MPEG-4 for interactive low-delay real-time communication

Olaf Landsiedel

Master's Thesis Defense

Committee:
Prof. Dr. G. Minden (Chair)
Prof. Dr. J. Gauch
Prof. Dr. J. James

Overview

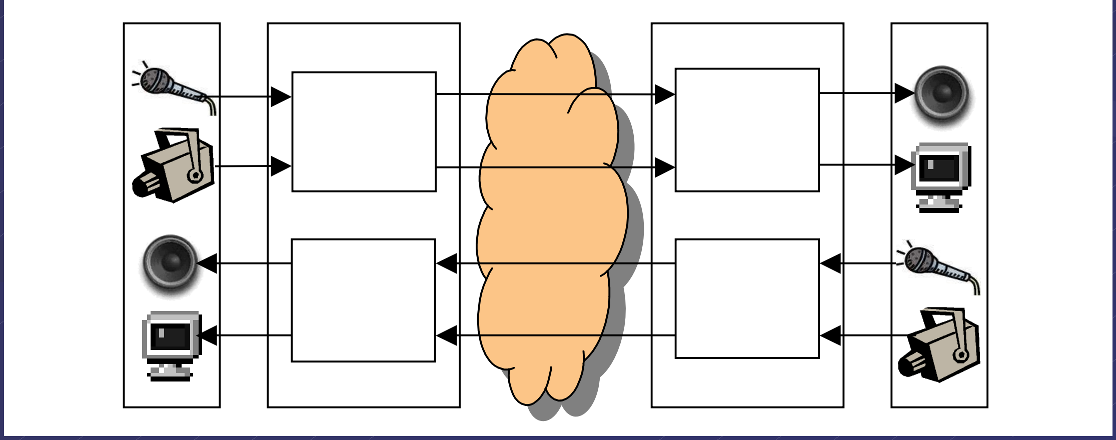
- motivation
- related work
- introduce models
  - an adaptive intra-frame insertion algorithm
  - binominal bandwidth adaptation
  - security
- implementation
- evaluation

Motivation

- internet broadcasting techniques are wide spread and have been the focus off many research projects
- today’s systems lack
  - interactivity, due to high transmission and coding delays
  - security
  - bandwidth adaptation
  - transmission error recovery
  - compression
- many applications can benefit from high interactive streaming
  - video conferencing
  - e-learning
  - remote control, example: robot control via an on-board camera or supervising laboratory experiments

Motivation (cont.)

- we introduce a streaming system
  - to deliver real-time video and audio data with low-delays
  - support high compression
  - secure transmissions
- we introduce and evaluate two techniques to reduce the impact of the changing bandwidth conditions
  - adaptive intra-frame insertion
  - binominal congestion control for RTP
- a MPEG-4 low-delay real-time streaming system for low bit-rates was implemented to incorporate the introduced techniques.

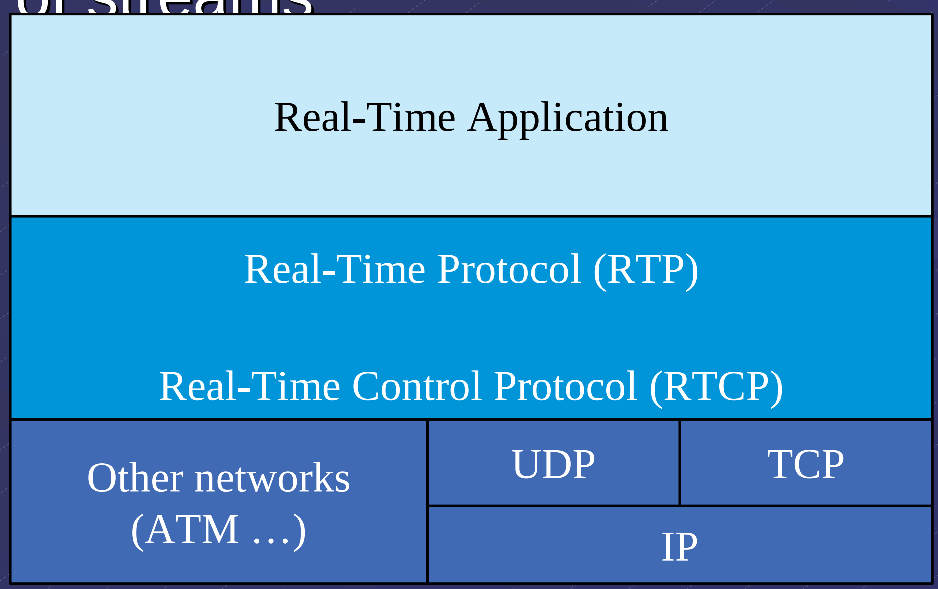


Challenges

- user studies indicate that users consider delays larger than 300ms not suitable for interactive conversation
- the Internet is a best-effort network
  - does not guarantee quality of service
- challenges for interactive communication
  - bandwidth variation
  - packet loss
  - packet errors
  - delay variation

Real Time Transport Protocol

- Real-Time Transport Protocol (RTP)
  - provides functionality to transport real-time data, including audio and video
- Real-Time Transport Control (RTCP)
  - control of streams



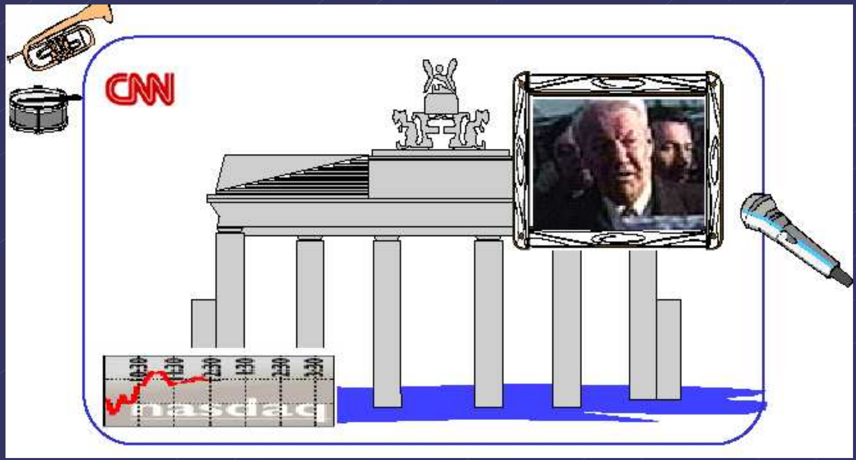
Real Time Transport Protocol (cont.)

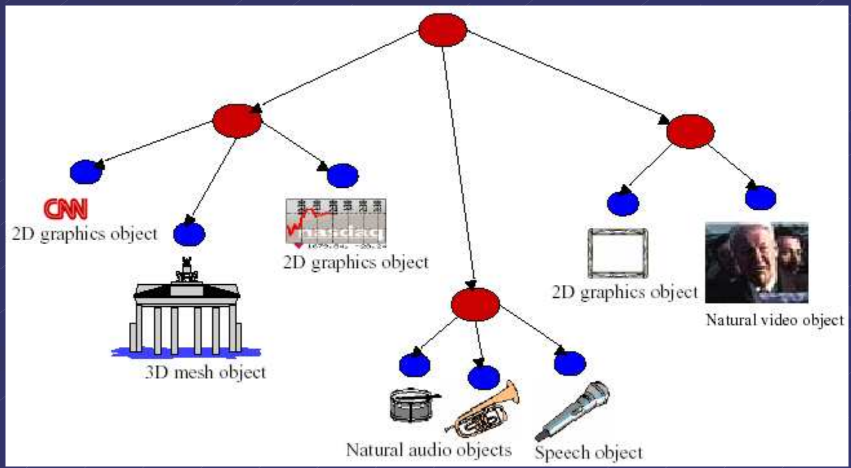
- transport protocol RTP
  - low-overhead protocol
  - control media timing
  - detect data loss
  - identify the content of a particular stream
  - achieved by an additional header:
    - timestamps
    - sequence numbers
    - source identification
- control protocol RTCP
  - management of real-time sessions
    - for multiple participants and streams
    - participants can join and leave a transmission and identify each other
  - quality of service feedback from receivers
    - packet loss

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
|            contributing source (CSRC) identifiers             |
|                             ....                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

MPEG-4 overview

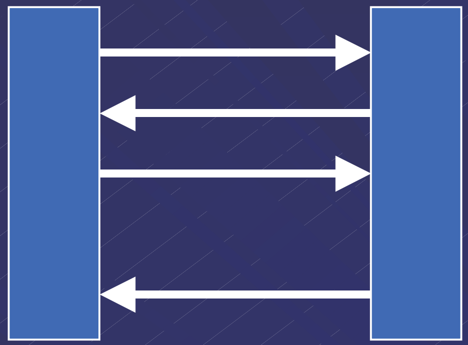
- standard by Moving Picture Experts Group (MPEG)
  - generic coding methods for moving images and voice for various applications
  - wide scope of applications
    - digital storage
    - communication
  - unlike H.261 and H.263
  - compare MPEG-1, MPEG-2: high bit-rate video and audio
    - MPEG-2: for DVD and HDTV
  - low bit-rate streams of audio and video data
  - error resilience capabilities
  - audio: AAC (Advanced Audio Coding)
  - transport of MPEG-4 streams via RTP
  - scene description for high compression





Analysis

- encoded MPEG-4 bit-stream contains limited redundant information
  - need to analyze the impact of
    - packet / frame loss
    - bit errors
- internet protocols like TCP guarantee delivery
  - recovery from packet loss by retransmission
    - introducing buffering and transmission delay
- retransmission is not suitable for high interactive streaming
  - to allow high interactivity: end-to-end delay < 300ms
  - assume transmission delay 80ms
  - with retransmission: 320ms delay
  - capture, playback, encoding, encryption and jitter add delays (about 100ms)



