International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.1080

(12/2008)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

Multimedia quality of service and performance – Generic and user-related aspects

---

Quality of experience requirements for IPTV services

Recommendation ITU-T G.1080

ITU-T

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	**G.1000–G.1999**
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000–G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000–G.8999
ACCESS NETWORKS	G.9000–G.9999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T G.1080

Quality of experience requirements for IPTV services

Summary

Recommendation ITU-T G.1080 defines user requirements for quality of experience (QoE) for Internet protocol television (IPTV) services. The QoE requirements are defined from an end user perspective and are agnostic to network deployment architectures and transport protocols. The QoE requirements are specified as end-to-end and information is provided on how they influence network transport and application layer behaviour. QoE requirements for video, audio, text, graphics, control functions and meta-data are provided. Compression coding schemes addressed in this Recommendation are examples, and detailed numeric values as performance targets, e.g., bit rate, packet loss rate, are also examples. The readers may appropriately choose or replace these parameter values in order to be consistent with the requirements of each IPTV service context to which they are intended to be applied.

Source

Recommendation ITU-T G.1080 was approved on 7 December 2008 by ITU-T Study Group 12 (2009-2012) under Recommendation ITU-T A.8 procedure.

Keywords

IPTV, QoE, QoS.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

CONTENTS

Page

Recommendation ITU-T G.1080

Quality of experience requirements for IPTV services

1 Scope

This Recommendation defines user requirements for quality of experience (QoE) for IPTV services. The QoE requirements are defined from an end-user perspective and are agnostic to network deployment architectures and transport protocols. The QoE requirements are specified for the end-to-end service and information is provided on how they influence network transport and application layer behaviour. Compression coding schemes addressed in this Recommendation are examples, and detailed numeric values given as performance targets, e.g., bit rate, packet loss rate, are also examples. The readers may appropriately choose or replace these parameter values in order to be consistent with the requirements of each IPTV service context to which they are intended to be applied.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T E.800] Recommendation ITU-T E.800 (1994), *Terms and definitions related to quality of service and network performance including dependability*.

[ITU-T F.700] Recommendation ITU-T F.700 (2000), *Framework Recommendation for multimedia services*.

[ITU-T H.262] Recommendation ITU-T H.262 (2000) | ISO/IEC 13818-2:2000, *Information technology – Generic coding of moving pictures and associated audio information: Video*.

[ITU-T H.264] Recommendation ITU-T H.264 (2005), *Advanced video coding for generic audiovisual services*.

[ITU-T P.10/Amd.1] Recommendation ITU-T P.10/G.100 (2006), Amd.1 (2007), *New Appendix I – Definition of Quality of Experience (QoE)*.

[ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.

[ITU-T T.140] Recommendation ITU-T T.140 (1998), *Protocol for multimedia application text conversation*.

[ITU-T Y.1541] Recommendation ITU-T Y.1541 (2006), *Network performance objectives for IP-based services*.

[ITU-R BT.500-11] Recommendation ITU-R BT.500-11 (2002), *Methodology for the subjective assessment of the quality of television pictures*.

[ITU-R BT.601-6] Recommendation ITU-R BT.601-6 (2007), *Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios*.

[ITU-R BT.1359-1] Recommendation ITU-R BT.1359-1 (1998), *Relative timing of sound and vision for broadcasting*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

3.1.1 Quality of experience (QoE) [ITU-T P.10/Amd.1]: The overall acceptability of an application or service, as perceived subjectively by the end-user.

NOTE 1 – Quality of Experience includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc.).

NOTE 2 – Overall acceptability may be influenced by user expectations and context.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 channel zapping: The act of quickly changing from one channel to another.

3.2.2 clean audio: Audio track of an IPTV service with background sounds removed.

3.2.3 group of pictures: The group of pictures (GOP) is a group of successive pictures within a MPEG-coded film and/or video stream. Each MPEG-coded film and/or video stream consists of successive GOPs. From the MPEG pictures contained in it, the visible frames are generated.

