An Overview of Internet Multimedia Networking

Prof. Reza Rejaie

Computer & Information Science

University of Oregon

http://www.cs.uoregon.edu/~reza

Winter 2003

Motivation

- There is a wide range of applications for streaming (audio/video) over the Internet such as
  - Tele-conferencing, Distance learning, Internet- telephony, Media on-demand
- Our main focus is on transport mechanisms for streaming applications
  - How to stream mm objects through the network/Internet in a large scale

Multimedia streams

- Digitized mm streams have high bandwidth, example:
  - 30 frames/sec * 512*512 pix/frame * 16 bits/pix
  - => BW ~ 125 Mbps
- Other examples:
  - Telephone quality audio: 64 Kbps
  - CD quality audio: 1.5 Mbps
  - Bcast quality NTSC: 120 Mbps
  - Studio quality NTSC: 200 Mbps
  - HDTV: 1-2 Gbps
- mm streams are often compressed/encoded

Audio Taxonomy [from K. Meyer Patel @ unc]

Compressed

Speech-based

LPC-10E

CELP

GSM

General

ADPCM

Dolby AC-3

MPEG-1 L3

MPEG-2 AAC

DTS

SDDS

THX

MiniDisc (ATRAC)

Uncompressed

μ-law

A-law

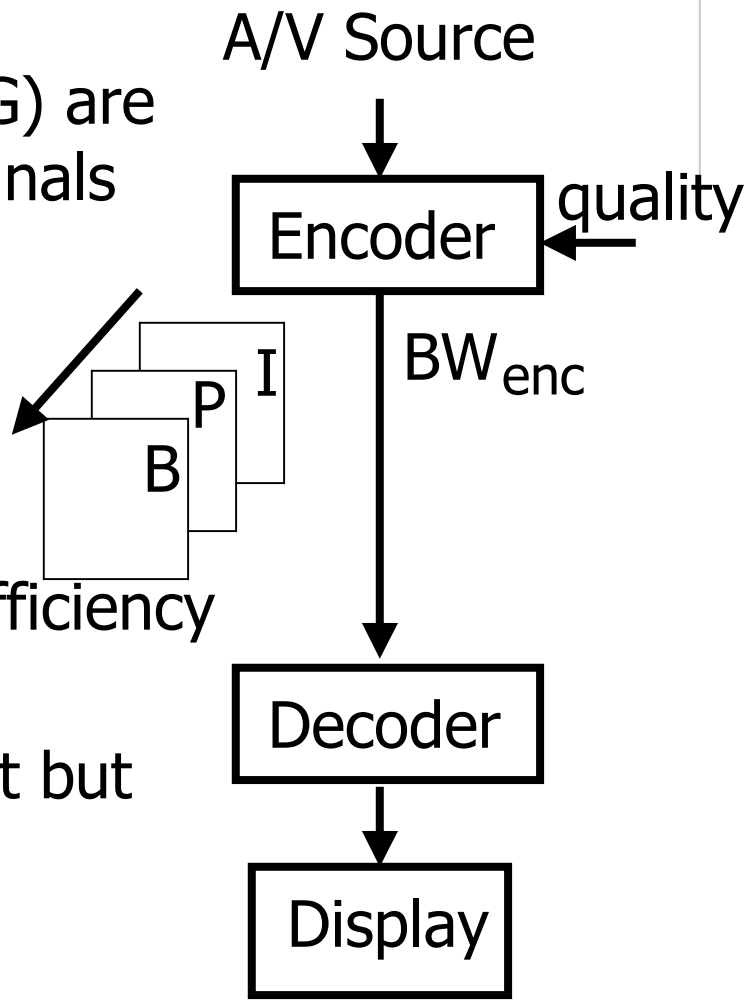
Linear PCM

Video Taxonomy [from k. Meyer Patel @ unc]

Compressed		Uncompressed	
Tape	Streaming	Digital	Analog
Digital Betacam	MPEG-1	D-1 (CCIR 601)	Betacam
DVCPro (D-7)	MPEG-2	D-2	VHS
DV	MPEG-4	D-5	S-Video
DVCAM	M-JPEG		Video-8
Digital-8	H.261		Hi-8
D-9	H.263+		Betamax
DVCPro50	Real		
	Sorenson		
	Cinepak		
	Indeo		
	Video for Windows		