80ms + 80ms + 80ms + 80 ms = 320 ms

→ even without retransmission the goal of high interactivity is a challenge

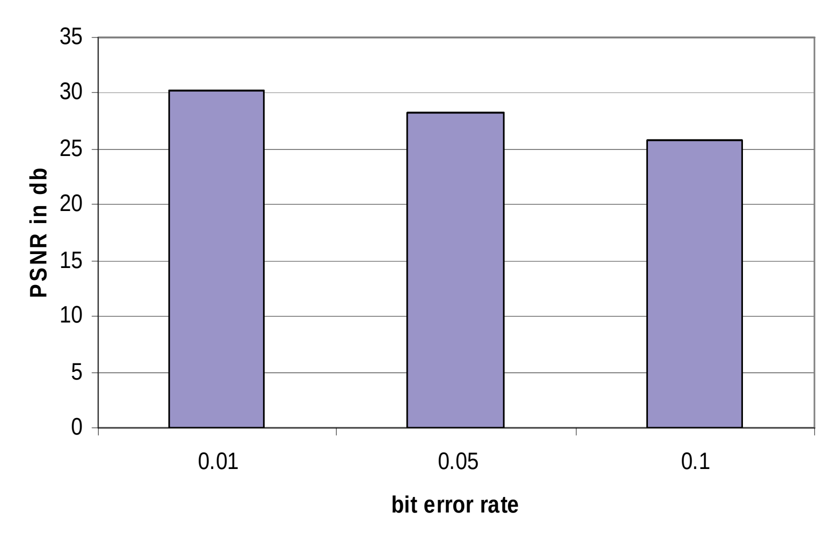
- RTP usually encapsulated in an UDP-protocol
  - no guaranteed delivery of packets
  - signals the loss of a packet based on receiver reports

Bit errors

- MPEG-4 specifies techniques to improve robustness of the audio and video stream
  - error resilience capabilities to detect and localize errors
  - recover after errors
  - visually conceal the effect of errors
- re-synchronization markers
  - ignore invalid or identify lost data
- data partitioning and header protection
  - redundancy for important fields of the bit stream
- MPEG uses variable length coding
  - high compression rates
  - sensitive to bit errors
- MPEG-4: reversible variable length coding (RVLC)
  - decodable in forward and reverse direction
  - without significant impact on the coding efficiency
  - a single bit error would not make the coded data unusable
  - ensures that bit errors do not propagate in a frame
- result: minimal quality degradation by bit errors
- internet streaming has a low bit error rate

$$RMSE = \sqrt{\frac{1}{N * M} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \left(f(x,y) - f'(x,y) \right)^2}$$

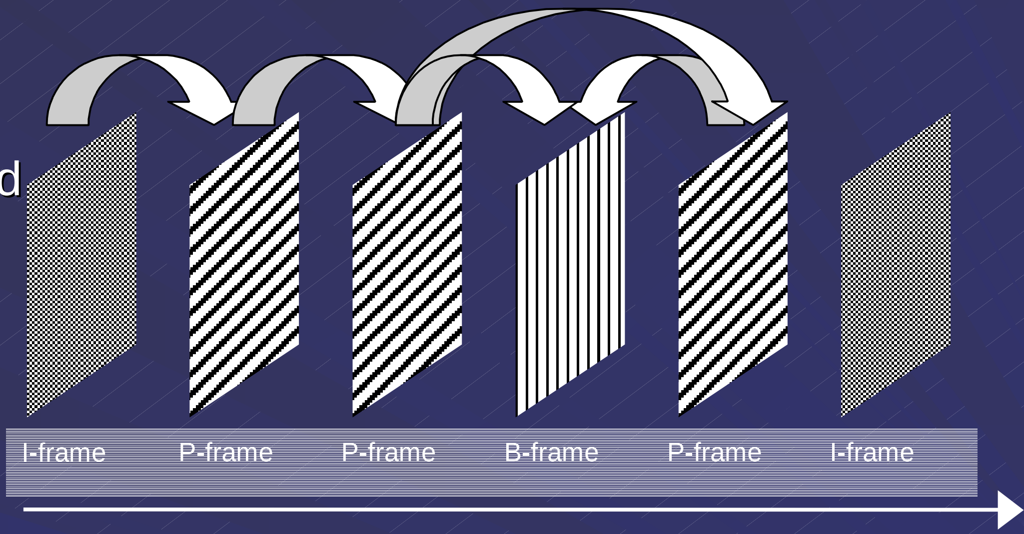
$$PSNR = 20 \, Log_{10} \frac{255}{RMSE}$$



→ bit errors are not a concern

MPEG-4 video packet types

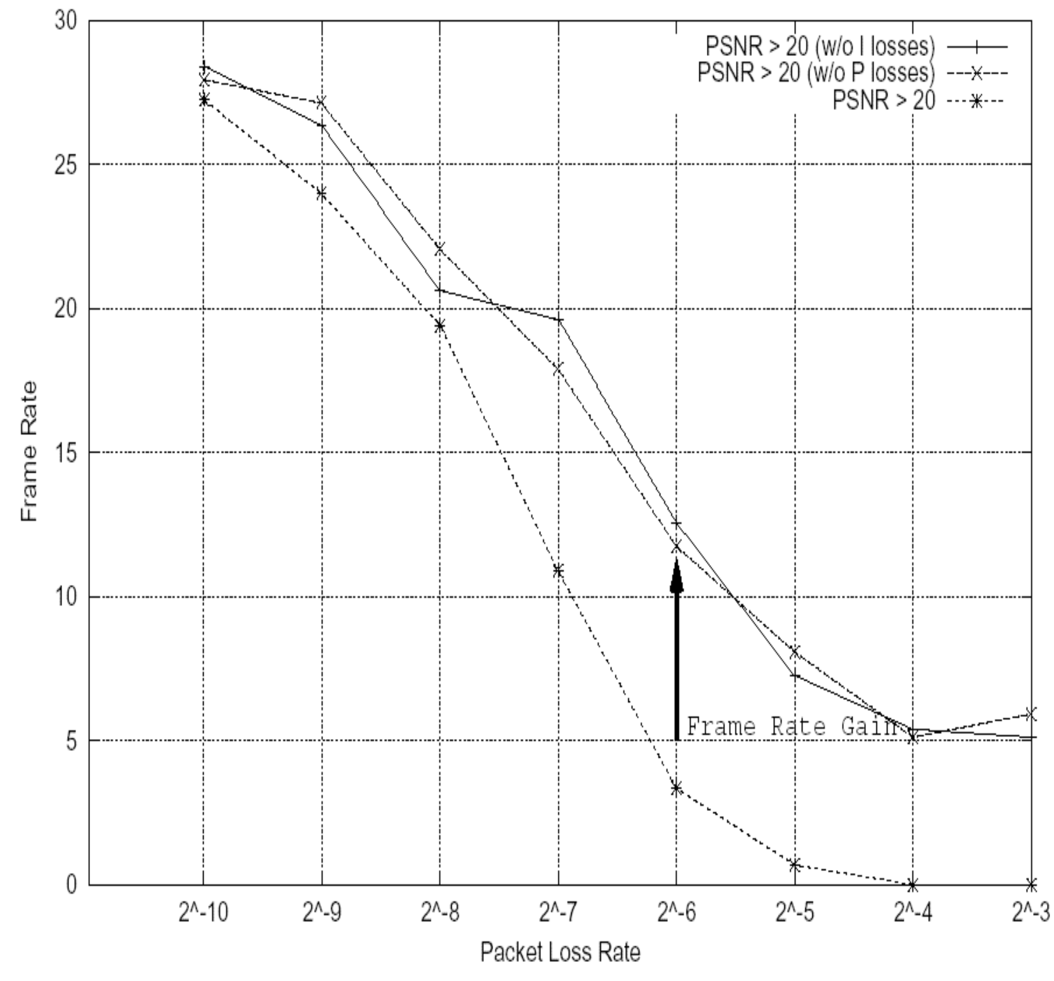
- MPEG-4 video uses inter frame prediction
  - decoding of some frames based on previous or succeeding frames
  - errors caused by lost or corrupted frames propagate between frames
  - recovery is guaranteed at reference frames (I- or intra-frames)
- reference-frames contain a fully encoded image
  - do not depend on a previous or later frame
  - significantly bigger
  - after a cut or fade
  - typical upper bound for the reference-frame interval: 10 to 20 seconds



bit-rate (kbps)	Frame type	avg. frame size (bytes)	min frame size (bytes)	max frame size (bytes)
60	prediction	239	57	1705
	intra	1998	1249	7108
300	prediction	1228	319	4076
	intra	4012	2835	8350

Intra-frame recovery

- lost Predictive-Frame
  - minor impact on the video quality
- lost Intra-Frame
  - high quality decrease (why?)
  - long error propagation (why?)
- main target: video conferencing
  - no high demanding environment
- loss of a P-frame can be accepted
- loss of a I-frame cannot be accepted
  - high quality decrease
- approach
  - add header to RTP-packet
    - intra-frame sequence number
  - resynchronization message (only) when an I-frame has been lost.