3.2.4 triple play services: Services that include IPTV, VoIP, and Internet access.

3.2.5 VoD trick modes: Download and streaming video on demand (VoD) systems provide the user with a large subset of VCR functionality including pause, fast forward, fast rewind, slow forward, slow rewind, jump to previous/future frame, etc. These functions are usually referred to as "trick modes".

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

A/V	Audio Video
AAC	Advanced Audio Coding
AC-3	Dolby digital audio (Advanced Codec 3)
ARQ	Automatic Repeat reQuest
AVC	Advanced Video Codec
AVS	Audio and Video coding Standard (Chinese)
BER	Bit Error Rate
BML	Broadcast Markup Language
CBR	Constant Bit Rate
CPU	Central Processing Unit
DCT	Discrete Cosine Transform
DSCQS	Double Stimulus Continuous Quality Scale
DSL	Digital Subscriber Line
DVB	Digital Video Broadcast

DVD	Digital Video Disk
ECG	Electronic Content Guide
EPG	Electronic Program Guide
FEC	Forward Error Correction
FFW	Fast ForWard
fps	frames per second
GOP	Group of Pictures
GWR	GateWay Router
HDTV	High Definition TeleVision
HG	Home Gateway
HTML	HyperText Markup Language
IGMP	Internet Group Management Protocol
IGP	Interactive Gateway Protocol
IP	Internet Protocol
IPG	Interactive Program Guide
IPTV	Internet Protocol TeleVision
MOS	Mean Opinion Score
MP	Measured Point
MP3	MPEG-1 audio layer 3
MPEG	Moving Pictures Expert Group
MPLS	MultiProtocol Label Switching
NICAM	Near Instantaneous Companded Audio Multiplex
PAL	Phase Alternating Line
PC	Personal Computer
PDV	Packet Delay Variation
PHB	Per-Hop Behaviour
PLR	Packet Loss Ratio
PTD	Packet Transfer Delay
QoE	Quality of Experience
QoS	Quality of Service
RFC	Request For Comments
SDH	Synchronous Digital Hierarchy
SDTV	Standard Definition TeleVision
SECAM	SEquentiel Couleur A Mémoire (sequential color with memory)
SMPTE	Society of Motion Picture and Television Engineers
SONET	Synchronous Optical NETwork
STB	Set-Top Box

TS	Transport Stream
VBR	Variable Bit Rate
VCR	Video Cassette Recorder
VoD	Video on Demand
VoIP	Voice over IP

5 Introduction to QoE

QoE is defined in [ITU-T P.10/Amd.1] as the overall acceptability of an application or service, as perceived subjectively by the end-user. It includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc.) and may be influenced by user expectations and context. Hence in principle, QoE is measured subjectively by the end-user and may differ from one user to the other. However, it is often estimated using objective measurements.

Differences in perceptual acuity and preference mean that QoE judgments obtained from different people may vary. Therefore, measurements of QoE are generally made using group data. Where the necessary studies have been done to calibrate the relationship with QoE, it may also be estimated using objective measurements.

A number of system performance characteristics contribute to QoE of the media stream. For example, the codec and the encoding bit rate used, media resolution in the source and at the display, corruption or loss of information, and delay. Interactions among video content, the codec and bit rate used, and the specific bits corrupted and/or packets lost contribute to a high variability in the perceived quality of the video output.

There are additional factors that can influence the viewer's response. Some of these affect the perception of quality, such as the context of the judgment (a particular image will be rated one way in the context of standard definition TV (SDTV), another in the context of high definition TV (HDTV), and still differently in the context of a video clip on Internet), cultural background, motivation, attention related factors, emotional state, and so on. (Direct evaluations of QoE are designed to exclude these factors, since they are not generally under the control of a network operator and so do not contribute to equipment requirements.)

Other factors influence the viewer's judgment of acceptability. These include things like previous experience with the specific communication mode or related modes (for instance, experience with DVD quality will influence how acceptable one finds IPTV or VoD), how much one is paying for the service, and what special benefits the service provides (mobility, time independence, exceptionally large program library), and so on.

