Multimedia Networking and Quality of Service

Mario Baldi

Technical Univeristy of Torino
Computer Engineering Department

mario.baldi@polito.it
+39 011 564 7067
staff.polito.it/mario.baldi

Nota di Copyright

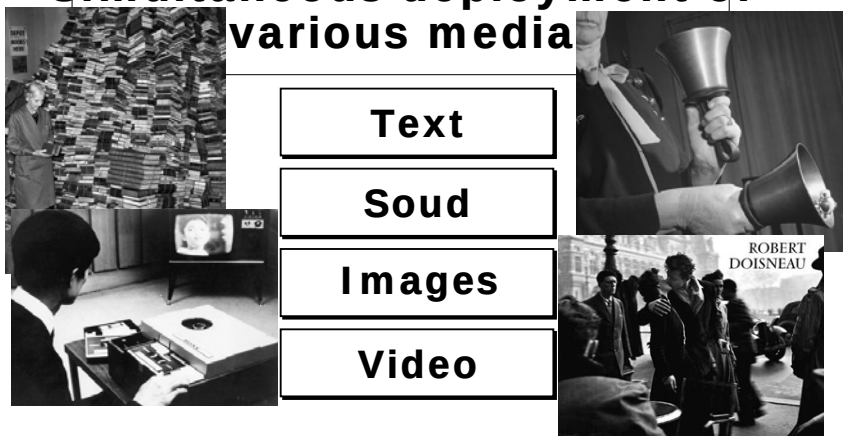
Outline

- Multimedia applications
  - Requirements and consequences on the network
- Quality of service techniques
  - Queuing
  - Access control
- Quality of service approaches
  - IntServ and DiffServ

Multimedia Applications over Packet Networks

What is Multimedia?

Simultaneous deployment of various media



- Text
- Soud
- Images
- Video

Media Coding

Sampling and quantization

Loss of quality

↓

Un-noticeable

Coding of samples

Possibility of playing back without loss of quality

Compression

Redundancy elimination

- → Spatial
- → Temporal
- → Possible information loss
  - → Loss of quality

Image Coding

- JPEG
- JPEG2000
- GIF
- TIFF

Video Coding

MPEG1

- Low quality
- 1.5 Mb/s

MPEG2

- High quality (DVD)
- 3.6 Mb/s

Video Coding

MPEG4

- Coding and compression based on identification of objects

H.261

- Videoconference
- Low quality
- Low bandwidth

Voice Coding

PCM (64 Kb/s 56 Kb/s)

GSM (13 Kb/s)

G.729 (8 Kb/s)

G723.3 (6.4 Kb/s & 5.3 Kb/s)

Audio Coding

AVI

MP3 (MPEG layer 3 - audio)

- 128 Kb/s or 112 Kb/s
- CD equivalent quality

Audio Coding

MP3

- Particularly robust
  - Each fragment can be played back independently from the others
- Suitable to the Internet

Coding Standards: Which one to choose?

The coding choice depends on

- Processing capability of terminals
- Resource availability in the network
- Type of application
  - Live (real-time)
  - Store&retrieve

Networked Multimedia Applications

- World Wide Web
- Video broadcasting
- Video on Demand
- Telephony
- Radio
- Jukebox service

Networked Multimedia Applications

- Teleconference
- Interactive distributed games
- Distance learning
- Virtual reality

Networked Multimedia Applications

Maybe just one media, but …

different features with respect to traditional applications

Distinctive Features: Streaming

- **Continuous flow of data**
- **The profile of generated flow must be the same as the profile of the flow to be played back**
  - → **Continuous playout**
- **Very different from traditional applications**

Distinctive Features: Interactivity

- **With another human**
- **With a computer**
- **Short response time**

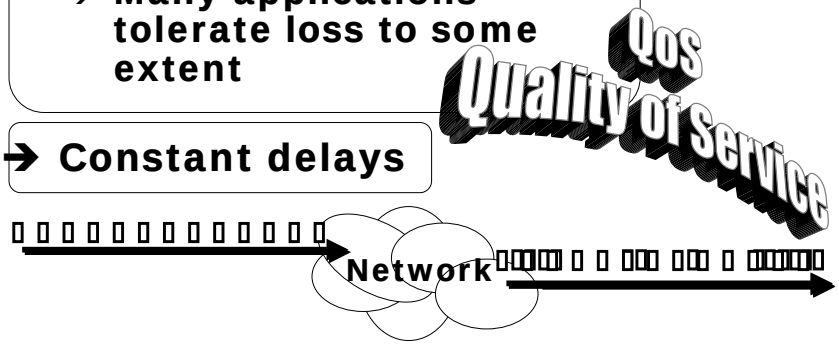
Distinctive Features

- **Large transmission bandwidth**
- **Group communications**
  - → **Many to many communications**

