Networking QoS Basics

François Fluckiger
CERN, Geneva

- Setting the Scene
- Internet QoS Options
- TCP and Congestion Control
- Multimedia over the Internet

Why is QoS Important in GRID Environments?

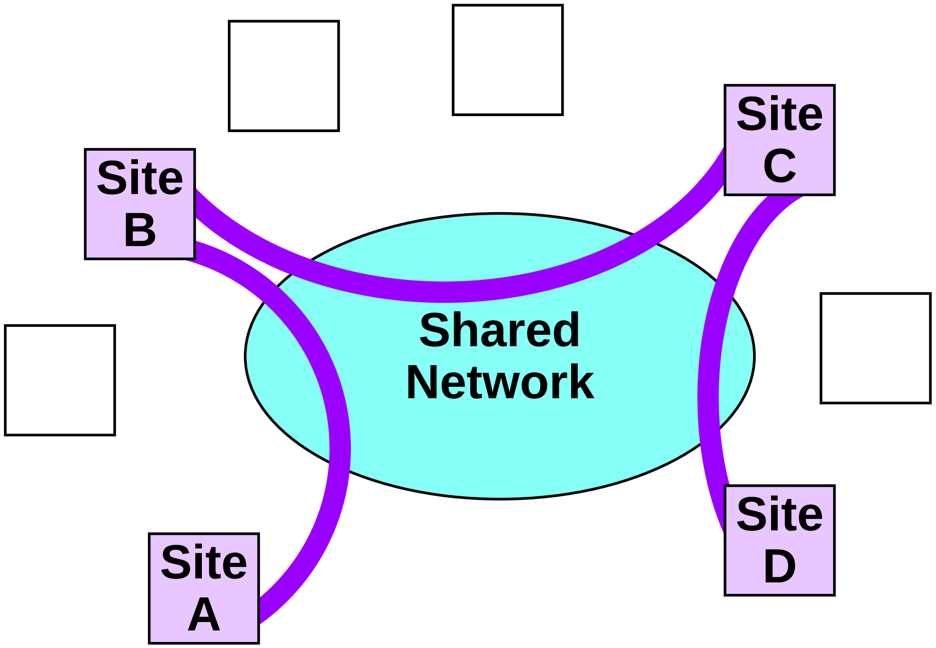
- **GRID of systems**
  - **Fast transfers => High Bit rate connections**
  - **Predictable behavior**
  - **Availability**

- **Grid of people**
  - **New type of traffic for collaborative activities**

Why improving QoS Guarantees?

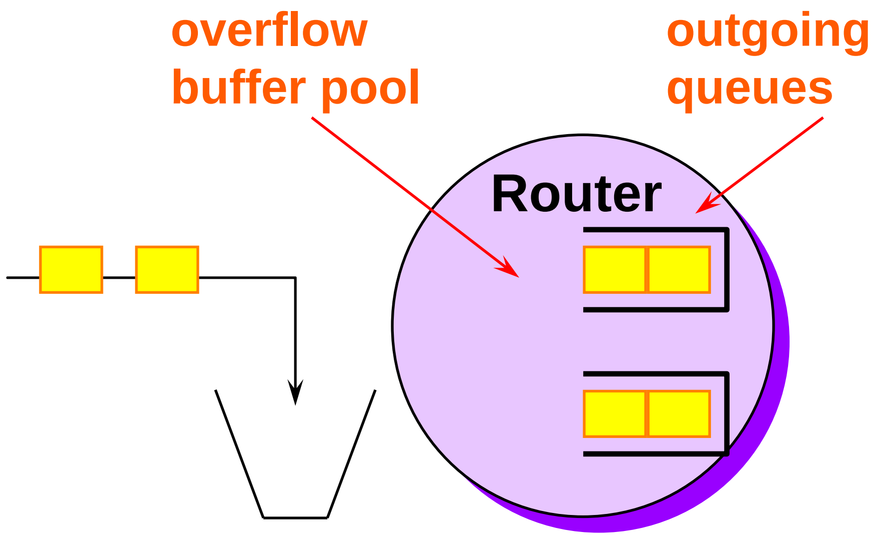
- Create **"virtual private networks"** with performances guarantees

- Requires set of QoS-guaranteed **point-to-point** pipes

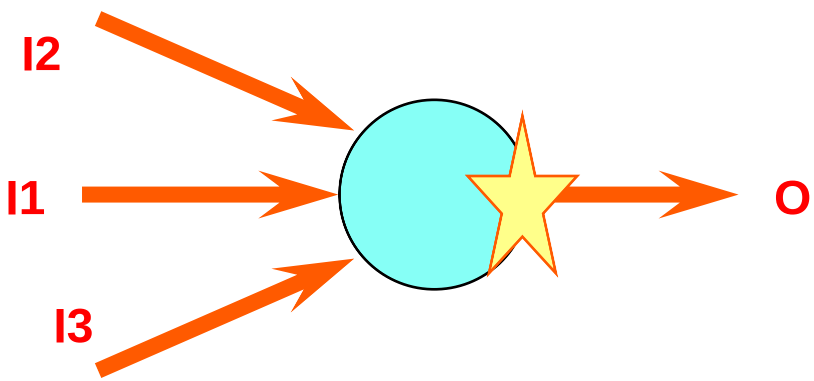


Why improving QoS Guarantees? (cont)

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



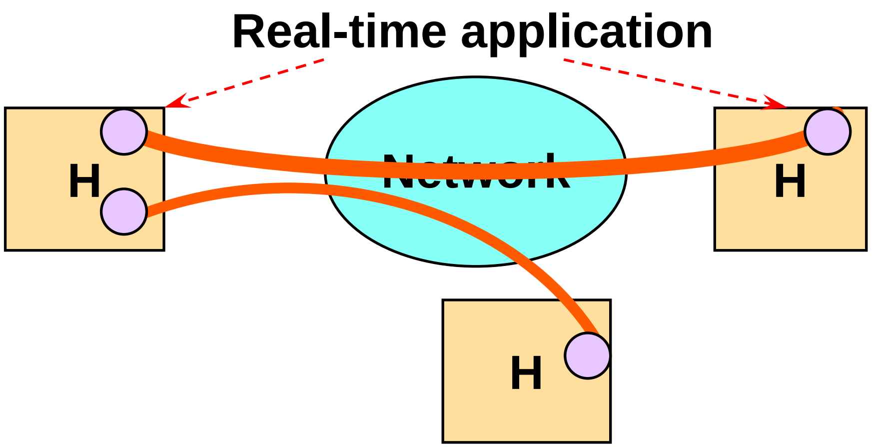
  - **N-to-1** problem



Σ Input traffic > Output capacity

Why improving QoS Guarantees? (cont)

- 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, …

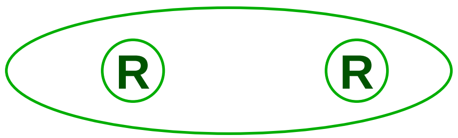
What the end-system promises to respect

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

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

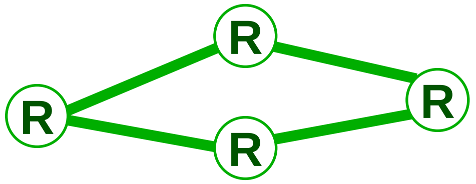
Internet: Back to Basics

- Formed of switches called routers

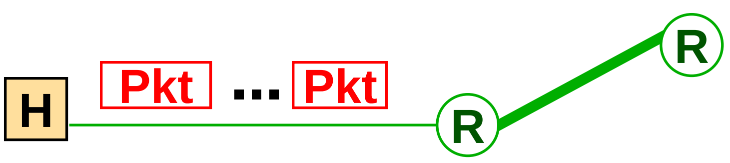


- Routers interconnected by links

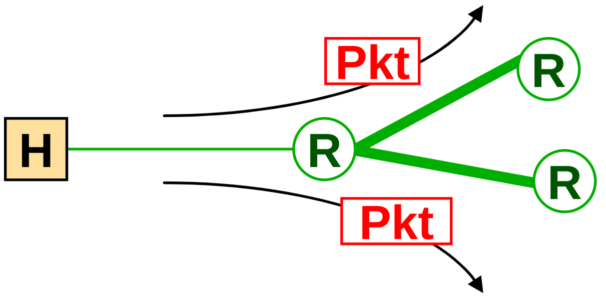
- Topology usually meshed



- Hosts chop data stream into blocks called packets

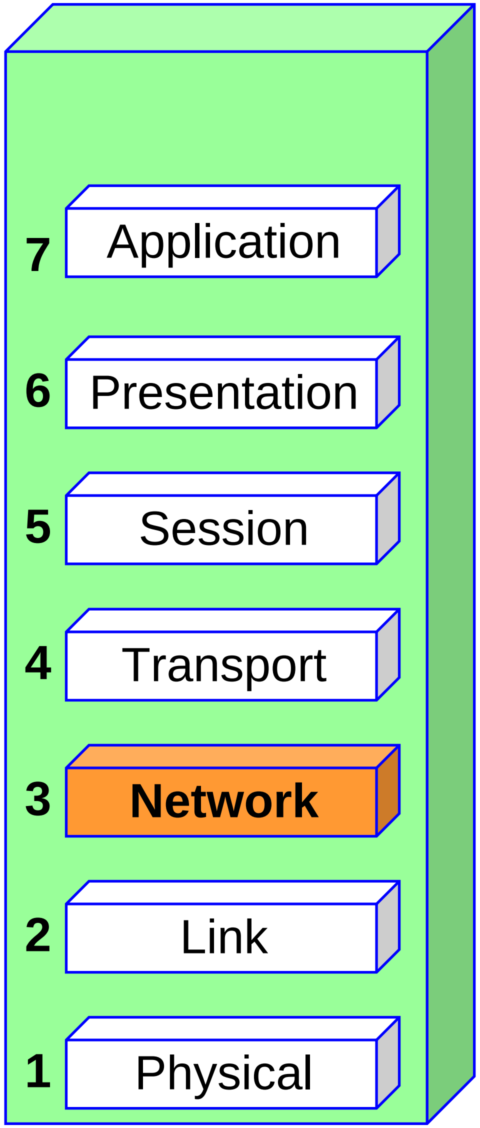


- Routers switch individual packets

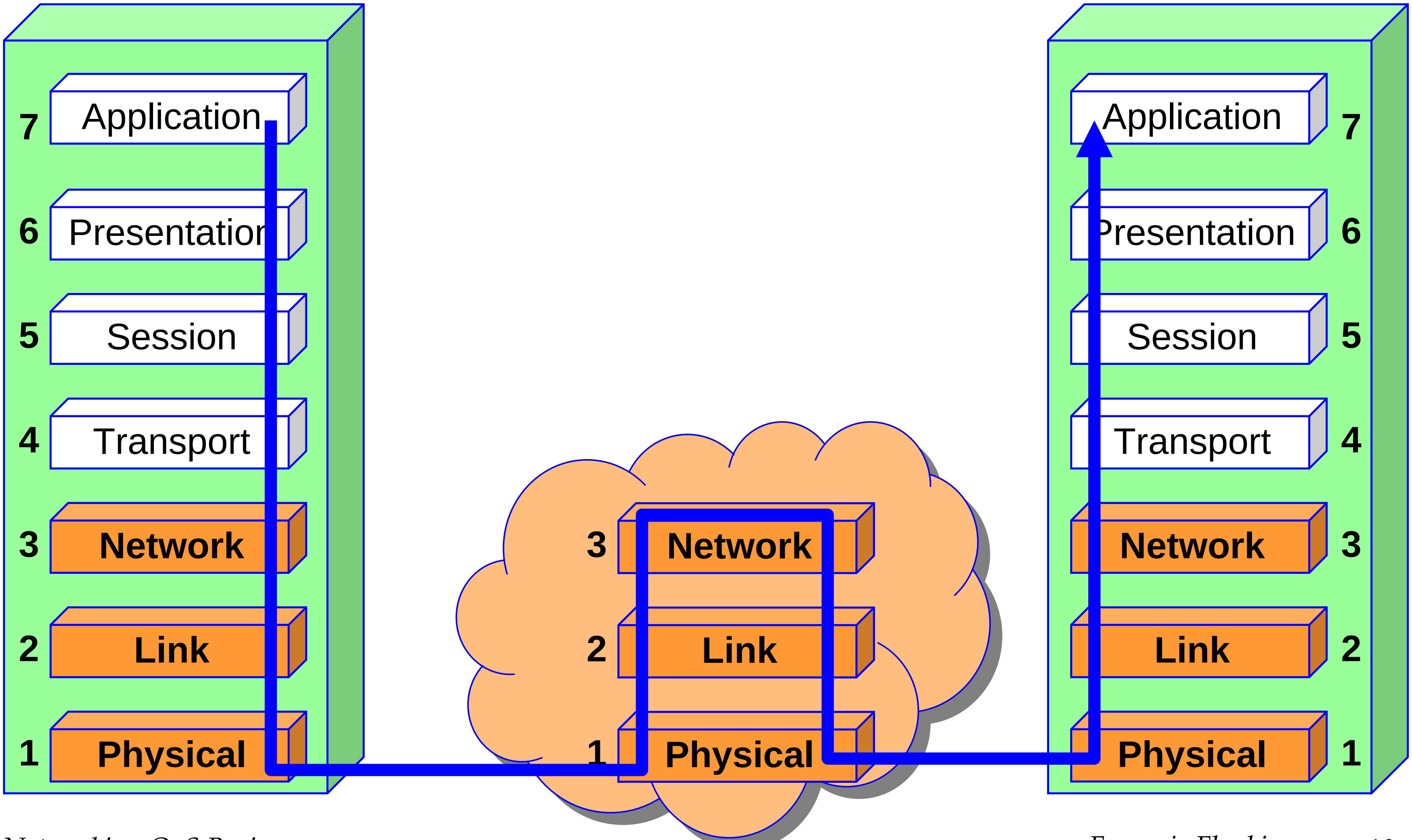


IP

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

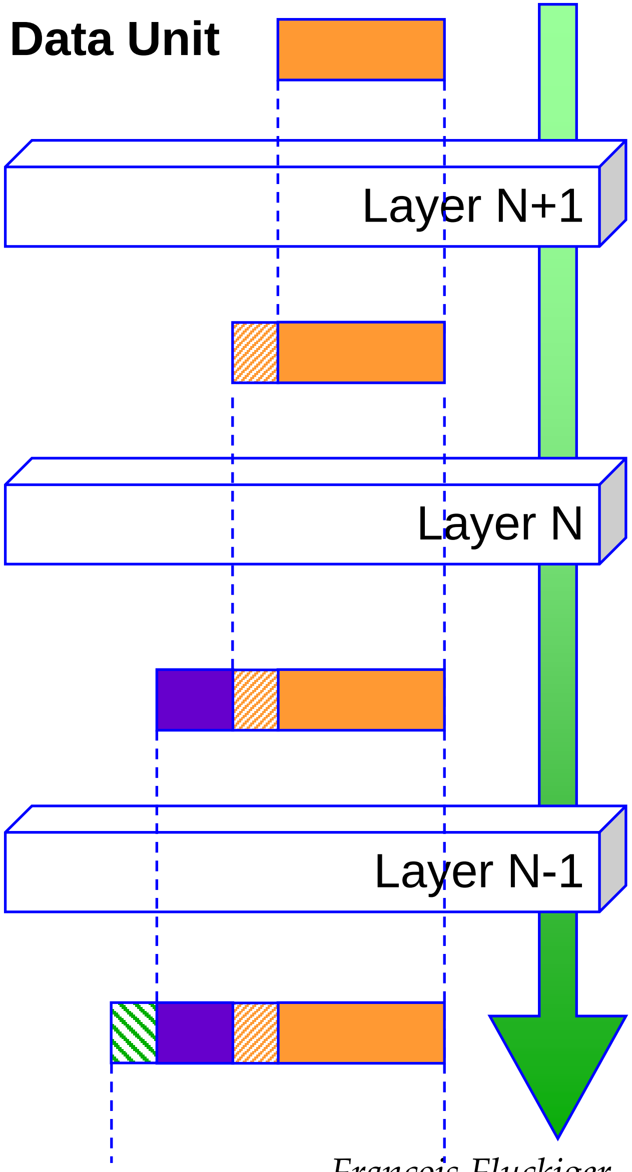


Protocols understood by routers



Layering principle (emission)

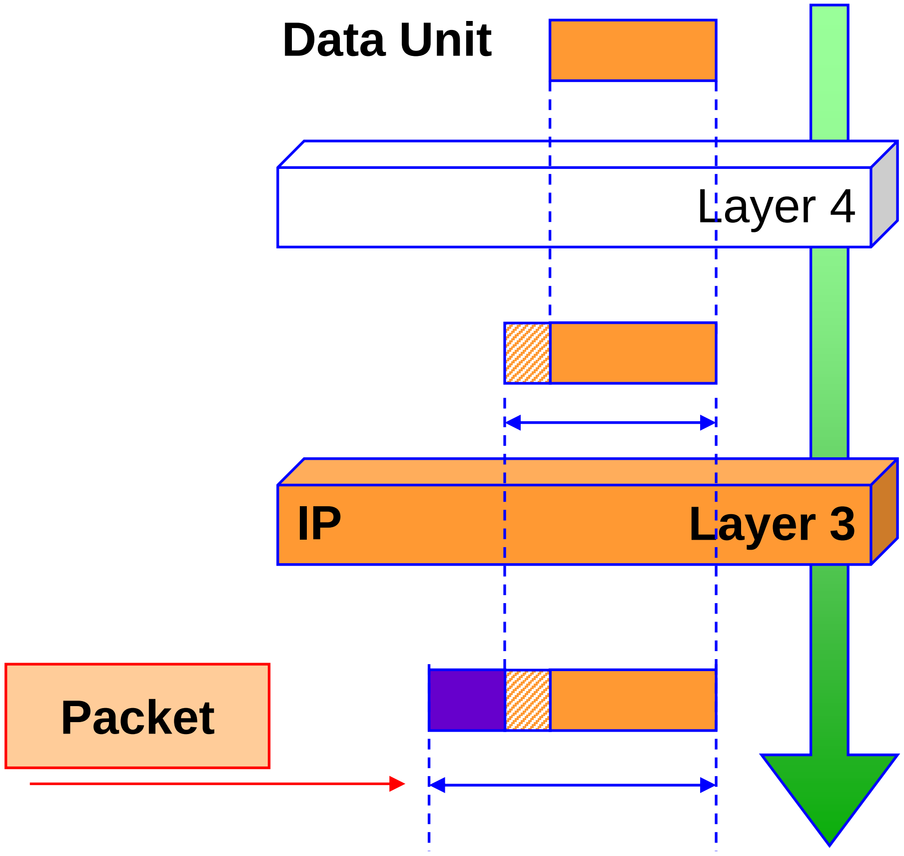
- Each protocol layer **N** adds a **Header** to the data unit received from layer **N+1** (1)
- Header contains **control** information; e.g. :
  - Numbering of the data unit
  - Coding of the destination
  - Codes for error detection
  - Priority of the data unit



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

CO and CL networks: Examples

Connection-oriented	Connectionless
■ Telephone	■ Post office ■ Road Network
■ “Lambda on-demand” ■ ISDN ■ ATM ■ Frame Relay ■ *SNA* ■ *X.25*	■ Ethernet LANS ■ Internet IP ■ *DECnet*