Acceptability is not equivalent to QoE. A low resolution video image will have a lower QoE than a high resolution image, but it may be completely acceptable for certain applications and services, depending on the end device, the physical size of the display, and the purpose for which it is being used.

Figure 5-1 shows factors contributing to QoE. These factors are organized as those related to quality of service and those that can be classified as human components.

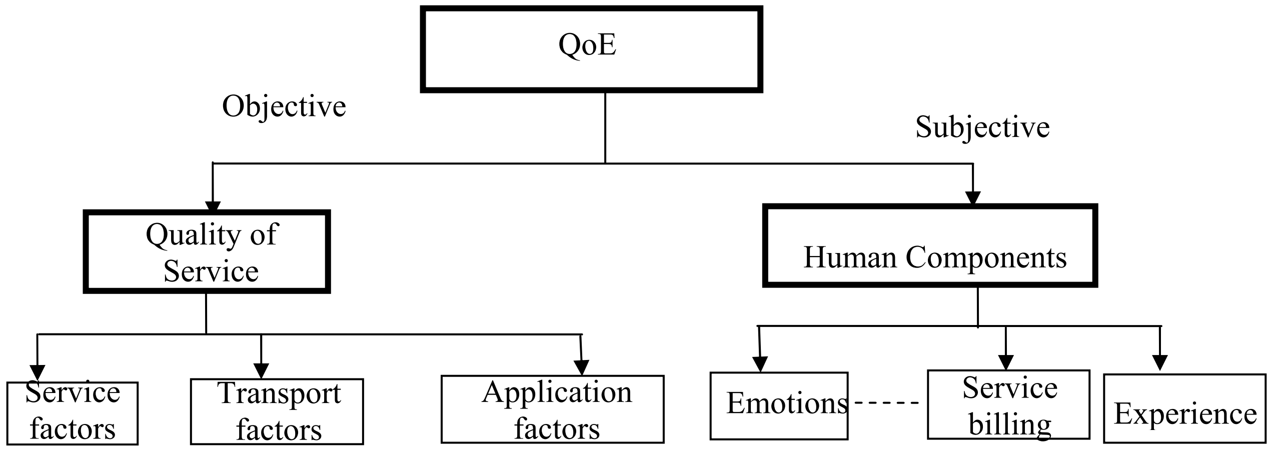


Figure 5-1 – QoE dimension

QoE for video is often measured via carefully controlled subjective tests [ITU-R BT.500-11], [ITU-T P.800], where video samples are played to viewers, who are asked to rate them on a scale typically consisting of five points. The ratings assigned to each case are averaged together to yield a mean rating or mean opinion score (MOS).

Quality of service (QoS, see [ITU-T E.800]) involves the totality of characteristics of a telecommunication service that bear on its ability to satisfy stated and implied needs of the user of the service. In general, network performance is a major component of QoS, and so network mechanisms for QoS are an important consideration. QoS mechanisms include any mechanism that contributes to the improvement of the overall performance of the system and hence to improving end-user experience. QoS mechanisms can be implemented at different levels. For example, at the network level it includes traffic management mechanisms such as buffering and scheduling employed to differentiate between traffic belonging to different applications. QoS mechanisms at levels other than the transport include loss concealment, application forward error correction (FEC), etc.

Related to QoS are the QoS performance parameters. Similarly to the QoS mechanisms, QoS parameters can be defined for different layers. At the network layer, those parameters usually include packet loss rate, delay and delay variation.

Note that in Figure 5-1, service billing relates to the "value for money" perceived by the user for the particular service.

Typically, there will be multiple service level performance (QoS) metrics that impact overall QoE. There are a number of service level performance characteristics (that is, objective parameters of service performance such as encoding bit rate, packet loss, delay, availability, etc.) that affect QoE. In general, these are correlated with QoE as measured by the MOS. The relation between QoE and service performance (QoS) metrics is typically estimated empirically. Once identified, the QoE/QoS relationship can be used in two ways:

a) Given a QoS measurement, one could in principle predict the expected QoE for a user, with appropriate assumptions.

b) Given a target QoE for a user, one could in principle deduce the net required service layer performance, with appropriate assumptions.