Requirements on the network

Streaming

- Limited loss
  - Many applications tolerate loss to some extent
- Constant delays



Interactivity

- Low delay
  - Verbal interaction: below 100 ~ 150 ms one way

Large Transmission Bandwidth

High resource availability

- Transmission capacity
- Memory in network nodes (buffers)
- Processing power (routing, etc.)
- Switching

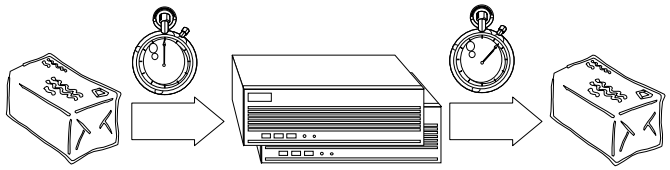
Technology advances help

Group Communications

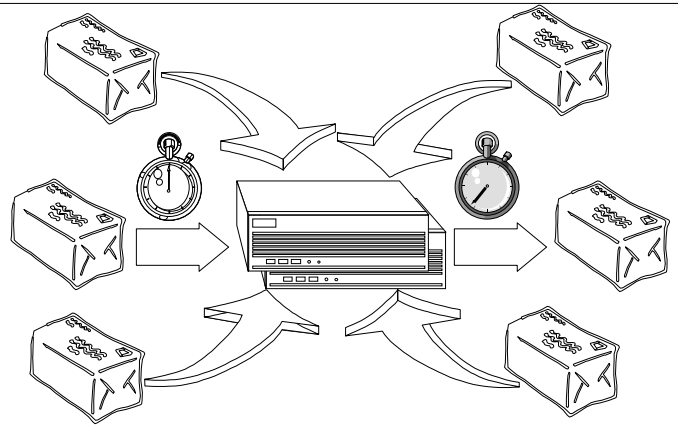
- Multicast transmission services
  - → IP multicasting
  - → MBone
- Server with *reflector* functionality or *multiconference unit*

Delay: That's the Problem!!

Multimedia applications are also generally called *real-time* applications



Delay: What's the Problem?



It is different depending on the *instantaneous* load on network nodes

Countermeasures in End-Systems

Compensate delay variations

- Replay buffer
- Fixed dimension for non-interactive applications
- Adaptive for interactive applications

Countermeasures in End-Systems

The only way to compensate for delay variations is to conform each data unit to maximum delay

"End-to-end" delay increase

Critical for interactive applications (telephony, conferencing, games, virtual reality)

Countermeasures in End-Systems

Adaptation to network conditions

Decrease traffic generated when quality of session decreases

Adaptability of applications is not unlimited

Tools

Information Exchange

- Time data
  - → RTP - Real-time Transport Protocol
    - → Time stamp
- Communication state
  - → RTCP - RTP Control Protocol

Adaptive Coding

- Quantization granularity
- Compression parameters
- Feedback
  - → E.g., RTCP

Layered Coding

- Base Layer
  - → Transmitted at higher priority
  - → Resources are possibly reserved
- Additional layers enable quality improvement
  - → Transmitted at lower priority
  - → Possibly as best-effort

Countermeasures in the Network

- Traffic classification
- Sophisticated scheduling algorithms
  - → WFQ, RR, WRR, CBQ
- Control on traffic entering the network
  - → At various levels
- (QoS routing)

Control on Traffic

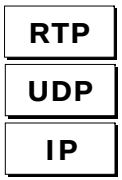
- Packet level
  - → Shaping / policing
- Call / flow level
  - → Signalling with resource reservation
  - → RSVP
    Resource reSerVation Protocol (IP)
  - → UNI
    User Network Interface (ATM)

Control on Traffic

- A priori
  - Network engineering
    - Network dimensioning according to expected traffic
    - Limit on the number of users
  - Traffic engineering
    - Controlled distribution of traffic across the network

Countermeasures in the Network

Collateral Problems



- UDP at the transport layer
- Real-time requirements are usually incompatible with TCP retransmission timing
  - An originally lost or corrupted packet eventually reaching the destination is useless because it took too long
  - Lost packets prevent delivery of following packets to applications

Greedy Applications

TCP adapts to traffic conditions, UDP ignores them

- Multimedia applications can harm the others
  - Especially those based on TCP, that is “polite”
- Traffic segregation and policing (bandwidth shaper)

Tools for Quality of Service Support

Classification

Classification

Identifying packets to which quality is to be guaranteed

in other words

In which queue to store an incoming packet

Classification

Based on information contained in the IP and TCP/UDP headers (quintuplet)

- Destination IP address
- Source IP address
- Transport protocol
- Destination port
- Source port