CONS vs CLNS

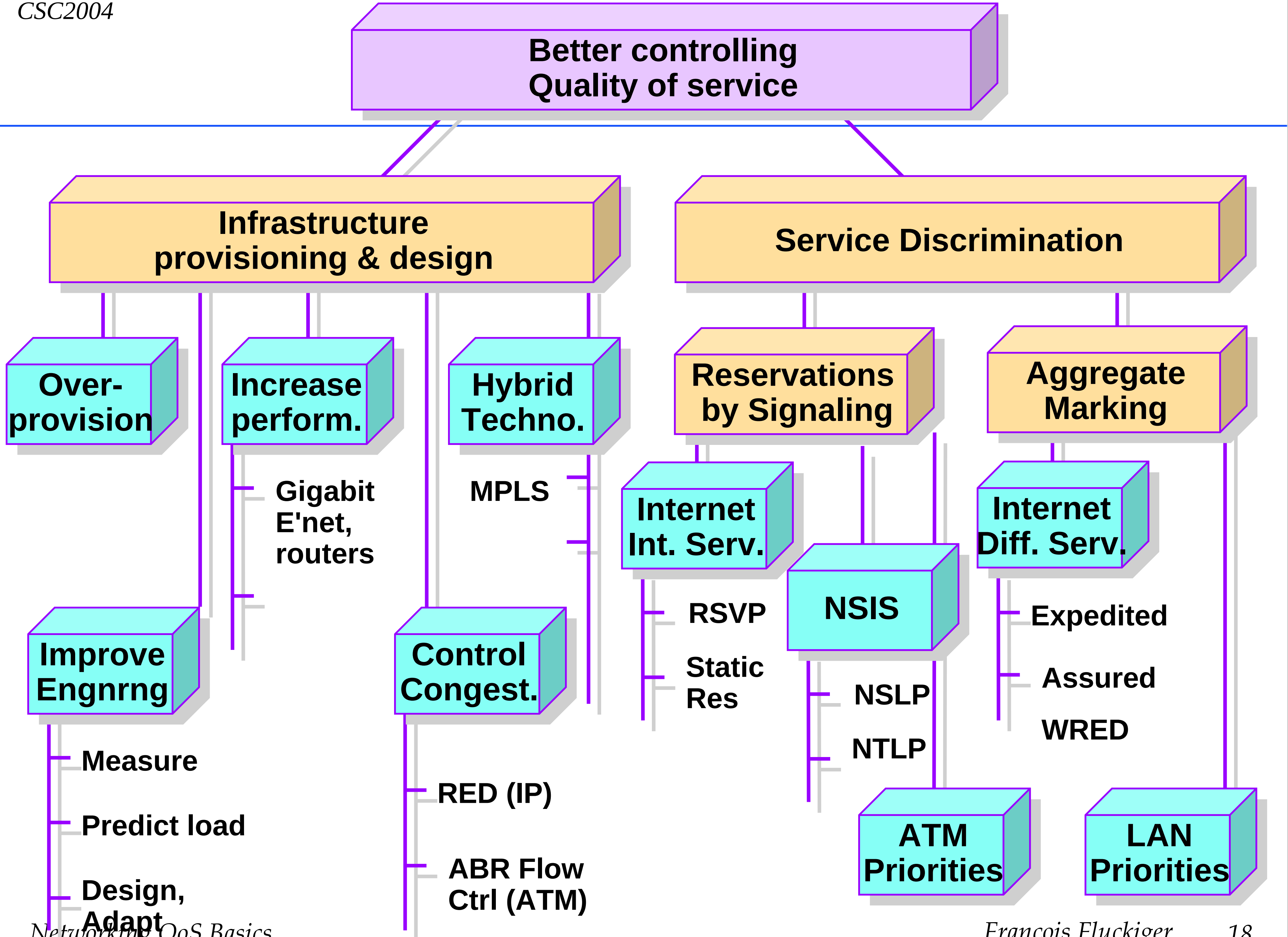
	+
CONS	**CLNS**
■ **Traffic more predictable**	■ **No call set-up delay before sending a packet**
■ **Easier for network to reserve resources**	■ **Routing possibly more dynamic**
■ **QoS guarantees easier to provide**	■ **Resilience**

- Setting the Scene
- Internet QoS Options
- TCP and Congestion Control
- Multimedia over the Internet

What to do to improve Network QoS?

- **Provisioning & design** of the infrastructure
  - Over-provision (often used over conventional IP or LAN nets)
  - Improve engineering
  - Enhance performance
  - Avoid/manage congestion
  - Combine technologies into hybrid solutions

- **Traffic discrimination**
  - Internet **packet marking** and resource **reservation**
  - LAN **frame marking**
  - ATM **prioritization**



Better controlling
Quality of service
Infrastructure
provisioning & design
Service Discrimination
Over-
provision
Increase
perform.
Hybrid
Techno.
Reservations
by Signaling
Aggregate
Marking
Gigabit
E'net,
routers
MPLS
Internet
Int. Serv.
Internet
Diff. Serv.
NSIS
RSVP
Static
Res
Expedited
Assured
WRED
Improve
Engnrng
Control
Congest.
NSLP
NTLP
Measure
Predict load
Design,
Adapt
RED (IP)
ABR Flow
Ctrl (ATM)
ATM
Priorities
LAN
Priorities

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)*

Internet service discrimination

Two proposed techniques

- ***Integrated Services***

- ***Differentiated Services***

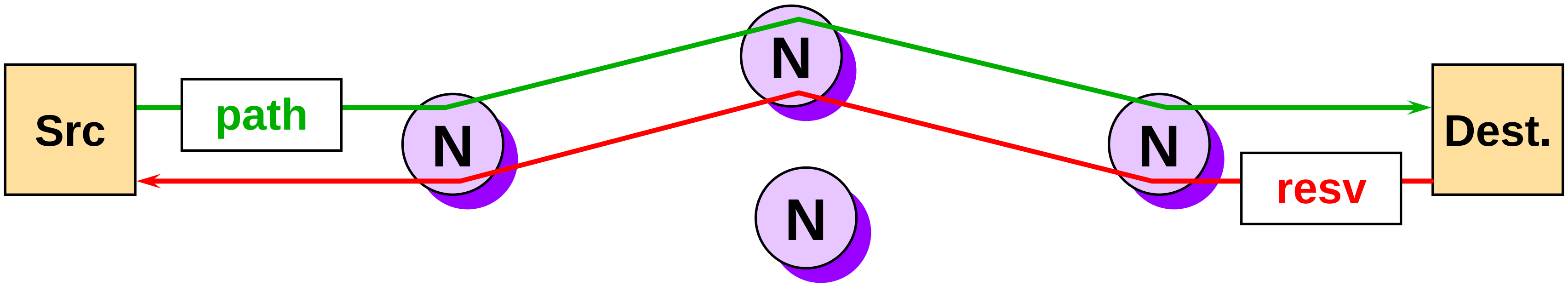
Integrated Services principles

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

RSVP: Resource Reservation Protocol

- **A <u>signaling</u> protocol**
- **No hard connections**
- **Unidirectional** reservations

RSVP protocol (simplified)



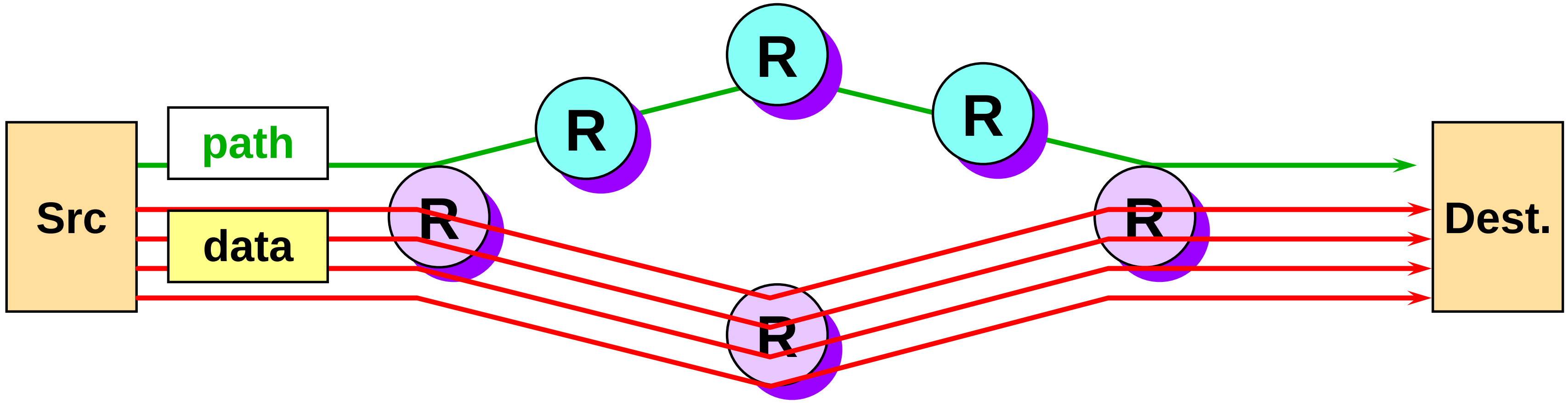
- "***PATH***" control message sent periodically by source
- "**PATH**" establish an RSVP state in intermediary routers
- dest replies with a "***RESV***" message
- "**RESV**" reserve resources on the route back
- if "**PATH**" not repeated after time-out, resources released

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



- "**path**" reserved over a **long** route
- **data** follow a **shorter** route

Concern #2: Scalability

- **States**
  - “How many soft-states can routers handle

- **Overhead**
  - Classifying packets between
    - Regular
    - Belonging to a flow with reservation

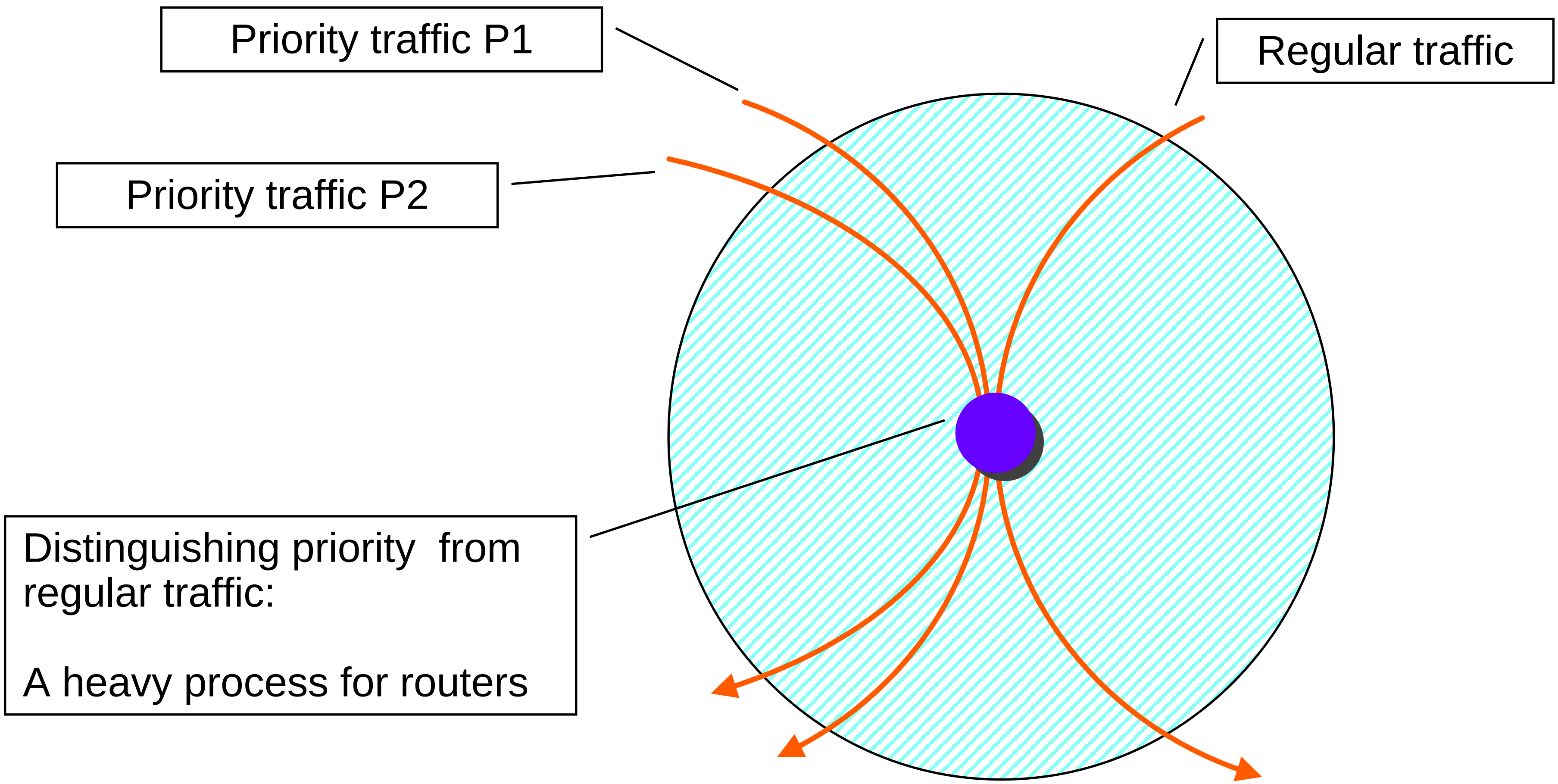
  - … an heavy process

Options for RSVP concerns

- **Concern #1: Route stability**
  - ***MPLS, Constraint-Based Routing***

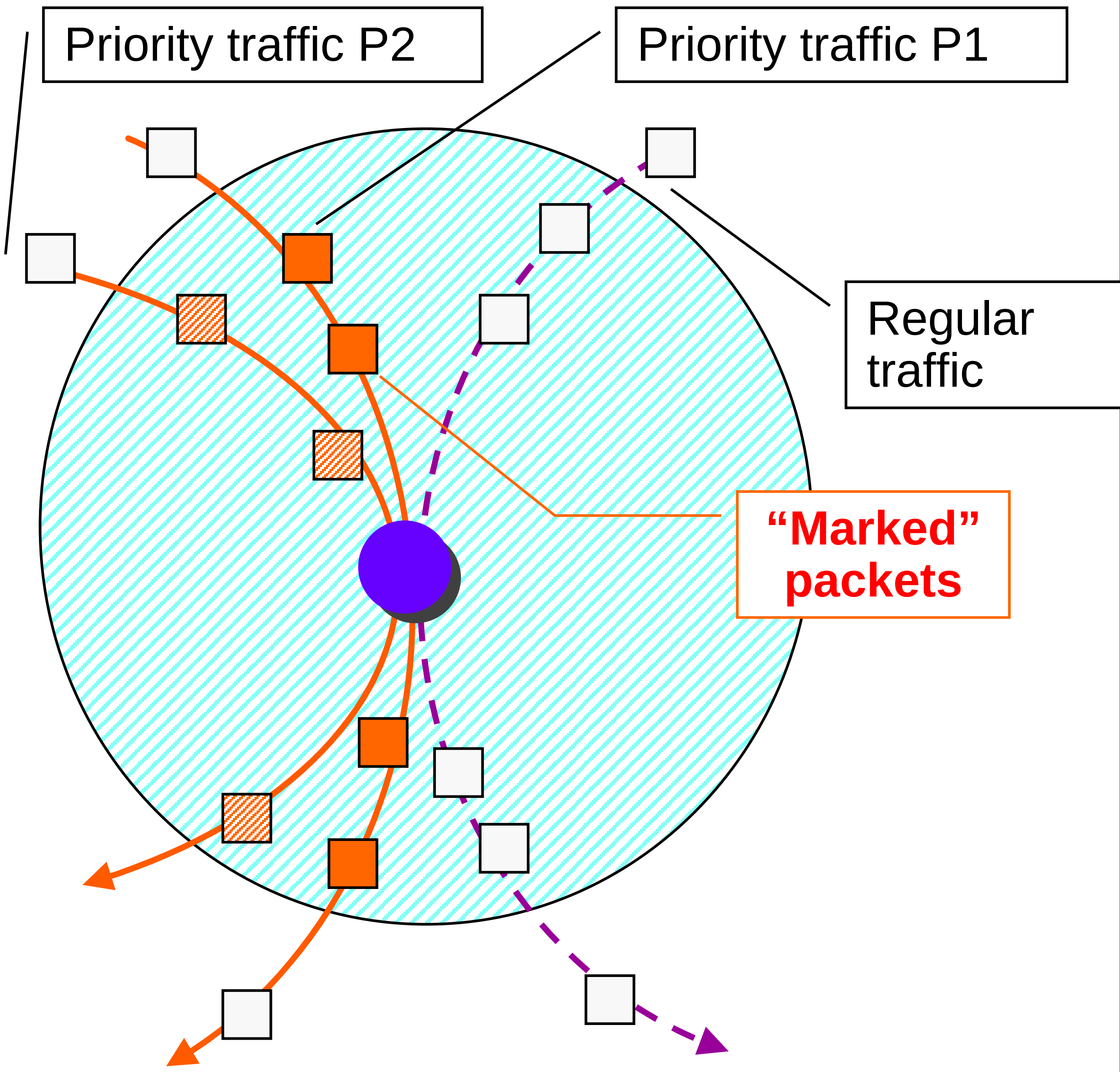
- **Concern #2: Scalability**
  - ***Diffserv***

Concern #2: Scalability



Diffserv Guiding Principles

- **A small number of priority levels defined**
- **Priority mark is inserted** before Pkts enter the “QoS improved core”
- Simple examination of the mark provides the priority