Encoding/Compression

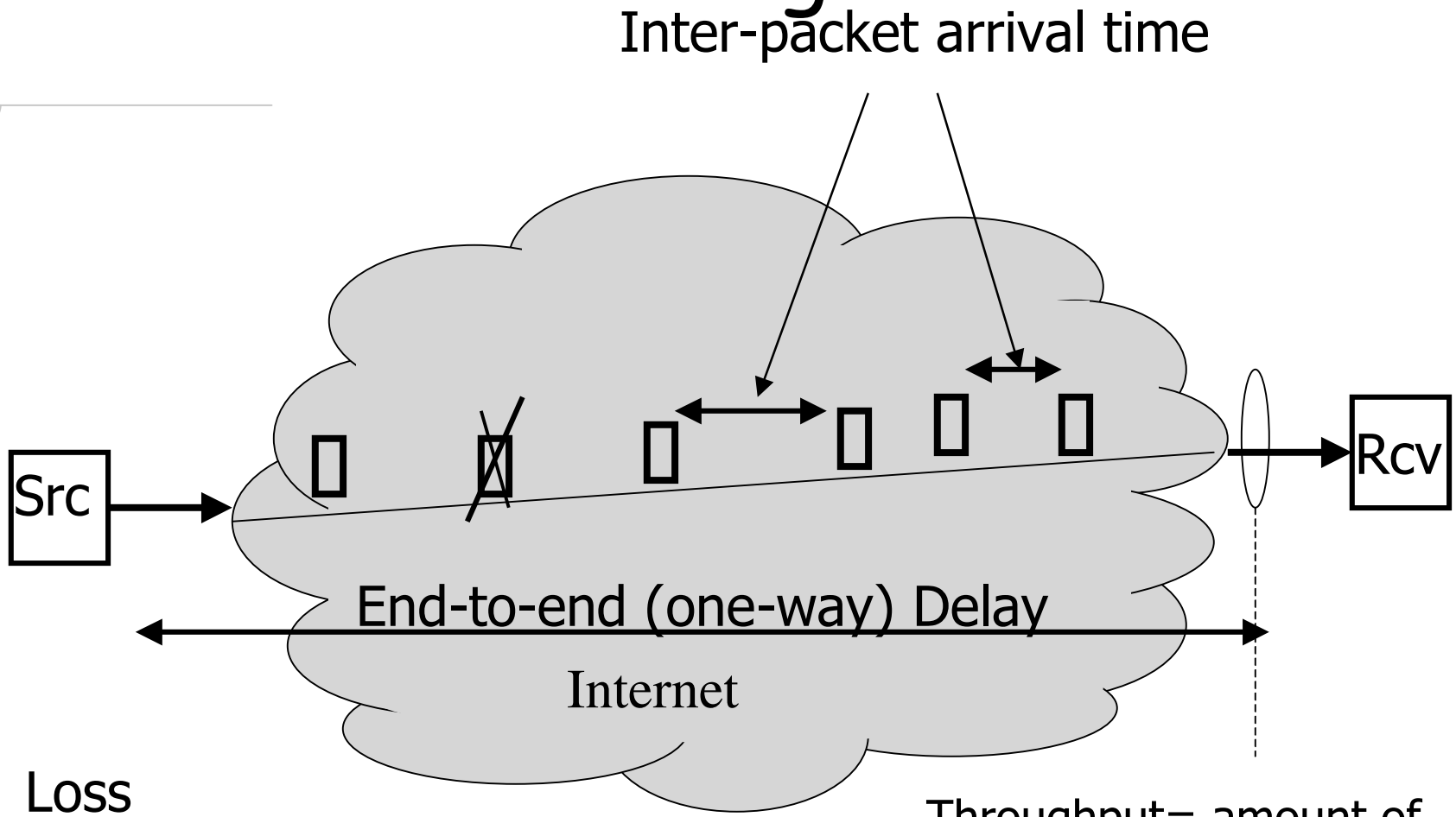
- Encoding mechanisms (e.g. MPEG) are used to compress audio/video signals
- Encoding audio & video
- Average BWenc depends on:
  - Encoding algorithm
  - Desired quality
- Tradeoff between compression efficiency and loss resiliency
- Average BWenc is rather constant but BWenc could be bursty



Streaming Applications

- Characterizations:
  - Continuous media (audio, video)
    - Periodic transmission
  - Timing dependency
  - Requires quality instead of reliability

Network Streaming: Basics



Loss
Delay
Bandwidth

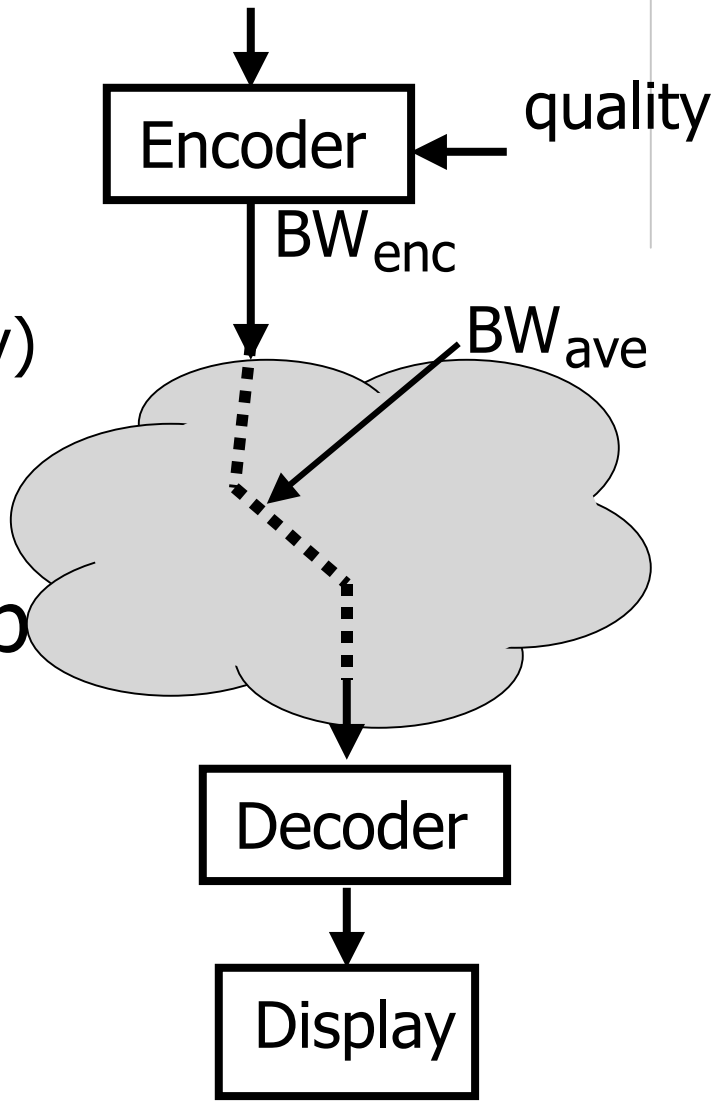
Throughput= amount of arriving data per unit of time