Classification

Complex algorithms

Hardware implementations

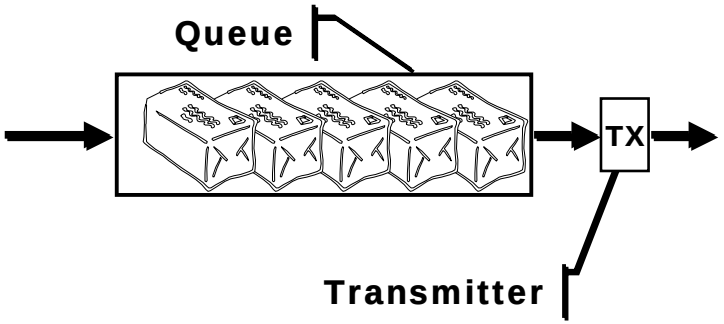
ASIC:
Application Specific Integrated Circuit

CAM:
Content Addressable Memory

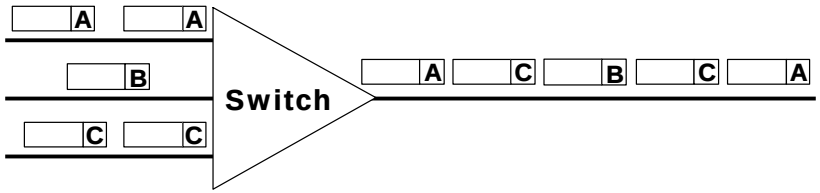
Tools for Quality of Service Support

Scheduling

Simple Queuing FIFO (First In First Out)

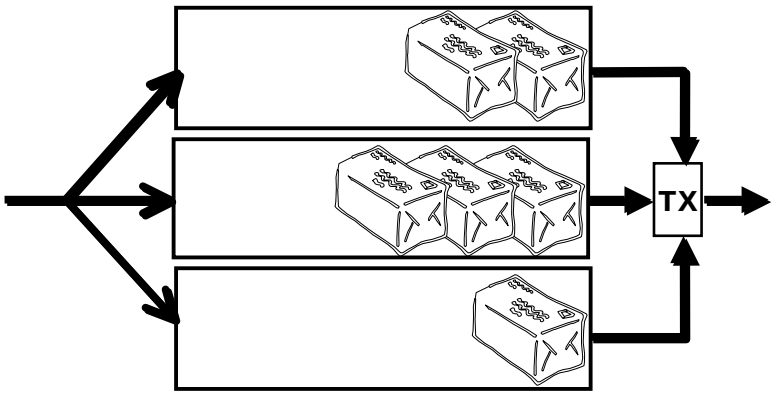


Statistical Multiplexing



Somebody will not be satisfied

Multiple Queues and Scheduling



Scheduling Algorithms

- Priority Queuing
- Round Robin
  - → Weighted Round Robin
- Class Based Queuing (CBQ)
- Weighted Fair Queuing (WFQ)
- Deadline queuing
  - → Jitter Earliest Due Date (non work-conserving)

Queuing and Switching

Output queuing

The "simplest" solution

but ...

Switching capacity is a limited resource

There is no guarantee for packets to be swithed as they arrive

Switching Capacity

- Guaranteed immediate switching requires speed-up
  - → The transfer speed of the switching fabric is higher than the input link speed
- Particularly critical when operating with high capacity links

Queuing and Switching

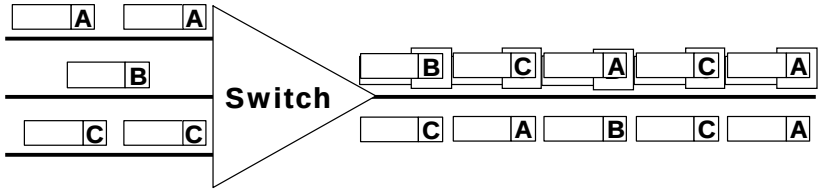
Input queuing

Distributed control (complex)

Virtual output queuing

Queues inside the switching fabric (*distributed queuing*)

Are We Set?



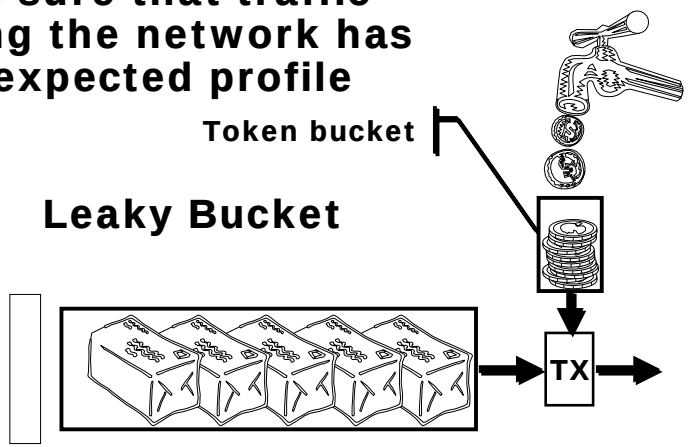
It is not possible to satisfy everyone!