Diffserv Fast packet classification

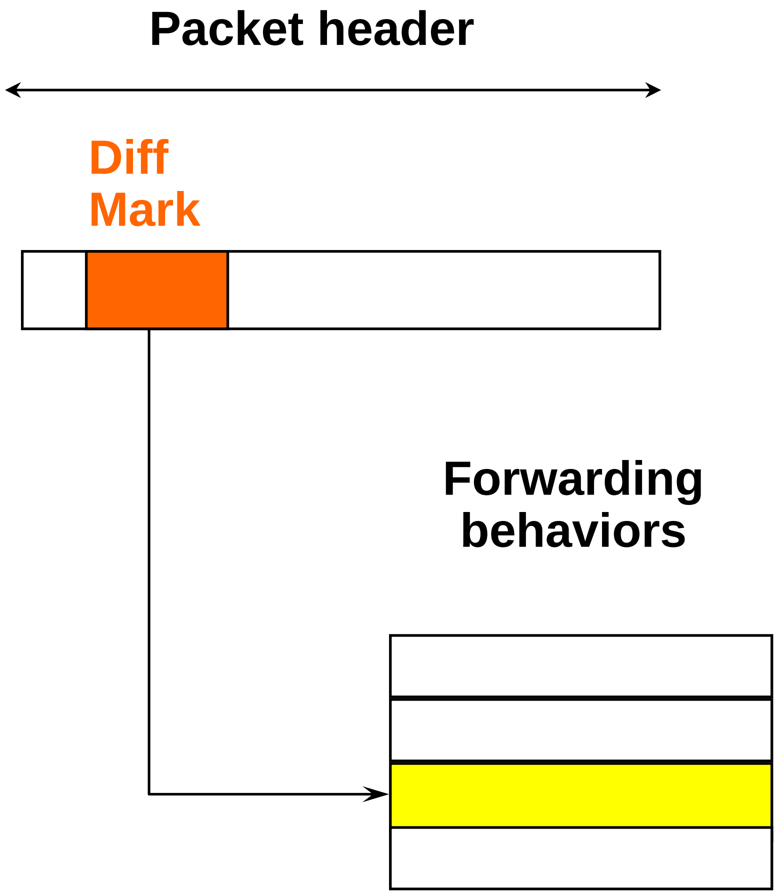
- At core routers

Differentiation mark

mapped

to node behaviors

behavior = how to proceed to guarantee the agreed quality of service



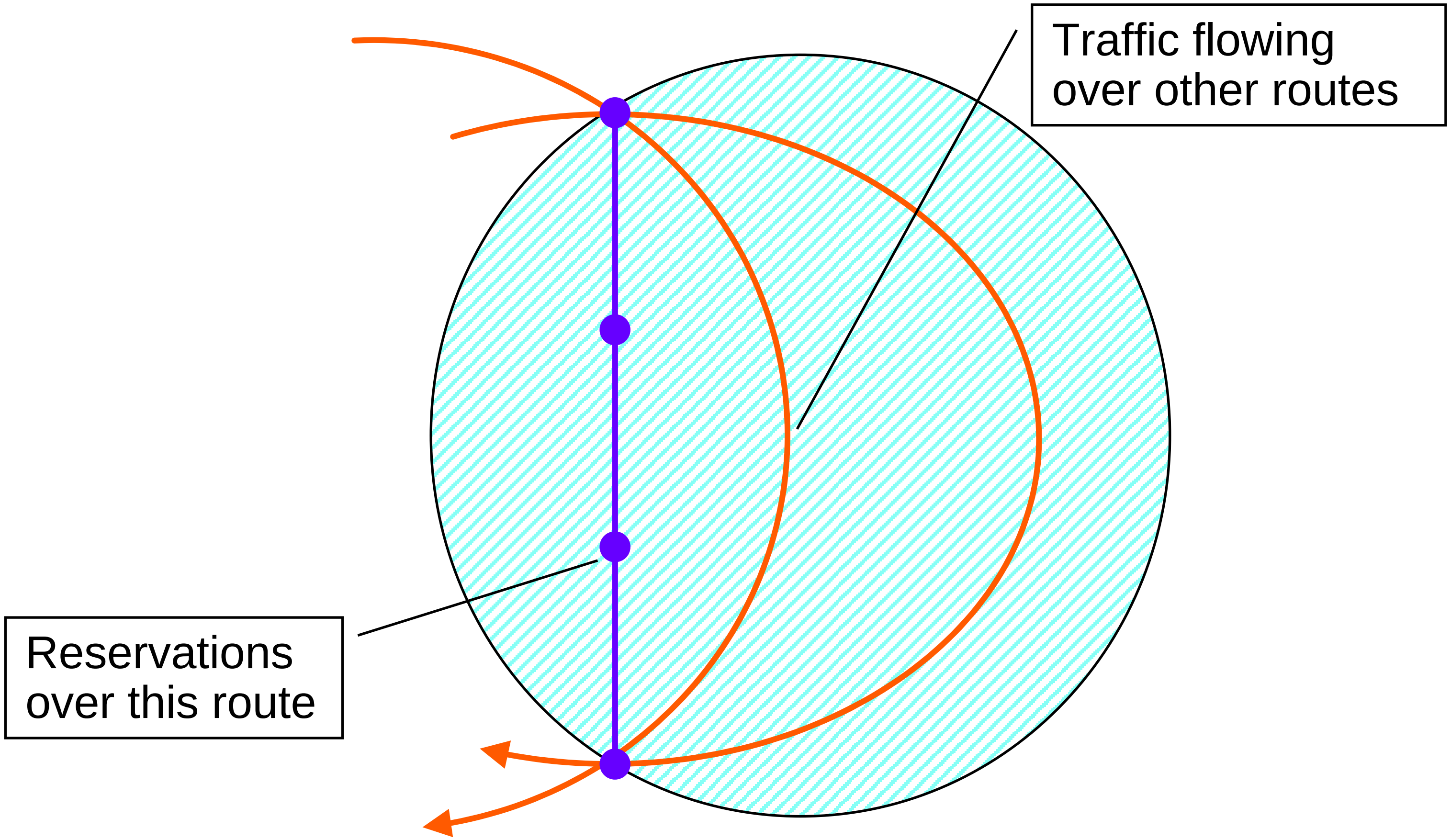
Implementing Packet Marking

- **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

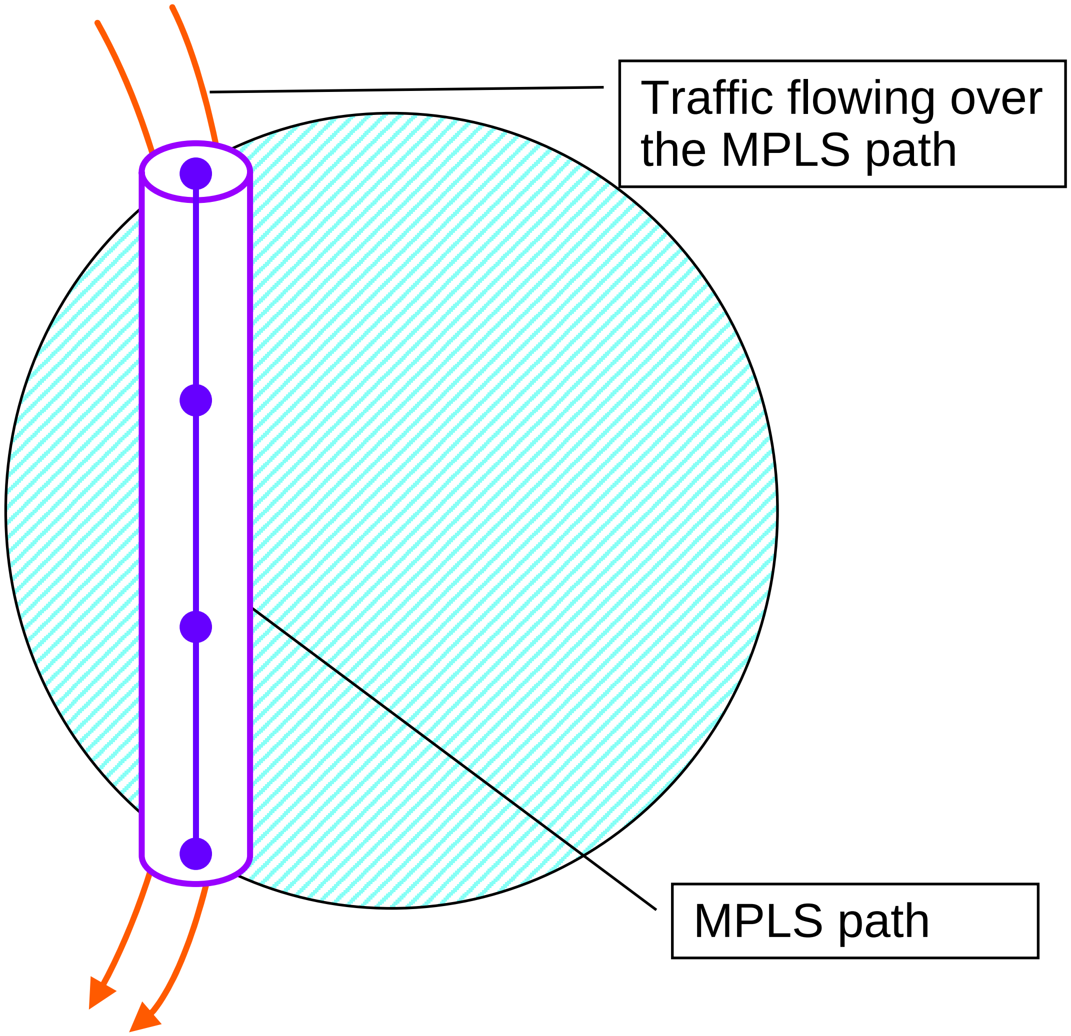
Concern #1: Route stability



MPLS

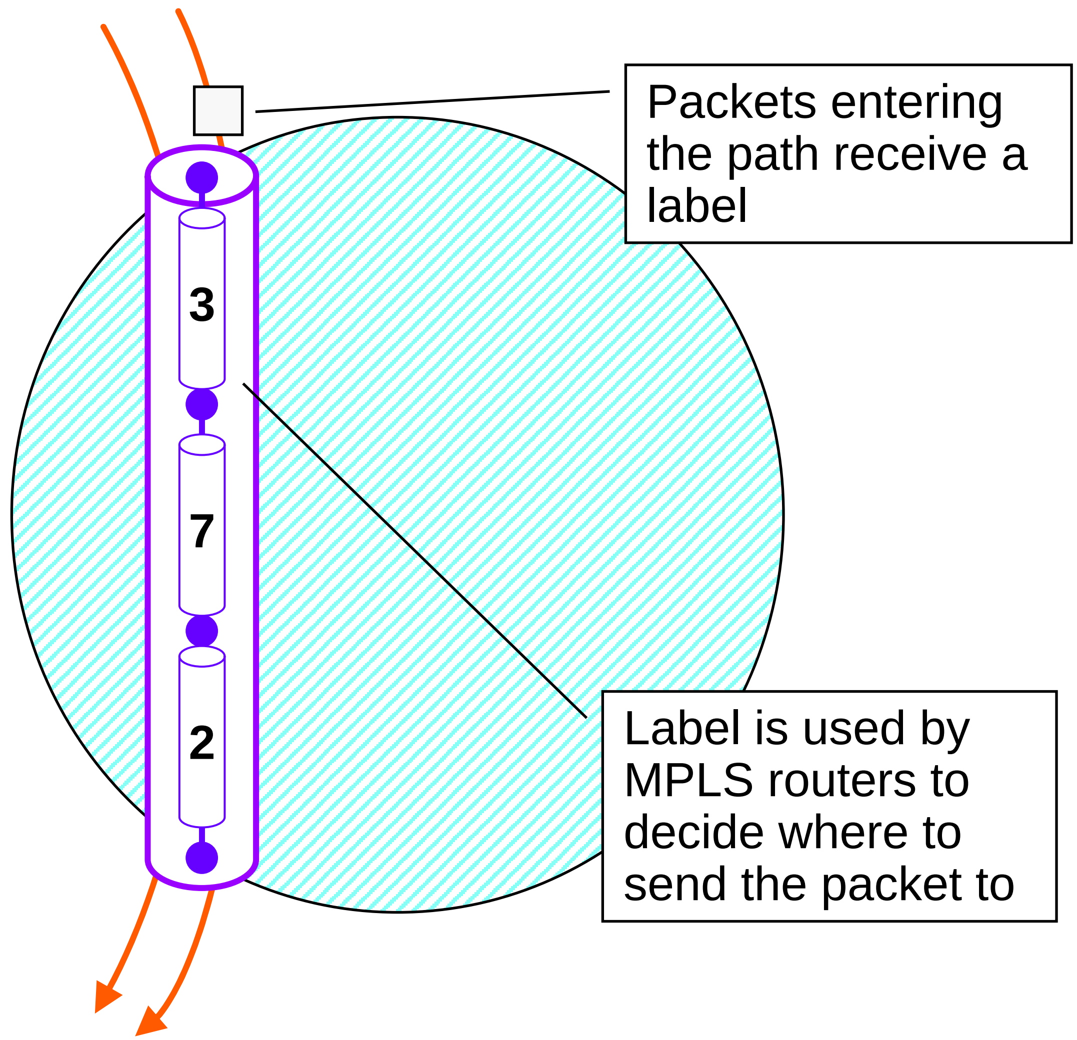
- **Create a “circuit” (called MPLS path)**
- **Force all traffic with**
  - **same destination**
  - **same Qos requirement**

to follow the path



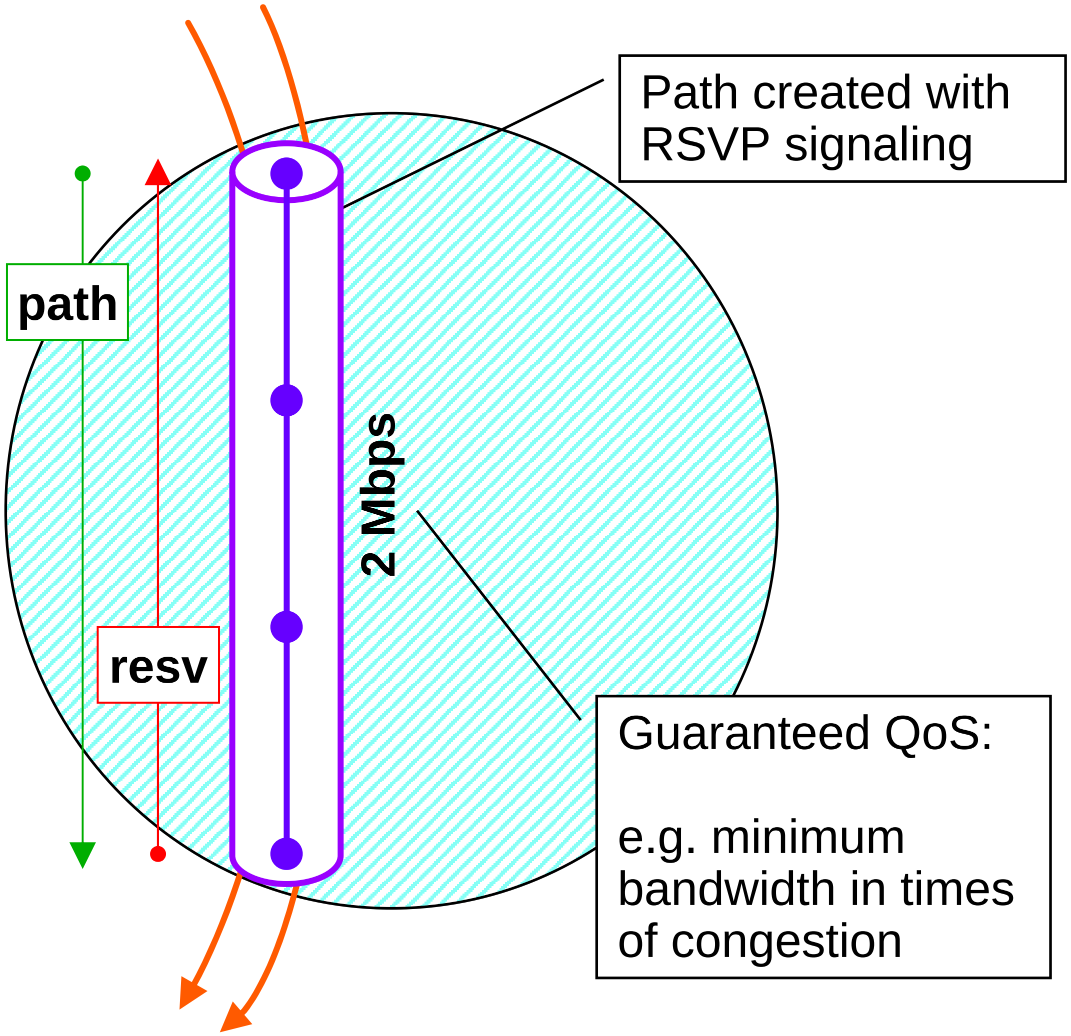
MPLS paths

- **Use “label switching” principles**
- **MPLS =**
  - concatenation of segments between routers
  - each identified by a numerical value (as with ATM, X.25)



MPLS and Reservation

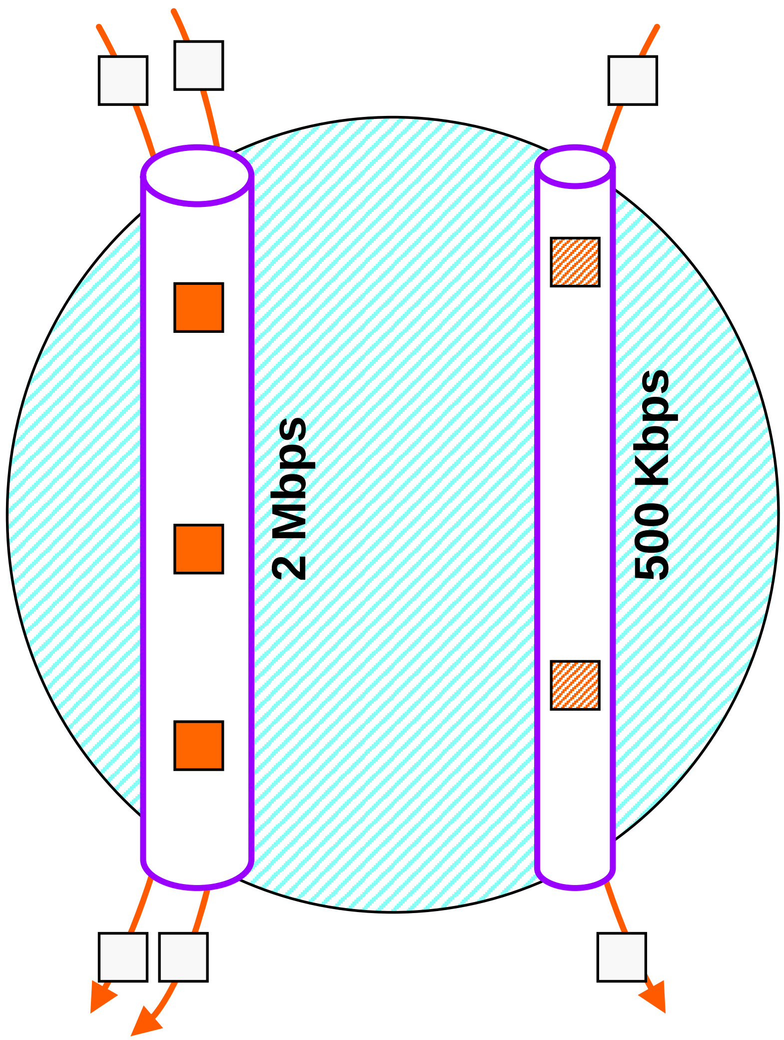
- **Use RSVP path/resv signaling to create the paths**
- **Reserve resources all over the path with RSVP**



MPLS and Diffsrev

- **For routers to know the priority and QoS attached to packets within paths:**

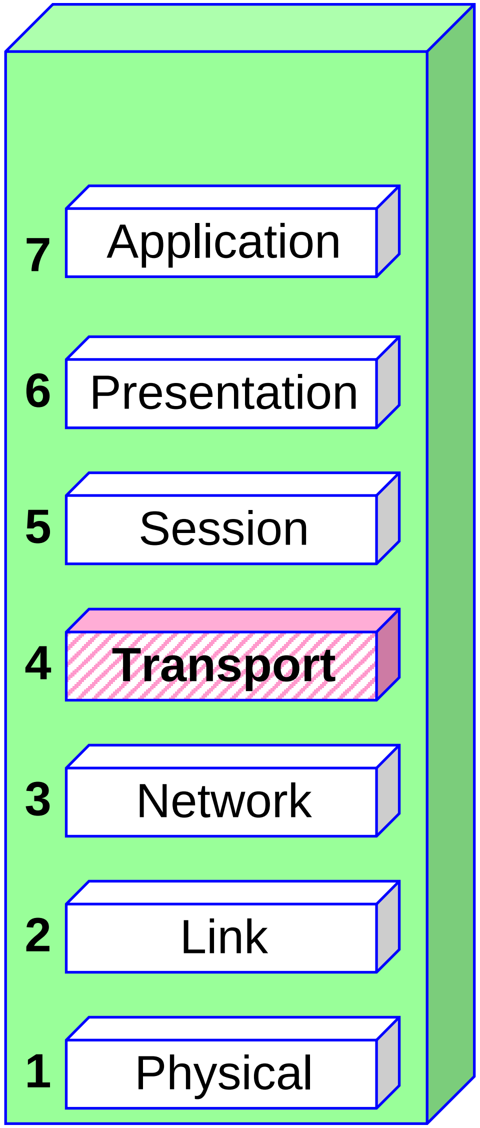
  Mark packet priority with Diffserv Mark