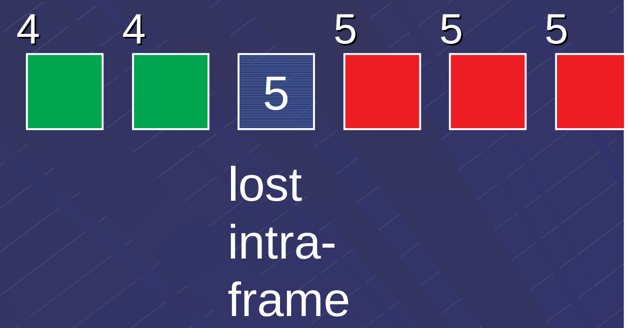


Immediate feedback

- recovery
  - by sending immediate feedback to sender
- encoding is done in real-time
  - allows sender to add key frame is to stream
  - minimal overhead
- approach does not change bit stream
  - can be used in combination with
    - error concealment
    - post processing tools
    - to limit the effects of bit stream errors or the loss of prediction frames.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|   FMT   |       PT      |          length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  SSRC of packet sender                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  SSRC of media source                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:            Feedback Control Information (FCI)                 :
:                                                               :
```

An adaptive intra-frame insertion algorithm




```
when a frame is received:
/* check whether the frame is an intra-frame */
if the frame is an intra-frame
    /* it is an intra-frame */
    /* log the number this intra-frame */
    lastReceivedIntraFrameNum = number of this frame
else
    /* it is not an intra-frame, so test for a lost intra-frame: */
    /* compare the log to the number of the intra-frame the */
    /* received frame is based on */
    if lastReceivedIntraFrameNum !=
    number of the intra-frame this frame is based on
          /* the last intra-frame was lost, so send feedback */
          send immediate feedback to sender to force an intra-frame
```

MPEG-4 Audio AAC

- AAC defines profiles with no inter frame dependencies
  - Long Term Prediction (LTP)
    - backward prediction to optimize compression
  - Low Delay (LD)
    - smaller frame size
  - quality difference to the Main profile is minimal
  - ideal for streaming