To ensure that the appropriate service quality is delivered, QoE targets should be established for each service and be included early on in system design and engineering processes where they are translated into objective service level performance metrics. Quality of experience will be an important factor in the marketplace success of triple-play services and is expected to be a key differentiator with respect to competing service offerings. Subscribers to network services do not care how service quality is achieved. What matters to them is how well a service meets their expectations for effectiveness, operability, availability, and ease of use.

6 QoE for video and audio

QoE requirements for video and audio may be based on QoE scales such as the mean opinion score (MOS) and double stimulus continuous quality scale (DSCQS) [ITU-R BT.500-11]. However, conducting subjective tests is difficult because they are time-consuming and expensive. Moreover, reliable objective quality assessment methods have not been established for transmitted video and audio. Therefore, this clause provides provisional QoE requirements on the basis of the objective parameters that are correlated to the subjective QoE.

This Recommendation addresses QoE targets and shows how to express QoE requirements in the context of numerical parameters such as bit rate or packet loss rate. The process of determining QoE performance targets must consider a number of issues, for example: the purpose of the IPTV service, QoE level of the current broadcasting systems (which sets user expectation), compression coding scheme to be used for the service, content characteristics, content provider requirements, customer satisfactions. While the requirement values shown in the tables in this clause are NOT generically applicable to any specific or all IPTV services, they are to be understood as provisional values which are subject to change. Readers of this Recommendation are invited to replace the numerical values shown in the tables in this clause with appropriate ones that conform to the requirements required by a specific IPTV service context.

6.1 Requirements for media compression and synchronization

One of the main components of QoE for video and audio is digitization and compression of video and audio source materials and the various settings and parameters selected. Since video compression schemes such as those defined by the moving pictures expert group (MPEG) are not lossless and an identical copy of the original source material cannot be recovered, there are potentially negative impacts on video picture quality and therefore on viewer QoE. The main factors influencing video QoE at the application layer due to compression are:

- Quality of source material.
  - The quality of the delivered media depends on the quality of the source material.
- The baseline quality (no network impairments) of the codec.

NOTE 1 – A partial list of video codecs is provided in Appendix II.

- Resolution.
  - Some systems reduce the horizontal resolution to achieve the target bit rates, for example in standard definition television (SDTV) the resolution may be reduced to 'half' or "three quarters", which produces a less sharp picture than 'full' resolution.
- Bit rate.
  - During periods of high complexity (entropy), compression may leave visible artefacts if the bit rate is not sufficient.

- Application layer video encoding – constant bit rate (CBR) vs. variable bit rate (VBR) at the encoder output.
  - Video encoding is naturally variable bit rate, but to simplify network engineering for Telco delivery systems, the video encoders are set to provide a constant bit rate (as averaged over some specified time period in the order of seconds).
  - VBR streams such as those used in DVD encoding have constant quality since the bit rate is allowed to vary to accommodate varying complexity of the source material.
  - CBR streams have variable quality since there may be times when the bit rate is insufficient to accommodate the video complexity, but CBR streams allow straightforward traffic engineering and system design.
- Group of pictures (GOP) structure.
  - Shorter GOPs improve quality in terms of performance of random accessibility and error recovery, but reduce the maximum compression ratio.
  - Longer GOPs improve maximum compression ratio, but increase channel change time and the amount of damage a lost packet will cause.
  - Dynamic GOPs can be used to better handle scene changes and other effects but are not always implemented on set-top boxes (STBs). In addition, dynamic GOPs can cause variability in zapping latency and may complicate mechanisms intended to increase zapping speed.
- Motion vector search range.
  - Wider searches provide improved quality but at increased complexity and encoder delay.
  - Large search ranges are required for high motion content such as sports.
- Rate Control.
  - Mode decisions greatly affect the bit rate.
  - Proprietary schemes are commonly used to gain competitive advantage.
- Pre-processing (such as noise reduction).
  - Usually proprietary and non-standard, but can improve bit rate/quality trade-off.
- Tandem encoding and rate shaping (e.g., digital turnaround).

