QoS/QoE Demands of Different Web Applications

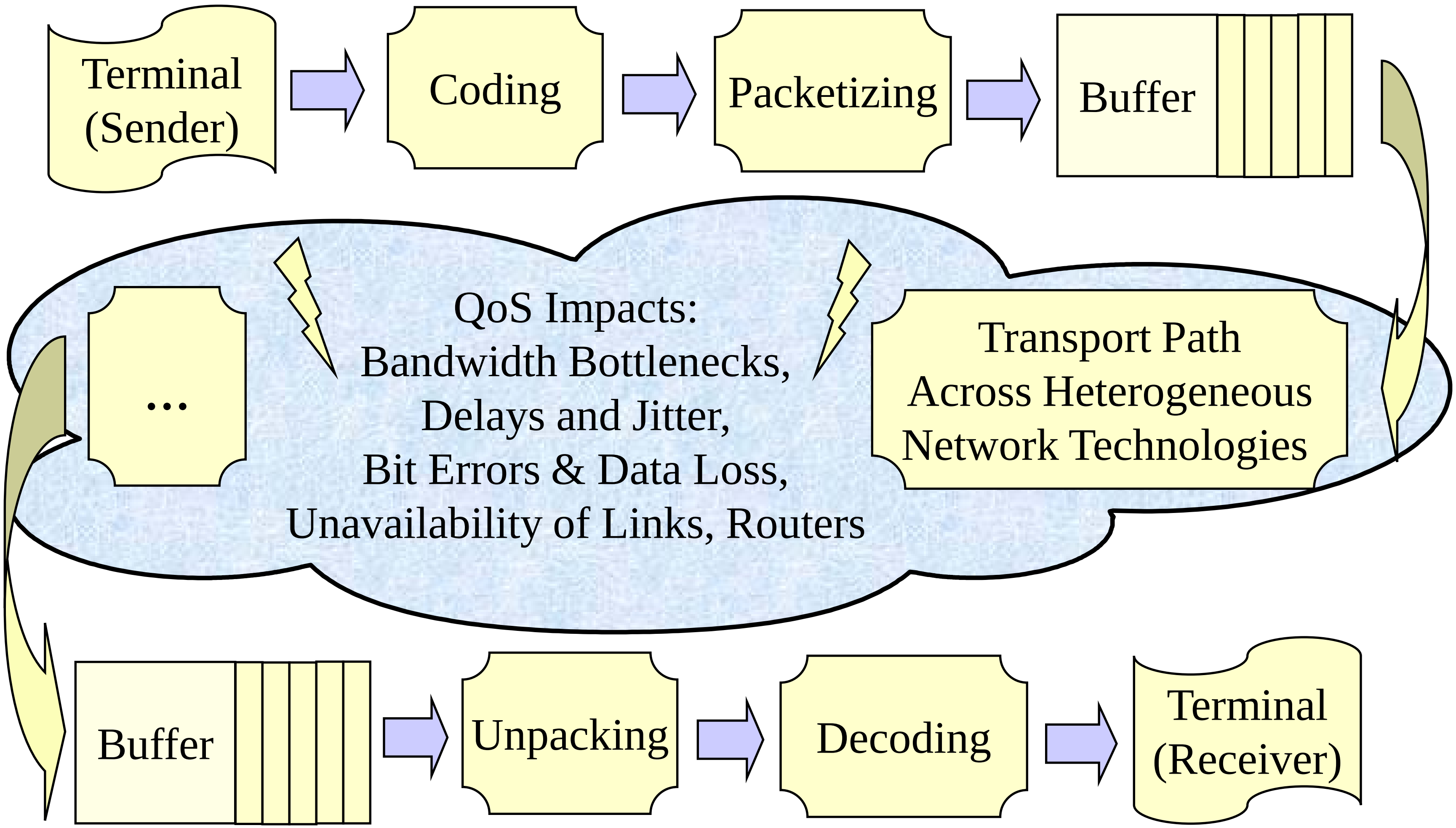
QoS Criteria & Key Performance Indicators for Application & Service Categories

QoS/QoE* for Real Time Services:

- Influence of Coding, Buffering
- Voice (ISDN, VoIP, Skype)
- Video (JPEG, MPEG coding)
- Gaming etc.

*Quality of Experience (QoE): User Perceived Quality

Data Transfer Chain Across Telecommunication Networks & QoS Impacts



Service-,Traffic- & QoS-Classes: Categories & Demand Profiles

Service / Traffic Classes	Conversational	Streaming	Interactive	Background
Applications	Voice Calls, Video Conferencing Online Gaming	Video / audio on demand, IP-TV	Web browsing, E-Commerce, E-Learning etc.	File Transfers, Downloads, P2P, E-Mail, SMS, …
Communication-Traffic Pattern	Human-to-Human Bidirectional Partly Multicast	Server → Human >90% Downstream Uni/Multi/Broadcast	Human ↔ Server Query/Response Pattern	Data transfers without human interaction
QoS Parameters: 1. Delay	< 0.1s: excellent >0.25s: inappropriate	Time sequence of data to be preserved	< 0.5s: excellent >4s: inappropriate	Not critical
2. Failure Rate	Tolerance up to a few % bit/packet errors		No failures in end-to-end transfers	
3. Bandwidth	Low & high data volume demands in each class (e.g. voice ↔ video)			
Source: UMTS Standardization by ETSI/3GPP (TS 27.107 V3.9.0, 2002)				

Future Demands for 5G Mobile Networks
Source: Next Generation Mobile Networks NGMN White Paper (March 2015)

Use case category	User Experienced Data Rate	E2E Latency	Mobility
Broadband access in dense areas	DL: 300 Mbps UL: 50 Mbps	10 ms	On demand, 0-100 km/h
Indoor ultra-high broadband access	DL: 1 Gbps, UL: 500 Mbps	10 ms	Pedestrian
Broadband access in a crowd	DL: 25 Mbps UL: 50 Mbps	10 ms	Pedestrian
50+ Mbps everywhere	DL: 50 Mbps UL: 25 Mbps	10 ms	0-120 km/h
Mobile broadband in vehicles (cars, trains)	DL: 50 Mbps UL: 25 Mbps	10 ms	On demand, up to 500 km/h
Airplanes connectivity	DL: 15 Mbps per user UL: 7.5 Mbps per user	10 ms	Up to 1000 km/h
Ultra-low latency	DL: 50 Mbps UL: 25 Mbps	<1 ms	Pedestrian
Resilience and traffic surge	DL: 0.1-1 Mbps UL: 0.1-1 Mbps	Regular communication: not critical	0-120 km/h
Ultra-high reliability & Ultra-low latency	DL: From 50 kbps to 10 Mbps; UL: From a few bps to 10 Mbps	1 ms	on demand: 0-500 km/h

ngmn the engine of broadband wireless innovation

QoS/QoE Demands of Different Web Applications

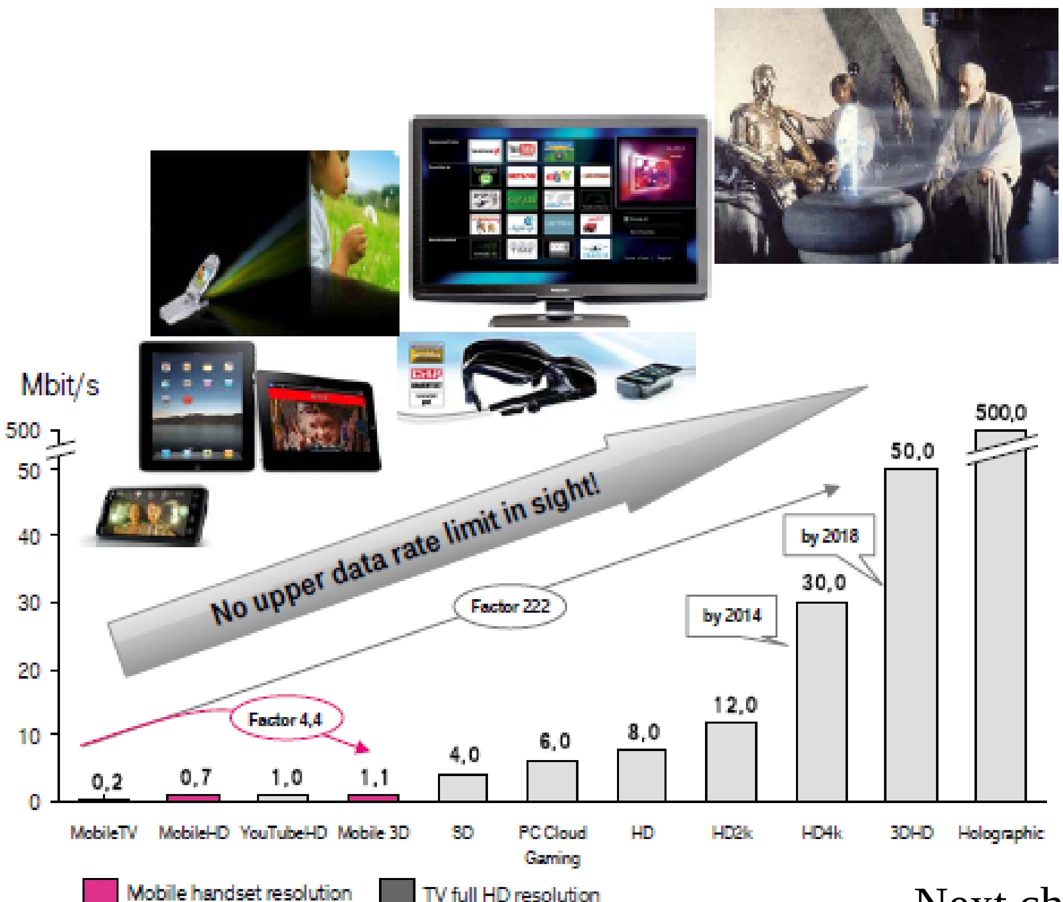
QoS Criteria & Key Performance Indicators for Application & Service Categories

