Internet QoS and Network Performance

François Fluckiger
CERN, Geneva

Outline of Lecture Series

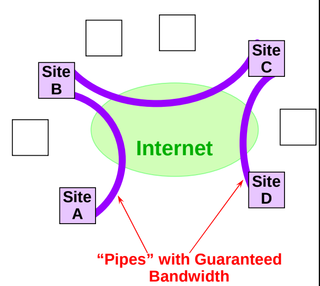
1. Internet QoS Options
2. TCP and Congestion Control
3. Multimedia over the Internet

Why is QoS Important in GRID Environments?

- GRID of systems
  - Fast transfers => High Bit rate connections
  - Predictable behavior
  - Availability
- Grid of people
  - New type of traffic for collaborative activities

Why improving QoS Guarantees? (cont.)

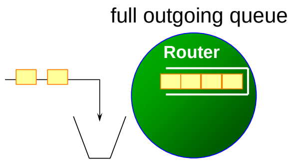
- Create **"virtual private networks"** with performances guarantees
- Requires set of QoS-guaranteed **point-to-point** pipes



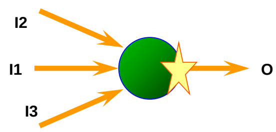
Site *x* Part of the Corporate Network

Why improving QoS Guarantees? (cont.)

- Avoid **congestion**:
  - lack of resources in network elements (e.g. buffers in switching nodes)



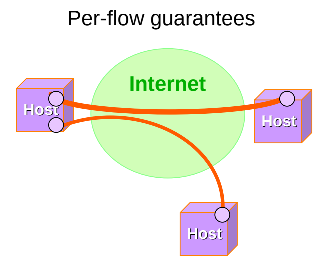
  - **N-to-1** problem



Σ **Input traffic > Output capacity**

Why improving QoS Guarantees?

- Support of **real-time** **(e.g multimedia)** applications



Expressing the Quality of Service?

A "contract" between the end-systems and the network

- ***Traffic*** profile
  - Sustained ***Data Rate*** (e.g. bit rate)
  - Possibly, peak Rate, Data Burst size, …

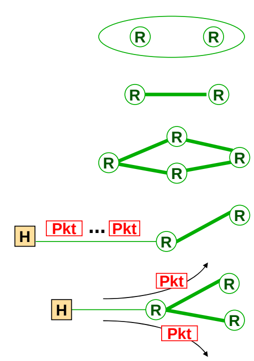
What the end-system promises to respect

- ***Quality*** Parameters
  - Packet ***Transit Delay***
  - Packet ***Loss Ratio***

What the network promises to guarantee if the end-system respect its traffic profile

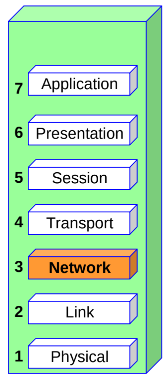
Internet: Back to Basics

- Formed of switches called **routers**
- Routers interconnected by **links**
- Topology usually **meshed**
- Hosts chop data stream into blocks called **packets**
- Routers switch individual packets

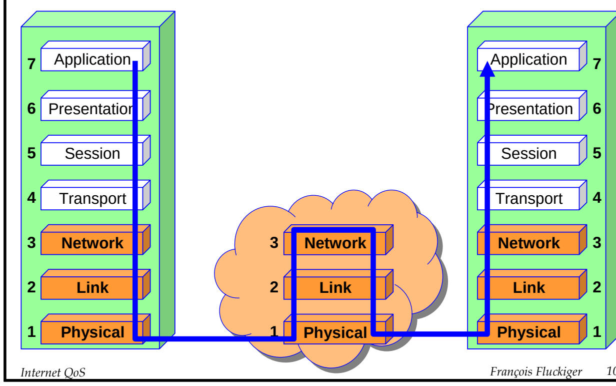


IP

- **Layer 3 (Network) protocol**
- **Specifies**
  - format of packets (size, header, …)
  - mechanism for routing
  - resulting service
- **The highest level protocol understood by routers**



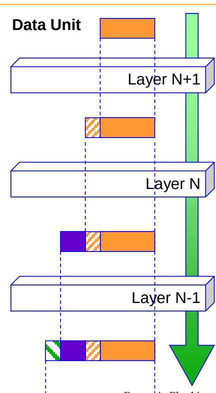
Protocols understood by routers



Layering principle (emission)

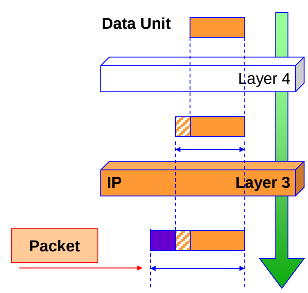
- Each protocol layer **N** adds a **Header** to the data unit received from layer **N+1** (1)
- Header contains **control** information; e.g. :
  - Numbering of the data unit
  - Coding of the destination
  - Codes for error detection
  - Priority of the data unit

(1) and segment the data unit if necessary



Packets

- Each **data unit** generated at a given layer has a specific name
- Data Unit at layer 3 (IP) is called a **Packet**



Main IP features

- **IP is a connectionless (CL) protocol *(stateless)***
  - all packets independently routed
  - packets carry full destination address
  - packets may be lost, miss-ordered
  - **all packets have same priority**
- **Opposite = connection-oriented (CO) *(stateful)***
  - no information sent before a hard connection is set up

Stateful / Stateless Networks

Stateful	Stateless
Telephone	Post office
	Road Network
"λ on-demand"	Ethernet
ISDN	IP
ATM	
Frame Relay	
SNA	*DECnet*
X.25	

CONS vs CLNS

Stateful	+ Stateless
Traffic more predictable	**No call set-up delay before sending a packet**
Easier for network to reserve resources	**Routing possibly more dynamic**
QoS easier to guarantee	**Resilience**

IP, HTTP Stateless Behavior

- **IP**
  - take a packet, forward it, forget it …
  - take a packet, forward it, forget it …
- **HTTP**
  - take a request, serve it, forget it
  - take a request, serve it, forget it

Predicting Load?