Networks: Packet- vs Circuit-switching Paradigm

- Circuit switch networks
  - + Performance guarantees
  - - Lower network utilization
  - Appropriate for homogeneous flow, e.g. telephone network
- Packet switch networks
  - + Statistical multiplexing => Higher network utilization
  - - Unpredictable performance (bw, loss, delay)
  - Appropriate for heterogeneous flows, e.g. Internet

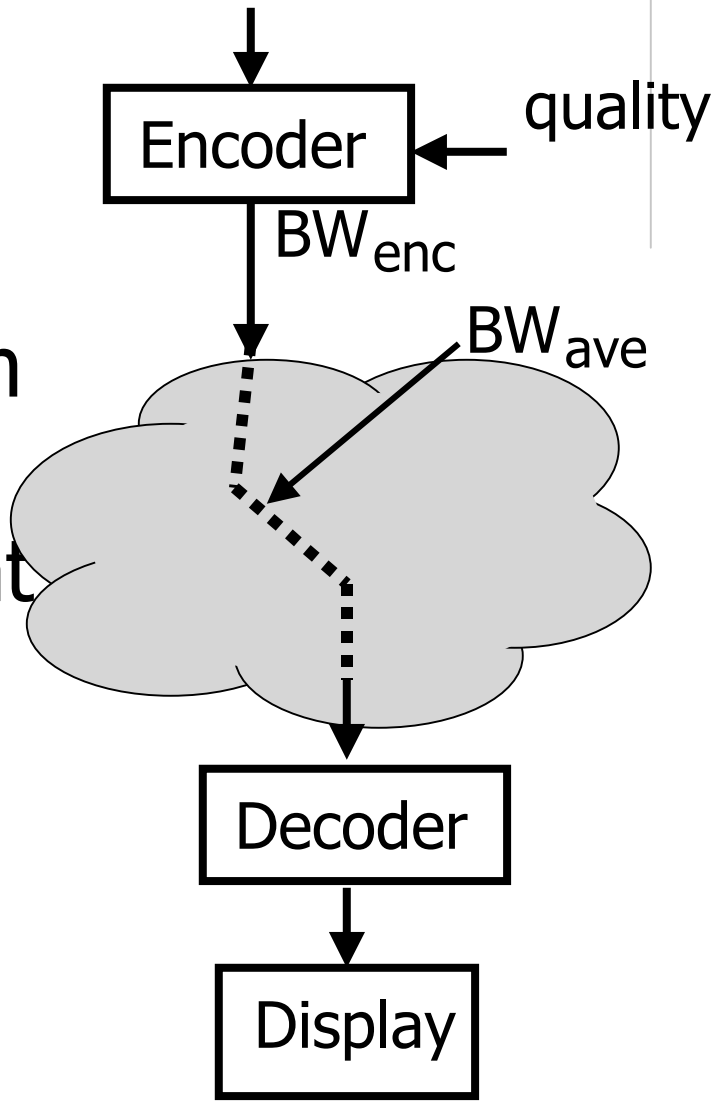
Multimedia Networking: Alternative solutions

1) Add new services to the network (QoS)
   - Integrated Services (IntServ)
   - Differentiated Services (DiffServ)
2) Enable applications to cop with best-effort service
   - **Adaptive streaming applications**