QoS/QoE* for Real Time Services:

- **Influence of Coding, Buffering**
- Voice (ISDN, VoIP, Skype)
- Video (JPEG, MPEG coding)
- Gaming etc.

*Quality of Experience (QoE): User Perceived Quality

Bandwidth is most important for QoS !
Growth demand driven by video streaming



Findings

- **Video on mobile devices become daily life service,** supporting also HD and 3D TV.
- Further dimensions like 3D and holography.
- Hi-Resolution screen already in the market (e.g. iPhone 4G).
- Micro projectors for mobile devices to appear end of 2010.
- New technologies: Flexible, head mounted, inorganic and sensing displays.
- **Large screens demand for multiple of bandwidth** than mobile devices with limited screen size.
- **It needs a management decision** whether fix net devices should be enabled to stream over wireless networks: huge traffic impact.
- Also multiples of streaming are necessary for some use cases:
  - Gaming as a streaming scenario (picture computing in the cloud).
  - Stream browsing (browsing in virtual rooms).

Next challenge beyond: Virtual reality

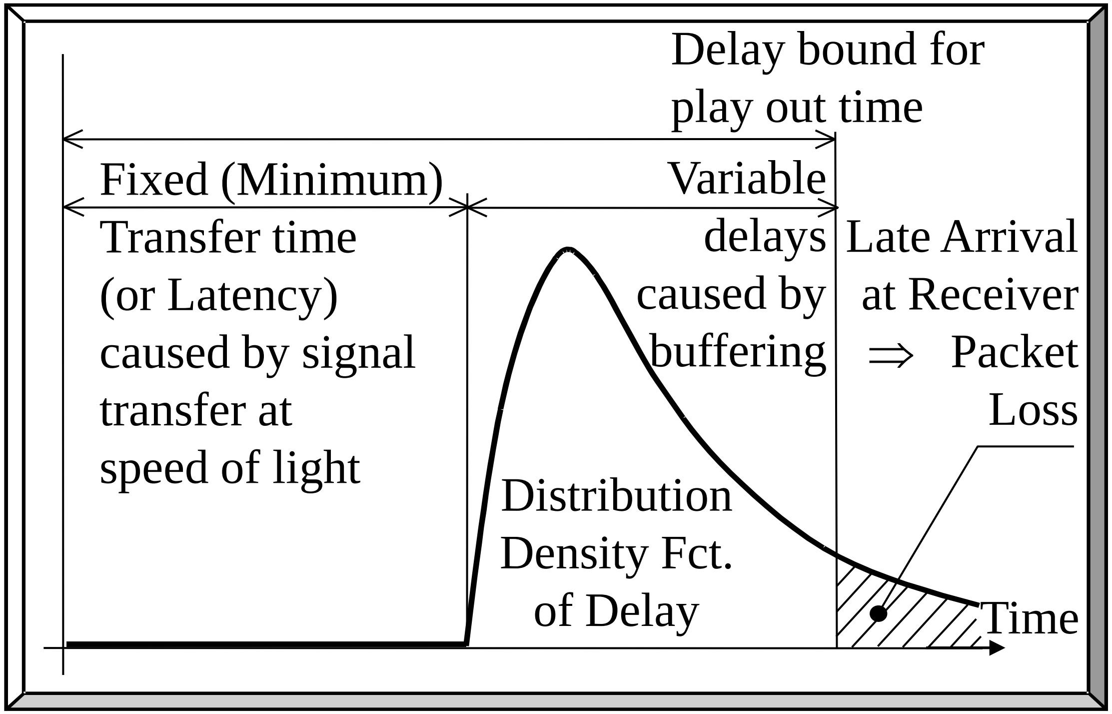
Source: Presentation by G. Kadel, Telekom Innovation Labs

Data Compression: Effect on Traffic Variability & QoS

- Data compression saves bandwidth, but …
- If compression schemes reduce data, e.g. for voice and video codecs (JPEG, MPEG), then the QoE is affected
- The effect of packet loss and bit errors is increased, i.e. spread over a larger portion of data after decoding
- Coding & decoding introduces additional delays
- Compression often turns constant into variable bit rate traffic
  - Voice, data: On-Off traffic pattern
  - Video: compression rate depends on dynamics in scenes
    MPEG introduces periodically changing compression rates over time
- Voice & video coding can adapt to available bandwidth

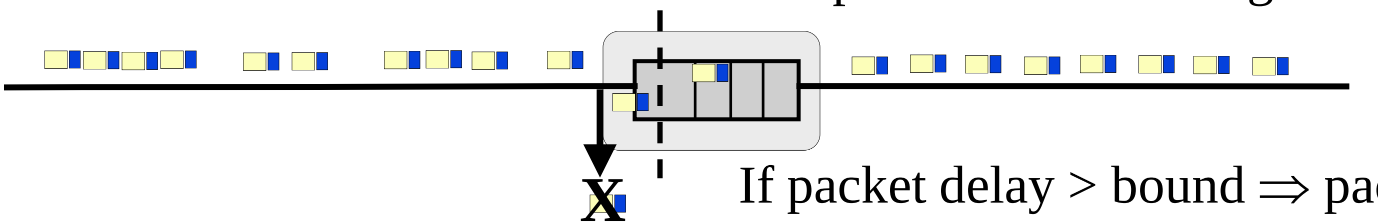
Delay ↔ data loss for real time services & streaming

Components of fixed & variable delay



The probability of a packet missing the delay bound is in fact the packet loss probability

De-Jitter buffer at receiver: Packet is processed at sending time + delay bound



= playout time

If packet delay > bound ⇒ packet is dropped

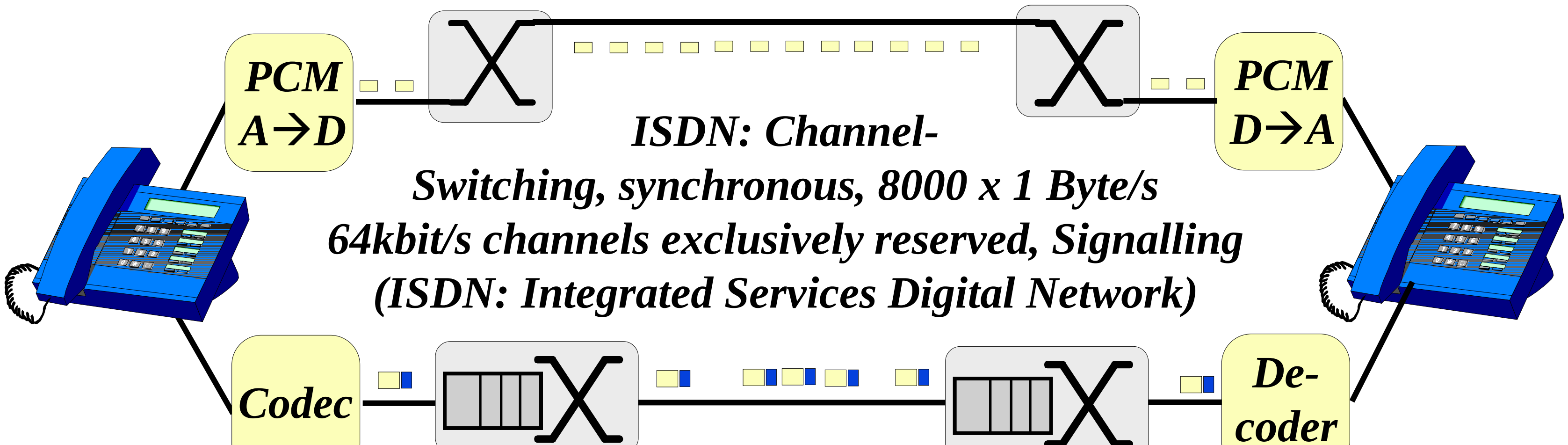
QoS/QoE Demands of Different Web Applications

QoS Criteria & Key Performance Indicators for Application & Service Categories

QoS/QoE* for Real Time Services:
- Influence of Coding, Buffering
- **Voice (ISDN, VoIP, Skype)**
- Video (JPEG, MPEG coding)
- Gaming etc.