Types of Applications

- ***Constant Bit Rate (CBR)***
  - **Traditional real-time applications -** e.g. PABXs
- ***Available Bit Rate (ABR)***
  - **Traditional bulk data applications -** e.g. file transfer
- ***Variable Bit Rate (VBR)***
  - **Modern real-time applications -** e.g. compressed audio, video

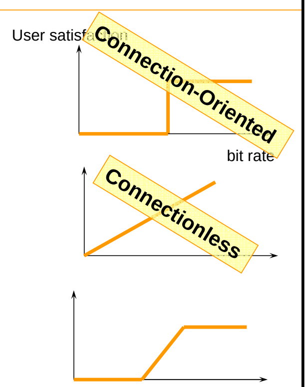
IP, HTTP Stateless Regular Behavior

- **IP switch**
  - take a packet, forward it, forget it …
  - take a packet, forward it, forget it …
- **HTTP server**
  - take a request, serve it, forget it
  - take a request, serve it, forget it

Predicting Load?

Quality of Service and bit rate

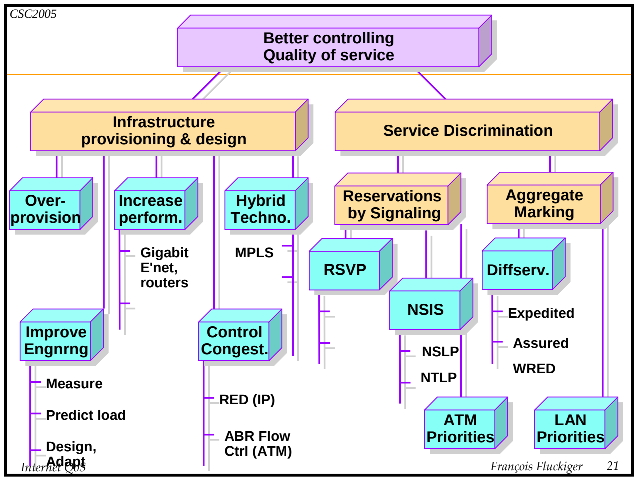
- CBR **applications**
- ABR **applications**
- VBR **applications**

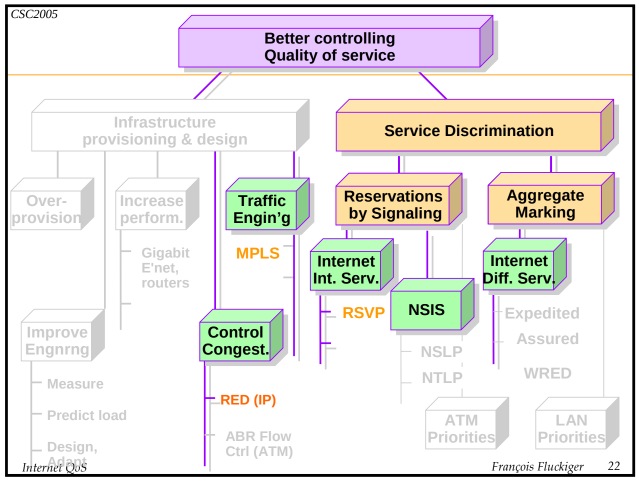


From Scott Shenker,
Fundamental Design Issues for the Future Internet

Outline of Lecture Series

1. Internet QoS Options
2. TCP and Congestion Control
3. Multimedia over the Internet





Internet Base IP service

- **Initial Internet**
  - single class of service:"***best-effort* service**"
  - packet forwarding completely **egalitarian**
  - **No service guarantee**
- **How to better guarantee end-to-end throughput, delays?**
  - ***"How to have packets more equal than others?"***

On service discrimination ...

- **Objective**
  - Give **better** service to some traffic
- **Consequence**
  - **...** at the expense of giving **worse** service to the rest

(hopefully in times of congestion only)

CSC2005

Internet service discrimination

Two proposed techniques

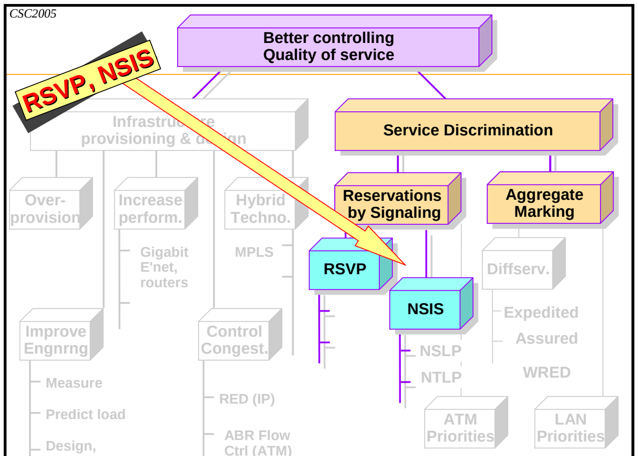
- ***Reservation by signaling***
  - Protocols: Next Step in Signaling (NSIS)
    Resource Reservation Protocol (RSVP)
- ***Aggregate marking***
  - Protocols: Differentiated Services (Diffserv)

CSC2005

Reservation by signaling: Principles

- ***Resource reservation*** is necessary
  - Reservations on a **per-flow** basis
- Routers have to maintain **flow-specific *states***
- *Protocol: **NSIS** (recent), **RSVP** (older)*

CSC2005

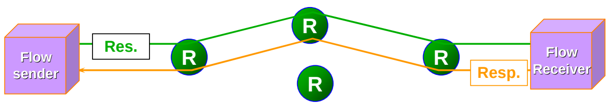


CSC2005

RSVP and NSIS Protocols

- **Signaling protocols**
- **no hard connections**

NSIS protocol (simplified)



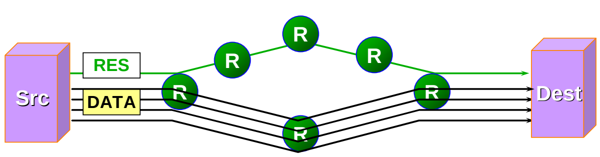
- **RESERVE** control message sent periodically by ***source***
- ***receiver*** replies with a **RESPONSE** control message
- **RESPONSE** reserve resources on the route back
- if **RESERVE** not repeated after time-out, resources released

Which resources are reserved?