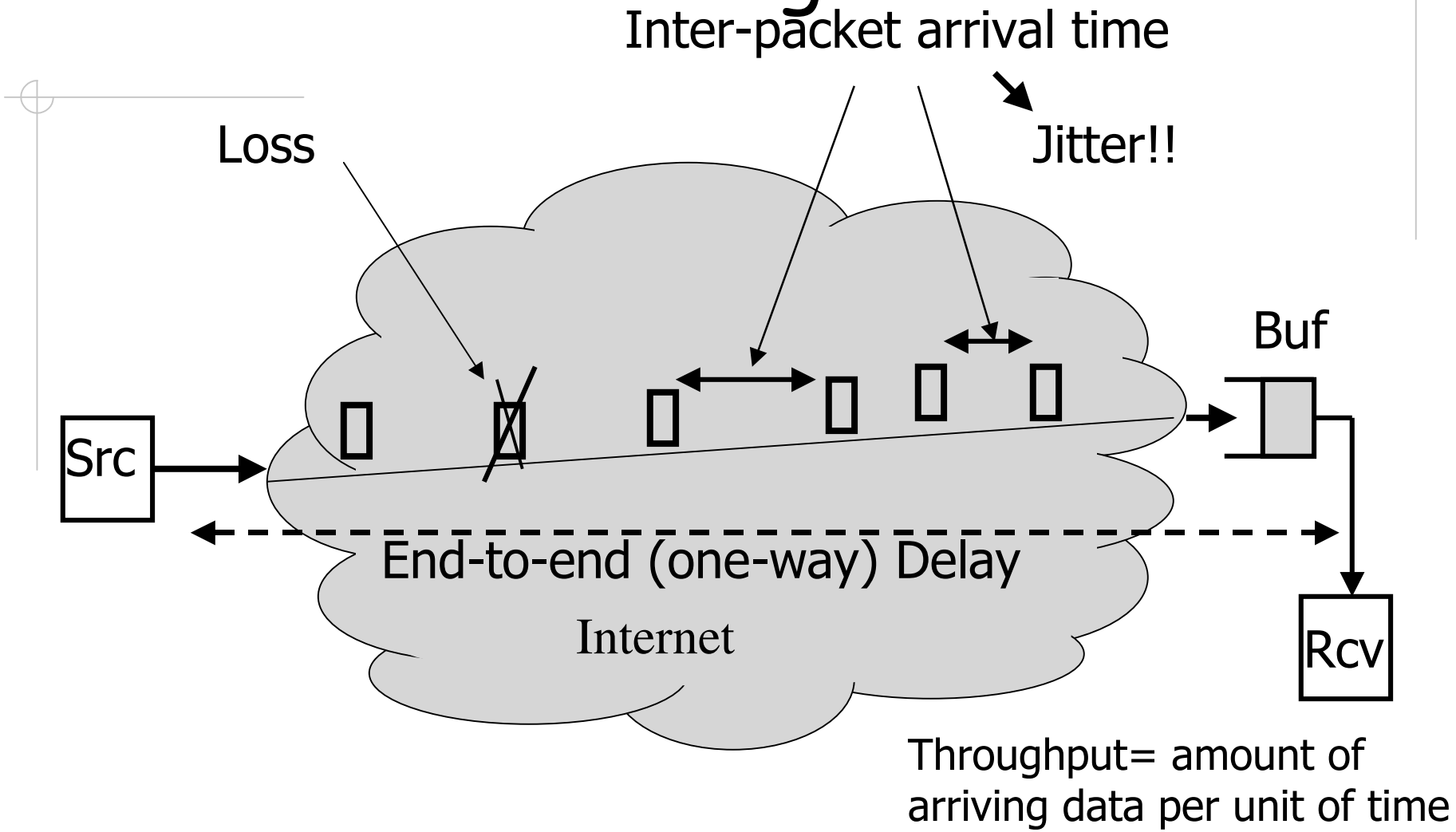


Streaming over IntServ networks

- IntServ (i.e. RSVP)
  - **Performance guarantee**
- Smooth multimedia stream to deliver through a CBR channel in order to prevent
  - Buffer overflow
  - Buffer underflow
- Smoothing



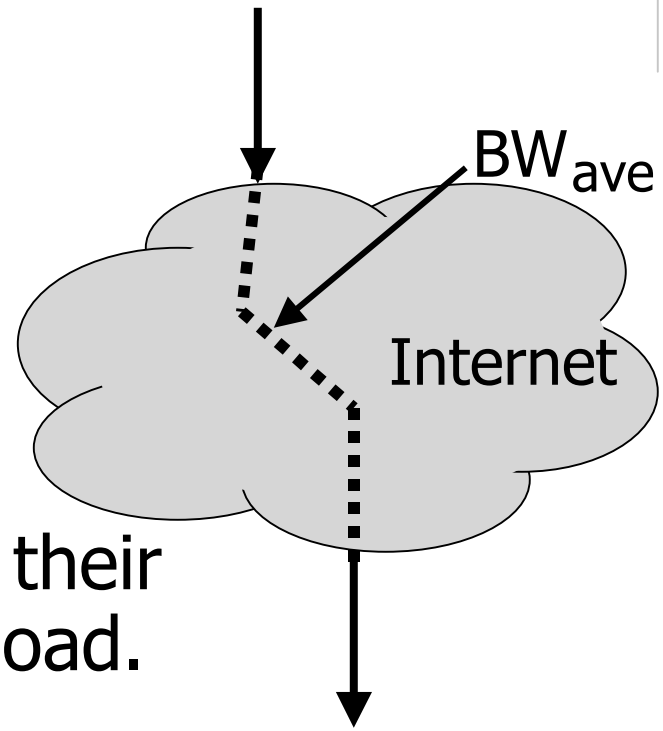
Internet Streaming: Basics



Throughput= amount of arriving data per unit of time

Streaming over best-effort networks (Internet)

- Best effort service
  - Available BW?
  - Loss rate and pattern?
  - Jitter?
  - Out of order delivery
- Resources are shared
  - Data sources should adjust their transmission with network load.
    - Congestion Control



Different Types of Streaming Applications

1) Live vs Pre- Recorded (on-demand)
   - Availability of future data
   - Ex: watching a live vs rebroadcast concert
2) Interactive vs Non-interactive (lecture-mode)
   - Level of interactivity is limited by e2e delay
   - Interactive applications can not tolerate more than 250 ms delay

- Interactivity and Liveliness are orthogonal issues
  - Any combination is possible, but some may not be as useful, e.g. pre-recorded & interactive