Video compression artefact examples

Figure 6-1 illustrates several kinds of compression artefacts that are largely due to insufficient bits allocated, resulting in a too coarse quantization of DCT coefficients or motion vectors and/or poor motion estimation. Additional details of compression artefacts may be found in [b-NTIA264].

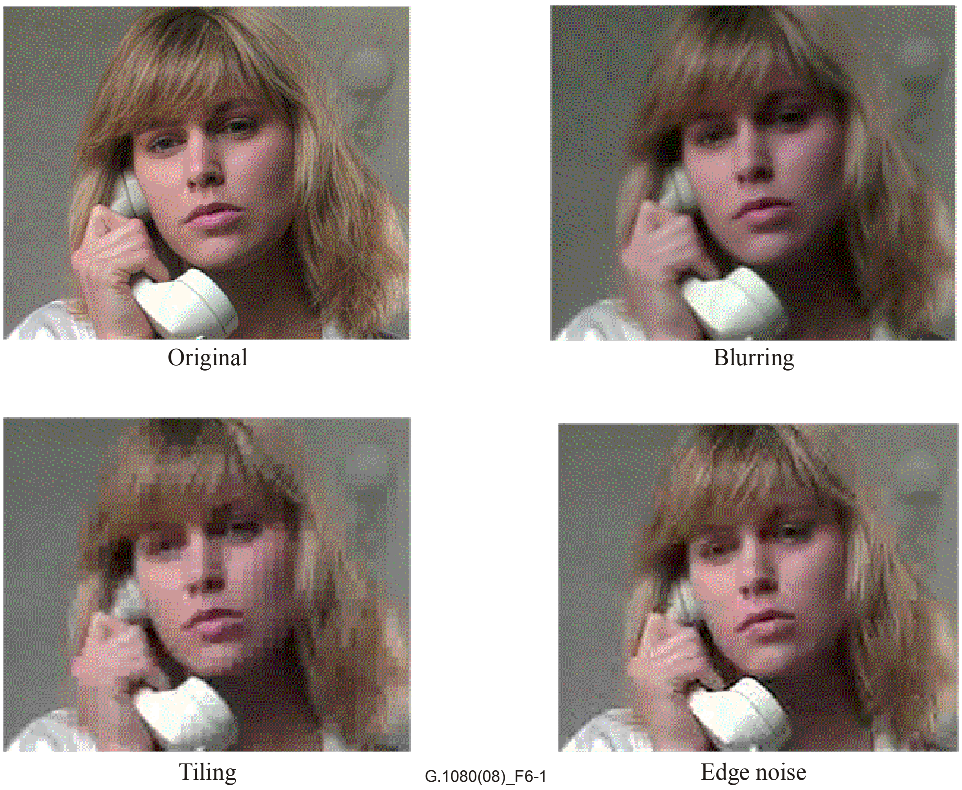


Figure 6-1 – Compression artefacts[1]

Similarly, there are similar parameter implications on the audio side.

NOTE 2 – A partial list of audio codecs is provided in Appendix II.

In addition to the separate audio and video application layer impairments, the synchronization between audio and video components must be maintained to ensure satisfactory QoE. There has been a great deal of research on A/V synchronization requirements in video conferencing and analogue broadcast systems, and specifications have been developed in such bodies as ITU-R ([ITU-R BT.1359-1]).

Because audio that appears before video is very unnatural (sound takes longer to propagate than light, so sound lagging visual is normal), some bodies specifying television specific A/V synchronization have recommended tighter tolerances than those typically used for video conferencing applications.

Recommended provisional minimum engineering objectives for application layer and data plane parameters are presented in the following clauses for various video services.

NOTE 3 – Appendix III provides supplementary information related to the following clauses.