- **"Resource" is implementation dependent:**
  - In practice, with today's routers, reservation of:
    - a slice **of link bandwidths**
    - a fraction of the **buffers**
- **Note: reservation ≠ allocation**

  Reserving a flight ticket is different from getting the seat allocated

Concern #1: Route stability



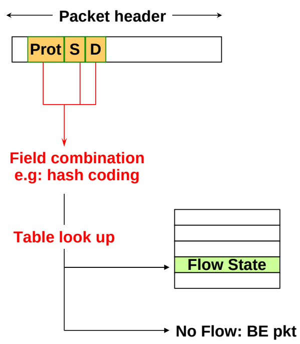
- **RESERVE** reserved resources over **long** route
- **DATA** follow **shorter** route

Concern #2: Scalability

- **States**
  - "How many soft-states can routers handle?"
- **Overhead**
  - **Classifying** packets between
    - Regular
    - Belonging to a flow with reservation
  - … is an heavy process

Multi-field (MF) classification

- ***Multi-field (MF) classifier***
  - classifies on combination of **header fields**
  - occurs on **every** data pkt



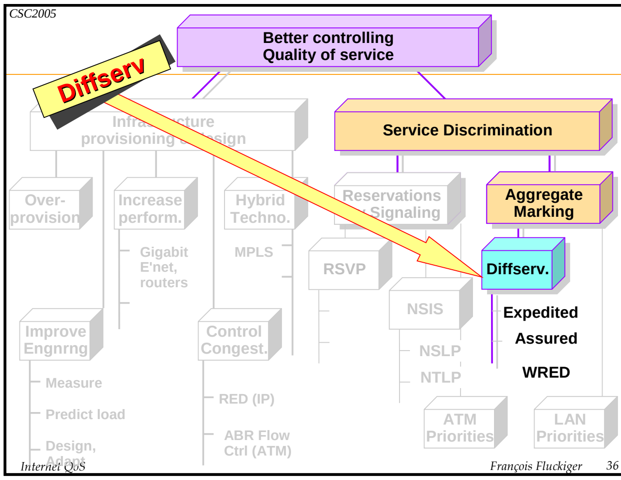
(1) Up to 6 fields may be inspected:
dest_add, dest-port, prot_number, src_add, src_port, ToS

(NSIS/RSVP) - Diffserv - MPLS merits

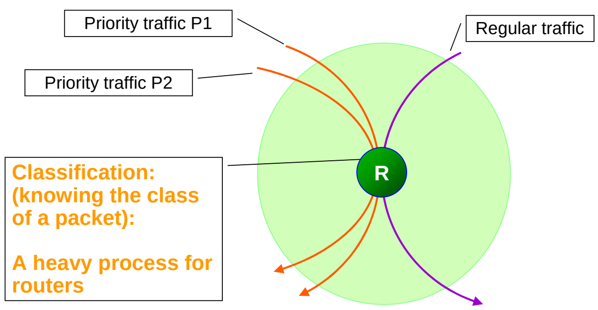
	Capacity Admission	Scalability	Route Stability
NSIS/RSVP	Yes	No	No
Diffserv			
MPLS			

Options for NSIS / RSVP concerns

- **Concern #1: Route stability**
  - ***MPLS***
- **Concern #2: Scalability**
  - ***Diffserv***

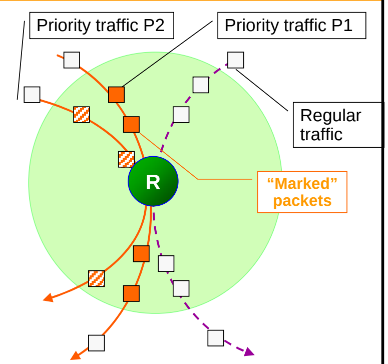


RSVP/NSIS Concern #2: Scalability



Diffserv Principle

- **Priority Mark inserted** before Pkts enter the "QoS core"
- Simple examination of mark provides priority

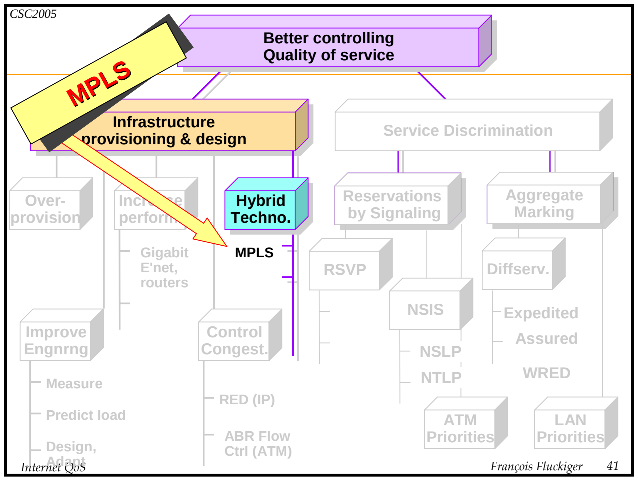


Implementing Packet Marking

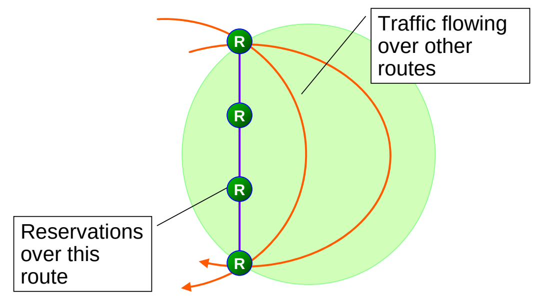
- **No need to change IP packet header, just refine meaning of existing fields**
- **IPv4**
  - Provided with a mechanism for packet priority marking, the ***Type of Service*** (***ToS***) octet
- **IPv6**
  - Provided with ***Traffic Class*** octet