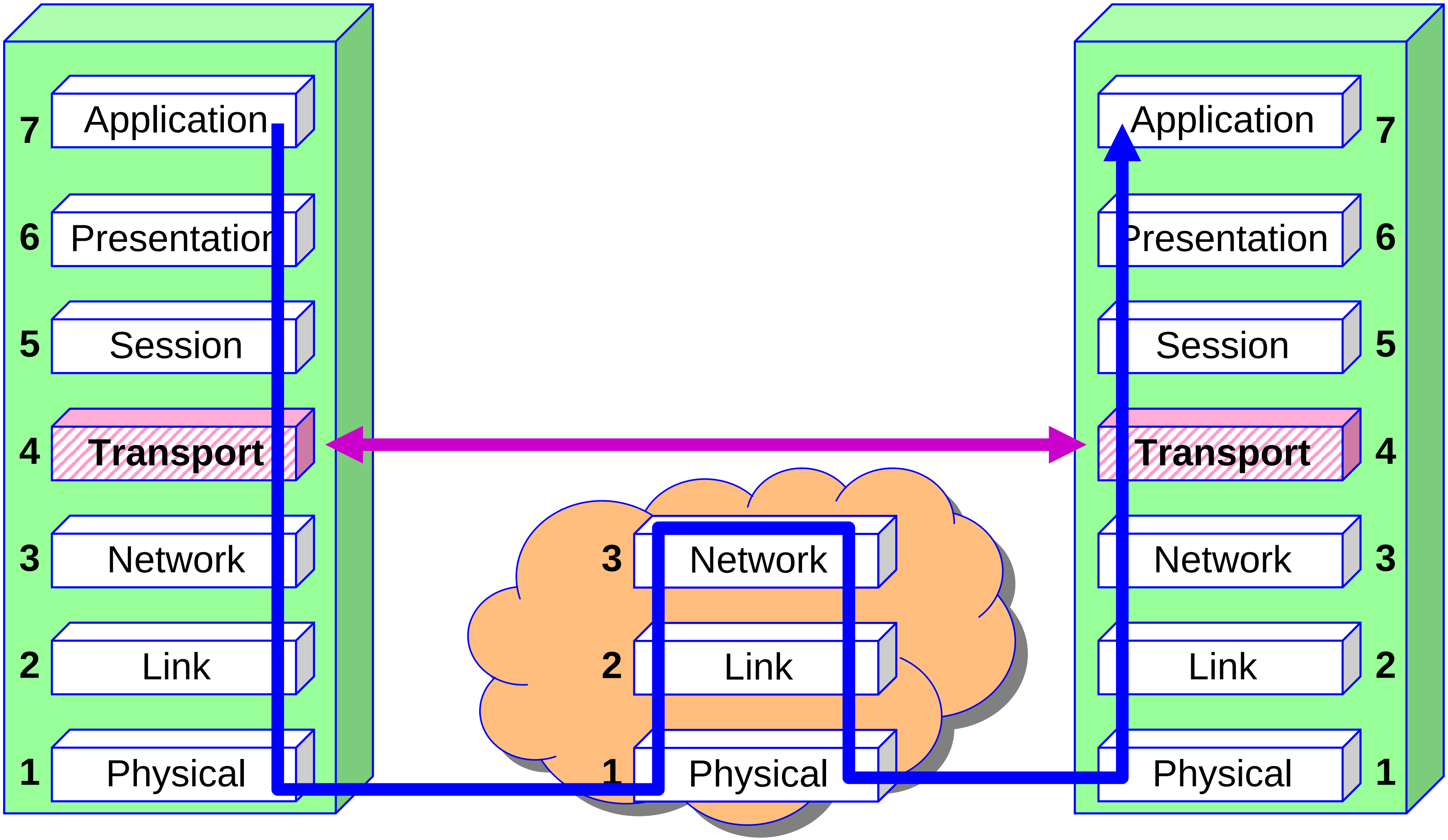
- Setting the Scene
- Internet QoS Options
- TCP and Congestion Control
- Multimedia over the Internet

TCP

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



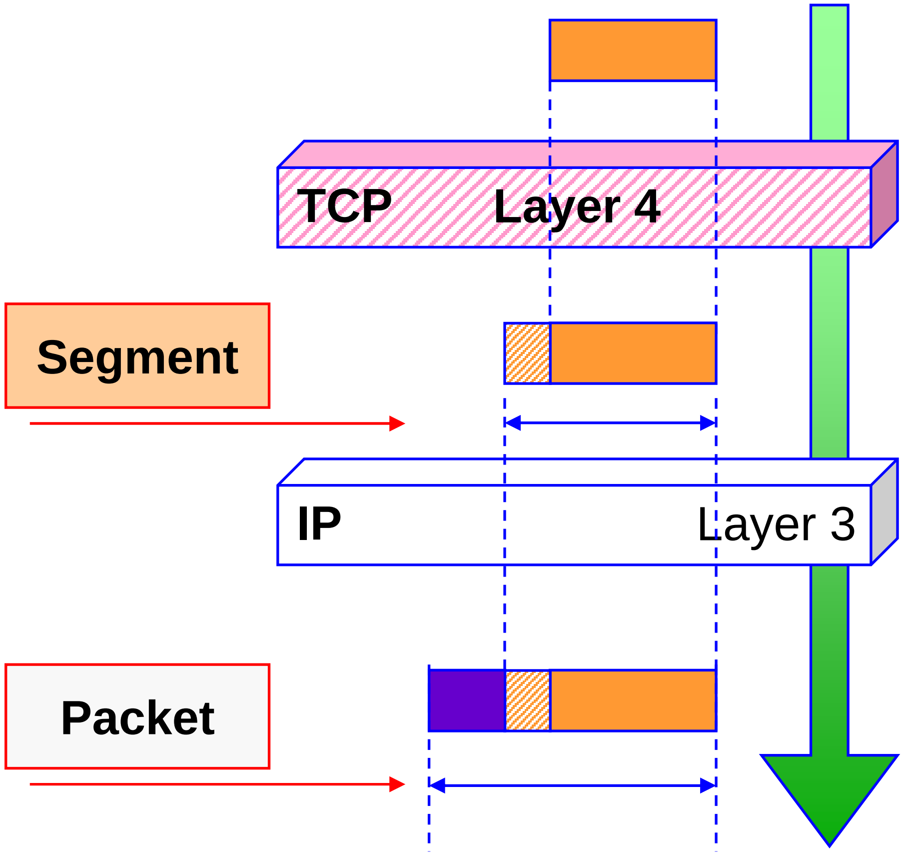
Protocols understood by hosts only



TCP Segments

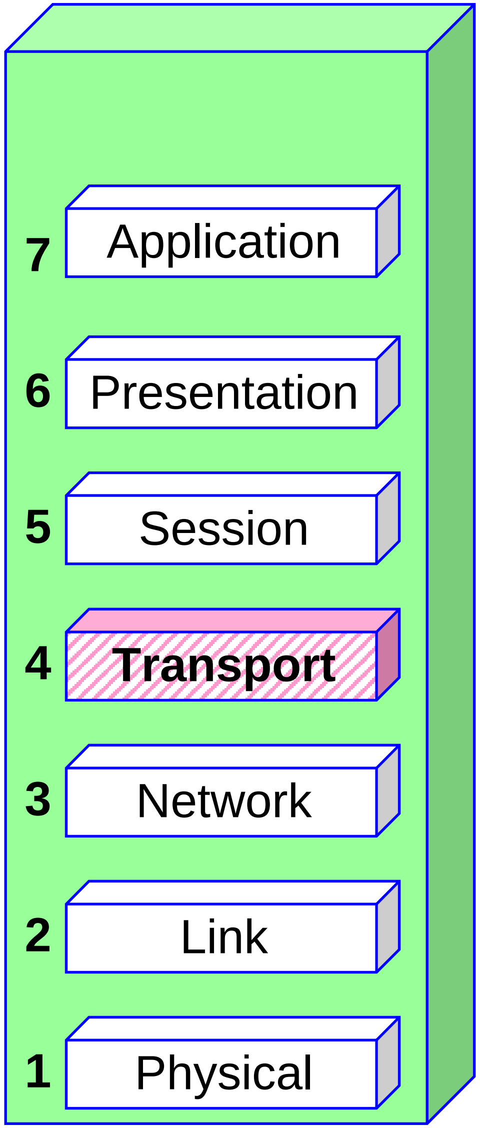
- Each **data unit** generated at a given layer has a specific name
- Data Unit at layer 4 (TCP) is called a **Segment**

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



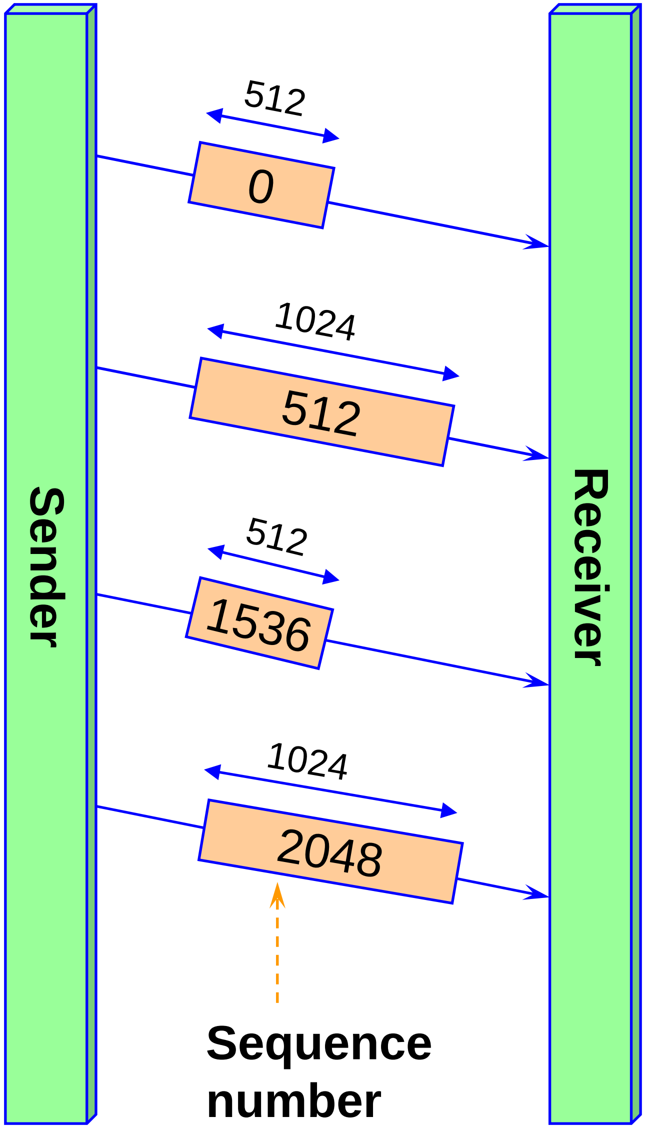
What does TCP Provide

- **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

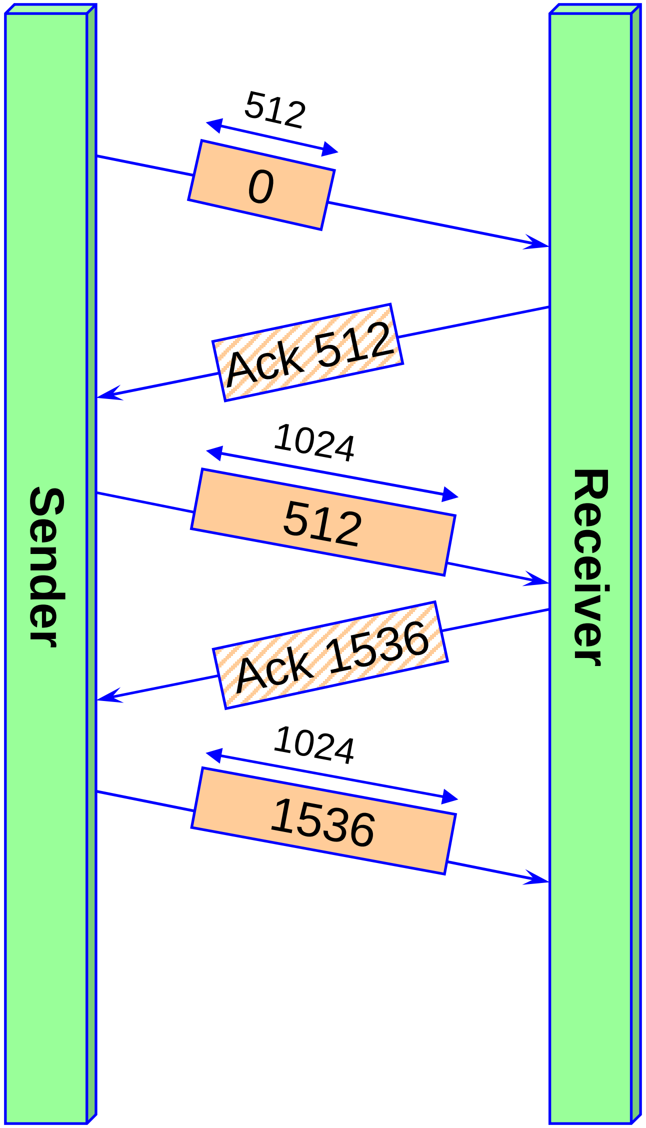


TCP Mechanisms for Errors

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

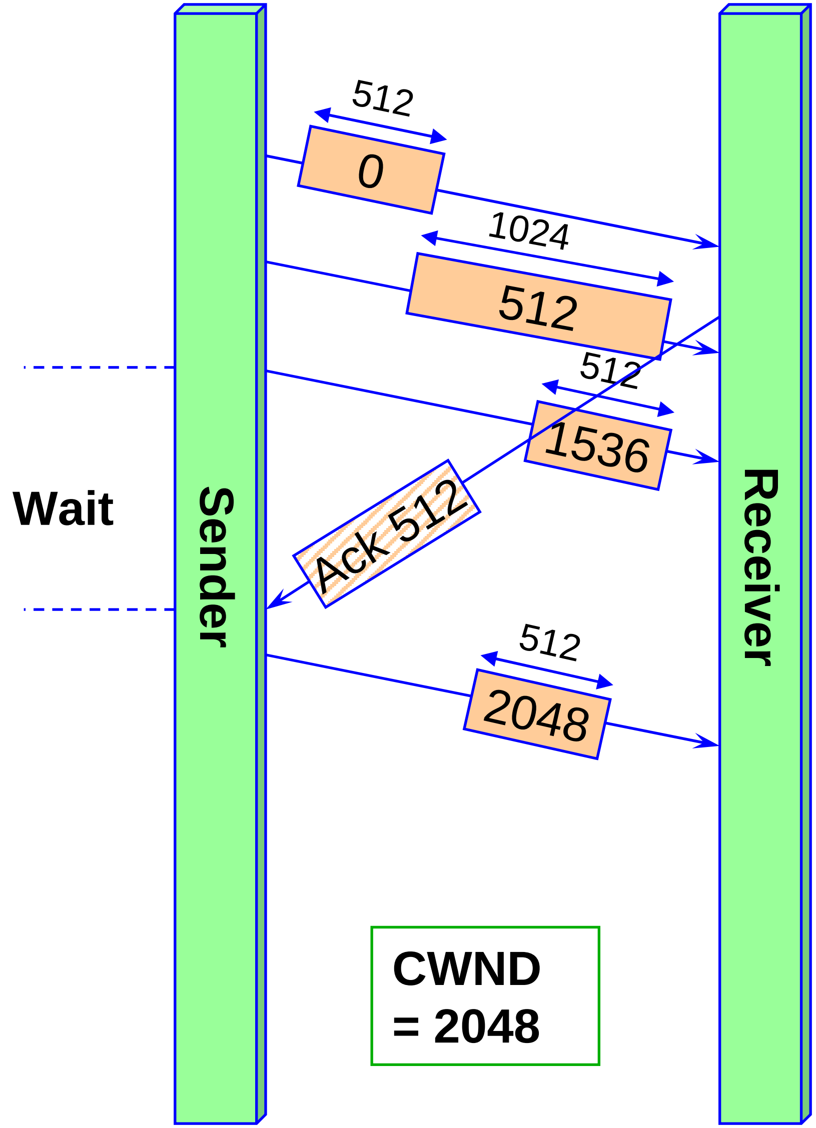
    " *I am now ready to receive octet #512 and beyond, because I correctly received all octets up to #511*"

- **Problem with this**
  - Sender needs to wait for ACK of segment 1 before sending segment 2



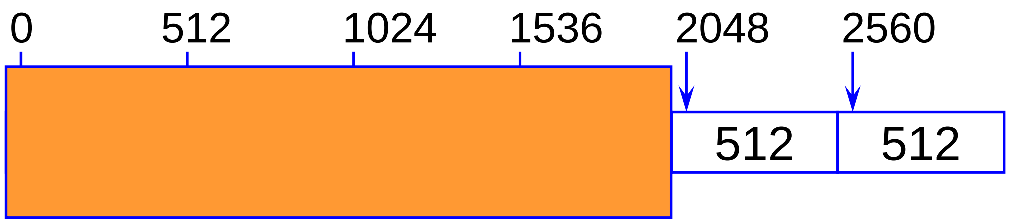
TCP Windowing

- **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**)