- error protection to reduce effect of bit errors
  - techniques like video (resych marker, RVLC…)
- conclude:
  - no error propagation (since no inter frame dependencies)
  - error robustness
  - retransmission not feasible for high interactivity

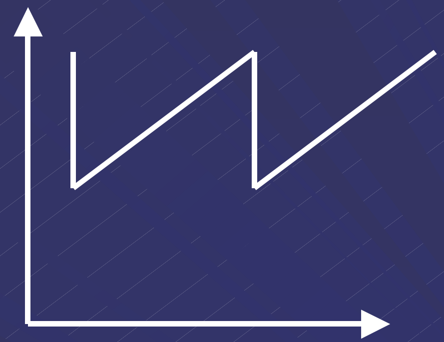
→ No recovery model for MPEG-4 AAC needed

Bandwidth adaptation

- the conditions of the internet regarding available bandwidth change with time
- not adjusting results in
  - random packet loss
  - non friendly behavior to other applications
- to deliver a stream in the highest possible quality for a given bandwidth: adjust bit rate
- to adjust properly:
  - adapt to the loss of packets: For example by adding a key frame for resynchronization
  - adjust the quality (bit-rate) of the stream to the available bandwidth
- for applications with no real-time constraints the solution is easy
  - they can afford a long buffer time
    - just buffer up a huge amount of data during start
  - this buffer is used to smoothen bandwidth oscillations
- Note: MPEG-4 allows variable bit-rate encoding

Congestion control

- TCP: successful delivery of a packet is acknowledged by the receiver
  - based on this feedback the transmitter will reduce or increase its bit-rate
  - packet loss: reduce bit-rate by half
  - linear increase to probe for available bandwidth
  - called additive increase / multiple decrease (AIMD)
  - results in extreme rate oscillations



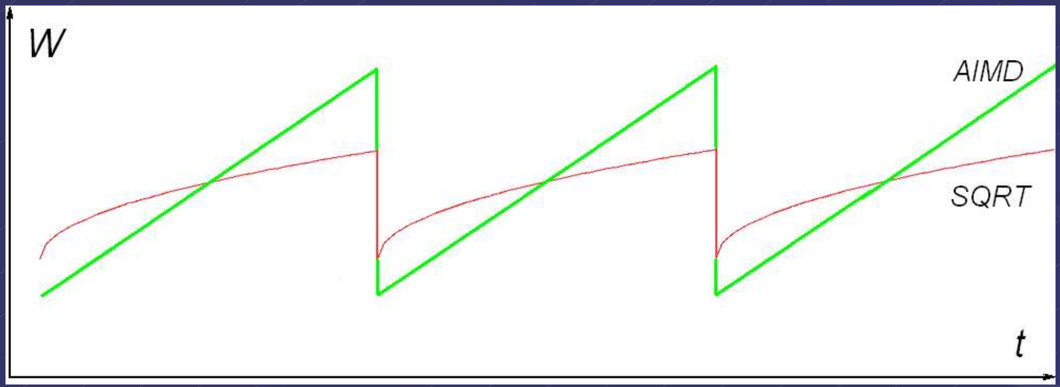
- streaming is based on UDP:
  - no acknowledgements
  - RTCP control messages sent by the receiver can be used for packet loss detection
- we need an algorithm that
  - reduces its throughput up on congestion
  - probes for bandwidth
  - not introduce extreme rate oscillations
  - TCP-friendliness to guarantee fair bandwidth sharing
  - handle "compound" feedback from RTCP reports