(NSIS/RSVP) - Diffserv - MPLS merits

	Capacity Admission	Scalability	Route Stability
NSIS/RSVP	Yes	No	No
Diffserv	No	Yes	No
MPLS			



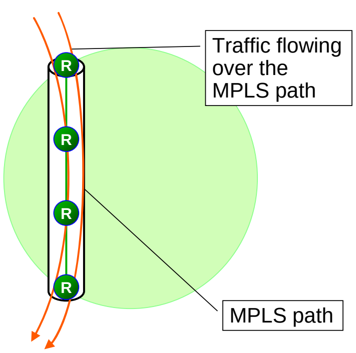
Concern #1: Route stability



MPLS

- **Create a "circuit" (called MPLS path)**
- **Force all traffic with**
  - **same destination**
  - **same Qos requirement**

to follow the path



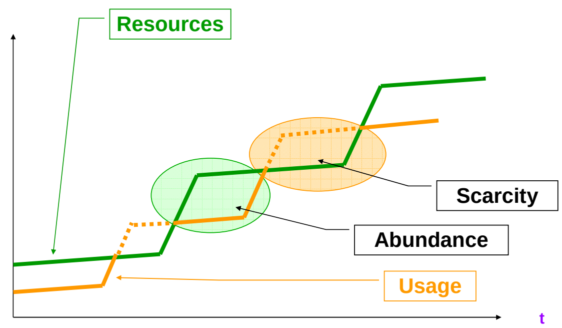
(NSIS/RSVP) - Diffserv - MPLS merits

	Capacity Admission	Scalability	Route Stability (*)
NSIS/RSVP	Yes	No	No
Diffserv	No	Yes	No
MPLS	No	No	Yes

(*) + Traffic Engineering

Where is all this going?

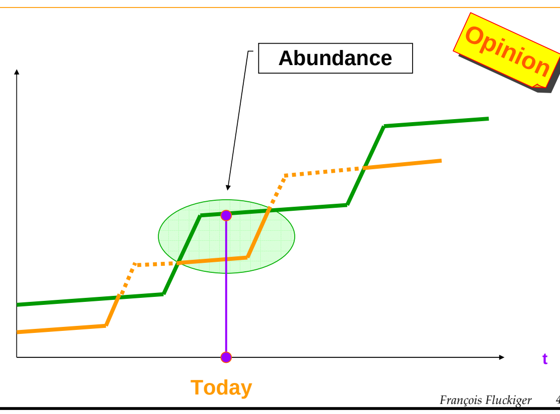
Resources and Usage



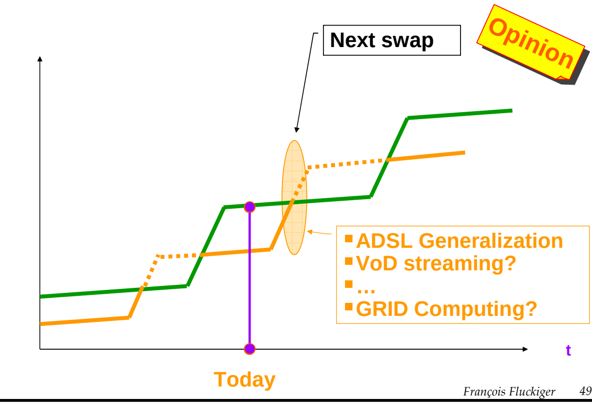
Observations

- **Core ISPs**
  - Capacity utilized at 12%
- **LANs**
  - Over-provisioned
- **Real problems**
  - Some LAN-WAN boundaries
  - Wireless **Internet Telephony**

Resources and usage



Resources and usage



So what?

Opinion

- **Full QoS complexity not justified in core**
- **Specific applications require guarantees**
  - VPNs
  - Internet Telephony (jitter, packet loss guarantees)
- **Limited-bandwidth nets** (e.g. wireless) **need QoS**
- **Scarcity will be back**
  - Question is when

End of

Part 1

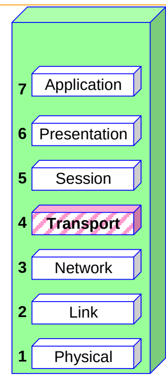
Internet QoS options

Outline of Lecture Series

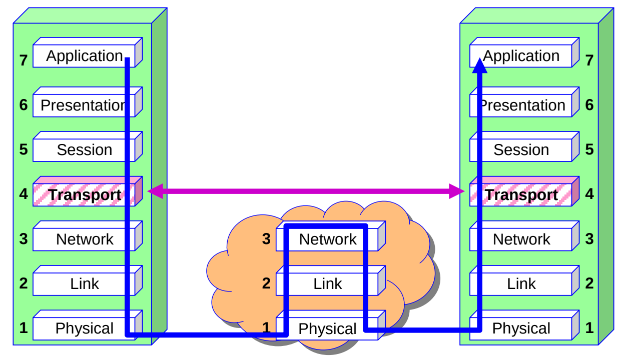
1. Internet QoS Options
2. TCP and Congestion Control
3. Multimedia over the Internet

TCP

- **Layer 4 (Transport) protocol**
- **Specifies**
  - Format of segment
  - Mechanisms for flow control, error detection, error recovery
- **The lowest protocol used by hosts**