6.1.1 Standard definition TV (SDTV): General minimum objectives

Table 6-1 provides provisional video application layer performance objectives at the MPEG elementary stream level, prior to IP encapsulation for broadcast SDTV (480i / 576i). The objectives for audio elementary stream bit rates are additionally specified below.

[1] Source ITS Video Quality Research (2003)

Table 6-1 – Provisional application layer performance requirements for standard definition broadcast program sources

Video codec standard (non-inclusive list)	Minimum bit rate (video elementary stream level)	Pre-processing enabled
H.262 – Main profile at main level (MP@ML)	2.5 Mbit/s CBR	Yes (if available)
H.264 (Main profile at level 3.0)	1.75 Mbit/s CBR	Yes (if available)
SMPTE 421M	1.75 Mbit/s CBR	Yes (if available)
AVS	1.75 Mbit/s CBR	Yes (if available)

Table 6-2 provides provisional audio application layer performance requirements for standard definition audio sources.

Table 6-2 – Provisional audio application layer performance requirements for standard definition sources

Audio codec standard (Non-inclusive list)	Number of channels	Minimum bit rate (audio elementary stream level, in kbit/s)
MPEG-1 Audio Layer II	Mono or stereo	128 for stereo
Dolby Digital (AC-3)	5.1 if available, else left/right stereo pair	384 for 5.1 ch / 128 for stereo
AAC	Stereo	96 for stereo
MPEG-1 Audio Layer III (MP3)	Stereo	128
MPEG-2 Audio Layer III (MP3)	Stereo	For further study

In general, the chosen audio codecs should align with industry standards in the geography of deployment to ensure maximum compatibility with consumer receivers. Bit rates should be aligned with original source material quality and transcoding between formats should be avoided, if possible.

Table 6-3 provides provisional audio-video synchronization requirements.

Table 6-3 – Provisional SDTV audio-video synchronization requirement

Audio-video synchronization	Audio leading video	Audio lagging behind video
	15 ms maximum	45 ms maximum

Inconsistent loudness levels between channels can negatively impact QoE. It is recommended that equipment be used in the service provider head-end to ensure similar loudness levels across the range of channels provided to the user.

Another audio quality issue, which is beyond the scope of this Recommendation is the dynamic range compression for links between the STB and TV.

6.1.2 Standard definition (SD) TV: VoD and premium content objectives

Video on demand (VoD) and other premium content such as pay per view in standard definition format will have application layer performance factors similar to those of regular broadcast materials. However, subscriber expectation may be higher because of the additional fees paid to access the content and comparison to alternative delivery options. In the case of VoD, users may compare to VoD materials delivered over digital cable systems or even contained in DVDs.

Table 6-4 provides recommended video encoding bit rates for standard definition, VoD and other premium content.

Table 6-4 – Provisional application layer performance requirements for H.262 standard definition, VoD and premium program sources

Video codec standard (Non-inclusive list)	Minimum bit rate (video elementary stream level)	Pre-processing enabled
H.262 – Main profile at main level (MP@ML)	3.18 Mbit/s CBR	Yes (if available)
H.264 (Main profile at level 3)	2.1 Mbit/s CBR	Yes (if available)
SMPTE 421M	2.1 Mbit/s CBR	Yes (if available)
AVS	2.1 Mbits/s CBR	Yes (if available)

Table 6-5 provides provisional recommended audio codec bit rates for VoD and premium content.

Table 6-5 – Provisional audio application layer performance for VoD and premium standard definition sources

Audio codec standard (non-inclusive list)	Number of channels	Minimum bit rate (audio elementary stream level, in kbit/s)
Dolby digital (AC-3)	5.1 if available, else left/right stereo pair	384 for 5.1 ch / 192 for stereo
AAC	5.1 if available, else left/right stereo pair	384 for 5.1 ch / 192 for stereo

6.1.3 High definition TV (HDTV)

Table 6-6 provides provisional video application layer performance objectives for broadcast HDTV (720p/1080i).