Tools for Quality of Service Support

Control on Traffic

Policing and shaping

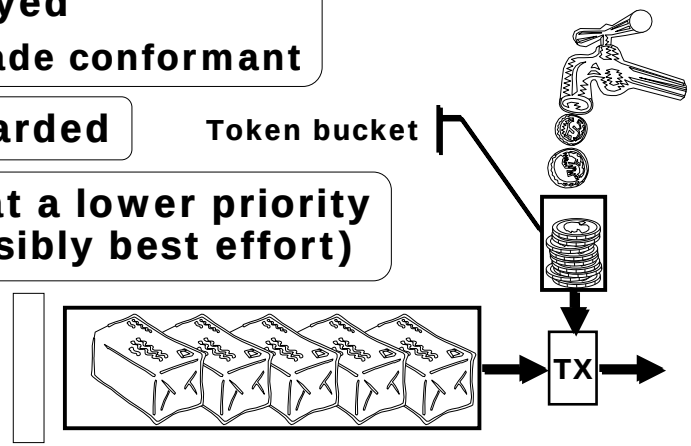
Make sure that traffic entering the network has the expected profile



Leaky bucket

Non conformant packets are

- → **Delayed**
  - → **Made conformant**
- → **Discarded**
- → **Set at a lower priority (possibly best effort)**



Call Admission Control (CAC)

- Signalling
  - → Description of generated traffic
  - → Description of required service
  - → Examples: RSVP e UNI ATM
- Resource Reservation

QoS routing

- Finding a route with required resources
- Routing protocols distribute in real-time information on resource availability
  - → Very dynamic information

QoS routing

- Routing decisions are based on resource availability information
  - → Not only on topological information
- Instability with connectionless data transfer
- E.g., PNNI (private network node interface) in ATM
  - → Cranckback

Network engineering Traffic engineering

Preventive actions

- Network is dimensioned for the (almost) worst case
  - → Statistics on user traffic
- Traffic matrix is determined
  - → Traffic distribution

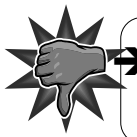
Network engineering
Traffic engineering

Actions throughout

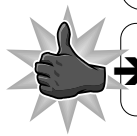
- Network state is continuously monitored
- Network dimensioning and traffic matrix can be chenged if needed

Network engineering
Traffic engineering

Distinctive properties



- Low efficiency in network resource utilization



- Simplicity and scalability

Tools for
Quality of Service Support
Complementary Issues

Policy

Defines general aspects of how the network functions

Determines specific aspects of how a device functions

- Queuing strategy
- Call admission control rules
- Leaky bucket parameters

Flexibility: Policy Management

- A policy might depend on
  - → Type of traffic
  - → Time of day
- You don't want to have to configure every single device ...
- ... and to periodically change its configuration

COPS (Common Open Policy Service)

Automatic policy distribution

- Network devices retrieve policies from a server
  - → Distribution model
- A server actively installs policies into devices
  - → Provisioning model

Internet Quality of Service Support Frameworks

IntServ and DiffServ

Internet's Ambitious Solution: Integrated Services (IntServ)

Features

- Per-flow resource reservation
  - RSVP: Resource reSerVation Protocol
- Guaranteed quality of service
  - Per-flow queuing inside routers

Internet's Ambitious Solution: Integrated Services (IntServ)

Limits

- High complexity
- Low scalability

Internet's Ambitious Solution: Integrated Services (IntServ)

State of the art

- Standard completed
- Implemented by router vendors
  - RSVP message handling
  - Queuing algorithms (?)
- Unusable on a large scale (public networks)

Lowering Ambitions: Differentiated Services (DiffServ)

- No quality of service guarantees
- No resource reservation
- Different service to different types (class) of traffic: class of service
  - DS (DiffServ) field
  - Per-class queuing

Lowering Ambitions: Differentiated Services (DiffServ)

How

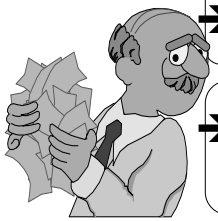
- Network engineering
- Traffic engineering
- Access control at the boudaries
  - → Policing

Lowering Ambitions: Differentiated Services (DiffServ)

Features

- Low efficiency
  - Large fraction of traffic is best- effort

- Simplicity and scalability



- Increasingly used
  - IP telephony