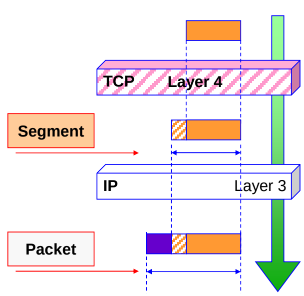
Protocols understood by hosts only



TCP Segments

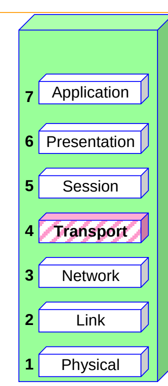
- Each **data unit** generated at a given layer has a specific name
- Data Unit at layer 4 (TCP) is called a **Segment**

  (sometimes also called **block)**
- Data Unit at layer 3 (IP) is called a **Packet**)



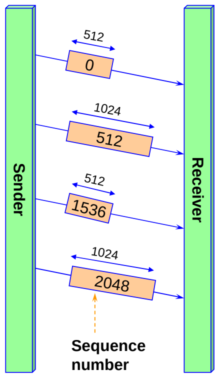
What does TCP Provide

- **Error services**
  - Detection of corrupted data
  - Detection of loss, duplicated, out of sequence packets
  - Correction of errors
- **Flow control between receiver/server**
- **Mechanisms to limit network congestions**



TCP Mechanisms for Errors

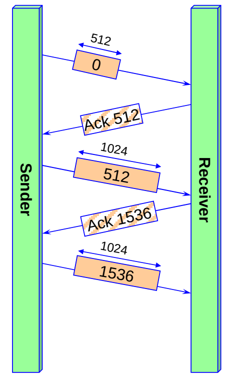
- **Segments**
  - Theoretical maximum= 65535 bytes
- **Error detection**
  - Segments carry a **sequence number**
  - Sequence number = the **order of the first octet of segment** in the data stream
  - Receiver can detect out-of-sequence packets



TCP Mechanisms for Errors

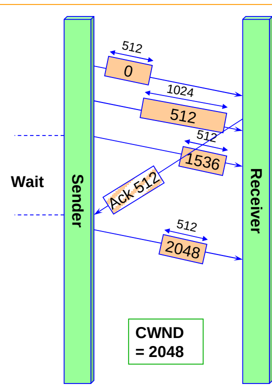
- **Error notification**
  - Positive ACK by receiver
  - e.g. "ACK 512 means":

    " *I am now ready to receive octet #512 and beyond, because I correctly received all octets up to #511*"
- **Problem with this**
  - Sender needs to wait for ACK of segment 1 before sending segment 2



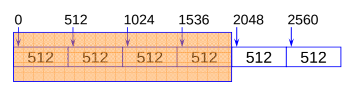
TCP Windowing

- **Windowing**
  - Mechanism to anticipate on the ACK
  - Sender manages a byte count which gives the limit of the highest octet that can be sent without being acknowledged
  - This is called the **Congestion Window** (**CWND**)

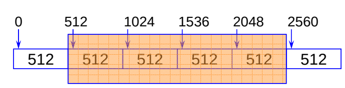


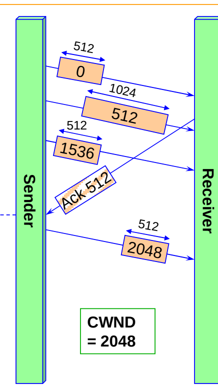
TCP Sliding Window

Initial window



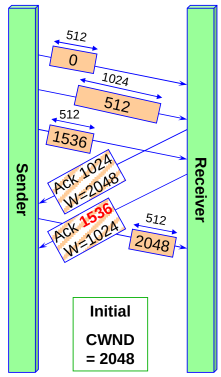
Window after 1st ACK



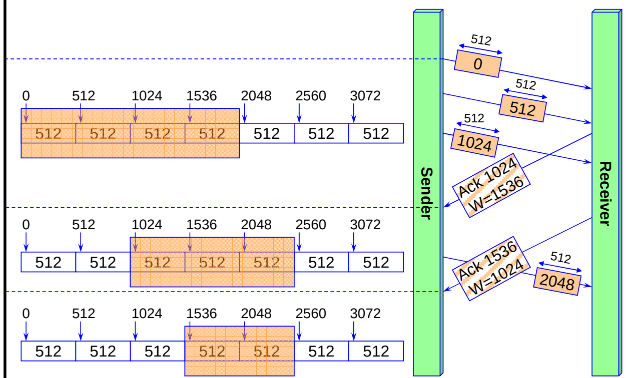


TCP mechanism for Flow control

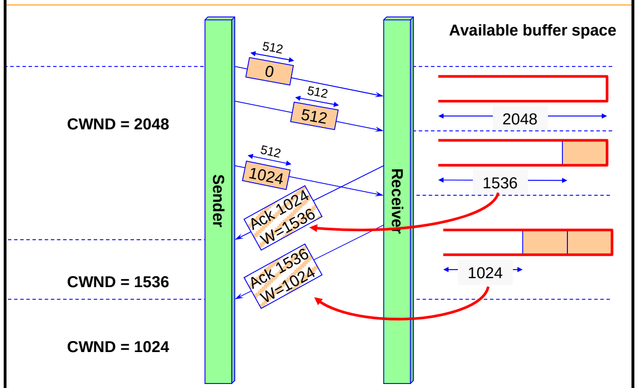
- **Dynamic Window size**
  - Window size CWND not fixed
  - Sender may reduce if detecting network congestion
  - Receiver indicates value of window in ACK
  - Value of Window = remaining incoming **buffer space** in receiver