  - video only
    - Audio is a fragment of the connection

Binominal congestion controls

- Binominal congestion control is a superset of TCP's AIMD congestion control algorithm
- generalization includes entire linear congestion algorithms
  - k=0, l=1: AIMD (additive increase, multiple decrease) used by TCP
- k=0.5, l=0.5: SQRT
  - increase is inversely proportional
  - decrease is proportional to square-root of the window size.
  - TCP-friendly
  - low bit rate oscillations

$$I : w_{t+rtt} \leftarrow w_t + \frac{\alpha}{w_t^k}; \alpha > 0$$

$$D : w_{t+\delta t} \leftarrow w_t - \beta w_t^l; 0 < \beta < 1$$



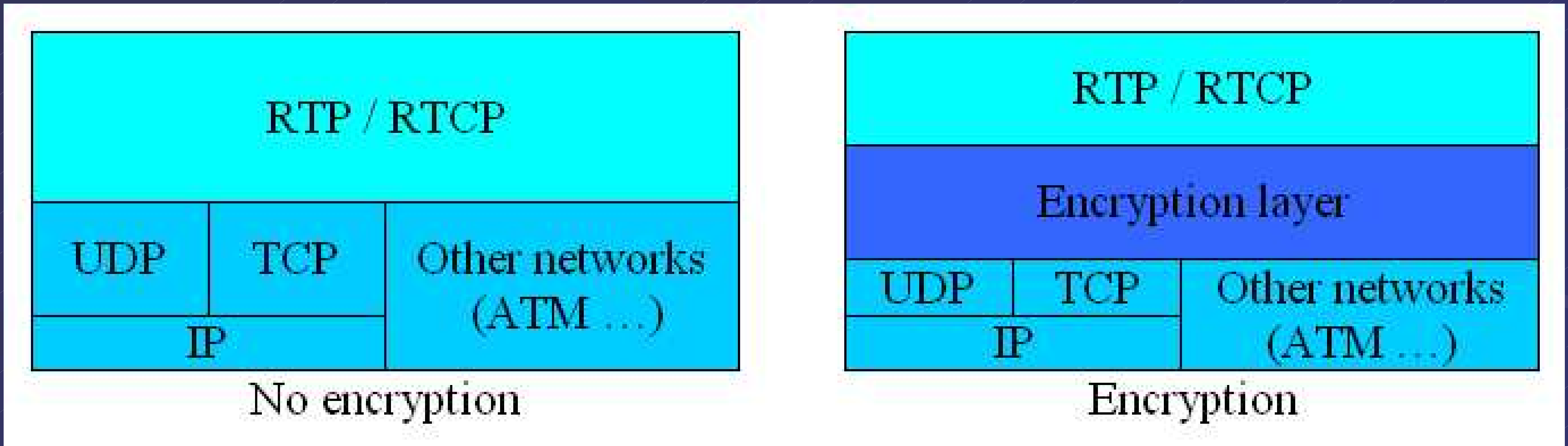
→ SQRT is a perfect candidate for media streaming

Security

- to ensure that the requirements everyone has for communication cannot be subverted by other people
  - data integrity
  - authentication
  - confidentiality
  - key distribution
- Ambient Computing Environment (ACE) provides
  - authentication
  - key distribution
- real-time constraints
  - the algorithms have to be fast
- bandwidth constraints
  - hashing, padding add overhead

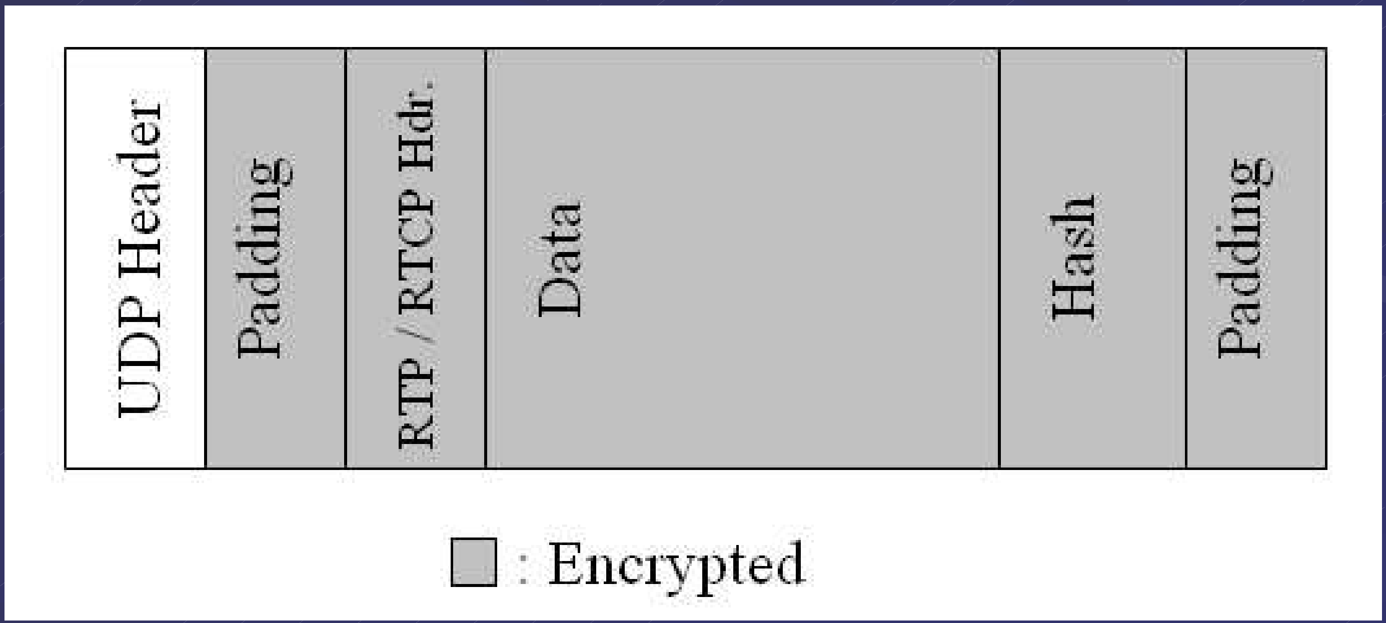
RTP and encryption

- all data has to be encrypted
  - RTP headers
  - RTCP control messages
- achieved by adding additional layer



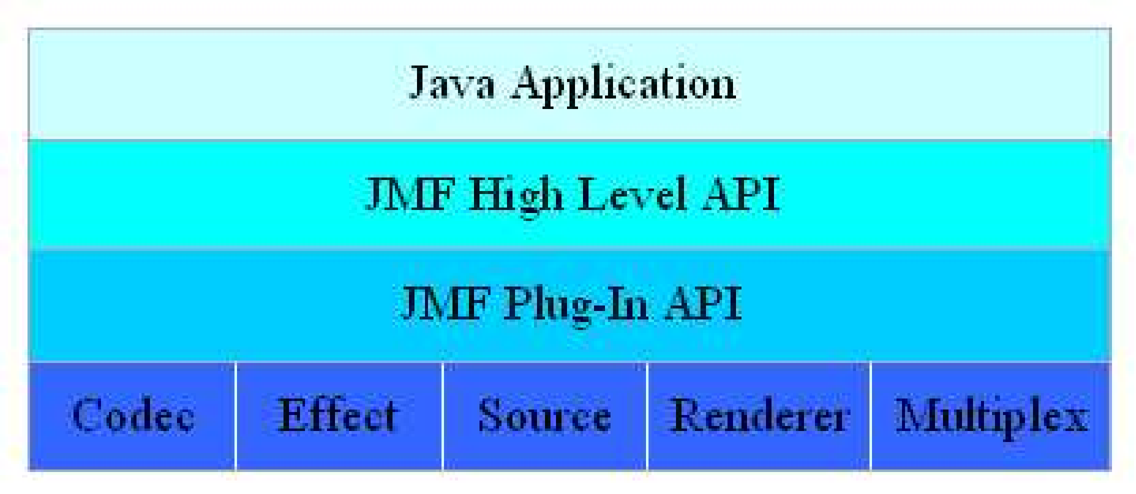
Algorithms

- Encryption / Decryption
  - Advanced Encryption Standard (AES)
    - 128, 192 and 256 bit keys
- Hashing
  - Secure Hash Algorithm (SHA-1)
    - 160 bit

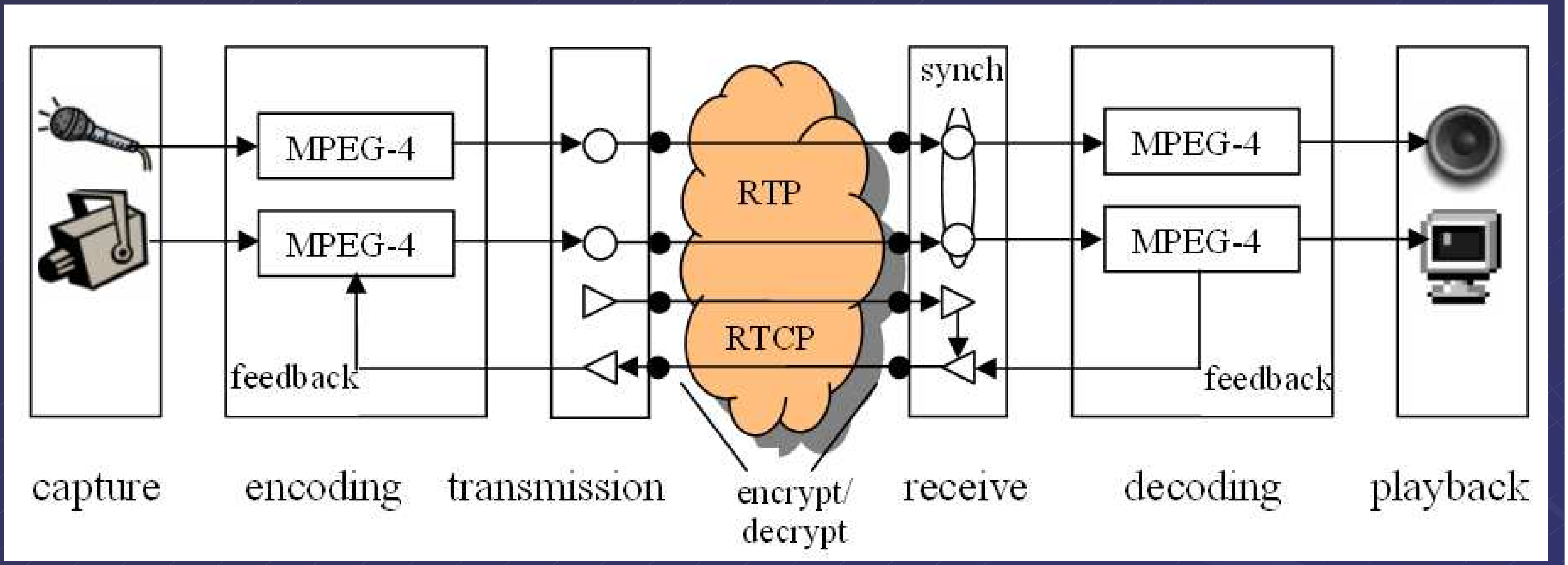


Implementation

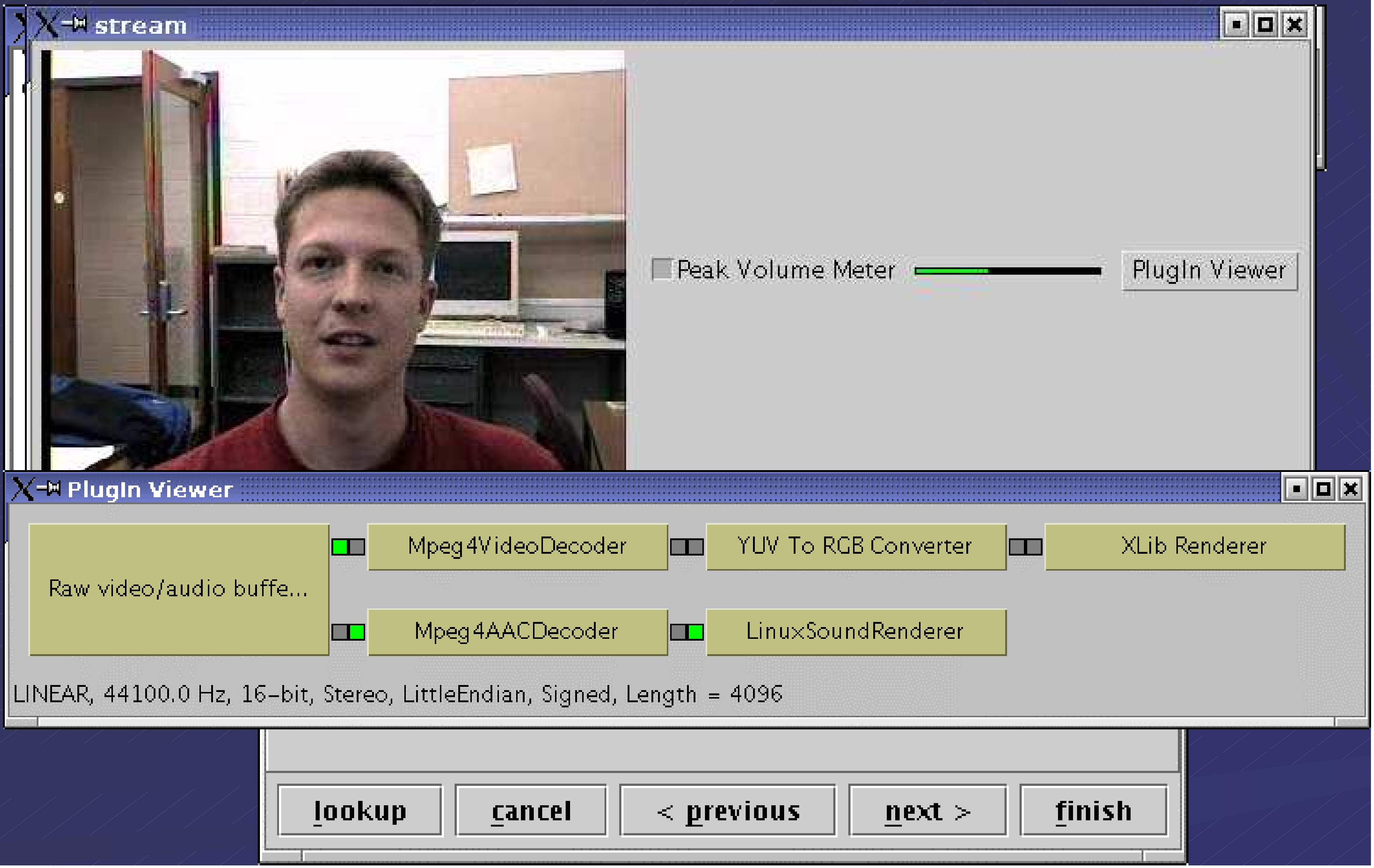
- Java
- Java Media Framework (JMF)
- low delay sound API
  - ALSA and the Java Native Interface (JNI)
  - 30 ms delay
- MPEG-4
  - video: XVID, FFMPEG
  - audio: FAAC
- low performance codecs
  - video: H.263
  - audio: G.723, GSM