*Quality of Experience (QoE): User Perceived Quality

Voice Transmission over ISDN ↔ VoIP
Integrated Services Digital Networks ↔ Voice over IP



VoIP
Paket switching with buffers in switches and end systems
asynchronous transfer; can tolerate delay jitter

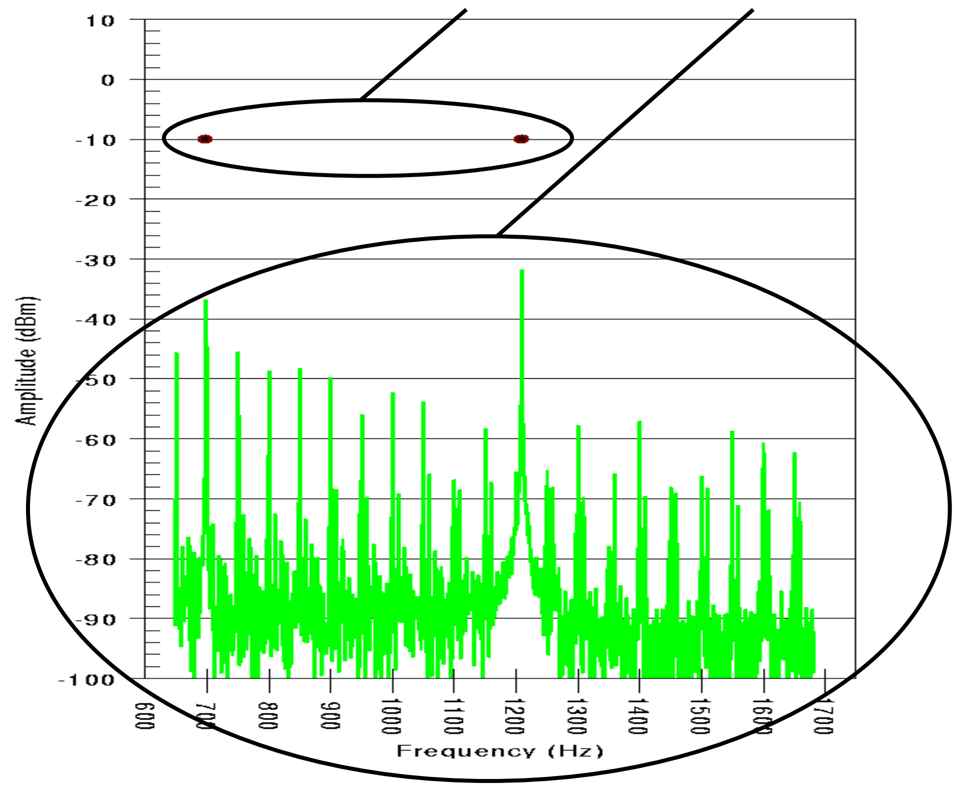
Subjektive & Objektive Voice Quality Measurement

Subjektive Methods are based on individual user experience (MOS*-Scale)

Objektive Methods are based on automated measmt. tools

- Impairments are evaluated (Signal ÷ Noise ratio, delay, jitter)
- A User Experience Model has been standardized to evaluate voice quality depending on impariments

Signal at Sender ↔ Receiver



Example: Spectrum analysis of a signal @ 700 & 1200 Hz with fading & superposed noise on the received side as part of a MOS analysis

*MOS: Mean Opinion Score

Voice QoS Measurement on MOS- or R-Skala

MOS-Scale: ↑
Mean Opinion Score
(ITU P.800)

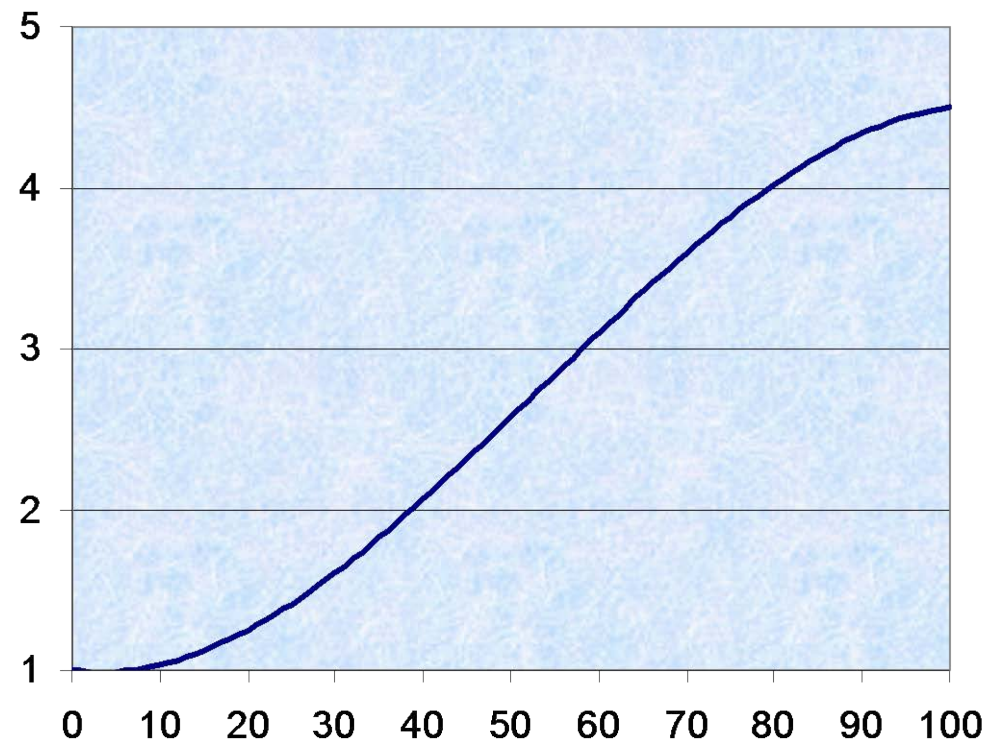
Excellent:

Good:

Sufficient:

Bad:

Unusable:

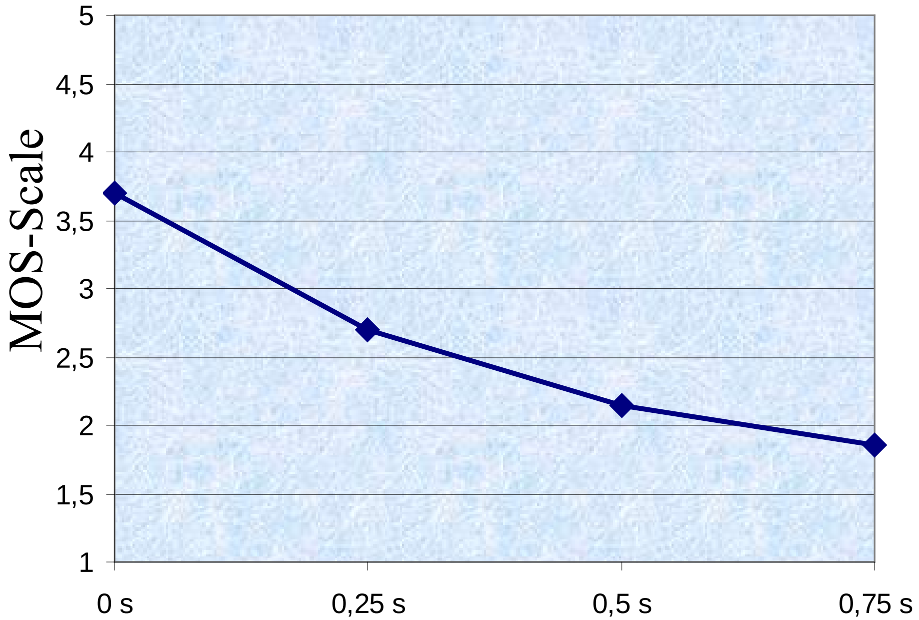


R - Skala (ITU G.107) →

100: No impairments
90: Hardly noticable impariments
80: Acceptable for all users
70: Acceptable for most users
60: Mixed user acceptance
<50: Inacceptable for most users