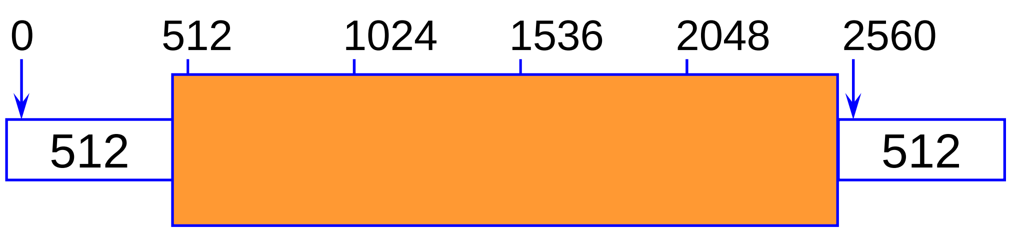


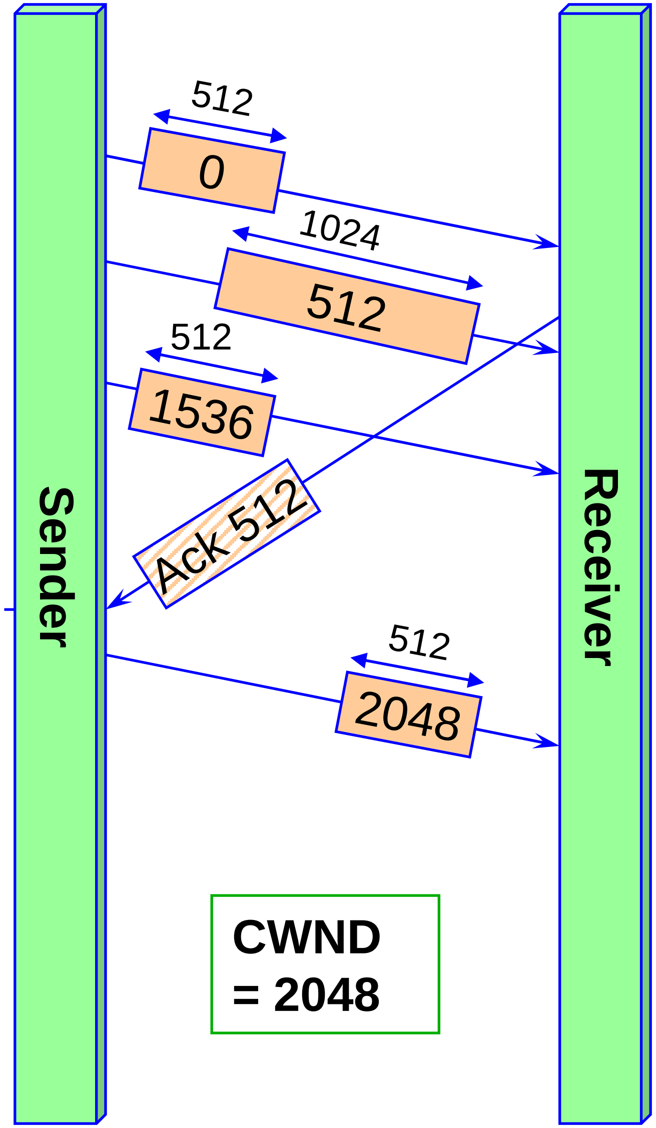
TCP Sliding Window

Initial window



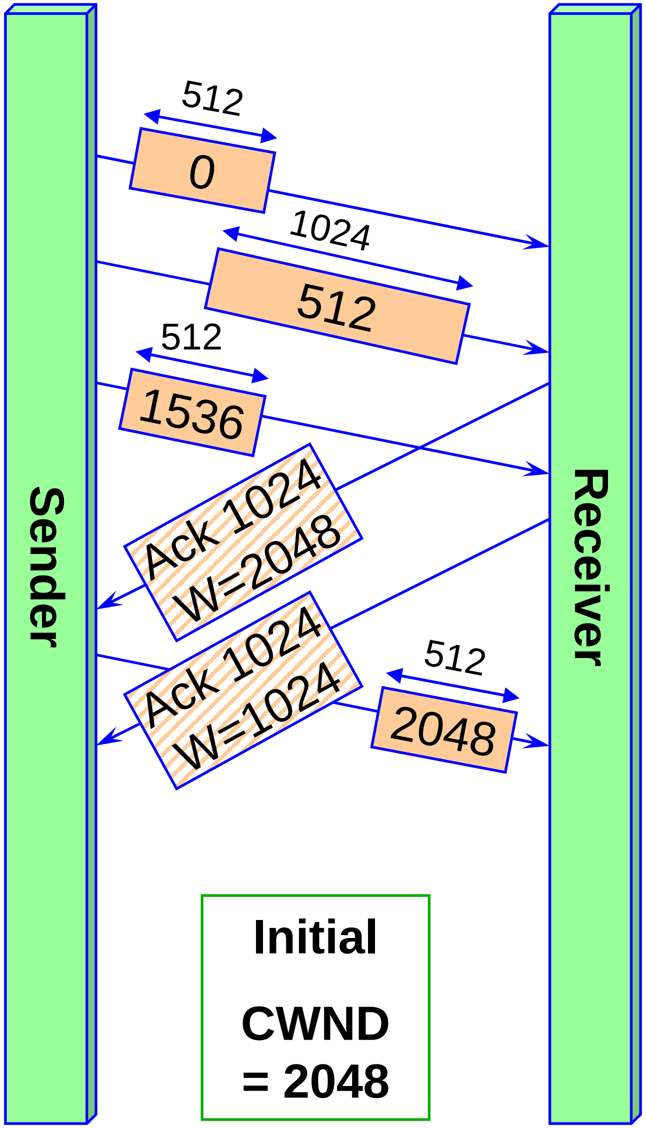
Initial window after 1st ACK



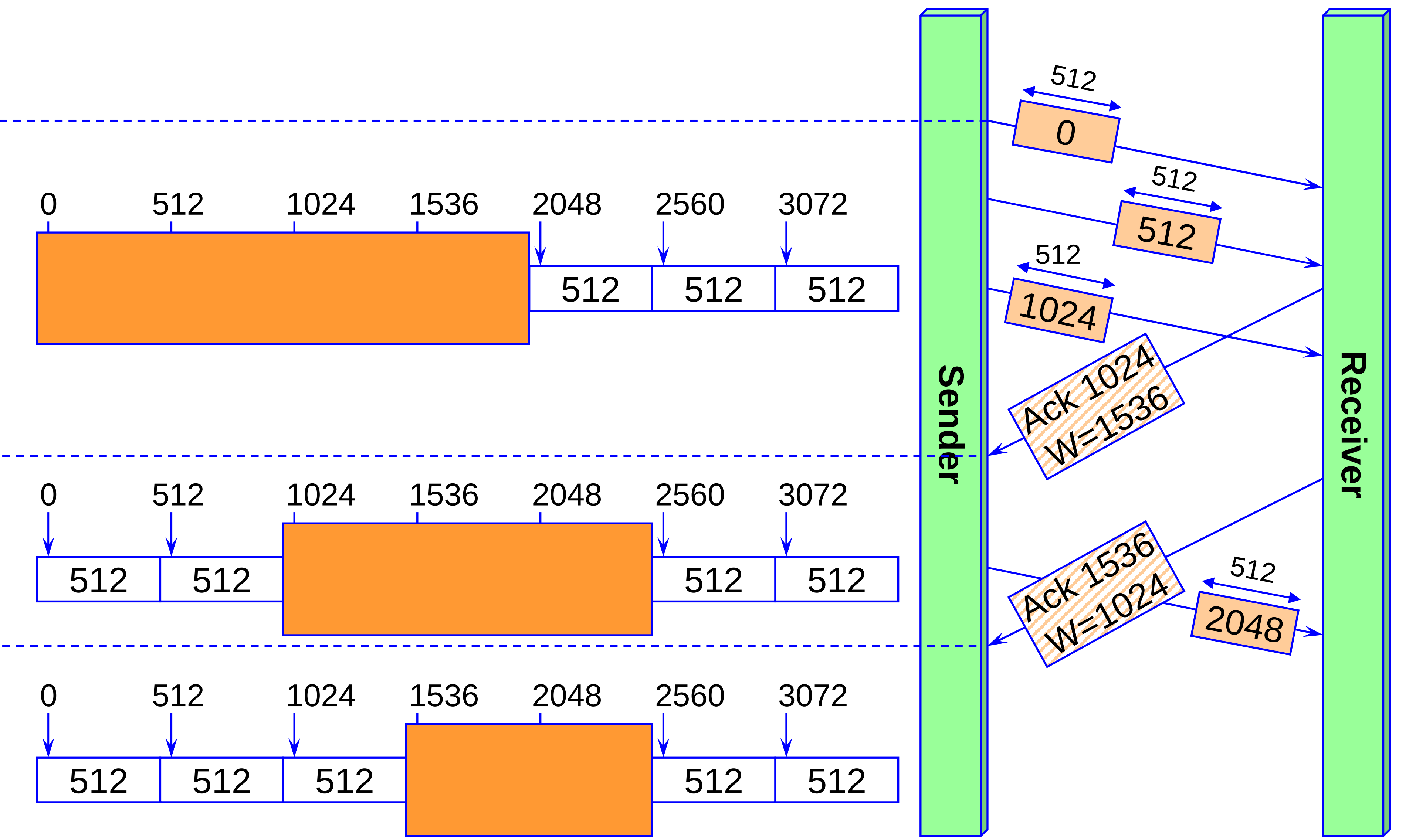


TCP mechanism for Flow control

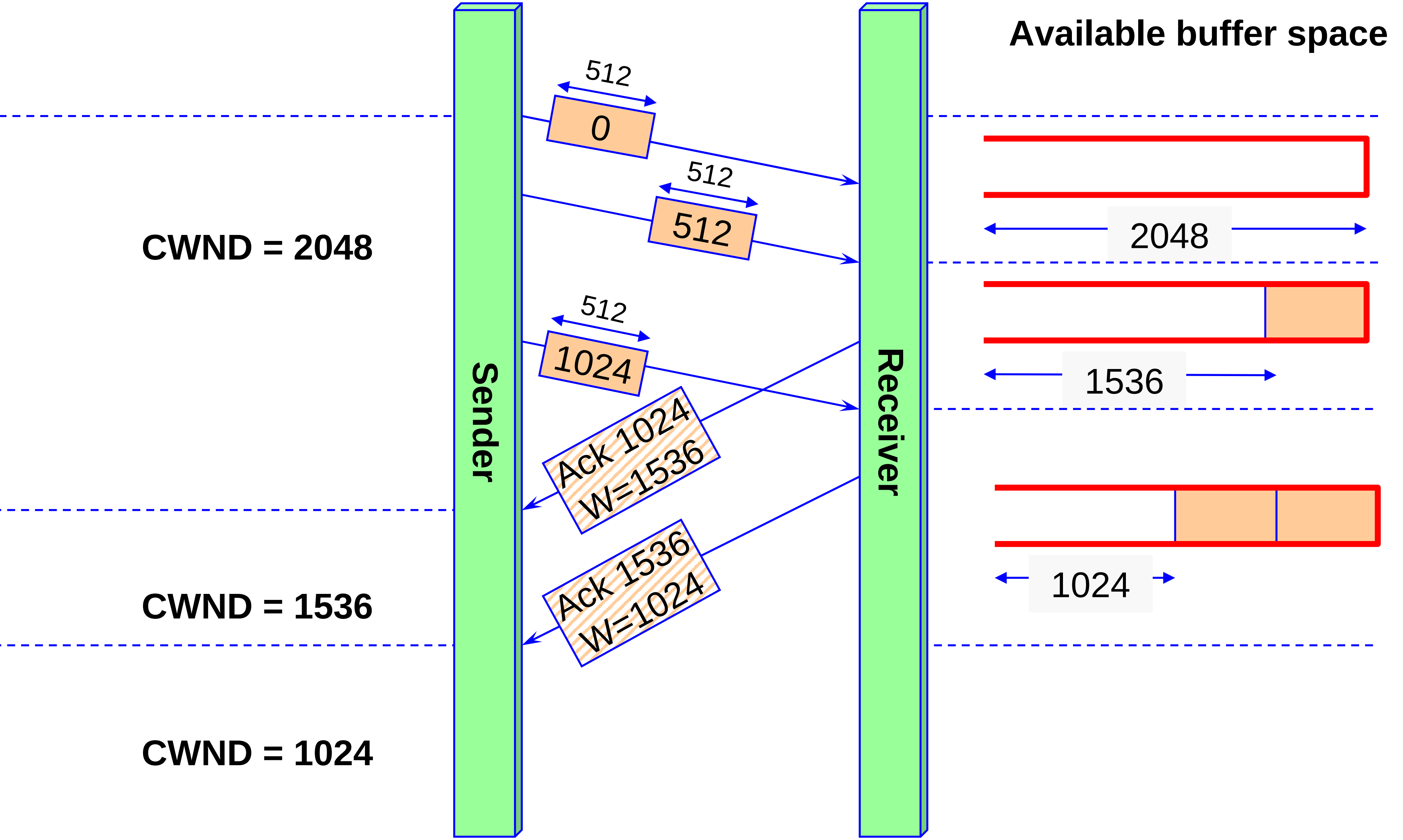
- **Dynamic Window size**
  - Window size CWND not fixed
  - Sender may reduce if detecting network congestion
  - Receiver indicates value of window in ACK
  - Value of Window = remaining incoming **buffer space** in receiver



TCP mechanism for Flow control

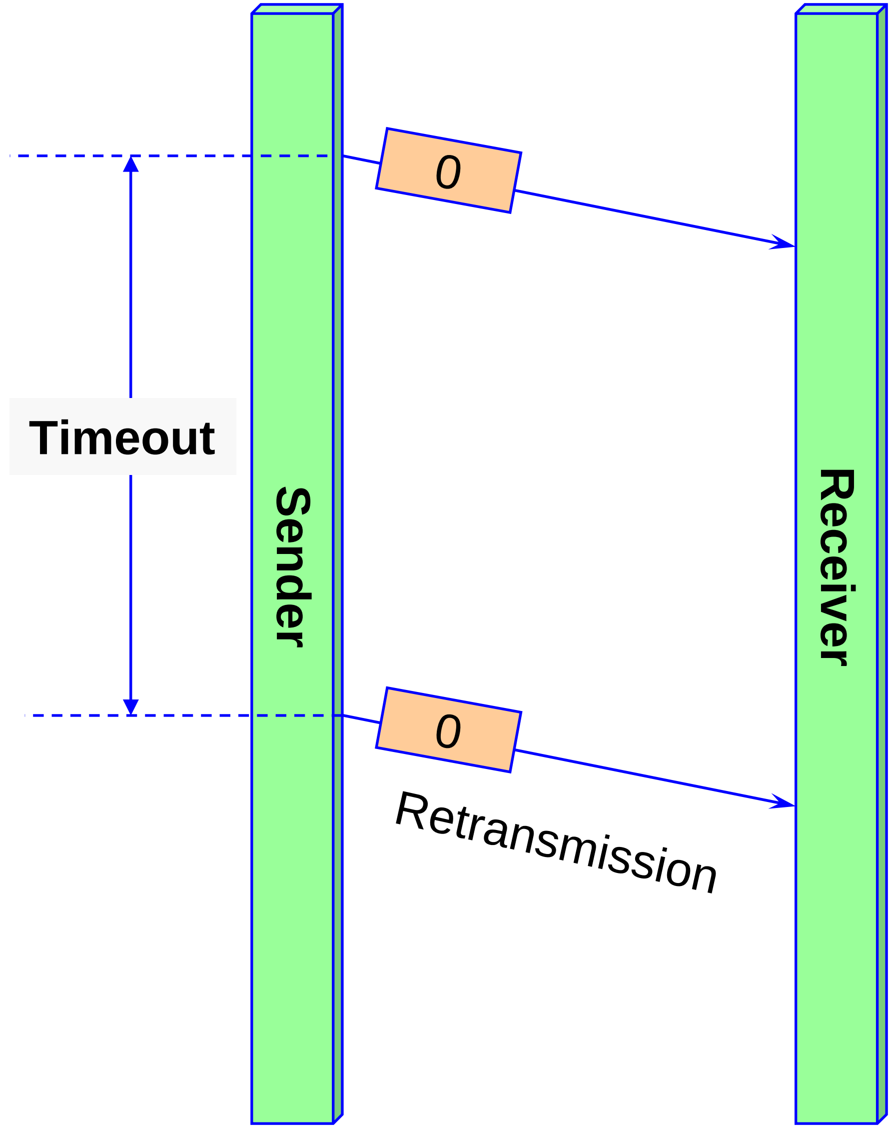


TCP buffer size



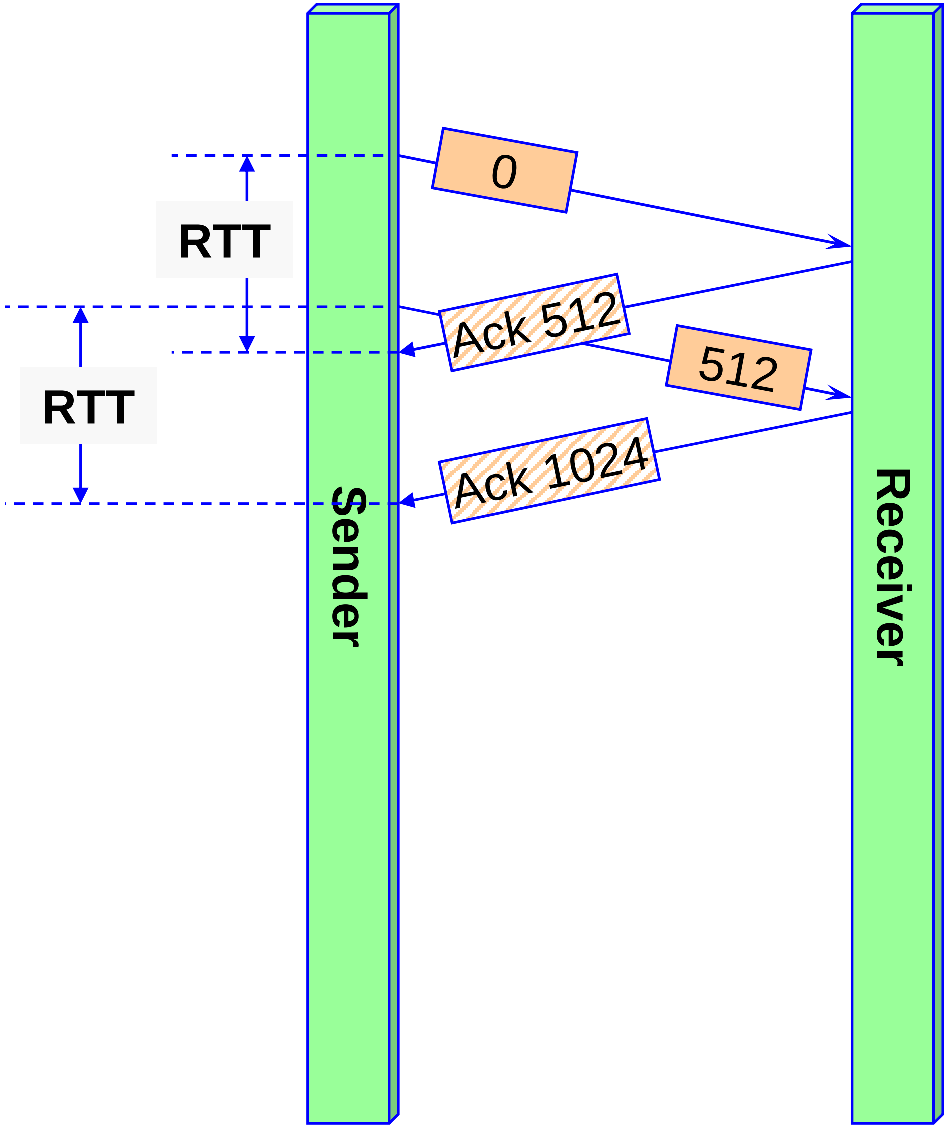
TCP mechanism error correction

- **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

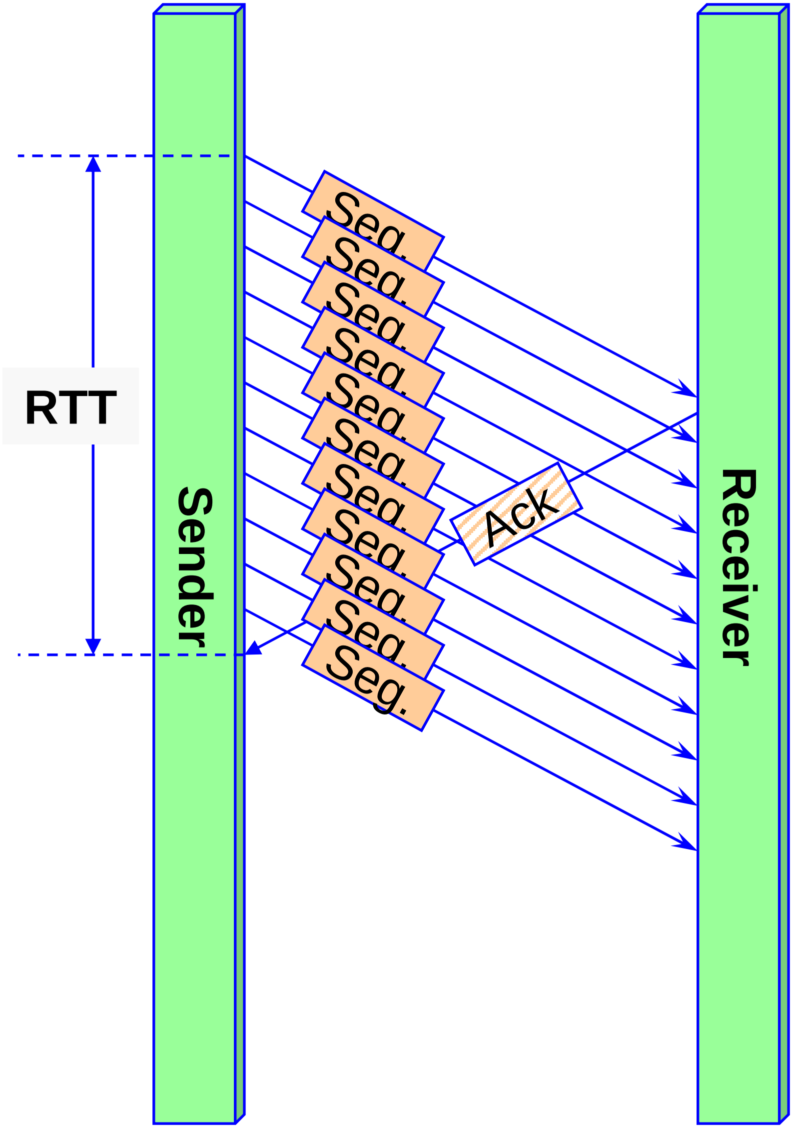
- **Question:** ***which value for the Timeout*****?**
- **TCP** measures permanently the **Round Trip Time** **(RTT)**
  - RTT = Average Time between segment emission and ACK reception
- **Timeout = β x RTT**
  - Choice of β delicate (simple choice: β = 2)