TCP mechanism for Flow control

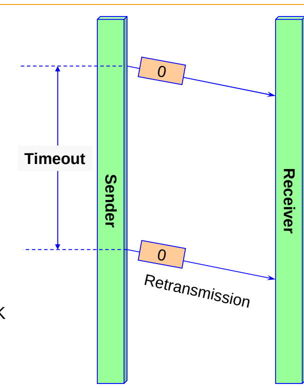


TCP buffer size



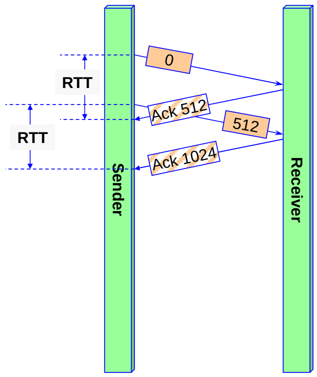
TCP mechanism error correction

- **TCP retransmits non-acknowledged segment**
- Question: ***when to retransmit***?
- **Timer: TCP …**
  - starts a timer on any transmission
  - waits for the ACK
  - retransmits segment in no ACK when timer expires



TCP Timeout

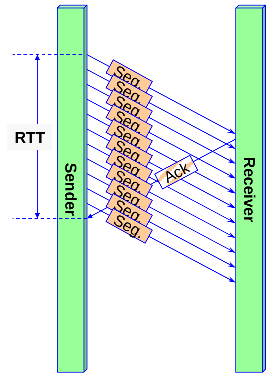
- **Question:** ***which value for the Timeout*?**
- **TCP** measures permanently the **Round Trip Time (RTT)**
  - RTT = Average Time between segment emission and ACK reception
- **Timeout = β x RTT**
  - Choice of β delicate (simple choice: β = 2)



Optimal buffer size

- **Sender must keep all sent segments until acknowledged**
- **Question:**

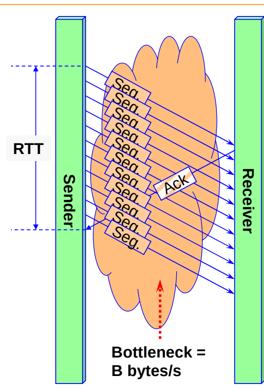
  What is the optimal buffer size to keep all segments?



Optimal buffer size

- If bottleneck of the Bandwidth between Sender and Receiver is **B** byte/second
  - Sender can send up to **B** bytes / second over **RTT** seconds
  - Maximum number of bytes to store = **B x RTT**
- Called the **Bandwidth*Delay Product** (**BDP**)

BDP = B x RTT

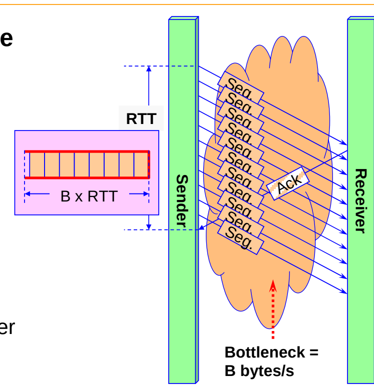


Optimal buffer size

- **Optimal buffer size** (1)

  =

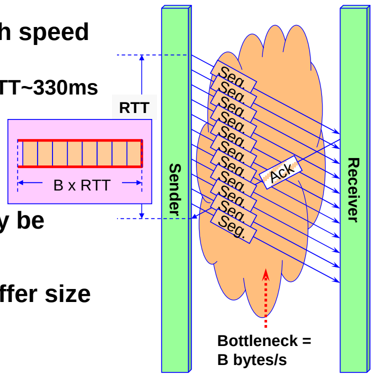
  Bandwidth*Delay Product
- **BDP = B x RTT**
- (1): sender and receiver



BDP over long distance links

- **Long and very high speed links. e.g.:**
  - **Geneva-Taiwan: RTT~330ms**
  - **B = 1 Gbps**
  - **BDP ~40 Mbytes**
- **In case packet may be dropped**

  Recommended buffer size = 2 BDP

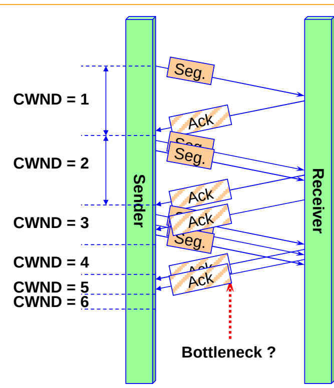


Outline of Lecture Series

1. Internet QoS Options
2. TCP and Congestion Control
3. Multimedia over the Internet

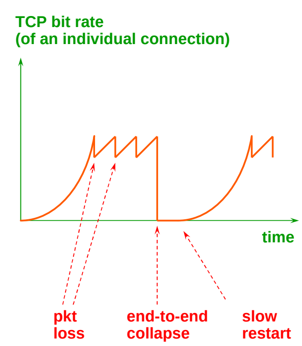
Avoiding congestions

- When TCP starts sending, ignores the network bottleneck (often at LAN – WAN interface)
- Can't send at full speed (would flood the network).
- Start with
  - maximum segment size, but
  - with minimum congestion window (CWND) = 1 segment
- On every ACK, increases CWND by 1



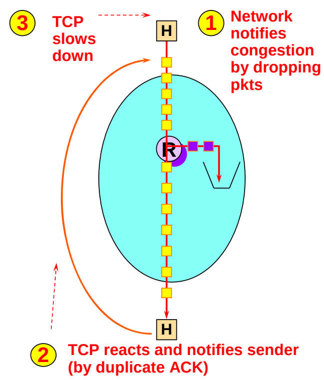
TCP congestion avoidance

- **Called slow start**
  - Rate **doubles** every round-trip time
- **If packet loss, sender halves window**
  - Then window increases **linearly**
- **In case of end-to-end collapse** (sender time out exhausted), **sender resumes with slow start**



Congestion Notification in Internet

- Current Internet, no **explicit** congestion notification from network to source
- Notification technique: *Network drops packets*
- TCP will react and slow down
- TCP is said to be *congestion-indication responsive*



End of

Part 2

TCP and Congestion Control

Outline of Lecture Series

1. Internet QoS Options
2. TCP and Congestion Control
3. Multimedia over the Internet

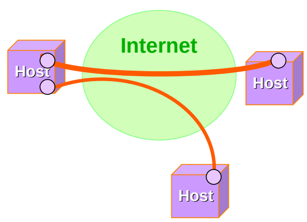
Why is QoS Important in GRID Environments?