Source: ITU Standardization

Quality on MOS-(R-)Scale



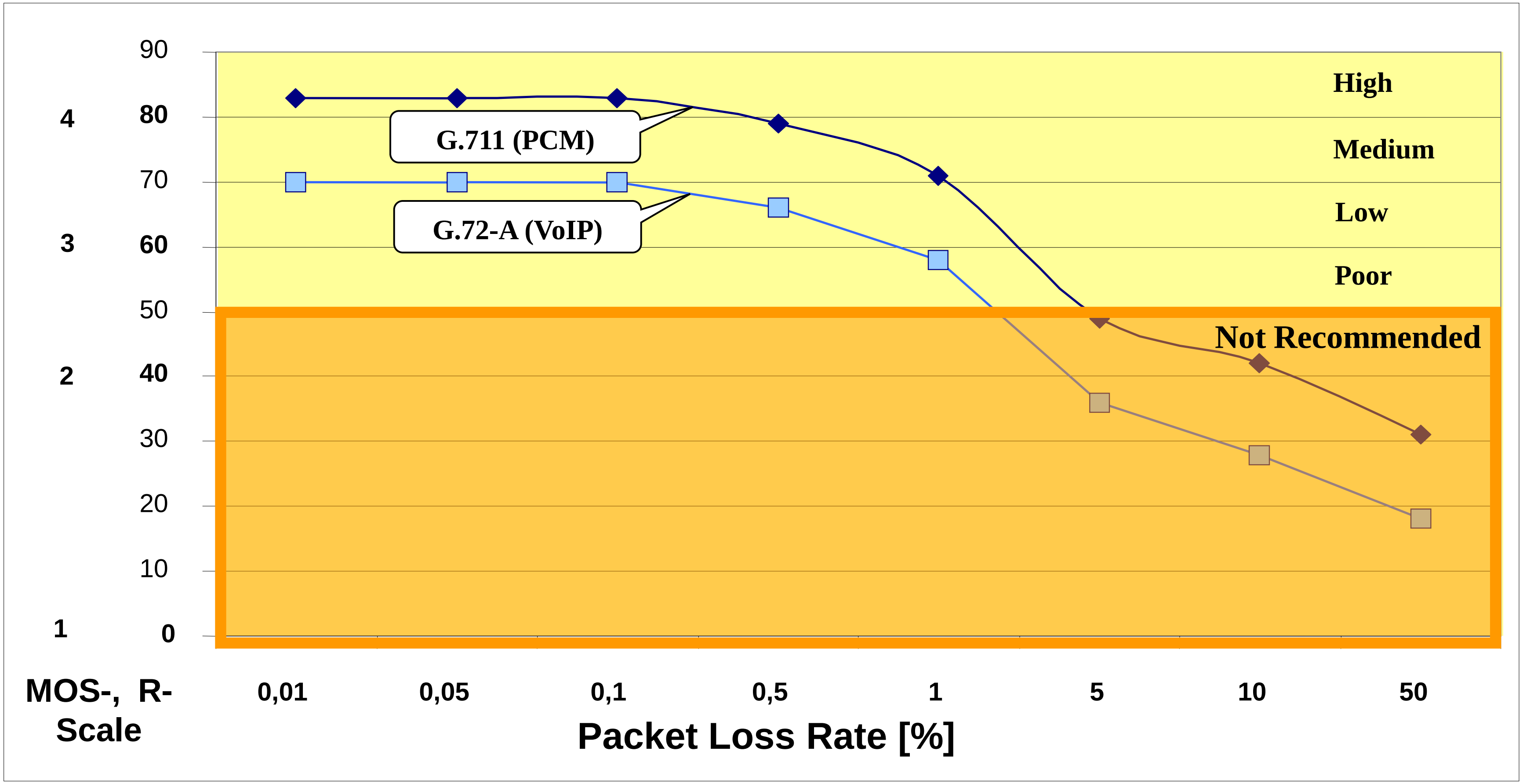
Impact of Delay
on Voice Quality

Examples of Voice Service Quality in Specific Network Environments	R-Scale	MOS-Scale
Local connection between two ISDN phones	94	4,4
Analog voice transmission with 20ms delay	82	4,1
Connection between a mobile and analog access (mobile side)	72	3,7
Connection between a mobile and analog access (analog side)	64	3,3
Voice over IP (G.729A–Coding) with VAD and 2% packet loss	55	2,8

Quality on MOS-/R-scale for
voice service use cases

Source: ITU Standardization

Examples for Voice Quality on the MOS-(R-)Scale: Impact of the Paket Loss Rate



Source: ITU Standardization

Delays & Coding Rates for Standard Voice Coding Methods

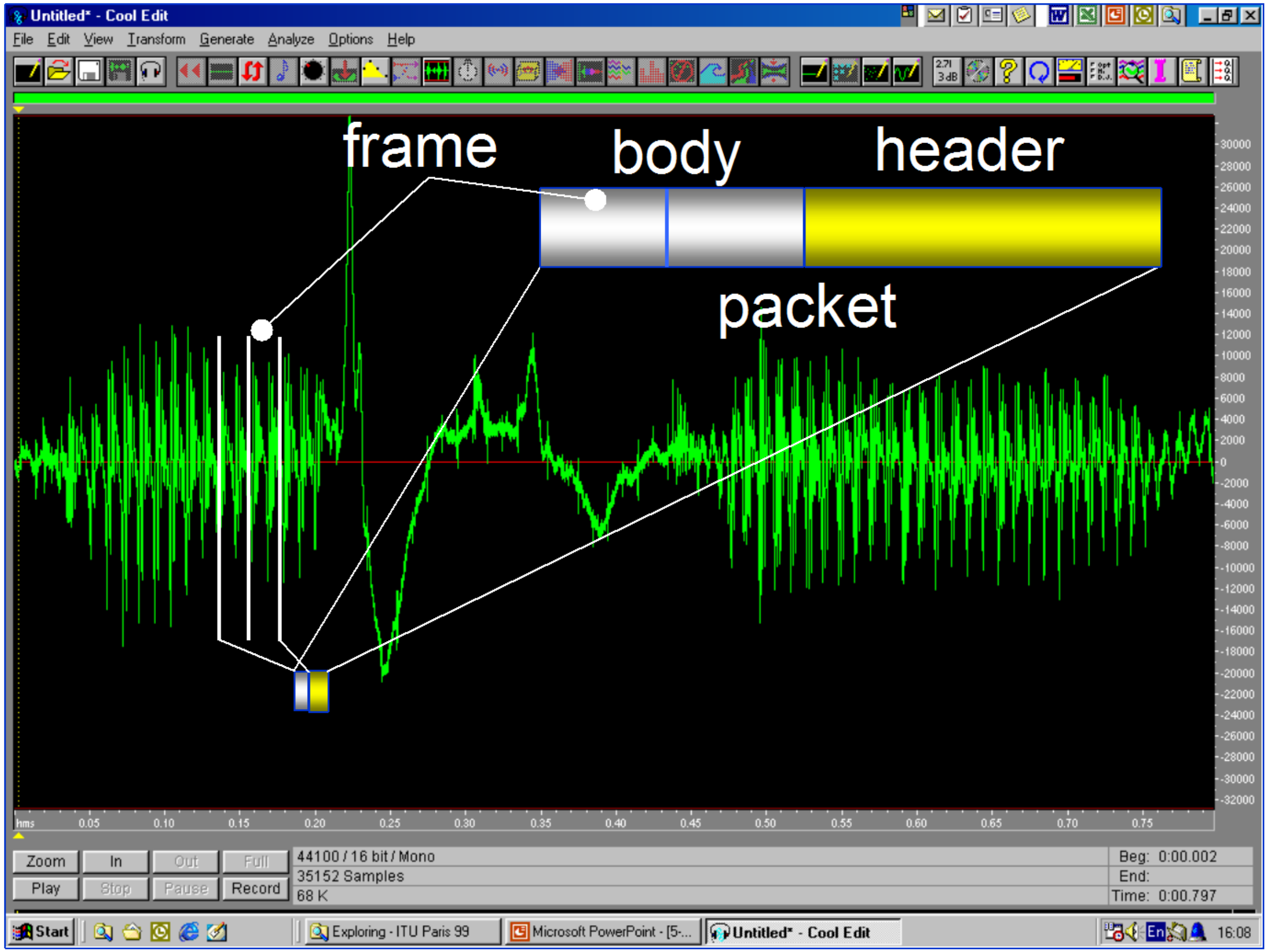
Coder Delay (Sender & Receiver Side)							
Coder Type	**Rate (kb/s)**	**Frame Size (ms)**	**Mean Framing Delay (ms)**	**Look-ahead (ms)**	**Air Inter-face Fra-ming (ms)**	**Mean one-way Delay (ms)**	**Reference to Standard**
PCM	64	0.125	0.25	0		0.25	G.711, G.712
ADPCM	32	0.125	0.25	0	13.625	13.875	G.721, G.726, G.727, DECT
ADPCM	16	0.125	0.25	0		0.25	G.726, G.727
LD-CELP	16	0.625	1.25	0		1.25	G.728
CS-ACELP	8	10	20	5		25	G.729
QCELP	8	20	40	0		40	IS-96-A
RCELP	8	20	40	10		50	IS-127
RPE-LTP	13	20	40	0	35	75	GSM 06.10, Full-rate
VSELP	5.6	20	40	0	35	75	GSM 06.20, Half-rate
MP-MLQ	6.3	30	60	7.5		67.5	G.723.1
ACELP	5.3	30	60	7.5		67.5	G.723.1
PCM: Pulse code modulation; AD: Adaptive differential; CELP: Code excited linear prediction							