Table 6-6 – Provisional application layer performance requirements for high definition (HD) broadcast program sources

Video codec standard (non-inclusive list)	Minimum bit rate (video elementary stream level)	Pre-processing enabled
H.262 – Main profile at main level (MP@ML)	15 Mbit/s CBR	Yes (if available)
H.264 (Main profile at level 4)	10 Mbit/s CBR	Yes (if available)
SMPTE 421M	10 Mbit/s CBR	Yes (if available)
AVS	10 Mbits/s CBR	Yes (if available)

Table 6-7 provides provisional audio application layer performance requirements for high definition audio sources.

Table 6-7 – Provisional minimum audio application layer performance requirements for high definition sources

Audio codec standard (non-inclusive list)	Number of channels	Minimum bit rate (Audio elementary stream level, in kbit/s)
MPEG-1 Audio layer II	Mono or stereo	128 for stereo
Dolby digital (AC-3)	5.1 if available, else left/right stereo pair	384 for 5.1 ch/128 for stereo
AAC	5.1 if available, else left/right stereo pair	384 for 5.1 ch/128 for stereo
MPEG-1/2 Audio layer III(MP3)	Stereo	128

6.2 Impact of network transmission on performance

Key criteria for network transmission include loss, latency and jitter (see Appendix I). In general, reasonable end-to-end delay and jitter values are not problematic due to STB de-jitter buffers, provided the de-jitter buffer size is provisioned to match network and video element performance. Video streams, however, are highly sensitive to information loss and the QoE impact is in turn correlated to a number of variables, including:

- Highly dependent on type of data loss.
  - System information and header losses produce different impairments.
  - Lost data from I and P frames produce different impairments than B frame packet losses due to temporal error propagation.
- Dependent on codec used.
- Dependent on MPEG transport stream packetization used.
- Loss distance and loss profile.
- With high encoding bit rates, the stream is more vulnerable to packet loss impairments.
  - For the same packet loss ratio, impairments due to loss on a higher rate video stream occur more frequently (i.e., there are more visible errors per unit time), simply because there are more packets per second transmitted and each one has the same probability to be affected.
- Decoder concealment algorithms can mitigate perceptual impact of some losses.

An error or sequence of errors in a video bit stream can cause effects ranging from no noticeable audio or video impact to the user to complete loss of the video or audio signal, depending on what was lost and the robustness of the implementation.

Appendix IV provides additional information on how transmission impairments can affect quality.

The video application should be able to operate normally in the presence of normal operational defects. One such normal operational consideration is the operation of protection switching mechanisms in the network. SONET/SDH protection switching mechanisms may result in a potential packet loss duration in the order of, for example, 50 ms. For some other protection mechanisms (e.g., MPLS fast reroute, fast IGP convergence) the potential packet loss duration can be longer, for example, on the order of 250 ms. Service providers are encouraged to add mechanisms to minimize or eliminate the visible effect of such protection mechanisms as these events cascade to a large number of subscribers.

In some other protection mechanisms the potential packet loss duration can be longer. For example, a complete reconvergence of the IP routing table (with IGP) would imply potential packet loss bursts of the order of 30 s. IPTV systems would not be expected to maintain normal service through such an event. Such events can be considered a service outage rather than a service quality defect.

The goal is to minimize visible artefacts to as few as possible using a combination of network performance requirements, loss recovery mechanisms (e.g., FEC, interleaver) and loss mitigation mechanisms (e.g., decoder loss concealment).

7 QoE for text and graphics

Information in this clause is taken from [ITU-T F.700].

7.1 Media component text

7.1.1 Definition

The media component text allows for the capture and representation of information, its transfer from originating user(s) to destination user(s), its presentation to human user(s), processing, filing and retrieval.

7.1.2 Description

7.1.2.1 General description

Text is a representation medium consisting of formatted characters. It is stored and transmitted as a sequence of codes. Although it may be displayed on the same screen as video and still pictures, it requires decoding into specific fonts for presentation to the user, whether on the screen or on paper. The input is through a keyboard. The output may be presented by a printer or on a screen.