System block diagram



Screenshots



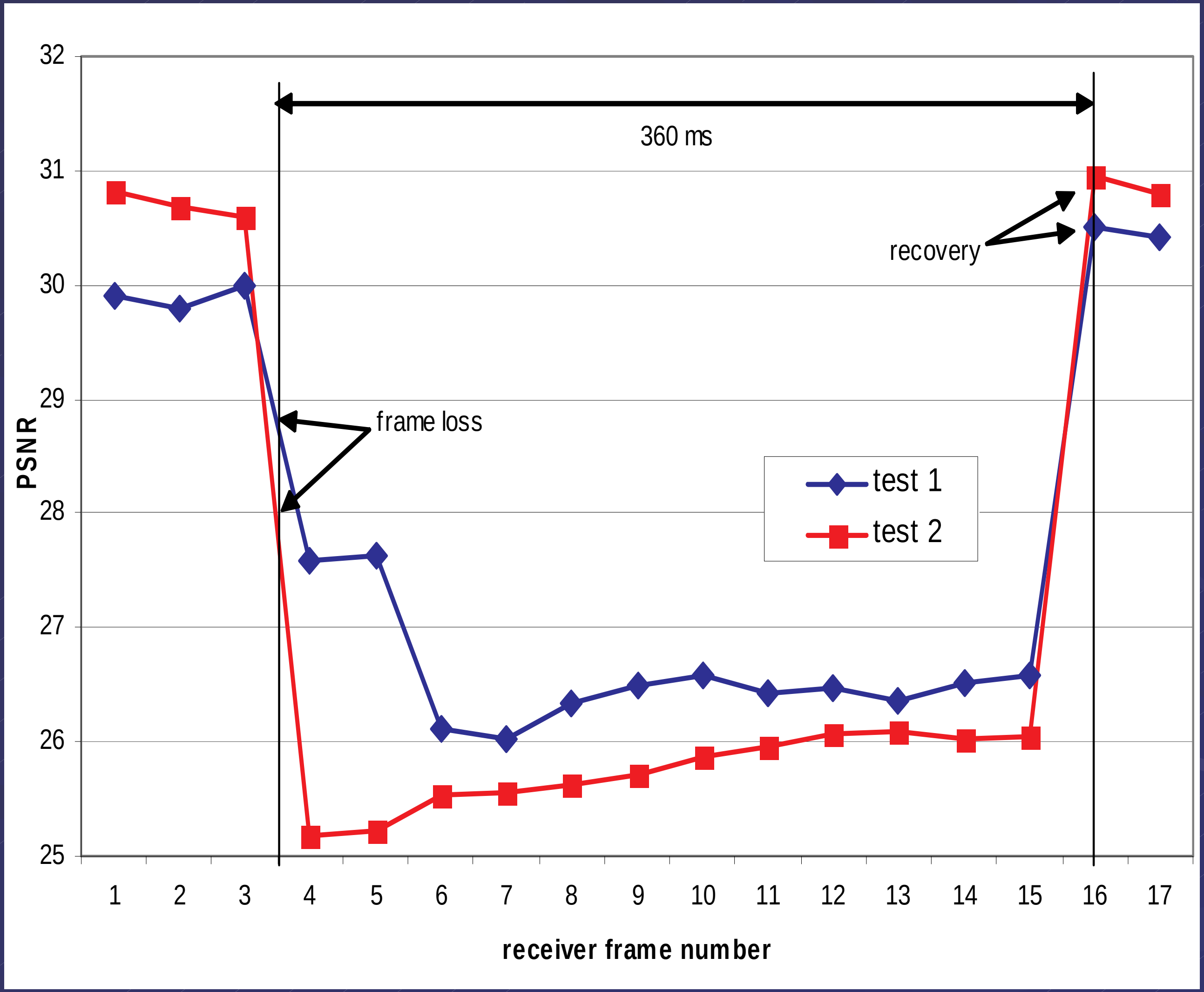
Evaluation: adaptive intra-frame insertion

frame 4	frame 9	frame 16

Test 2 (300 kbps, high motion):
upper row: receiver (first 2 pictures are degraded, the third picture is the recovery picture)
lower row: original uncompressed images

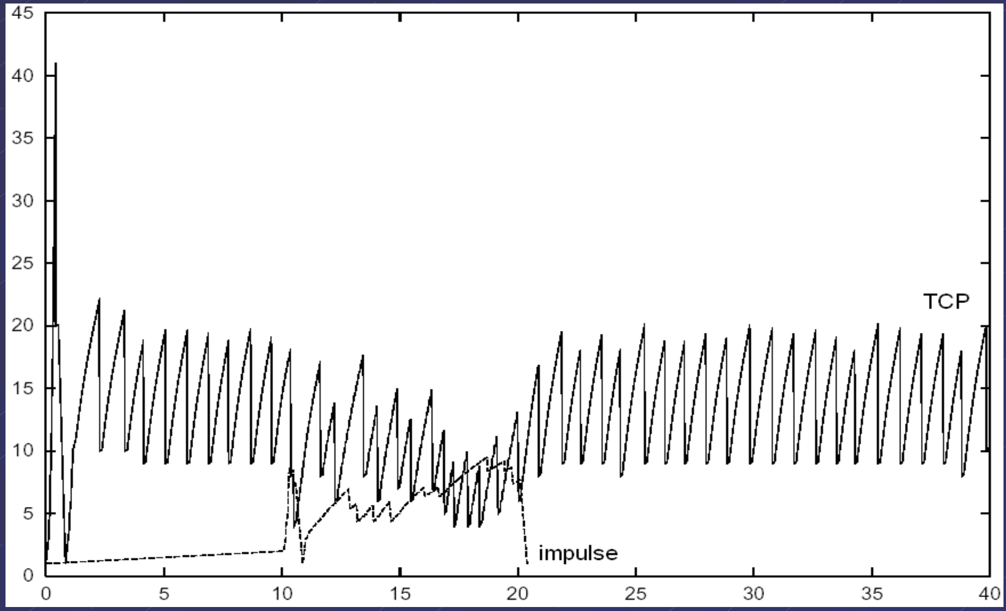
Adaptive intra-frame insertion

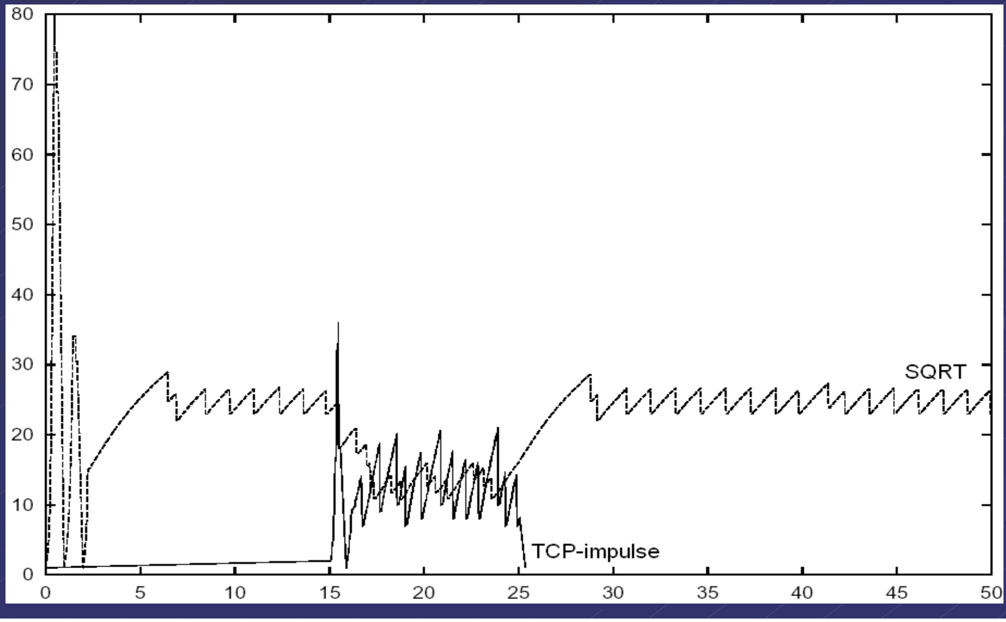
- round trip time 200ms
- recovery 360ms
- 160ms needed to react and handle the request



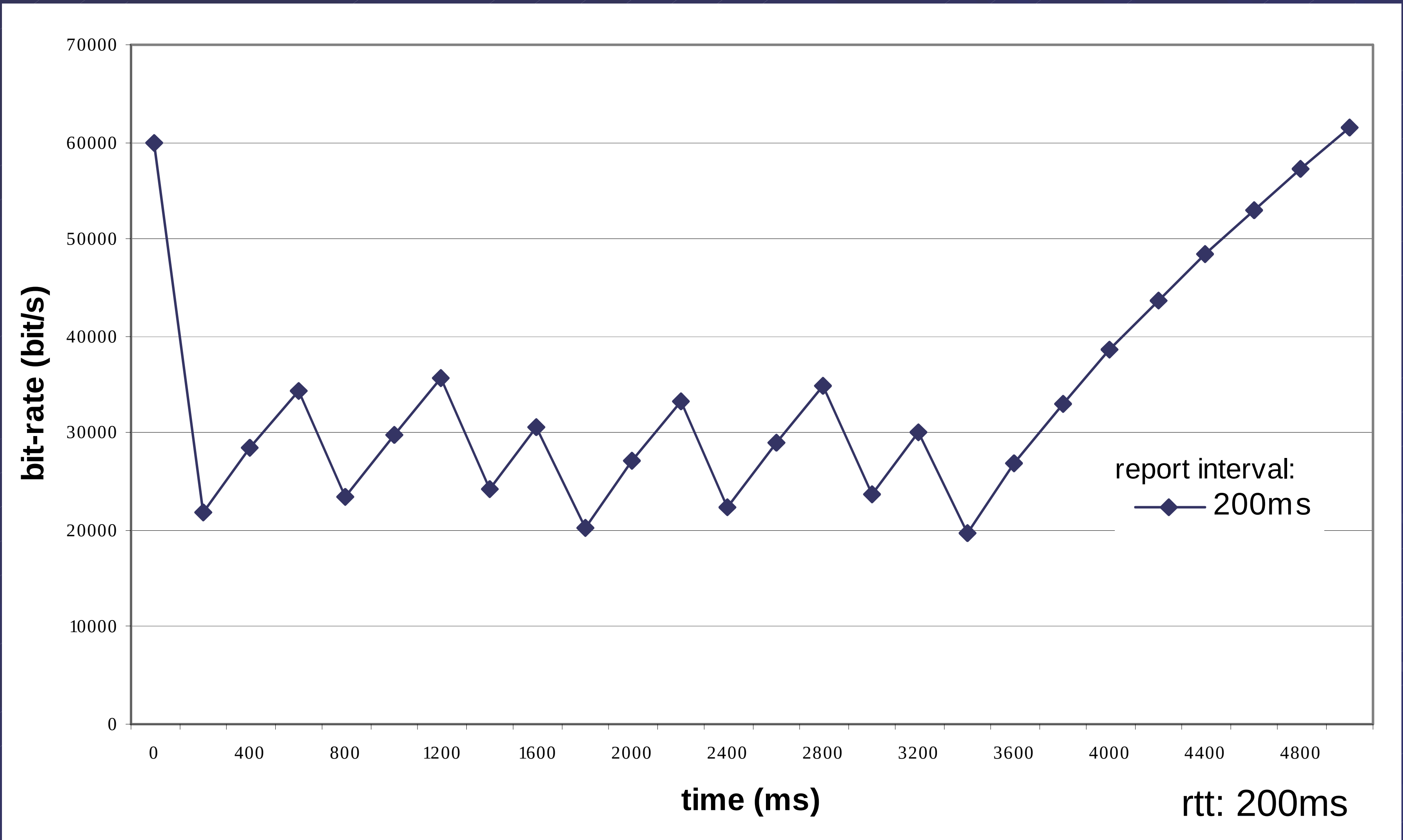
Binominal congestion control

- traditional TCP congestion control
  - too extreme rate oscillations
  - not suitable for streaming applications
- streaming requires a smooth playback
- binominal SQRT congestion control
  - less oscillation
  - TCP-friendly
    - throughput

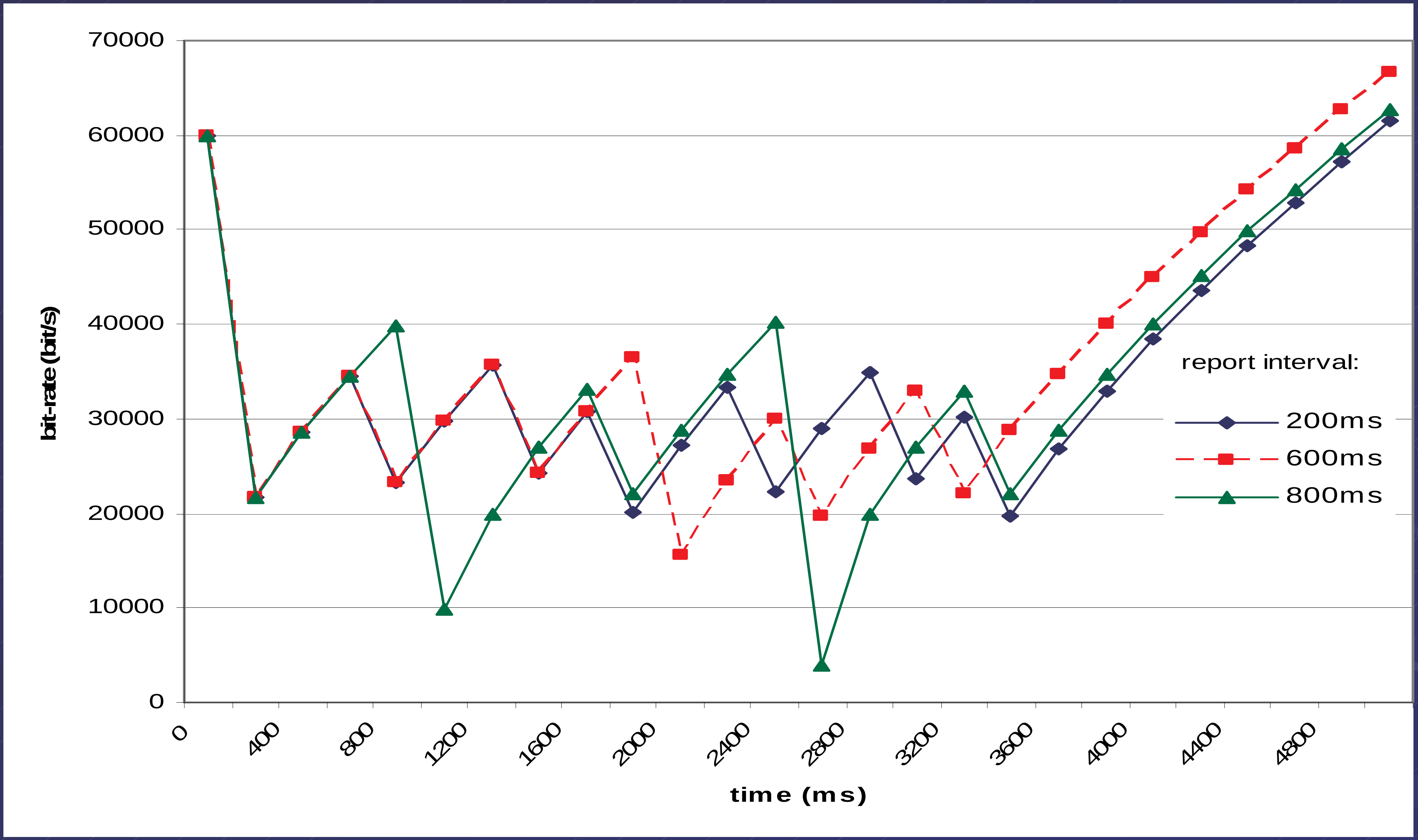




Adaptation with RTCP feedback

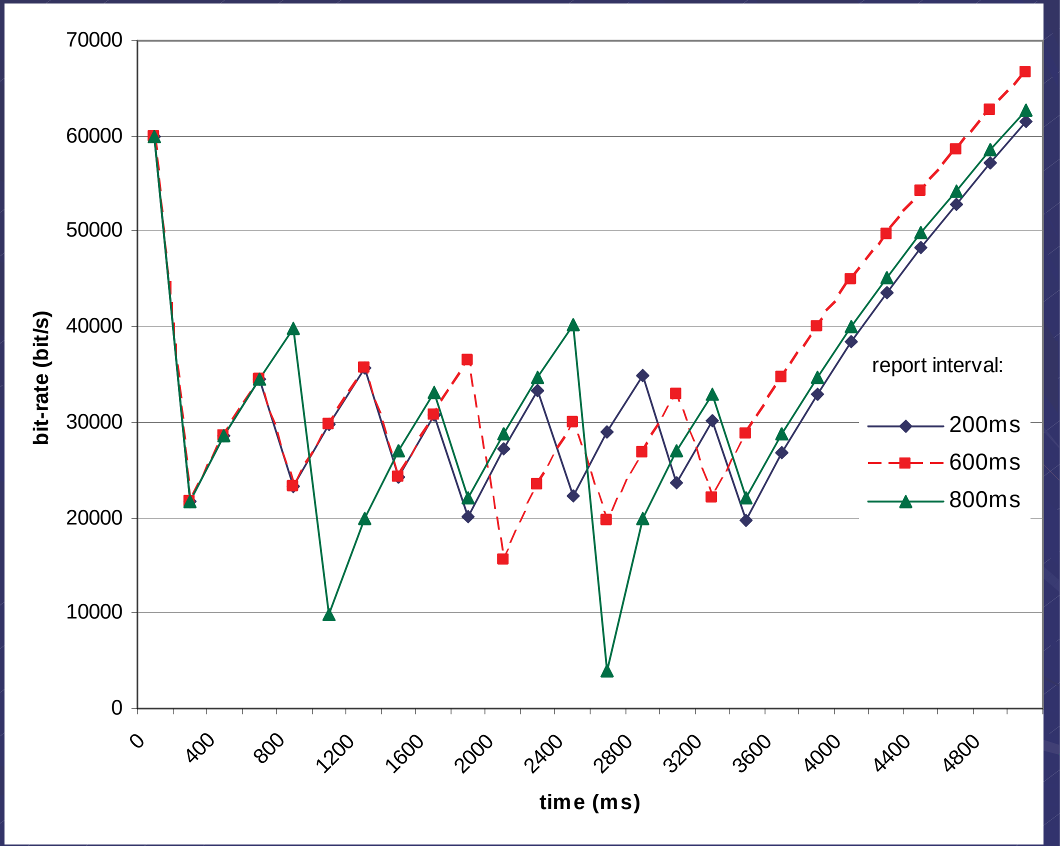


Adaptation with RTCP feedback (cont.)



Adaptation with RTCP feedback (cont.)

- adaptation quality depends on report intervals
  - 600ms and less
    - good adaptation
  - 800ms
    - too extreme bit-rate oscillation
    - offers a certain degree of adaptation.



Security

- bandwidth: padding and hashing adds overhead
- CPU load, delay:
  - No additional delays could be measured

Video Bit rate (kbit/second)	60	300	900
Frame rate (frames/second)	20	30	30
Avg. packet size (bit)	3000	10000	30000
SHA-1 hash size (bit)	160	160	160
Avg. padding (bit)	256	256	256
Overhead	13.9%	4.16%	1.39%

→ security overhead can be neglected

Future work

- high compression video and audio coding is under heavy development
  - example: new layer for MPEG-4 audio and video
    - its real-time coding will become feasible in the next years
- video bit-rate control
  - the quantization value for video has to be estimated, based on