Source: ITU Standardization

skype Voice over IP (VoIP)

- Since 2003, the first VoIP platform used on large / global scale even without any special QoS support
- Since 2009, SILK Super **Wideband Audio Codec** is made available via IETF: tools.ietf.org/html/draft-vos-silk-02
  Higher sampling rate >8kHz & broader frequency range
- Codec bit rate: 6 – 40 kbit/s
- Voice activity detection & silence suppression
- Packet loss resilience with Forward Error Correction (FEC) adaptive to bad transmission conditions by more control data
- Peer-to-peer network based with server support during setup

IP packets for voice (Voice over IP ...)



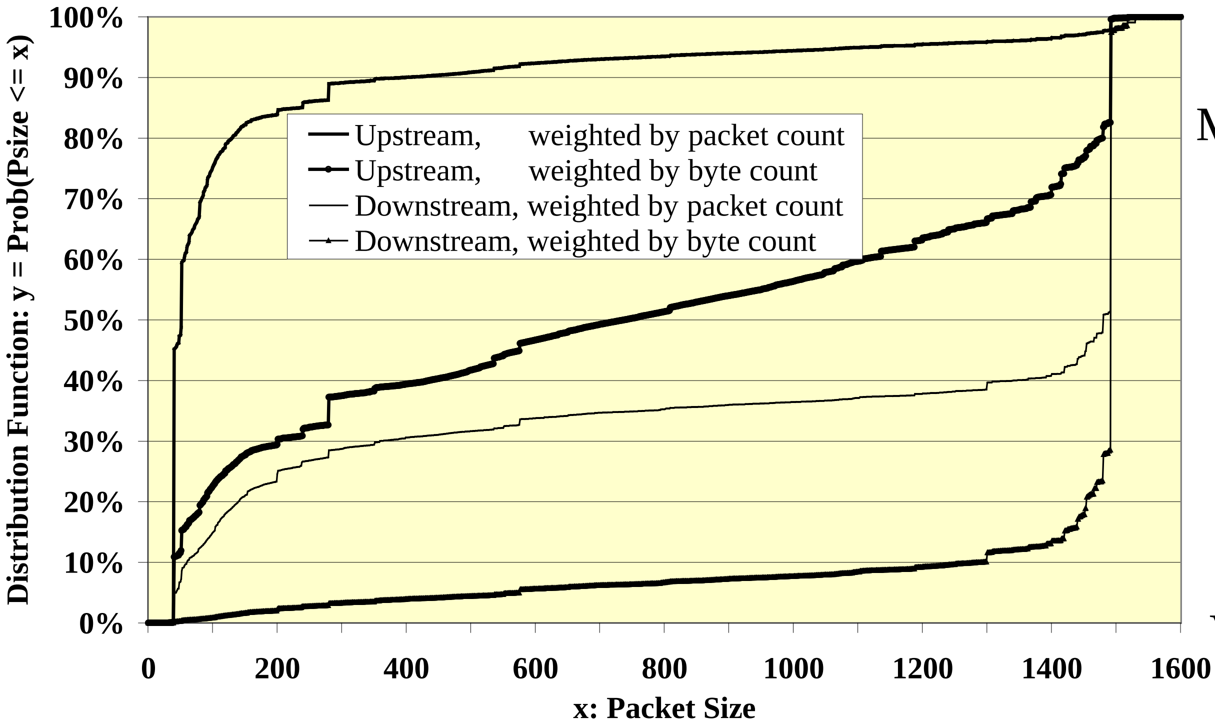
High overhead for VoIP packets

A voice codec generates only a few data bytes in a pecketization period of 20ms

Headers for IP, UDP, RTP , SIP... are required $\Rightarrow$

Overhead in VoIP packets exceeds payload data

Packet Size Distribution in Internet Traffic



Mean IP Packet Size:
~ 500 - 600 Byte

Most frequent packet size:
Upstream:
40, 44 Byte: up to ~50%
Downstream:
~1500 Byte: up to ~50%

Short 40 Byte Packets are prevalently TCP Acknowledgements
Voice Packets <300 Byte

Most IP traffic is transported in 1400 - 1500 Byte packets

Packet size for voice transmission

- Historic criterion: Low delay on access links
  A1500 Byte paket requires 0.1875s on a 64kbit/s ISDN access link
  ITU recommendation: <0.02 s delay per network element for real time data
  → solved by Byte-wise transmission via ISDN and SDH networks
- ATM Networks: Fixed packet (or cell) size of 53 Byte to undercut 0.02 s
  → Broadband access (> 1Mbit/s) now achieves 0.12 s delay for 1500 Byte
- Router data throughput [in Gb/s or Tb/s]
  → is proportional to packet size; Terabit routers go to their performance limits for bursts of small packets; energy consumption is affected
- Voice packets stay small:
  A usual 16 or 64 kbit/s codec generates only 40 or 160 Byte in 0.02s,
  → VoIP packets are ≤ 200 Byte
  → high overhead, but still not a severe problem, because VoIP generates < 5% of the total IP traffic with decreasing tendency

QoS/QoE Demands of Different Web Applications

QoS Criteria & Key Performance Indicators for Application & Service Categories

QoS/QoE* for Real Time Services:

- Influence of Coding, Buffering
- Voice (ISDN, VoIP, Skype)
- **Video (JPEG, MPEG coding)**
- Gaming etc.

*Quality of Experience (QoE): User Perceived Quality

JPEG & MPEG: Picture- & Video-Coding

JPEG: Compression of single pictures (Joint Photographic Expert Group)

- Transforms pixel data of 3 Byte per pixel for color or gray level) in blocks (8 x 8, 16 x 16) via discrete Cosinus Transform;
- JPEG is lossy depending on flexible compression rate (pixel resolution & Transform coefficient representation is variabel; Huffman coding is involved
- Decoding & retransformation on receiver side