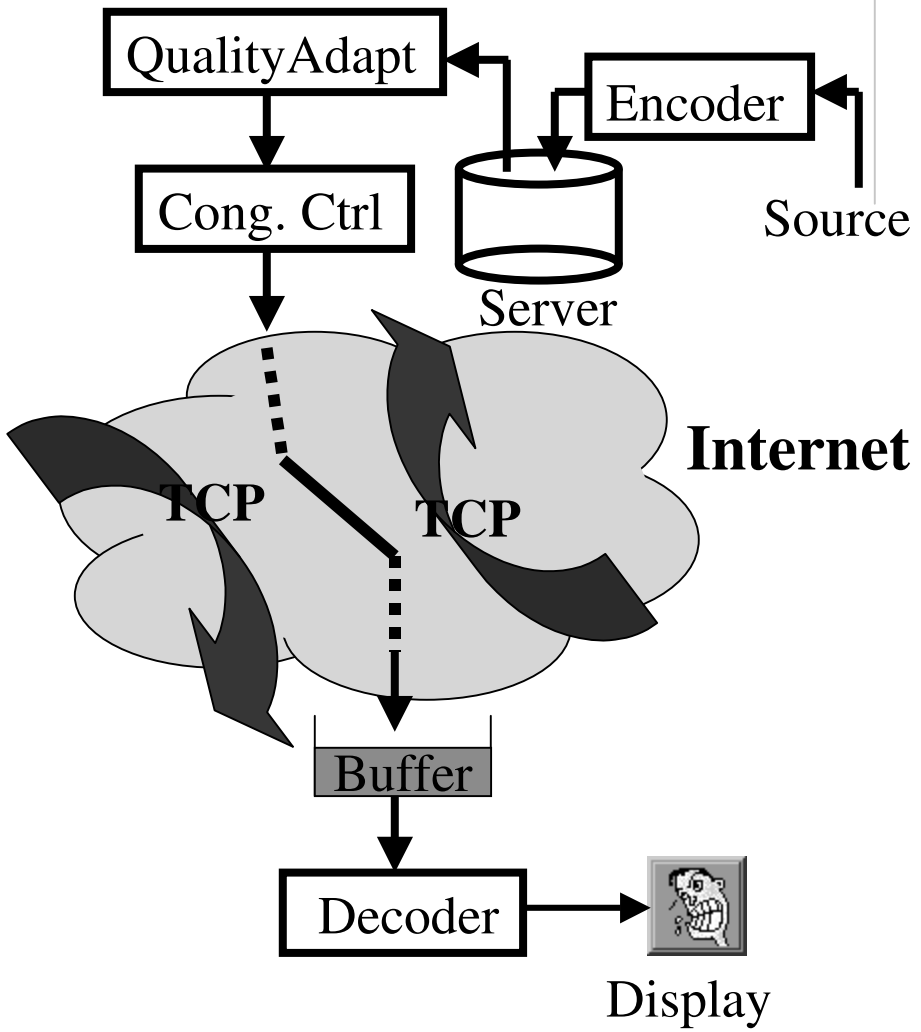
Delay sensitivity of Streaming App.

- The higher the level of interactions, the higher the sensitivity to delay, the lower the amount of buffered data

  - Internet telephony, Video conferencing
  - Video games
  - Live but non-interactive (i.e. lecture-mode)
  - On-demand playback of stored video

Adaptive Streaming Applications

- Basic issues
  1) Buffering
  2) Packetization
  3) Synchronization
  4) Signaling
- Transport Protocol
  - **Congestion control**
  - **Error control**
  - **Quality adaptation**
- Adaptive Encoding



1) Buffering

- Client Buffering is used to cope with network jitter
  - No upper bound for Jitter => late pkts
  - Avoid buffer overflow & underflow
- Buffering is also useful to absorb (short-term) variations in available BW
- Buffering adds to end-to-end delay
  - Inappropriate for interactive apps

2) Packetization

- Each packet includes:
  - Header: Sequence number, Time-stamp, …
  - Body: Payload
- Unique sequence number per packet
  - Used to detect packet loss, reordering
- Multiple packets may have same timestamp
  - A big frame is fragmented into several packets
- Application Level Framing (ALF) [Clark & Tennenhouse, 1990]
  - Design principle

Application Level Framing

- Data should be organized into units that is most meaningful for the app
  - Application Data Unit (ADU)
- Example
  - ADU for video or audio streams?
- Data is exchanged between app. and transport in terms of ADU
- App constructs an ADU that fits in network packets, i.e. Max. Transmission Unit (MTU)
  - A large video frame may be fragmented
  - Several small audio samples may fit in a packet

3) Synchronization

- Audio & video streams are separated!!
  - Need a way to couple audio/video
- Inter- vs Intra media synchronization
- Synchronizing audio and video streams, i.e. lip-sync
- Packet time-stamp may be used for inter- and intra-media synchronization

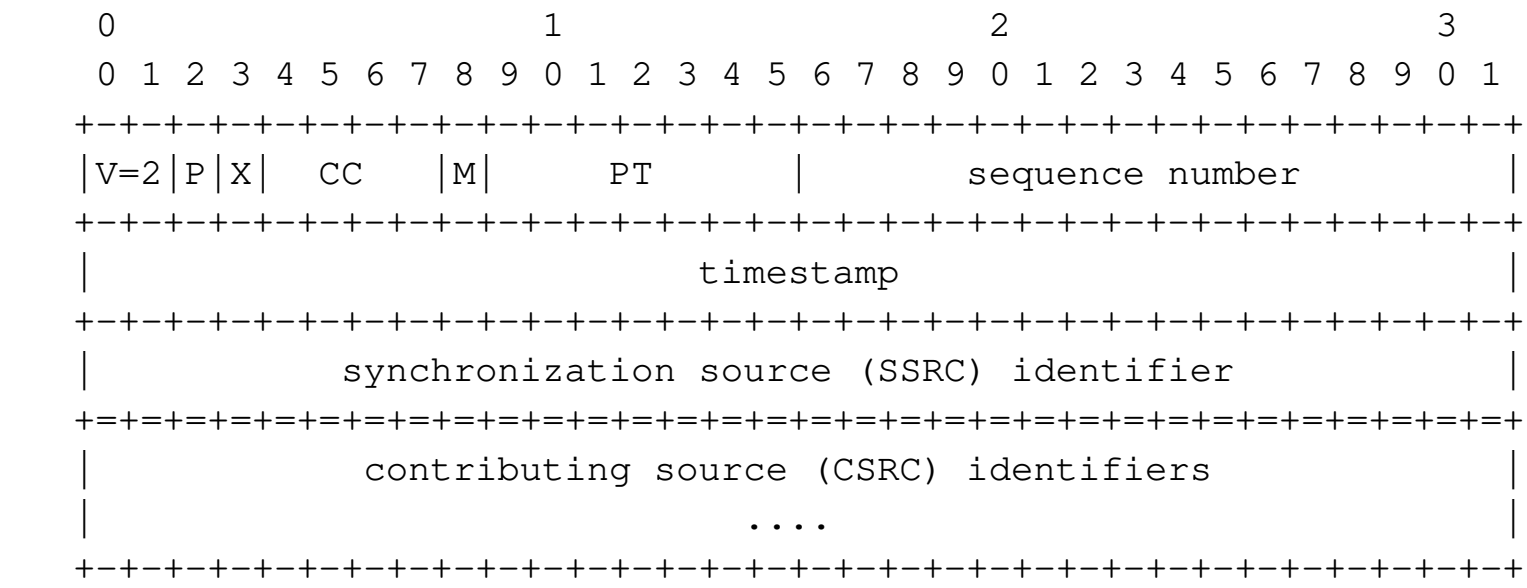
Real-time Transport Protocol(RTP)