Optimal buffer size

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

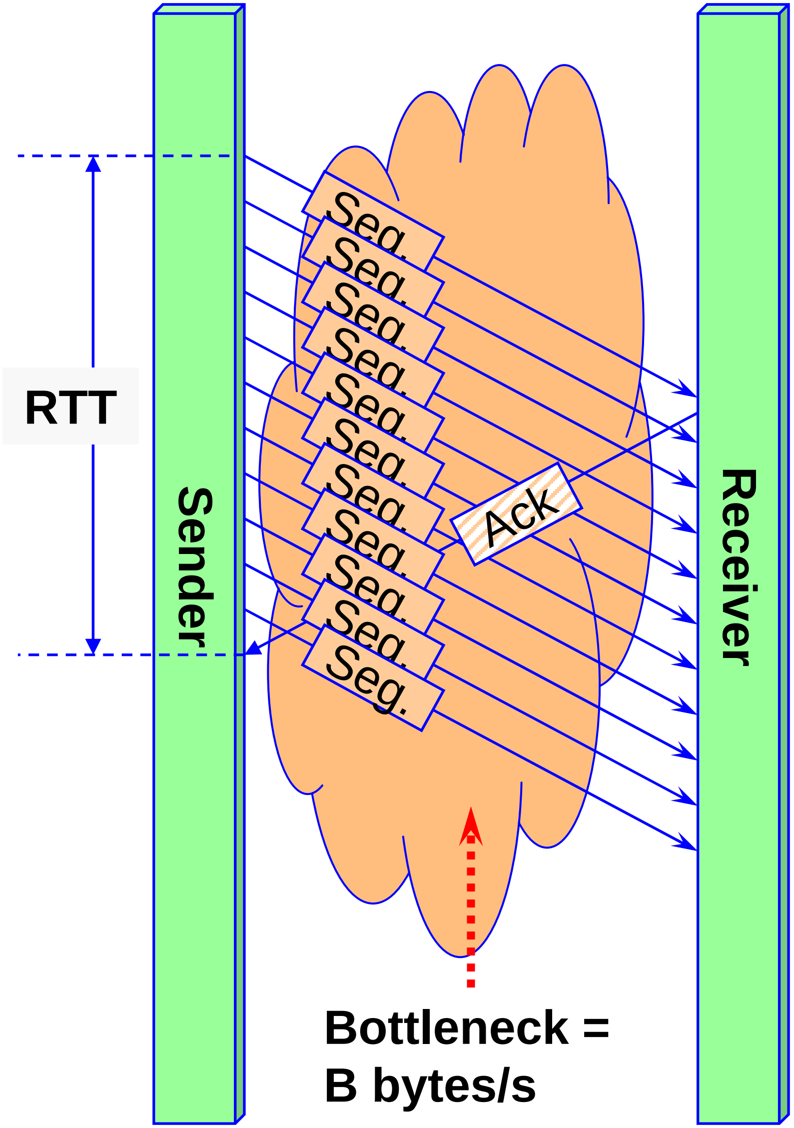
  What is the optimal buffer size to keep all segments?



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

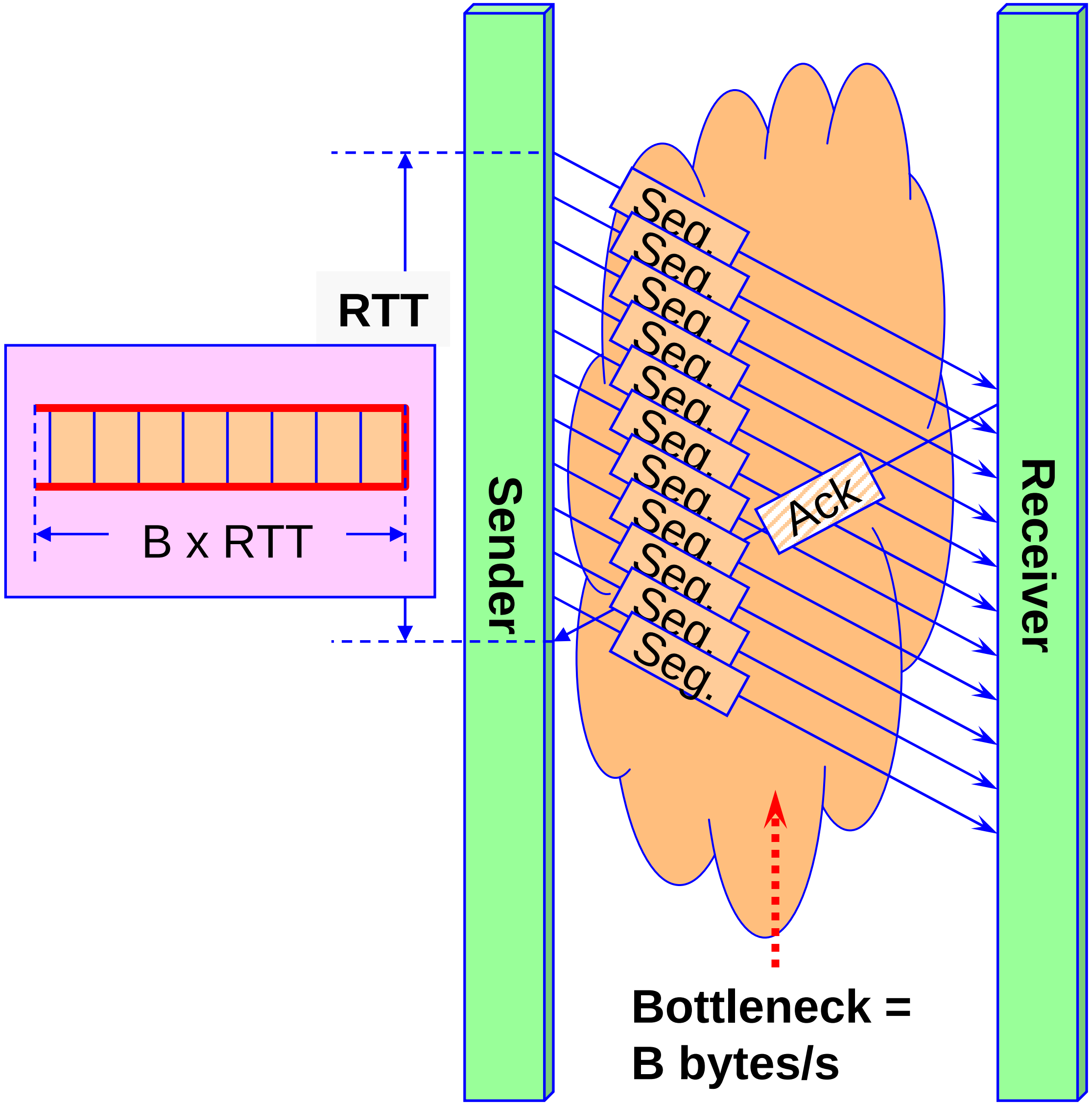
- **Optimal buffer size** (1)

  =

  Bandwidth*Delay Product

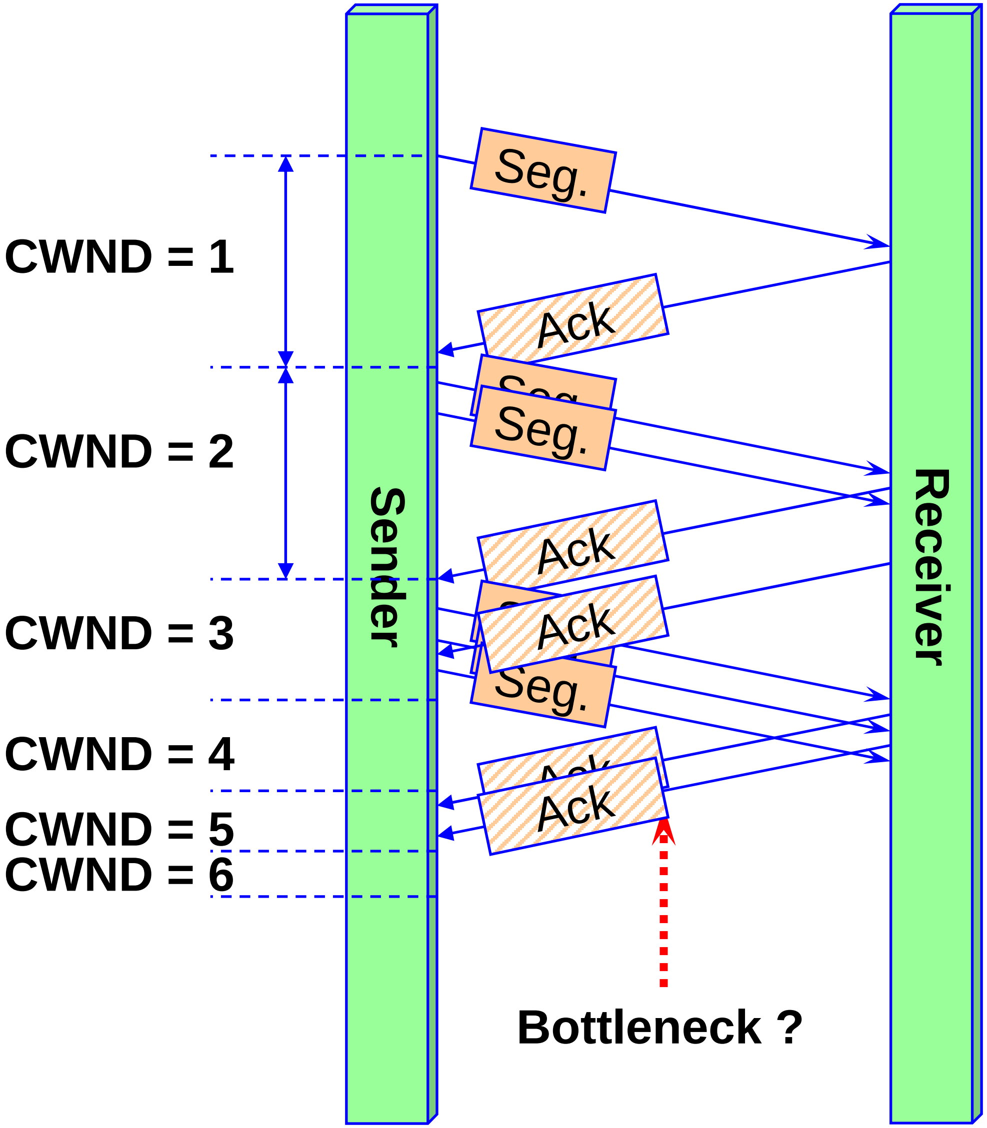
- **BDP = B x RTT**

- (1): sender and receiver



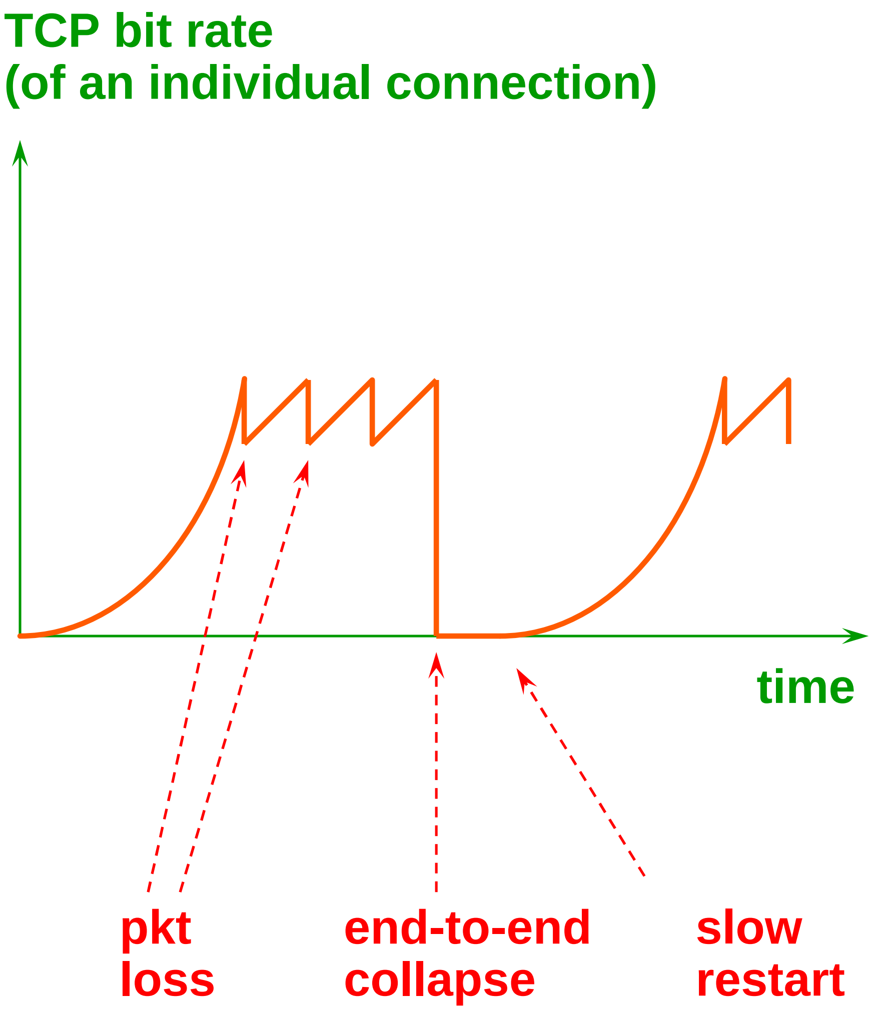
Avoiding congestions

- When TCP starts sending, ignores the network bottleneck (often at LAN – WAN interface)
- Can't send a 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



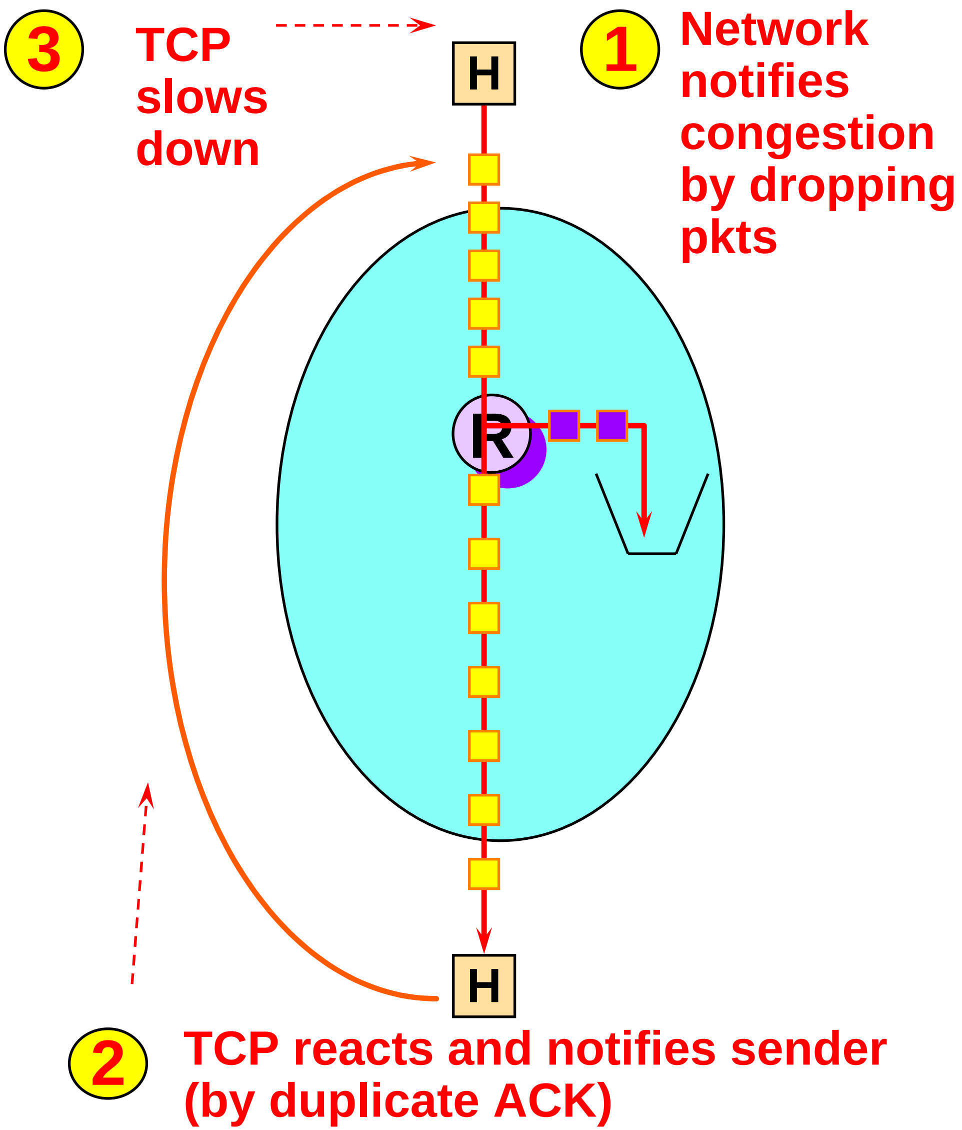
TCP congestion avoidance

- **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

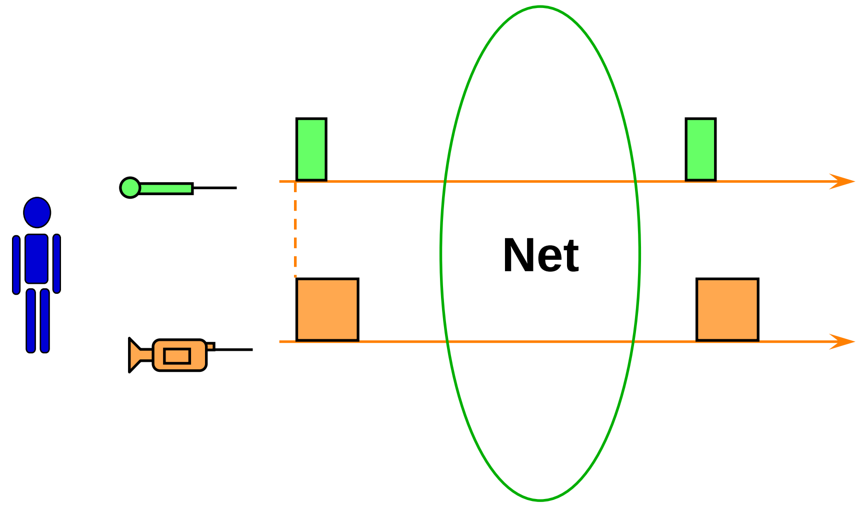
- Current Internet, no **explicit** congestion notification from network to source
- Notification technique: ***Network drops packets***
- TCP will react and slow down
- TCP is said to be ***congestion-indication responsive***