    - previous frames
    - motion
    - targeted bit-rate
  - recent research deployed new algorithms
  - their integration in today's codecs would
    - increase quality
    - reduce derivation from the target bit-rate

Acknowledgements

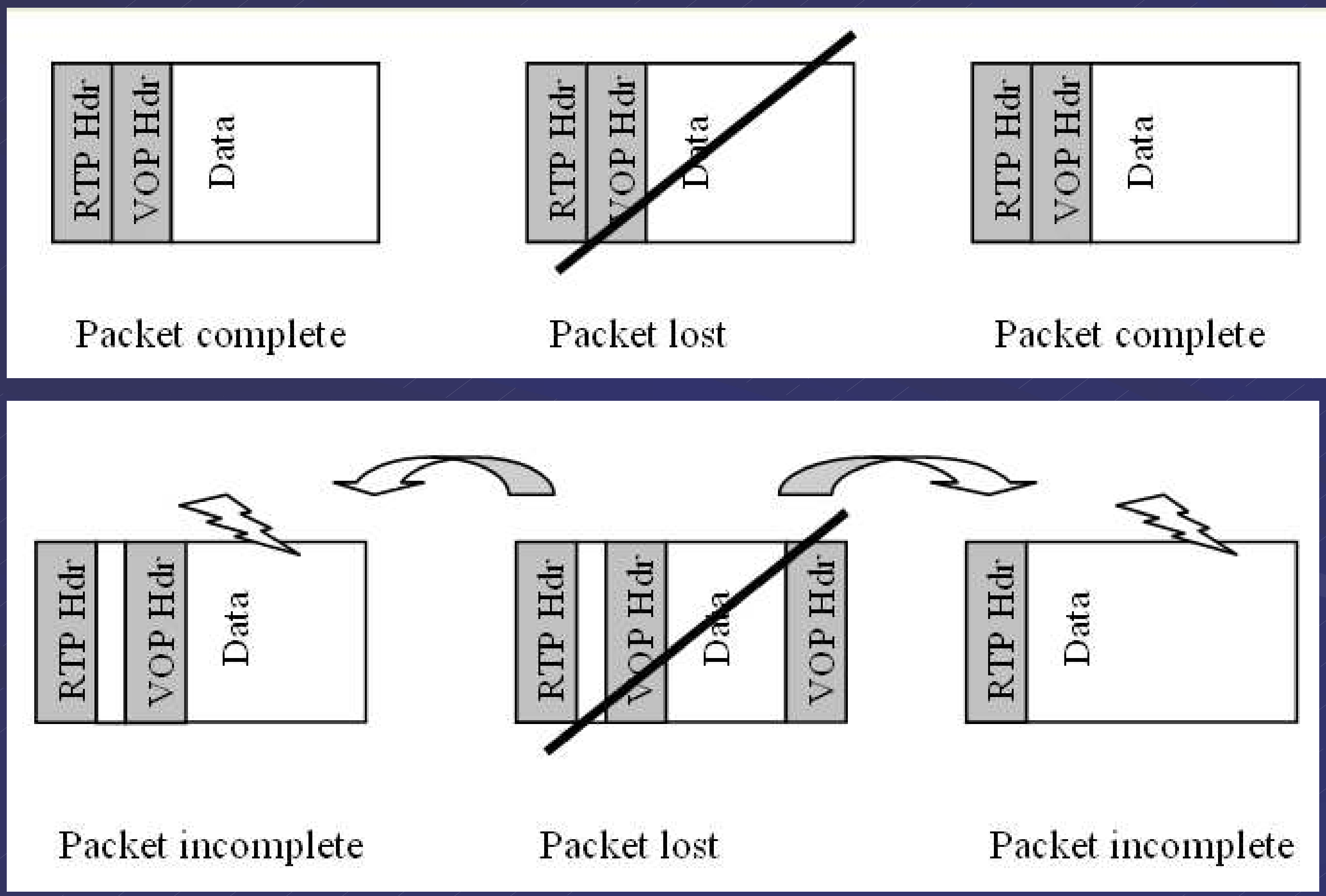
- I want to thank many people for
    - making my time in the US at the University of Kansas possible
    - for the chance to do my research here the Information and Technology Center ITTC
  - – Prof. Dr. Minden
  - – my Committee
  - – Leon Searl
  - – James Mauro
  - – The Department of State, Fulbright and IIE for the Fulbright Grant
  - – Christian Albrechts Universität zu Kiel, University of Kansas for the Direct Exchange Scholarships
  - – DARPA, NSF, Sprint and the University of Kansas for funding ACE

MPEG-4 for interactive low-delay real-time communication

Questions?

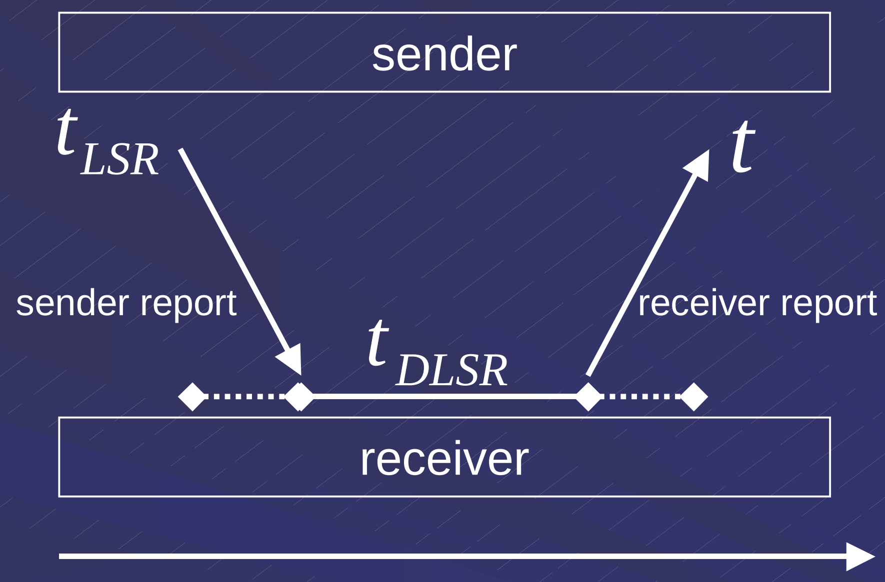
RTP and MPEG-4

- packet mapping
  - effect of packet loss when frames are mapped inappropriately



Round Trip Time RTT

- time of the last received sender report (t LSR)
- time elapsed between receiving the last sender report and sending this receiver report (t DLSR)
- receiver report arrival time (t)
- Note: calculation does not require clock synchronization of the participants
  - considered accurate



$$t_{rtt} = t - t_{DLSR} - t_{LSR}$$

RTP header

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
|            contributing source (CSRC) identifiers             |
|                             ....                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Intra-frame adaptation test

frame 7 | frame 9 | frame 16

Test 1 (60 kbps, low motion)
upper row: receiver (first 2 pictures are degraded, the third picture is the recovery picture)
lower row: original uncompressed images

Congestion control