- RTP provides end-to-end network transport functions for "transmitting real-time data"
  - Sequence numbering, time-stamping
- RTP does not provide any QoS guarantees
- RTP follows ALF principles
- RTP primarily designed for multicast, but it works with unicast
- Audio & video are sent through separate RTP sessions

RTP (Cont'd)

- RTP is an "incomplete" protocol framework
  - Only includes common functions across all the apps
  - RTP should be tailored through modifications and/or additions to the header => RTP profile
  - RTP profile includes application-specific info, e.g. payload type & mapping into payload formats
- RTCP is the Control Protocol
- RTP & RTCP are independent of underlying transport and network layer
  - Typically run over UDP/IP

Packet Format



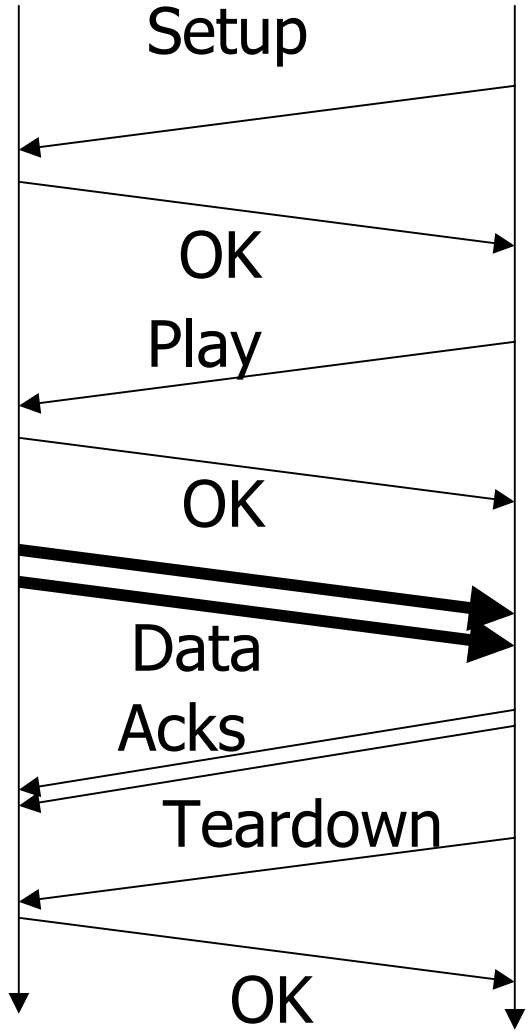
- V: Version/P: Padding/X: Hdr Extension/CC: no of CSRC/M: marker for frame boundary/PT: payload type/seqnum/ts/SSRC: unique src id/CSRC

RTCP

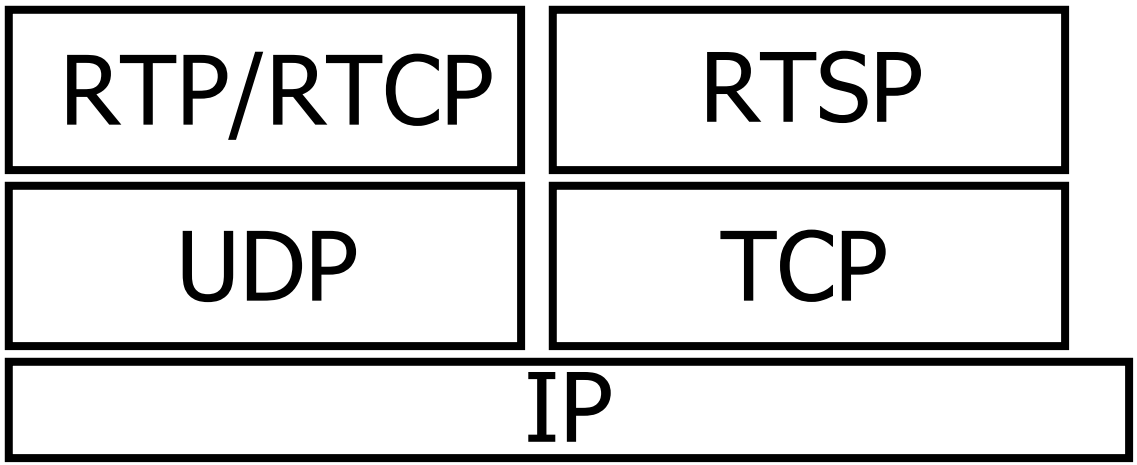
- Provides feedback on the quality of data delivery
- All participants send RTCP feedback to the entire group
  - Adjust feedback rate to scalable, i.e. bigger group => lower feedback rates
- RTCP packet might include:
  - SR: sender report
  - RR: receiver report, reception statistics
  - SDEC: member description
  - .. and more

4) Signaling, Control Protocol

- Provide remote control functions for a client
- RTSP provides these functionalities
- RTSP runs over TCP



Protocol Stack



Transport Issues

1) Congestion Control (CC)
   - How to determine available bw and adjust transmission rate?
2) Error Control (EC)
   - How to detect and correct lost packets?
3) Quality Adaptation (QA)
   - How to adaptively match quality (i.e. bw) of stream with available bw?