- Setting the Scene
- Internet QoS Options
- TCP and Congestion Control
- Multimedia over the Internet

Real-time media transmission (A / V)

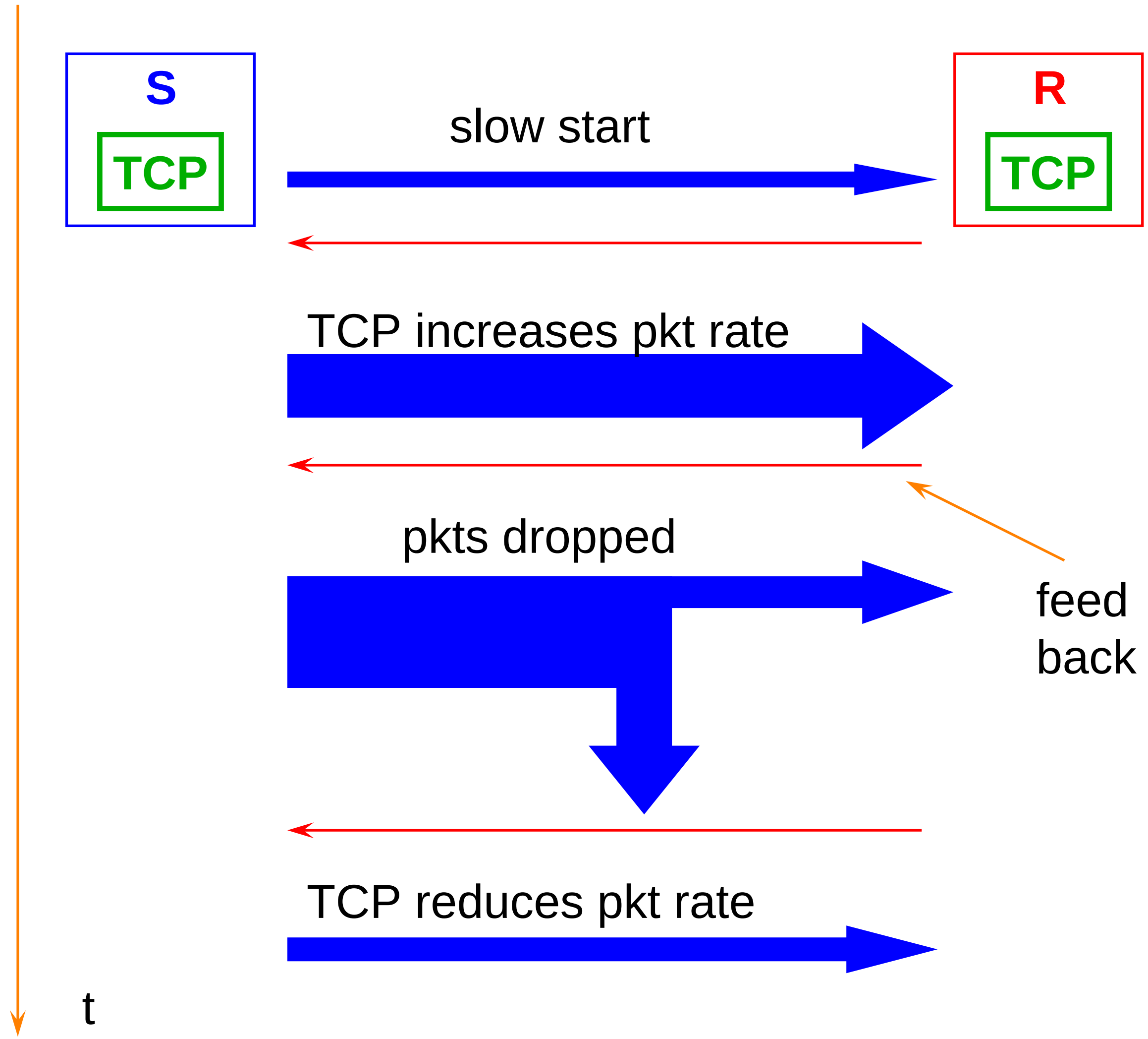
- Assume now the source of packets in an
  - analog digital converter
  - connected to a microphone and a movie camera
- Called **Streaming** audio/video



Audio and video packets

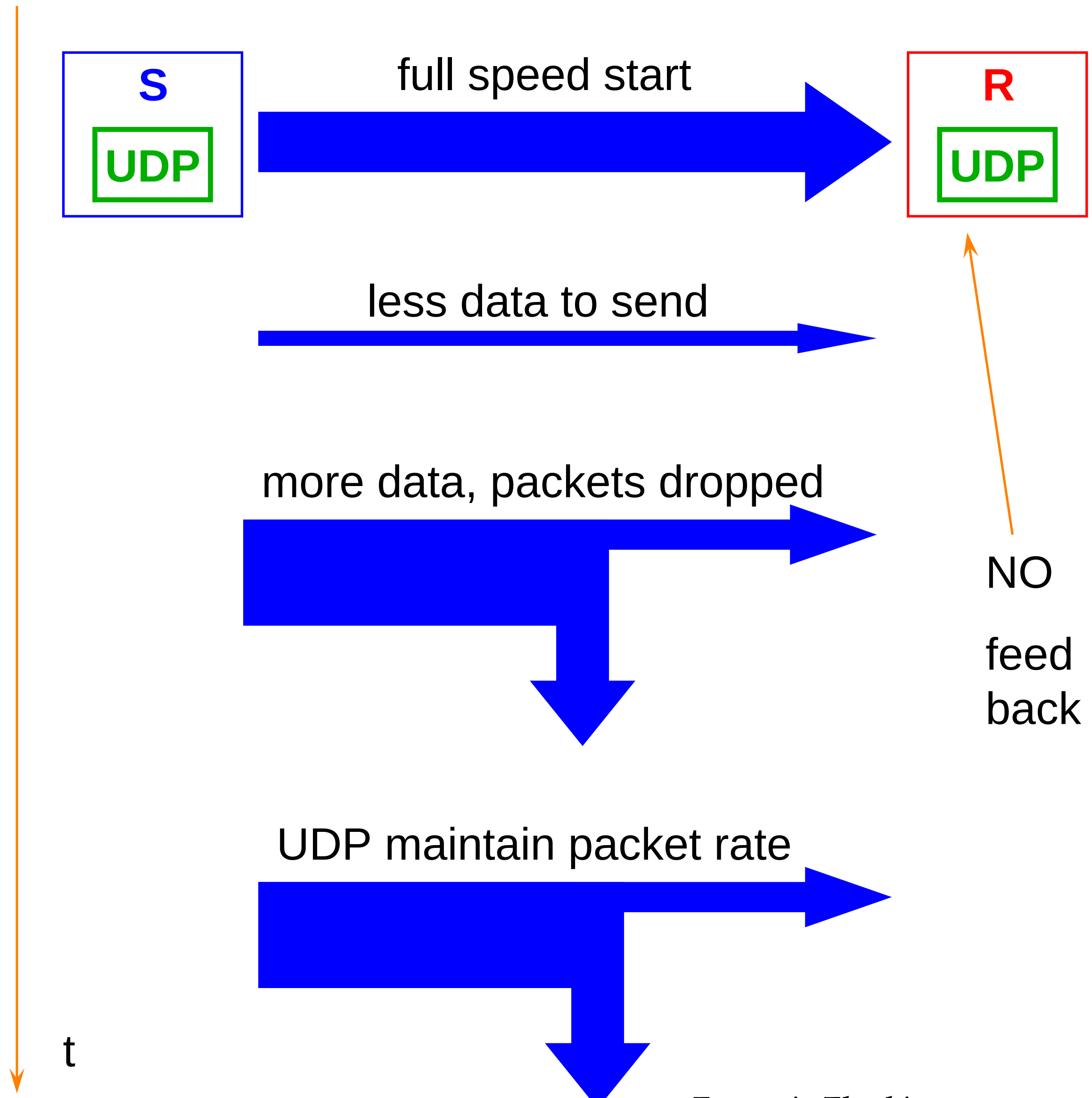
TCP behaviour

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



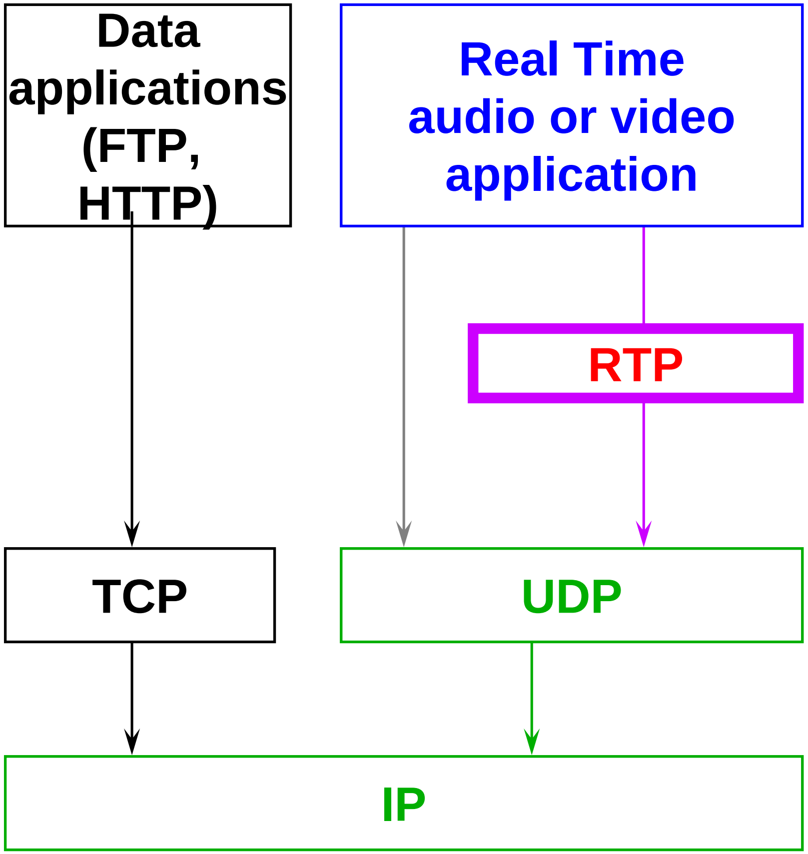
UDP behavior

- **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 feedback, ...
- **Most applications use RTP (Real-Time Transport Protocol)**
  - Packet loss detection (but not correction)
  - Some form of feedback

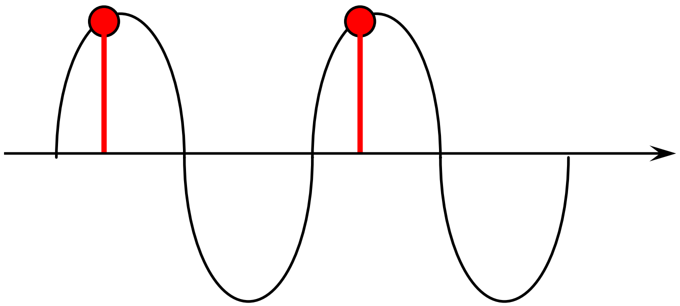


Audio/video network requirements

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

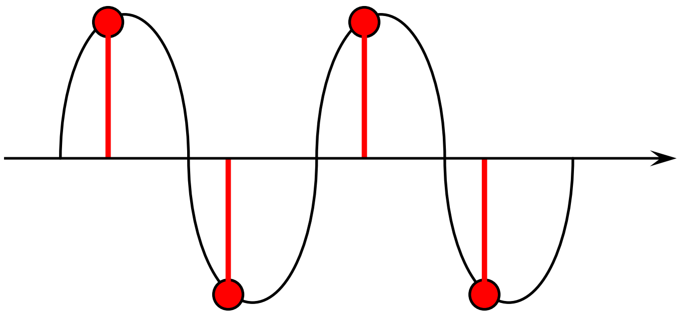
Nyquist theorem

- **To faithfully represent an analog signal**
  - if maximum frequency ***f***
  - sampling rate at least **2*f***



Sampling at *f* :
impossible to reconstruct

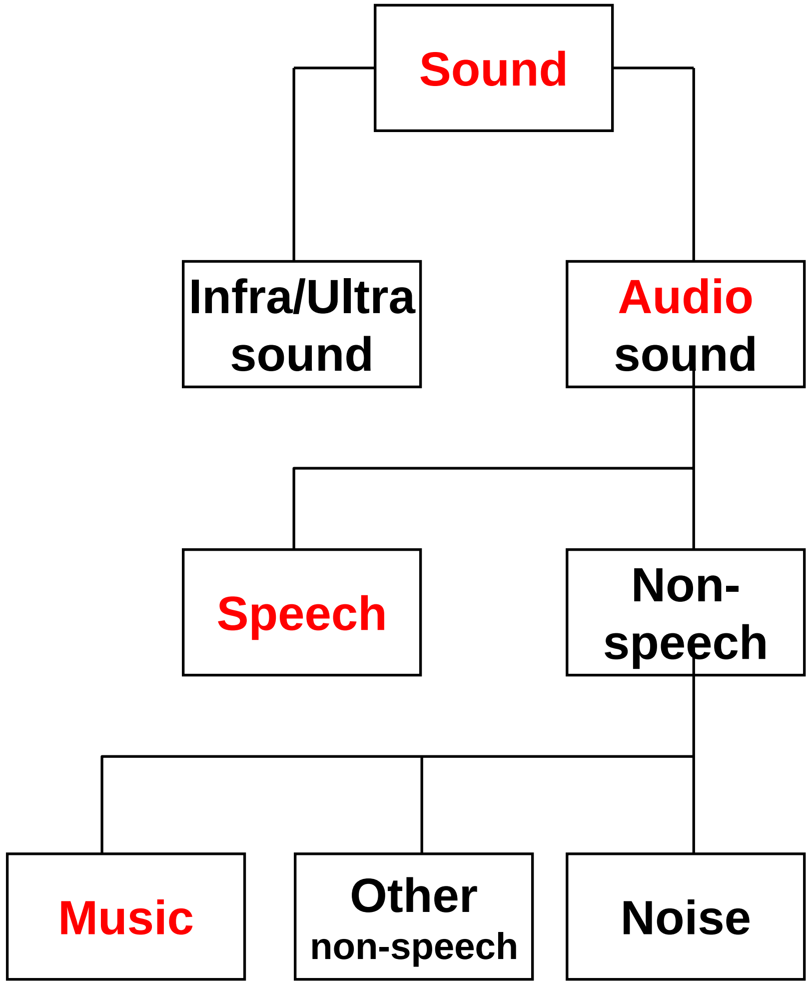
- **Application to audio**
  - if sampling rate is **8 kHz**
  - bandwidth is **3.4 kHz**



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 sound**
  - we hear better than we talk
  - speech spectrum: **50-10 kHz**
- ***Music***: a particular case of non-speech sound



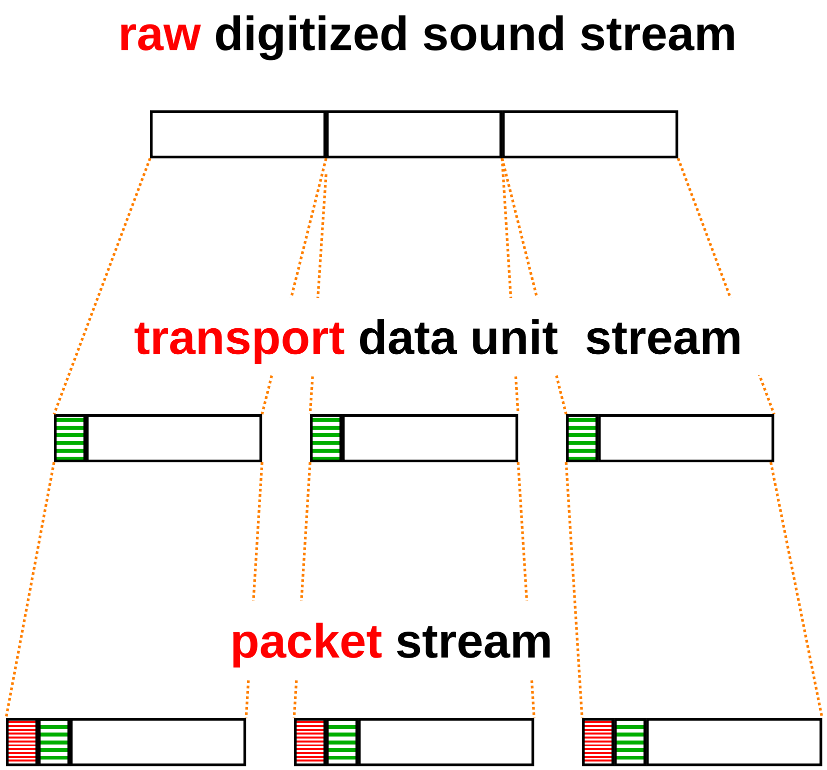
Audio bit rate requirements

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

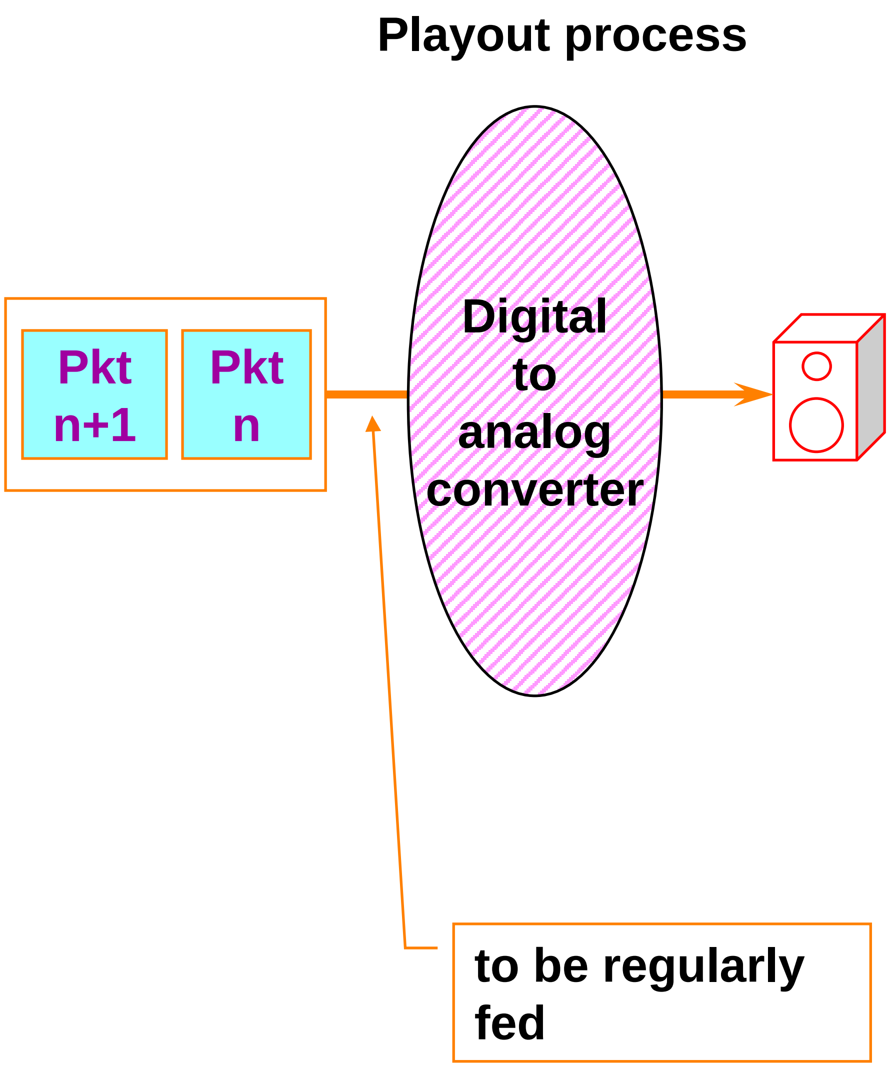
- **Monologue**
  - typically **15%** silence
- **Bi-party telephone conversation**
  - **20%** silence for overall conversation
  - **60%** silence for each party
  - If silence suppressed, required bit rate is in effect **<40%** of nominal raw bit rate

