that's the job of routing protocols
  - signaling decoupled from routing
- interact with forwarding of packets
  - separation of control (signaling) and data (forwarding) planes

RSVP: overview of operation

- senders, receiver join a multicast group
  - done outside of RSVP
  - senders need not join group
- sender-to-network signaling
  - *path message:* make sender presence known to routers
  - path teardown: delete sender's path state from routers
- receiver-to-network signaling
  - *reservation message:* reserve resources from sender(s) to receiver
  - reservation teardown: remove receiver reservations
- network-to-end-system signaling
  - path error
  - reservation error