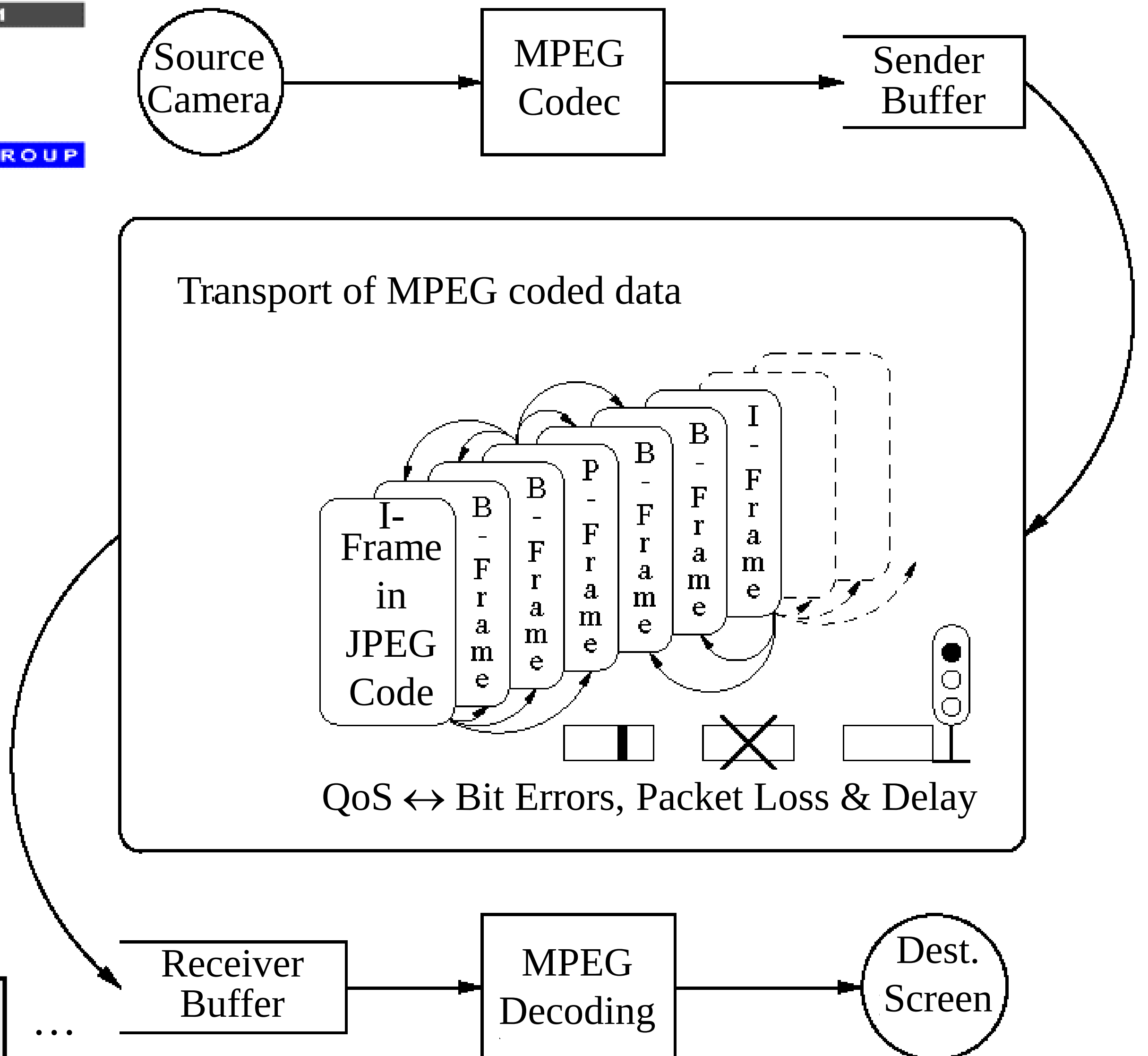
MPEG: Video compression exploits redundancy in a sequence of pictures (Moving Pictures Expert Group)

- 25, 30 or more pictures (frames) per secund due to television standards
- 3 frame typs: I-(Intracoded), P-(Predicted) & B-(Bidirectional) frames
- I-Frames: A picture is coded by JPEG without relationship to other pictures;
- P-Frames: The difference with regard to the last P-(I-)Frame is coded incl. pointers to blocks with only small differences;
- B-Frames: Difference to interpolation between previous & next P-(I-)frame;
- Periodical Group-of-Pictures(GoP)-Sequence z.B.: <u>IBBPBBPBB</u> IBBP...

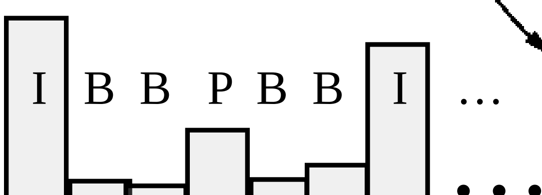
ISO/IEC JTC1/SC29 WG11
MPEG
MOVING PICTURE EXPERTS GROUP

MPEG-Video-Transfers: structured into Groups of Pictures (GoP) and Scenes

<www.mpeg.org>



Size of MPEG Frames



QoS for Video Services

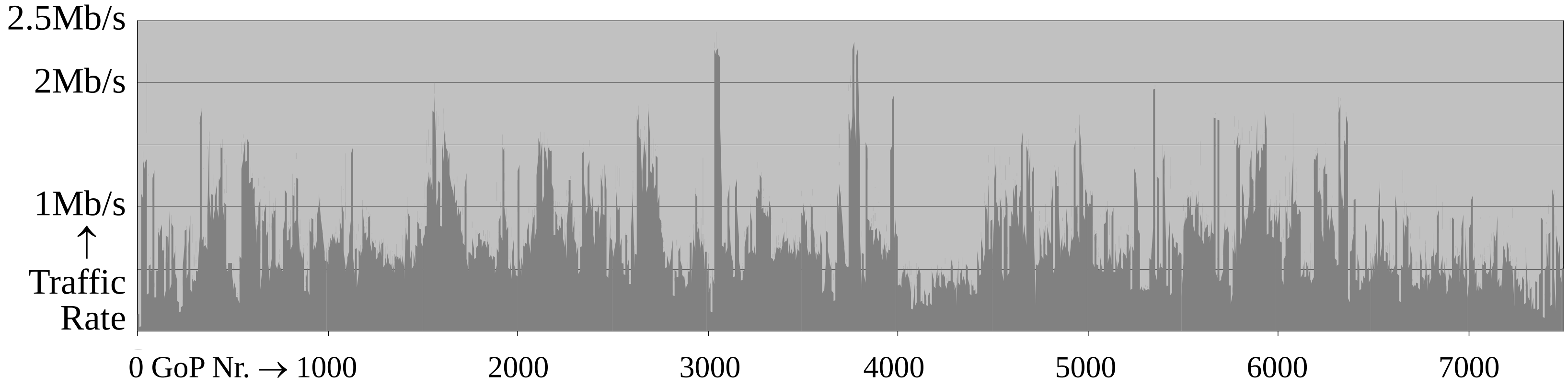
Main MPEG-Links & Standards

- <www.mpeg.org> MPEG-1, -2, -4, -7, A – E; dazu H.261 – H.265 bei ITU incl. video, audio & multimedia environment

MPEG QoS

- MOS scale (1-5) for video and audio measurement to be done by test persons standardization of measurement equipment like for voice QoS is under study
- Video conferencing & interactive video: Strict real time constraints (~ 0.2s maximum delay & synchronization of video / audio)
- Video on demand (VoD): Relaxed demands (some seconds delay, if not live)
- MPEG introduces ~ 0.1s delay for reordering B-frames with next P-(I-)frame
- Transmission error tolerance depends on compression
  Pixel error rate $<10^{-3}$ is hardly noticeable in uncompressed video
- Video is a broadband application
  Higher resolution is improving QoE up to data rates > 1Gbit/s per stream;
  Backbone capacities restrict most subscribers to a few Mbit/s (nowadays)

MPEG Video:
Variable Rate Coding e.g. for Jurassic Park



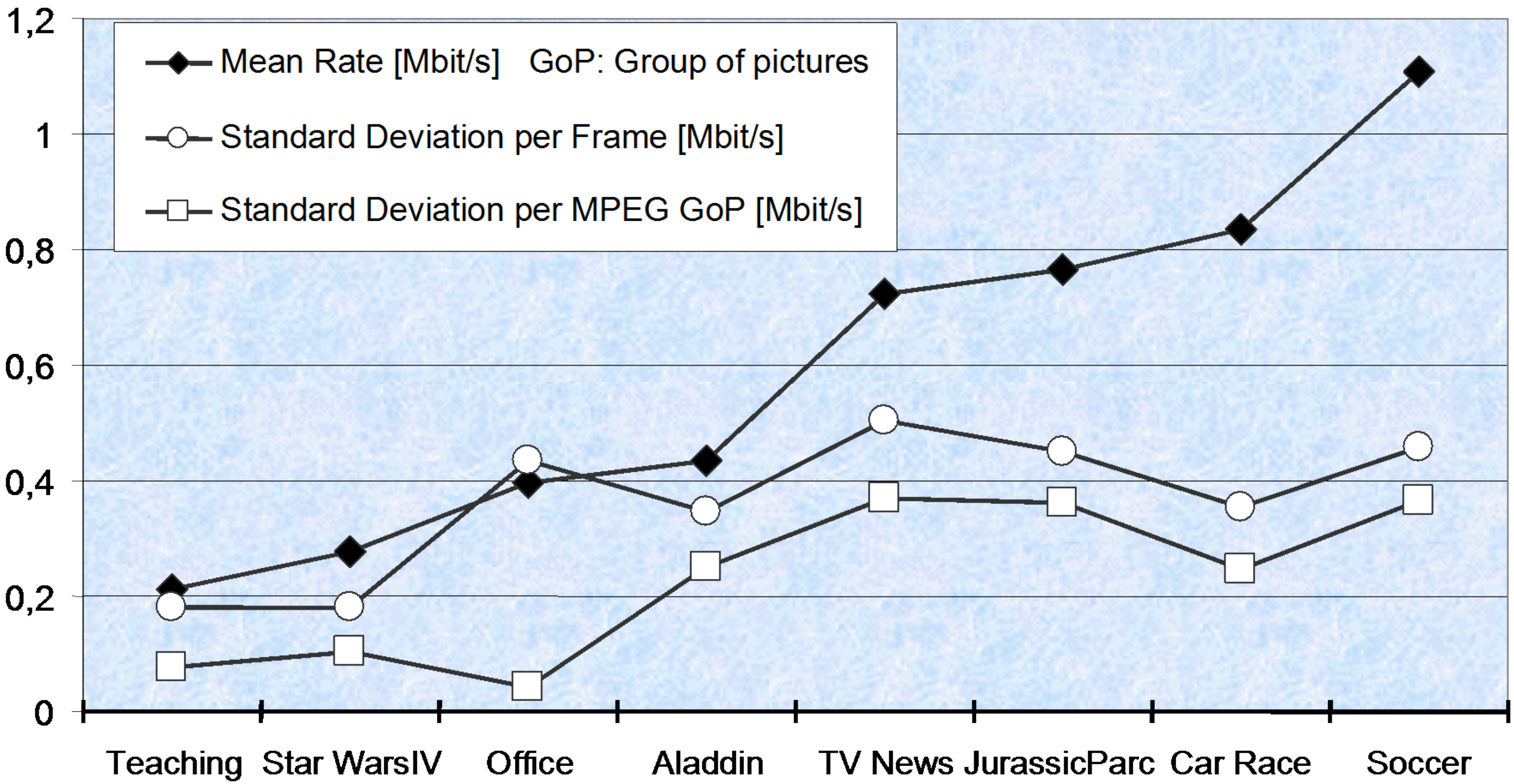
Traffic rate is varying depending on the size of GoPs