- k=0, l=1: AIMD (additive increase, multiple decrease)
  - used by TCP
- k=-1, l=1: MIMD (multiple increase, multiple decrease)
  - used by the slow start algorithm in TCP
- k=-1, l=0: MIAD (multiple increase, additive decrease)
- k=0, l=0: AIAD (additive increase, additive decrease)

$$I : w_{t+rtt} \leftarrow w_t + \frac{\alpha}{w_t^k}; \alpha > 0$$

$$D : w_{t+\delta t} \leftarrow w_t - \beta w_t^l; 0 < \beta < 1$$

Congestion control

$$I : w_{t+rtt} \leftarrow w_t + \frac{\alpha}{w_t^k} ; \alpha > 0$$

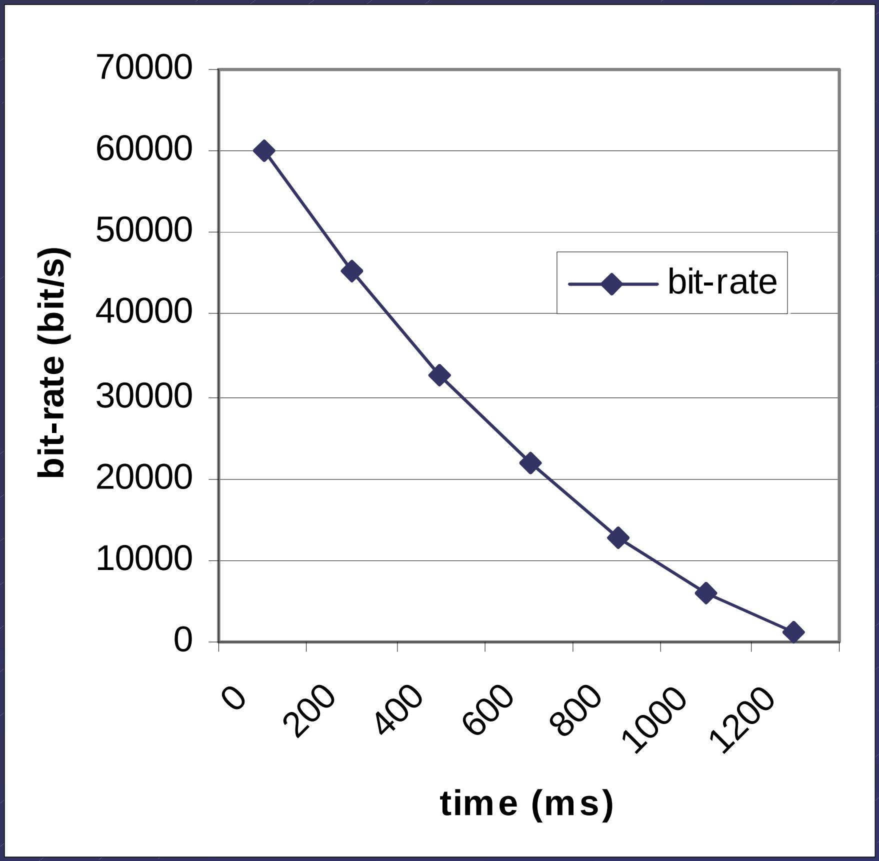
$$D : w_{t+\delta t} \leftarrow w_t - \beta w_t^l ; 0 < \beta < 1$$

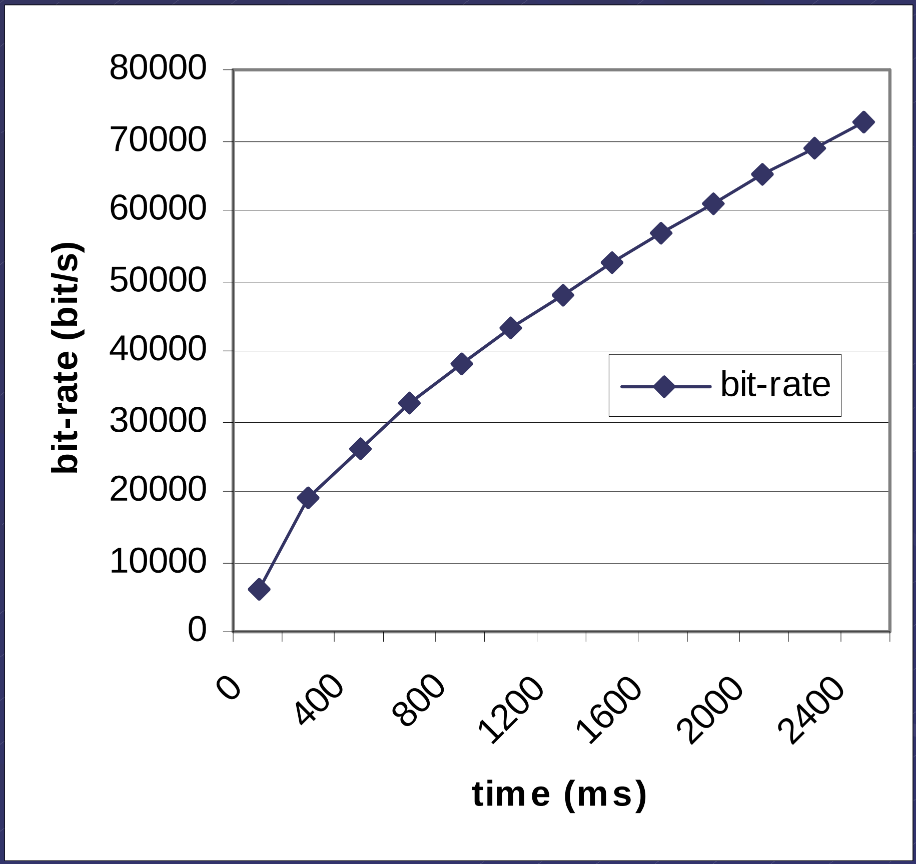
$$w = \frac{r * rtt}{MTU}$$

$$I : r_{t+rtt} = r_t + \left(\frac{MTU}{rtt} \right)^{\frac{3}{2}} \sqrt{\frac{1}{r_t}}$$

$$D : r_{t+\delta t} = r_t - 0.6 \sqrt{\frac{r_t * MTU}{rtt}}$$

SQRT increase / decrease





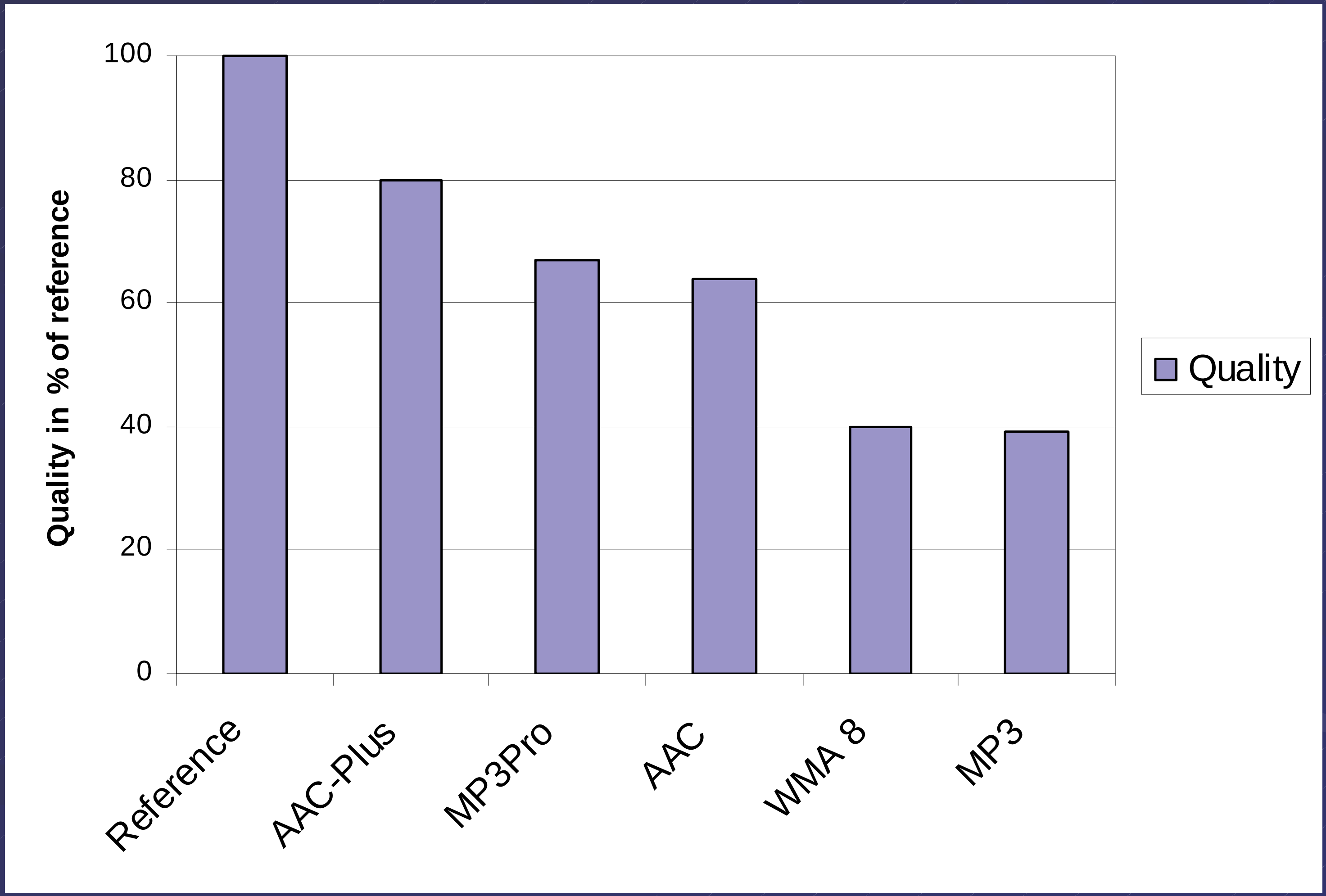
rtt = 200ms

New MPEG-4 layer

- MPEG-4 AVC
  - not yet suitable for real-time coding
  - example: resources for real-time video coding

Codec	CPU	Speed	Memory
Encoder	Pentium IV	3.0 GHz	512 MB
Decoder	Pentium IV	2.0 GHz	256 MB

MPEG-4 audio quality



vs. Microsoft NetMeeting

Feature	NetMeeting	Our work
Video		
Supported codecs	H.263	MPEG-4, H.263
Target bitrates	All	All
Variable bit-rate coding	No, two qualities offered for manual bandwidth adaptation	Yes, variable bitrates for automatic bandwidth adaptation
Recovery from frame loss	No	Yes, automatic key frame addition
Key frame interval	15 seconds	Variable, currently 10 seconds
Audio, fixed bit-rate		
Supported codecs	G.723	MPEG-4 AAC (16 and more kbits/s), G.723, GSM
Security		
Authentication	Certificates	Based on the ACE framework (including fingerprints sensors, certificates…)
Encryption	None for audio and video	AES and SHA-1
Session setup	H.323	ACE service directory