- Can we use TCP for streaming applications?

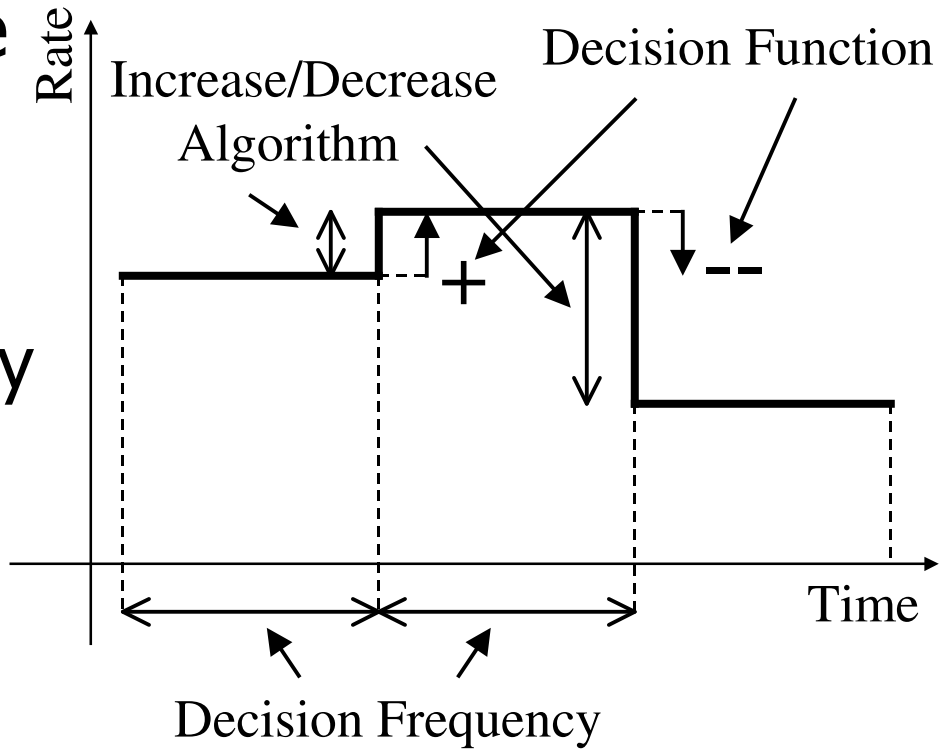
Congestion Control

- The Problem: how to determine available bw without any help from the network?
  - Why is it a hard problem?
- Basic idea:
  - Decrease tx rate when congestion occurs
  - Increase tx rate to probe for excess capacity
- How to detect a congestion and excess capacity?

Main Components

1) Decision Function
   - Incr. or Decr.?
2) Increase/Decrease Algorithm
   - How much to Incr./Decr.?
3) Decision Frequency
   - How often?

- Goals:
  - Fairness
  - Responsiveness



Decision Function

- Decr. tx rate when congestion occurs
  - Packet loss is a signal for congestion over the Internet (e.g. TCP)
  - React to congestion events rather than individual losses,
- Incr. tx rate periodically in the absence of congestion
  - This probes the network for excess capacity

Increase/Decrease Algorithm

- Responsiveness vs Smoothness
- Additive Inc., Multiplicative Dec.
  - Rapid reaction to a congestion results in a responsive flow but variable BW
- Smooth reaction to a congestion results in smooth variations in BW but it is not responsive to sudden changes in BW (e.g. TCP equation)