GoP: Group of Pictures includes 12 Frames (IBBPBBPBBPBB)

Frame rate: 25 frames/s; 7500 frames last for about 1 hour

Source: F. Fitzek and M. Reisslein, MPEG-4 and H.263 video traces for network performance evaluation, IEEE Network (Nov. 2001) 40-54; Traces: http://trace.eas.asu.edu

Mean Traffic Rate & Standard Deviation for 8 Video-Examples: Largely different depending on movement intensity



Source: http://trace.eas.asu.edu

Real Time Video Properties & Traffic Characteristics

MPEG generates bit rate (VBR) traffic

- The entropy for representing a picture varies over time (i.e. JPEG code size)
- The ratio of the size of P-(B-)frames compared to I-frames varies depending on the motion in a scene (fast motion: ~ 1; slow or no motion < 0.1)

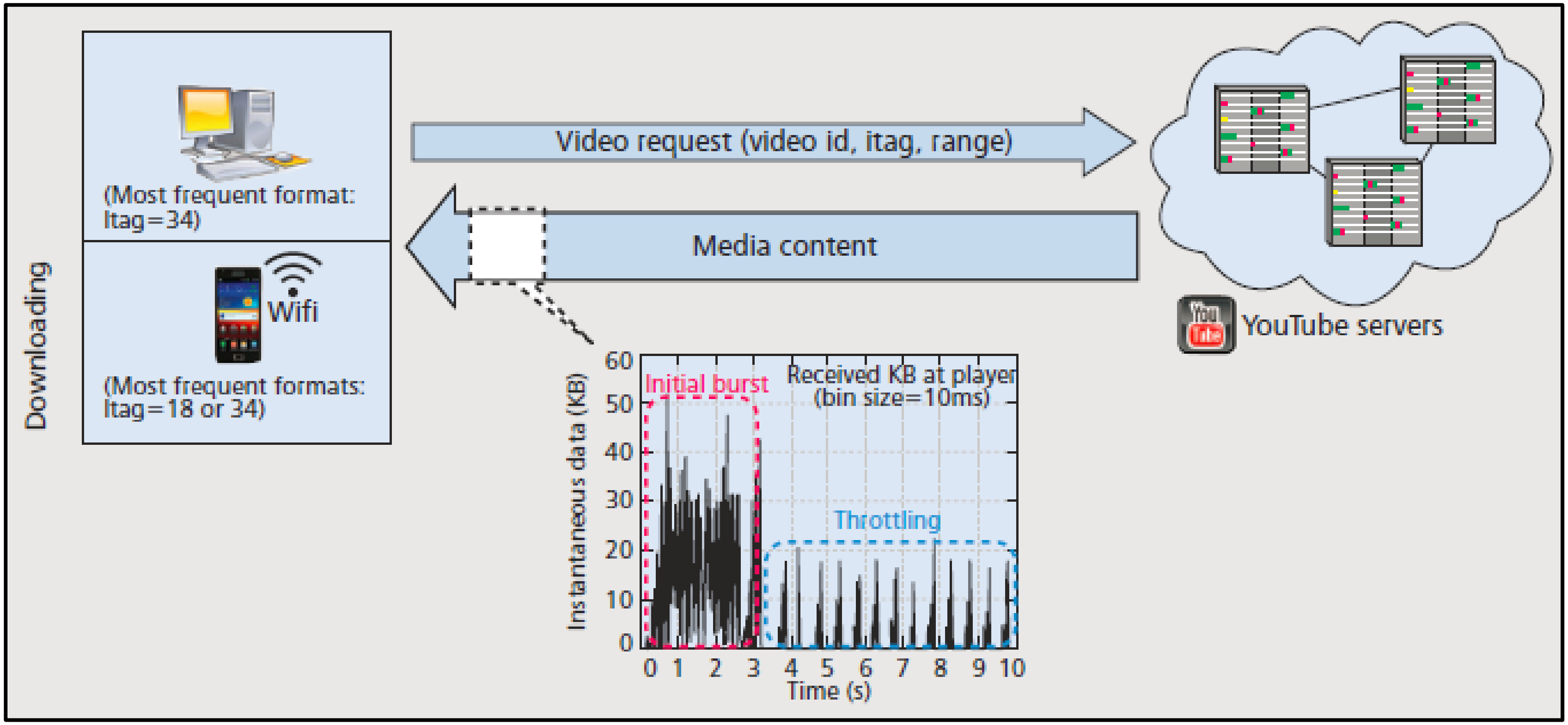
MPEG provides variable compression level & allows for adaptive rate control

- Coarse ↔ fine resolution of pixels; more & less significant JPEG info. parts
- MPEG coding can adapt to transmission capacity, e.g. for mobile devices
- Classification of data for multi-level quality e.g. into DiffServ code points

Distribution of traffic rate & long term correlation

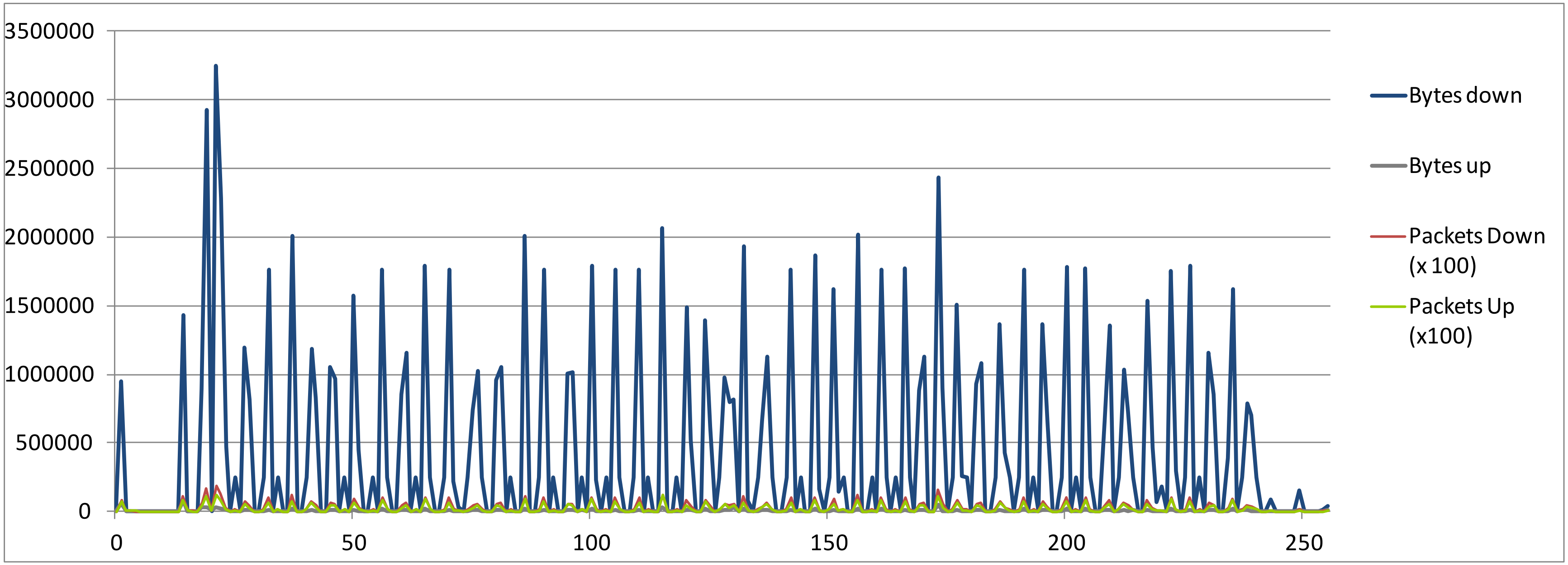
- Various shapes of rate distribution: Log-normal; dual peak & other examples
- High frame burstiness; smaller coefficient of variation for complete GoPs
- Intra-GoP periodicity (time scale ~ 0.5s)
- Strong exponentially decreasing short range dependency (time scale < 10s)
- Long range dependency within scenes (time scale: > 3s)
  What is a scene? E.g. new scene indicated by significant change in GoP size?