- GRID of systems
  - Fast transfers => High Bit rate connections
  - Predictable behavior
  - Availability
- Grid of people
  - New type of traffic for collaborative activities

Why improving QoS Guarantees?

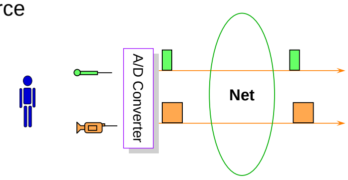
- Support of **real-time** (e.g multimedia) applications

Per-flow guarantees



Real-time media transmission (A / V)

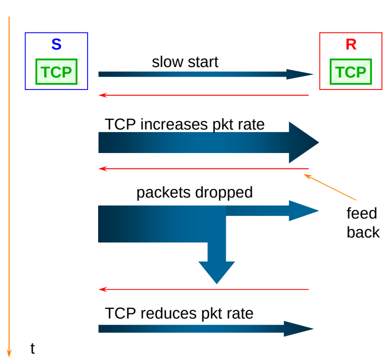
- Assume now the source of packets in an
  - analog to digital converter
  - connected to a microphone and a movie camera
- Called **Streaming** audio/video



Audio and video packets

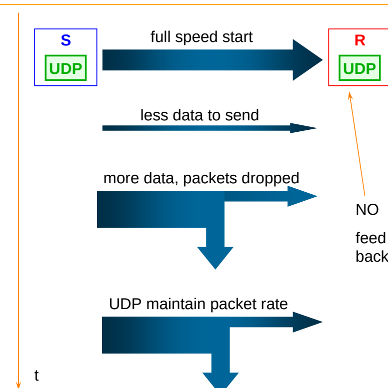
TCP behavior

- **Slow start**
- **Sender aware of packets dropped**
- **Sender decreases bit rate when packet dropped**



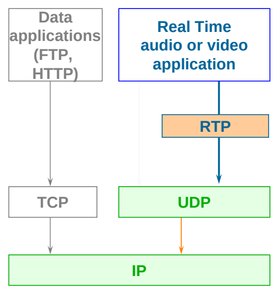
UDP behavior

- **UDP sends blindly to a receiver**
- **No feedback from the receiver**
- **Sender unaware whether packets are dropped/lost**



Protocols for real-time audio and video

- **Audio/video applications cannot operate over TCP**
  - slow start, congestion control, ...
- **They use UDP**
  - But UDP has no timestamp, no feedback, ...
- **Most applications use RTP (Real-Time Transport Protocol)**
  - time-stamp
  - packet loss detection

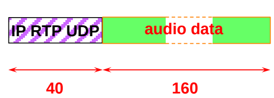


Real-Time Protocol

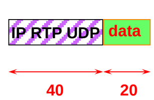
- **For support of applications with real-time properties**
  - loss detection
  - timing reconstruction (intra-media synchronization)
  - content (media) identification
- **Lighter than TCP**
  - no retransmission, no flow control
  - TCP header: **40 bytes**; RTP header: **12 bytes**

On header overhead

- **IP+UDP+RTP headers =** ***40 bytes***
- **At 64 Kbps PCM**
  - **20** ms = **160** Bytes
  - overall rate = **80** Kbps



- **At 8 Kbps** (e.g. with G.729) (e.g. over modem lines)
  - **20** ms = **20** Bytes



IP, UDP, RTP compression

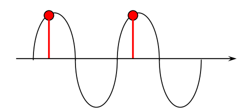
- **TCP/IP compression defined in RFC 1144**
- ***Robust Header Compression*** **(*ROHC*)**
- Can reduce to ***1 byte*** (best case, no UP checksum, …)
- **0-byte compression underway**
  - for IP cellular networks using existing air interfaces (such as GSM and IS-95)
  - most pkts will have no header

Audio/video network requirements

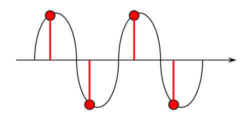
- **Key requirements**
  - Bit rates
  - Transit delay **variation**
- **Other requirements**
  - Transit delay
  - Error rate

Nyquist theorem

- **To faithfully represent an analog signal**
  - if maximum frequency ***f***
  - sampling rate at least ***2f***
- **Application to audio**
  - if sampling rate is **8 kHz**
  - bandwidth is **3.4 kHz**



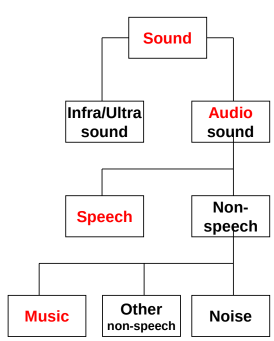
Sampling at *f* :
impossible to reconstruct



Sampling at *2f* :
easier to reconstruct

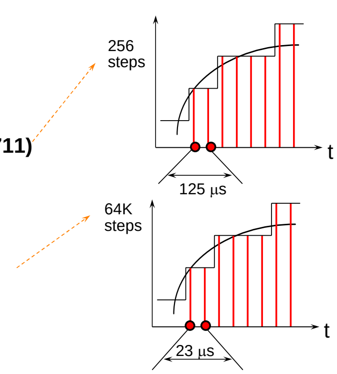
Sound, Audio, Speech, ...