Decision Frequency

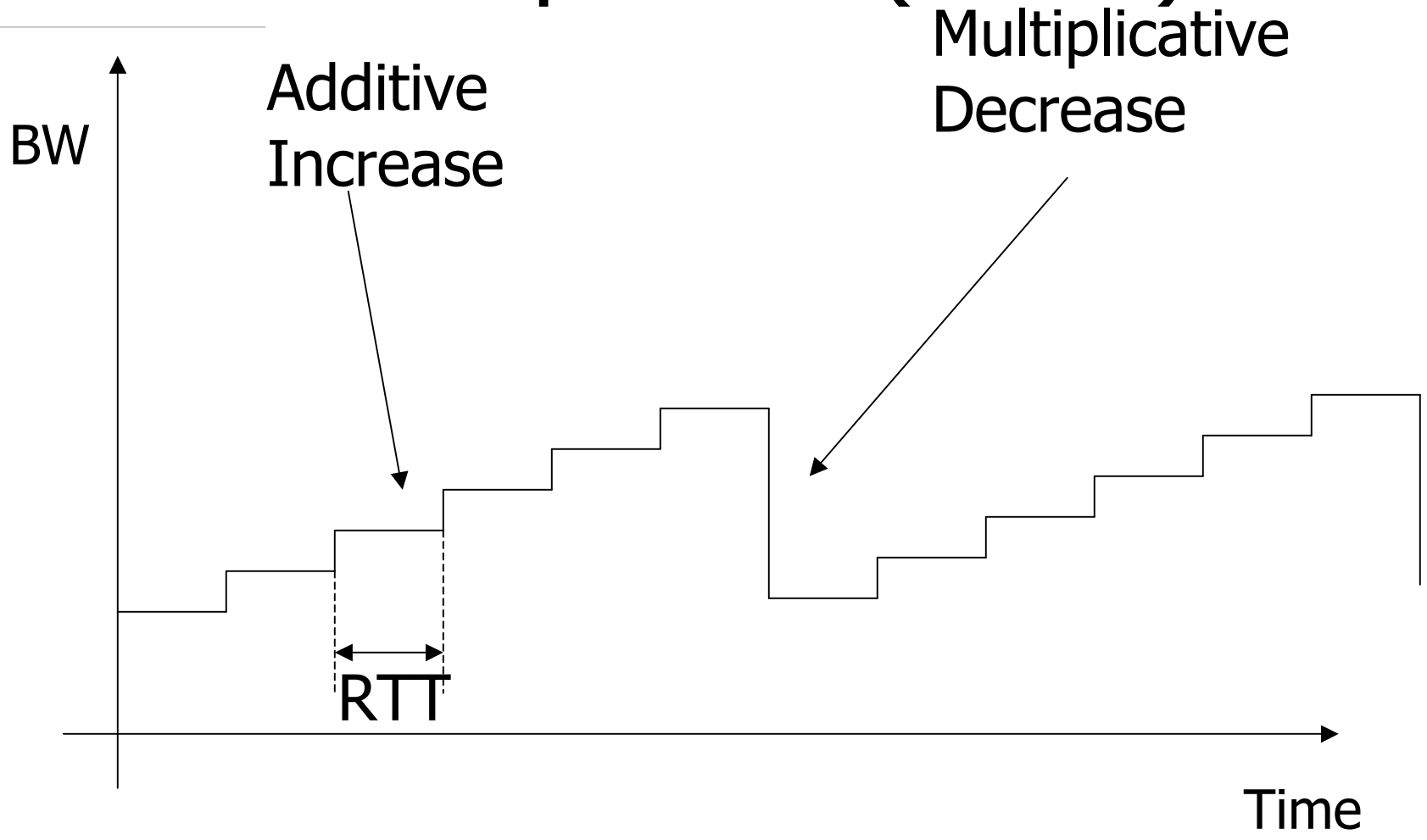
- Decision Frequency
  - Once per RTT, why?

1) When no congestion: increase the tx rate at most once per RTT

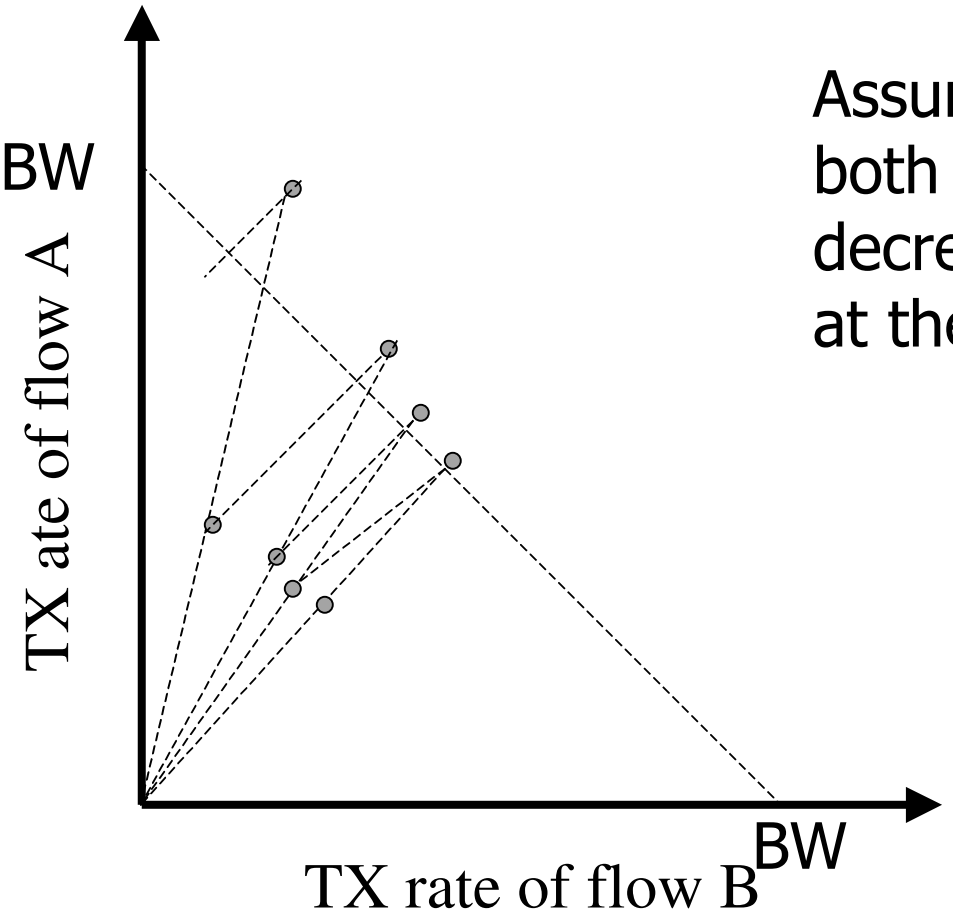
2)When congestion: decrease the tx rate at most once per RTT

- i.e. react to congestion event

Add. Inc. Exp. Dec. (AIMD)



AIMD Fairness



Assuming:
both senders increase &
decrease their tx rate
at the same rate

CC Taxonomy

- Where is the CC functionality implemented?
  - Sender-driven vs Receiver-driven CC
- How is tx rate regulated?
  1) Rate-based: controlling inter-pkt-gap
  2) Window-based: controlling no of pkt in flight
- Congestion Control vs Flow Control
  - CC: network is the bottleneck
  - FC: receiver is the bottleneck
  - Both mechanisms should work in parallel