YouTube Video Streaming:
Traffic Profile & Adaptation to End Devices



Source: J.J. Ramos-Munoz et al.: Characteristics of Mobile YouTube Traffic
IEEE Wireless Communications, Feb. 2014 pp. 18-25

YouTube Video Streaming: Traffic Profile from Measurement in DT Network



Stream characteristics (measurement in Oct. 2014):

Duration: 1328 s; Downstream Rate: 0.69 Mb/s; Upstream Rate: 0.09 Mb/s;

Traffic profile shows the byte count in 256 intervals of length 5.2 s;

High initial transfer volume and alternating load/stop phases are visible

Factors relevant for QoE for video streaming & conf.

- Encoding rate (evt. adaptive to network conditions)
- Start time / time shift for live videos
- Initial buffer filling rate / Average download rate
- Viewing quality (Resolution n x m pixels, frame frequency, QoE degradations: artifacts in pictures, jumps in motion progress, stallings)
- Audio quality
- Audio / video synchronisation
- Enhanced synchronization for 3D and other special video applications

QoS/QoE Demands of Different Web Applications

QoS Criteria & Key Performance Indicators for Application & Service Categories

QoS/QoE* for Real Time Services:

- Influence of Coding, Buffering
- Voice (ISDN, VoIP, Skype)
- Video (JPEG, MPEG coding)
- **Gaming etc.**

*Quality of Experience (QoE): User Perceived Quality

Extreme QoS/QoE Demands: Gaming, Thin Clients, …

Online Multi-User Games: >10ms of end-to-end delay mean a measurable handicap, if fast interactive reaction is relevant

Thin Clients & ***Cloud Computing***

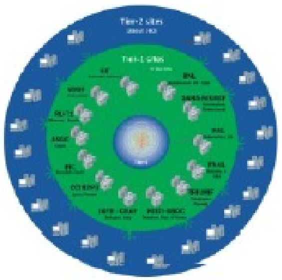
- PC functions are shifted to servers in a networks → cheaper terminals
  → Those thin clients have stringent QoS demands within the network
- Office functions are sensitiv already for small delays (< 0.1s)
  e.g. text editing with remote mouse/cursor control is similar to gaming …
- The availability of the network is also critical
- Subjective and objective tests to determine key performance parameters for QoS/QoE of thin clinets have been studies in recent work similar to QoS modelling for voice in ITU standardization

  Source: B. Staehle et al., *Quantifying the Influence of Network Conditions on the Service Quality Experienced by a Thin Client User* MMB Konf. 2008, Dortmund, Springer LNCS

Machine-to-Machine Control Applications (Industry 4.0; Car-to-car etc.)

Virtual / Augmented Reality Applications

CERN: High Energy Physics Workload in Distributed Cloud



Tier-0 (CERN):
- Data recording
- Initial data reconstruction
- Data distribution

Tier-1 (13 centres):
- Permanent storage
- Re-processing
- Analysis

Tier-2 (64 Federations, ~160 centres):
- Simulation
- End-user analysis

- 540,000 cores
- 450 PB

Run 1/LS 1

~1 PB/sec

25-30 PB/year

0.5 M cores

0.5 EB storage

2 M jobs/day

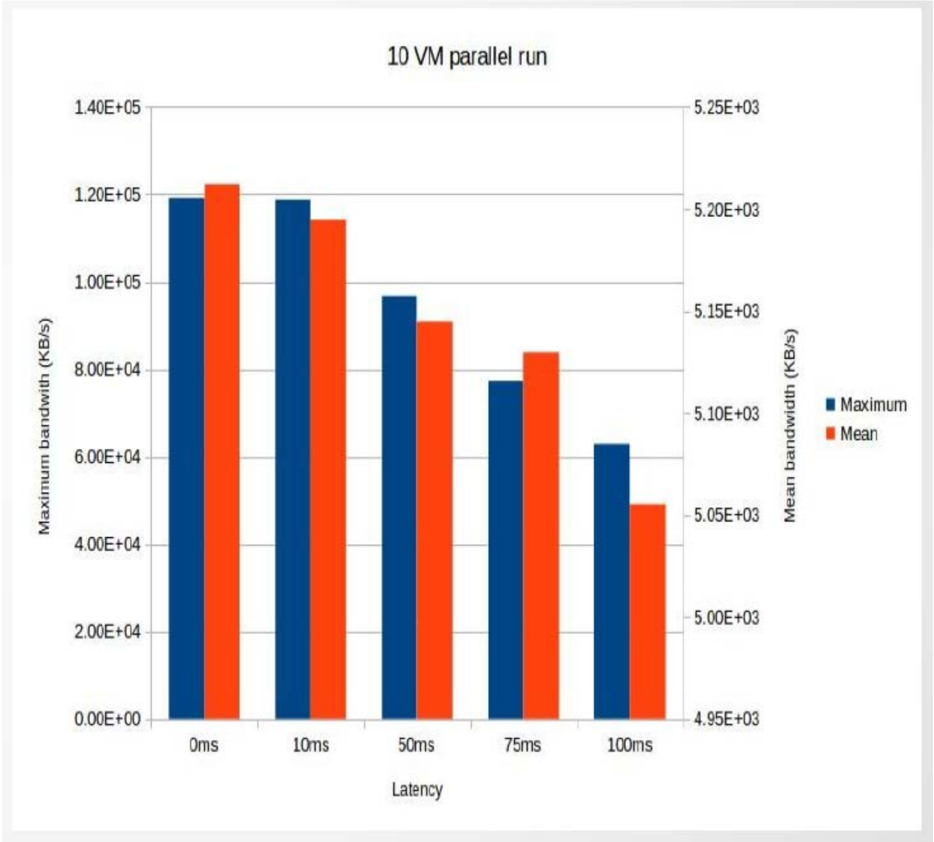
24/7 operations

CERN: High Energy Physics Workload in Distributed Cloud <home.cern>, <opennebula.org>

CERN cloud has massive computing demands with high data volumes required per computation job

Computation is distributed over a global cloud with most data centers in Europe

Data Throughput is affected already by small delays:



Large Scale Bandwidth & QoS Tests

- Tests of available bandwidth for Internet access by regulator BNetzA <www.initiative-netzqualitaet.de>
  First reports published in April 2013

Initiative Netzqualität

Der Breitband-Test

- Tests of the Connect journal since ~2009 on VoIP, HTTP-, FTP-Transfers etc.
  <www.connect.de> → Test Netzbetreiber →
- IETF Standard. Working Group on Large-scale MeAsurement of broadband Performance (LMAP)
- Network providers get precise feedback on QoS

TESTERGEBNISSE

ANBIETER	
ERGEBNISSE SPRACHE	
Rufaufbauzeit	(max. 30)
erfolgreiche Gespräche	(max. 120)
Sprachqualität	(max. 60)
Sprachlaufzeit	(max. 60)
SUMME SPRACHE	**(MAX. 270)**
HTTP-Messungen	(max. 45)
HTTP mit parallelem Upload	(max. 45)
FTP-Messungen	max. 45)
FTP mit parallelem Upload	(max. 45)
Website-Benchmarking	(max. 50)
SUMME DATEN	**(MAX. 230)**
connect URTEIL	max. 500

Conclusions: QoS demands for main IP services

- VoIP: 64kbit/s bandwidth; delay < 0.1-0.2s; < 10% packet & bit errors (depending on codec; high header overhead ↔ low mobile bandwidth)
- Video streaming: high bandwidth demand up to 1Gb/s; preserve time sequence between sender ↔ receiver; many QoE impact factors codecs with buffering and compression scheme are relevant for QoE
- Video conference: Combines demands of VoIP and video
- Audio: <1Mb/s bandwidth; synchronization with video
- Interactive and background data transfers:
  delay-tolerant elastic traffic; often high bandwidth;
  No error tolerance in data; enforced by TCP or appl. layer protocols
- Applications with strict delay tolerance: Online multi-user gaming (< 0.01s); remote desktop fct.; M2M e.g. car-to-car control