- ***Sound***: vibration of matter
- ***Audio***: audible sound (by humans)
  - human audible spectrum:
    ~20 Hz - 20kHz
- ***Speech*: a particular type of audio signal**
  - "we hear better than we talk"
  - speech spectrum:
    ~50-10 kHz
- ***Music***: a particular case of non-speech sound



Parameters of audio bit rates

- **Sampling rate**
- **Amplitude depth**
- **PCM telephony (ITU G.711)**
  - 8 kHz
  - 8 bit/sample
- **PCM CD-quality**
  - 44.1 kHz
  - 16 bit/sample

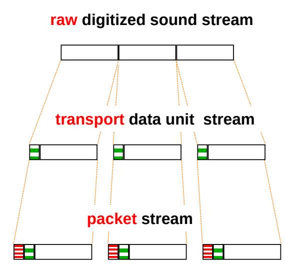


Audio bit rate requirements

Quality	Technique or standard	Kbps	Compr.
Telephone quality			
Standard	G.711 PCM	64	
Standard	G.721 ADPCM	32	Y
Lower	G.728 LD-CELP	16	Y
Lower	GSM	13	Y
Standard-	**G.729 LD-CELP**	**8**	**Y**
Lower+	CELP	5-7	Y
CD Quality			
Consumer CD-audio	CD-DA	1441 (stereo)	
Consumer CD-audio	MPEG with FFT	192-256	Y
Sound studio quality	MPEG with FFT	384	Y
Consumer CD-audio (MP3)	MPEG2.5 Layer III	128 (stereo)	Y

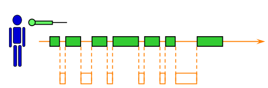
Which bit rate is actually needed?

- **Network overheads incl.:**
  - RTP header (12 bytes)
  - Transport Protocol header (usually UDP, 8 bytes)
  - IP header (20 bytes)
  - Example:
    raw G.711 **64** Kbps requires from **68** to 80 Kbps
- ***However, speech contains silence***

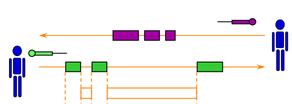


Silences in speech

- **Monologue**
  - typically **15%** silence



- **Bi-party telephone conversation**
  - **20%** silence for overall conversation
  - **60%** silence for each party
  - If silence suppressed, required bit rate is in effect **<40%** of nominal raw bit rate



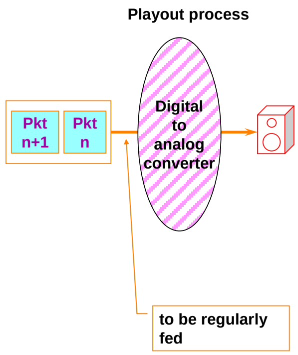
Video bit rate requirements

Quality	Technique or standard	Mbps	Compr.
Video conf. quality	**H.261**	**0.1**	**Y**
VCR quality	**MPEG-1**	**1.2.**	**Y**
Broadcast quality			
Compressed	**MPEG-2**	**2-4**	**Y**
Compressed	**MPEG-4**	**2**	**Y**
Studio-quality digital TV			
Uncompressed	**ITU-R 601**	**166**	
Compressed	**MPEG-2**	**3 to 6**	**Y**
HDTV			
Uncompressed	**CD-DA**	**2000**	
Compressed	**MPEG-2**	**25 to 34**	**Y**

Reproduced from "Understanding Networked Multimedia" by François Fluckiger, Prentice Hall 1995

Transit delay variation (Jitter)

- Receiver to wait a ***delay offset*** before playout
- **Called *delay equalization***
- **Increases overall end-to-end latency**

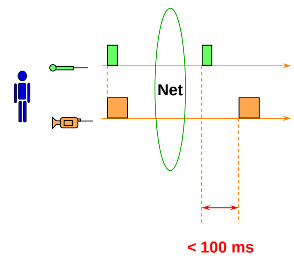


Inter-media synchronization

- **Particular case:**

  lip synchronization

  - A skew of **80-100 ms** is **generally tolerated**



Audio/video relative priorities

- The ear behaves as a ***differentiator***
- The eye behaves as an ***integrator***
- Toleration of transmission errors affecting sound much lower than for video
- **When audio and video streams part of the same application competing for network resources, audio stream should have priority**

Principle of Compression

1. **Remove non-perceived components from original signal**
2. **Remove redundancies from the original signal**

The Effect of compression

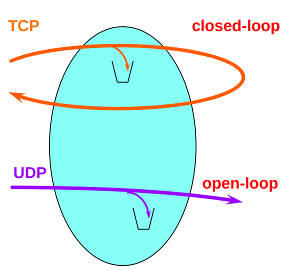
- **Redundancy essential for resistance to errors**
- **Compressed data more sensitive to errors**

In summary

- **Audio-video traffic place stringent requirements on**
  - **bandwidth guarantees**
  - **transit delay variation**
- **Techniques to support their QoS requirements:**
  - **Diffserv**
    - e.g. marking all **voice traffic** with the highest **absolute** priority, ahead of any other traffic
  - **MPLS with reservation over paths**
    - to carry aggregates of audio/video streams

Unresponsive flows

- **Unresponsive flows do not react to congestion indication (pkt loss)**
- **Can create**
  - bandwidth **starvation** inflicted to well-behaved responsive traffic



End of

Part 3

Transporting Multimedia over the Internet

Further reading ...

- **Internetworking with TCP/IP, vol 1**
  Douglas E. Commer, Prentice Hall, ISBN 0-130-183806
- **Computer Networks, Ed. 4**
  Andrew Tannenbaum, Prentice Hall, ISBN 0-130-661023
- **Understanding Networked Multimedia**
  Francois Fluckiger, Prentice Hall, ISBN 0-131-90992-4
- **Understanding Media**
  Marshal Mac Luhan, The MIT Press, ISBN 0-262-631159