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**

Transit delay variation (Jitter)

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

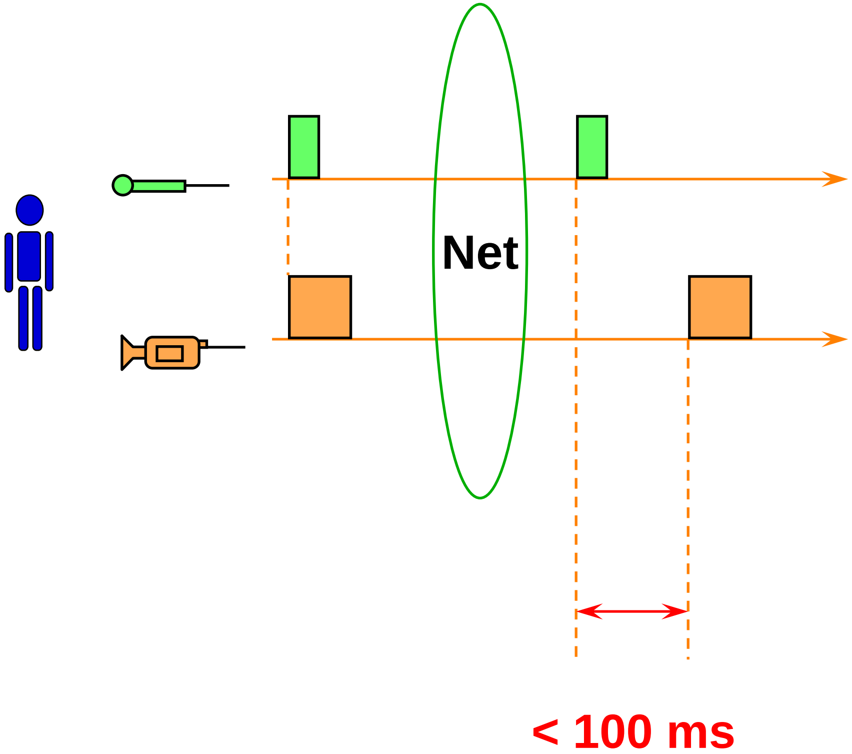


Inter-media synchronization

- **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**

Audio-compression techniques

Encoding techniques

- **e.g. DPCM:**
  code only differences between successive values, not values themselves

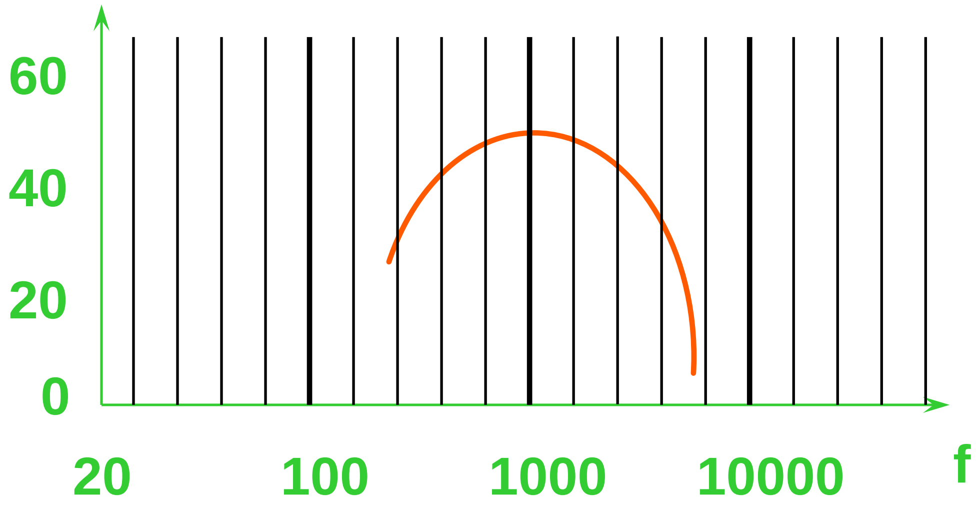
Source compression techniques

Based on psycho-acoustic model

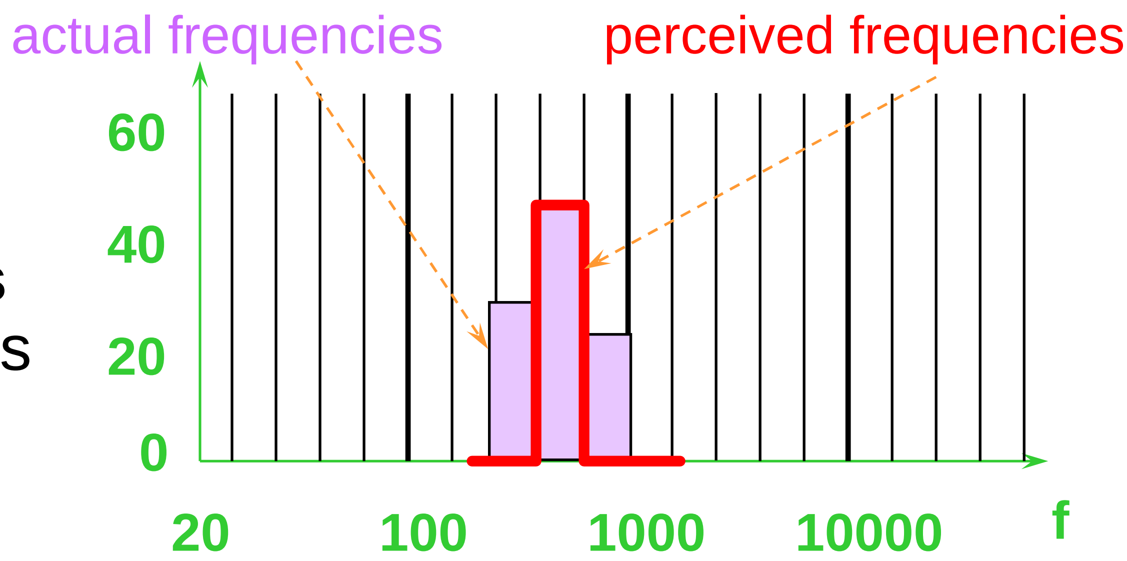
- **Transform encoding** (for all sounds)
  - e.g. Fast Fourier Transform (FFT)
- **Source modeling/synthesis coding** (for speech)

Psycho-acoustic and Masking

- **Response of ear to frequency:**
  - **ear most sensitive between 2 - 5 kHz**



- **Masking:**
  - ear does not register energy in **some frequencies band,** when there is more energy in a nearby band



Principle of Compression

- Remove non-perceived components from original signal

- Remove redundancies from the original signal

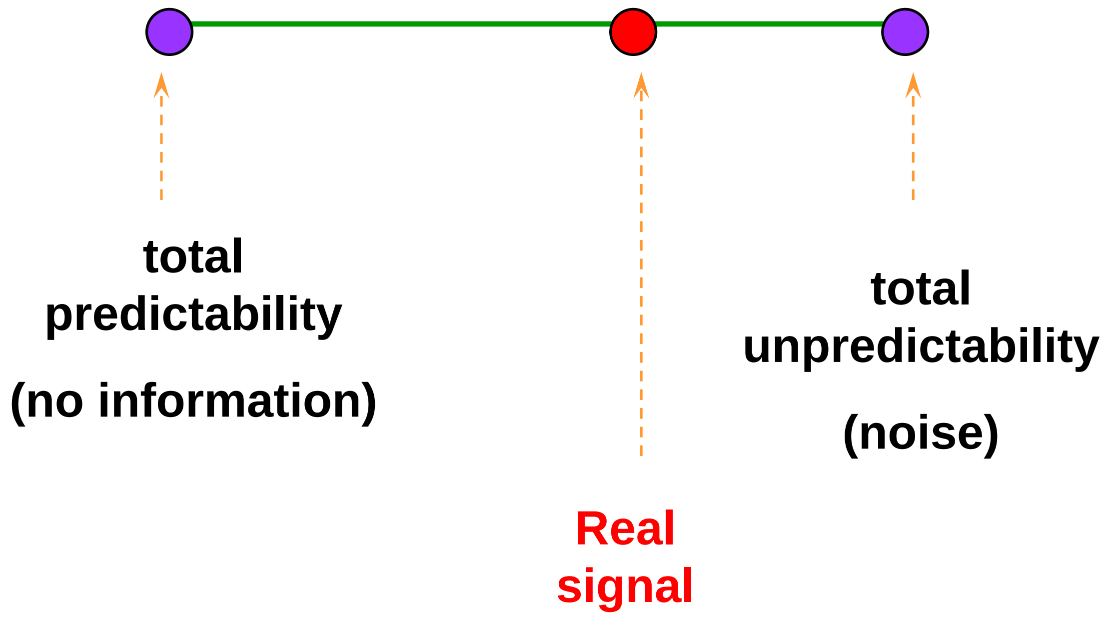
Information rate, bit rate, entropy

- **Information content or *entropy* of a sample:**
  - a function of how different it is from the predicted value
- **Shannon's theory:**
  - any signal which is **totally predictable**: carries **no information**
    - (e.g. a sine wave)
  - **noise** is completely **unpredictable:** high entropy

Real, contentless signals, noise

Real signals

lie somewhere between the two extremes



The Effect of Compression

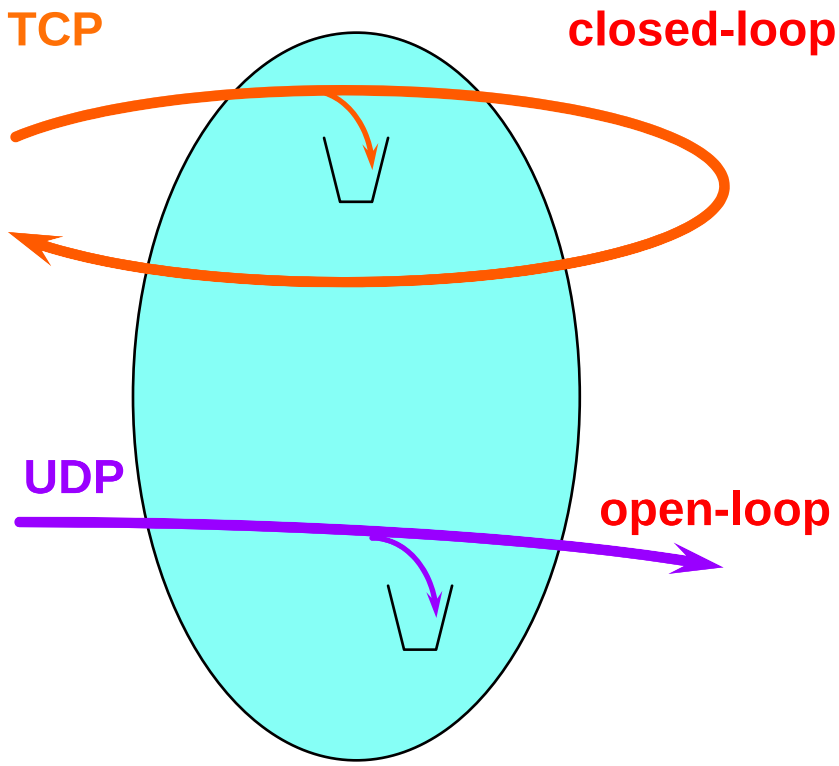
Compression removes redundancy ... *but*

The Effect of Compression

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

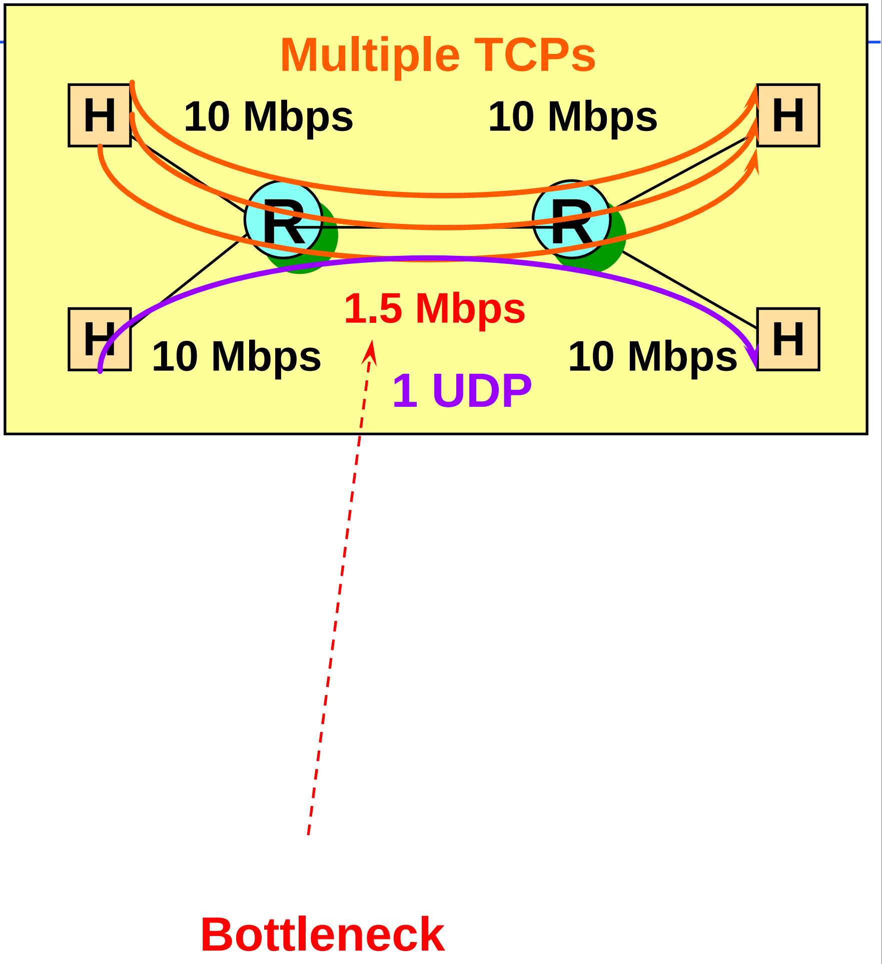
Unresponsive flows

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



Unfair competition

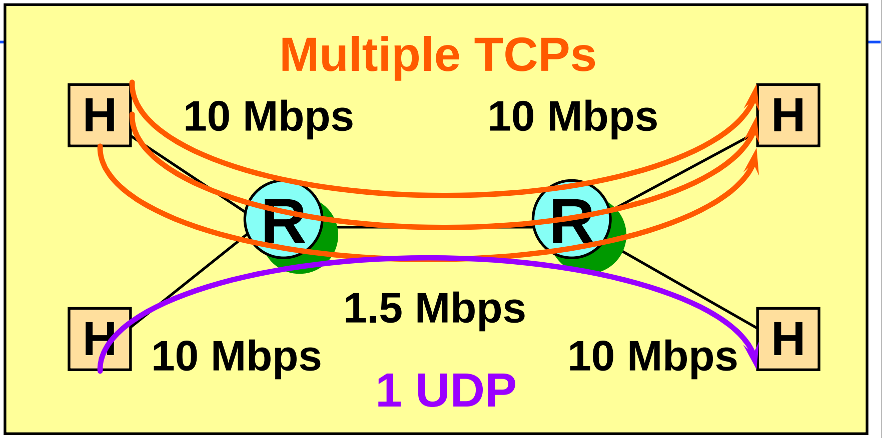
- **Case 1:**
  - **2 LANs (10 Mbps)**
  - **interconnected with T1 and a pair or routers**

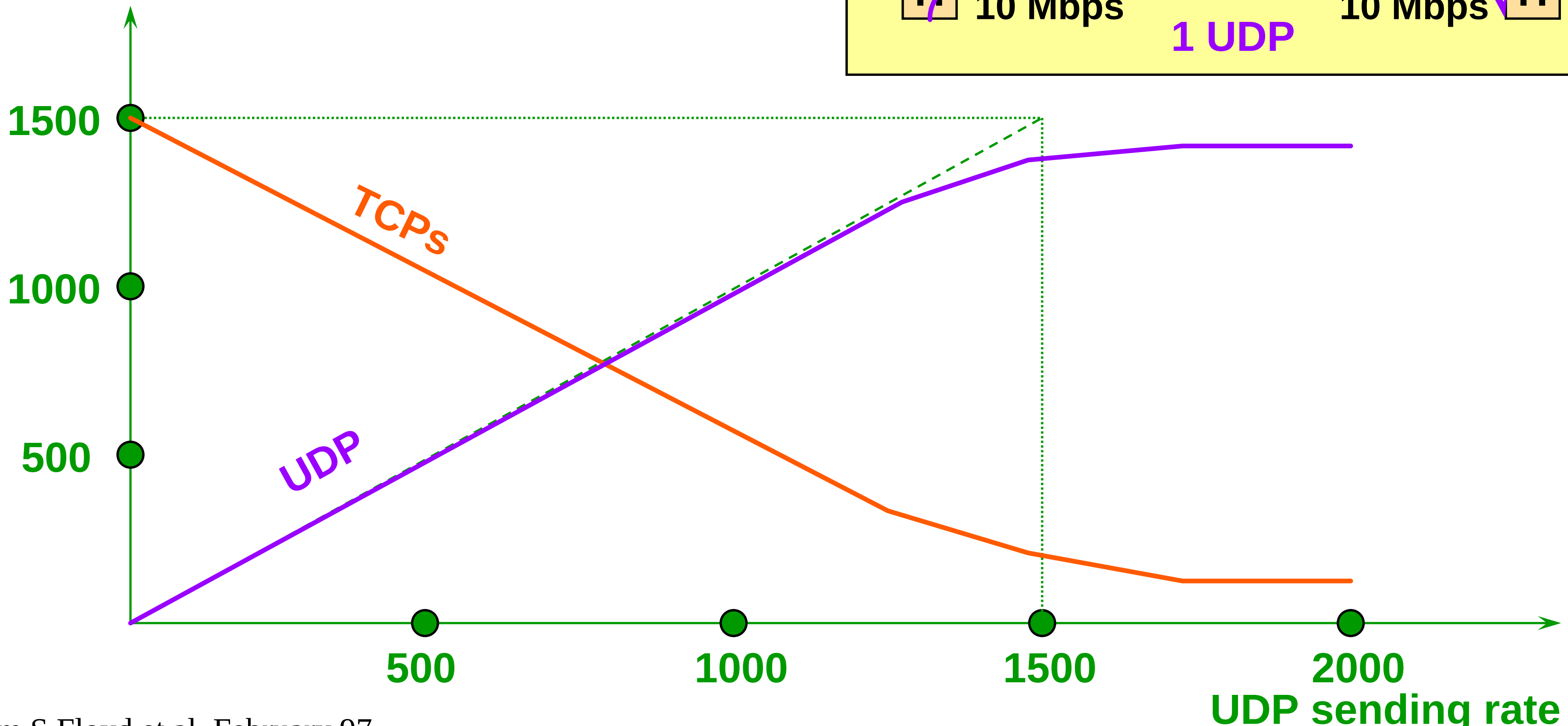


- **Competition between**
  - **3 TCP connections and**
  - **1 UDP connection**

From S.Floyd et al, February 97

Unfair competition





From S.Floyd et al, February 97