The following levels of quality are defined:

T0: minimum quality, basic alphabet and punctuation, no formatting or choice of font;

T0 *bis*: videotex quality, basic alphabet and punctuation, basic graphic character set, no formatting or choice of font;

T1: Usable text conversation quality characterized by:

- Font support for ISO 10646 Language area Latin-1 plus the target language area for the implementation.
- No more than 1 corrupted, dropped or marked missing character per 100.
- Delay from character input in the transmitter to display in the receiver shorter than 2 s.

T2: Good text conversation quality characterized by:

- Font support for all characters in [b-ISO 10646].
- No more than 1 corrupted, dropped or marked missing character per 500.
- Delay from character input in the transmitter to display in the receiver shorter than 1 s.

7.1.2.2 Additional facilities

The user may be given control over text through editing and presentation functions. He may also be able to insert graphics, still pictures or animated pictures within the text.

7.1.2.3 Requirements for various audiovisual services

When text is for the support of conversational services, the timing aspects of text entry and display are critical. Text may be transmitted and displayed in near real time, as text is entered. It may also be transmitted only after specific end-of-sentence action or on a specific send request. In a conversation between two users, the near real-time conversation is important for optimized benefit of the conversation. For multi-user conferences, a sentence based transmission may be more relevant in an open discussion, while for a subtitled speech, the real-time text transmission is preferred.

For retrieval services, it may be accepted to transmit and display a whole page of text in one operation.

For conversation, editing may be reduced to "new line", "erase last character", while the editing for information retrieval should contain a possibility to replace text anywhere on the page and add various formatting effects to any part of text. Annotations that stand out distinctly are also desirable.

The levels of text quality required for various services are the following (indicated with an X):

Table 7-1 – Levels of text requirements

Service	Quality level			
	T0	**T0 *bis***	**T1**	**T2**
Telex	X			
Videotex		X		
Text telephony			X	X
Total Conversation				X
Messaging services			X	X
Retrieval services			X	X

7.1.3 Quality aspects

The quality of text depends mainly upon the capabilities for formatting and using different types of fonts and special characters. When no error correction is made, for instance in conversation, text quality is also measured in terms of corrupted characters, dropped characters and characters replaced by the missing text marker (see [ITU-T T.140]).

7.1.4 Intercommunication

The characters with their formatting may be decoded and assembled into bit maps which can then be handled as still pictures, e.g., as facsimile pages.

7.2 Media component graphics

7.2.1 Definition

The media component graphics allows for the capture and representation of information, its transfer from originating user(s) to destination user(s), its presentation to human user(s), processing, filing and retrieval. This media component allows graphic pictures to be captured and transmitted as geometrical objects whose positions, shapes and colours are coded so that they can be reproduced in a distant terminal.

7.2.2 Description

7.2.2.1 General description

Graphics is a representation medium consisting of geometrical objects featured by their positions, shapes and colours. It is stored and transmitted as a set of codes and parameters. Although it may be displayed on the same screen as video and still pictures, it requires decoding into specific geometrical figures for presentation to the user, whether on a screen or on paper.

The input may be through a graphics tablet, an electronic pencil, some other two-dimensional transducer or dedicated graphic software on a microcomputer or workstation. The output device may be a printer or a screen.

7.2.3 Quality aspects

The intrinsic quality of the graphic depends on the number and the complexity of the objects that can be generated, the precision of their dimensions and positions, the number of possible colours. The overall quality perceived by the user depends also on the resolution of the input and output systems.

7.2.4 Intercommunication

The graphic objects may be decoded and assembled into bit maps which can then be handled as still pictures.

8 QoE for control functions

8.1 QoE requirements for channel zapping time

Channel zapping time (channel switching time) has strong relationship with end user experience of service quality. Generally, it is primarily determined by the time required to have a proper frame at the STB to start decode processing for the new channel. Channel zapping requests can occur when:

– there is a meta data request in the EPG or IPG;

– at random selection of a channel by entering the channel number using the remote control;

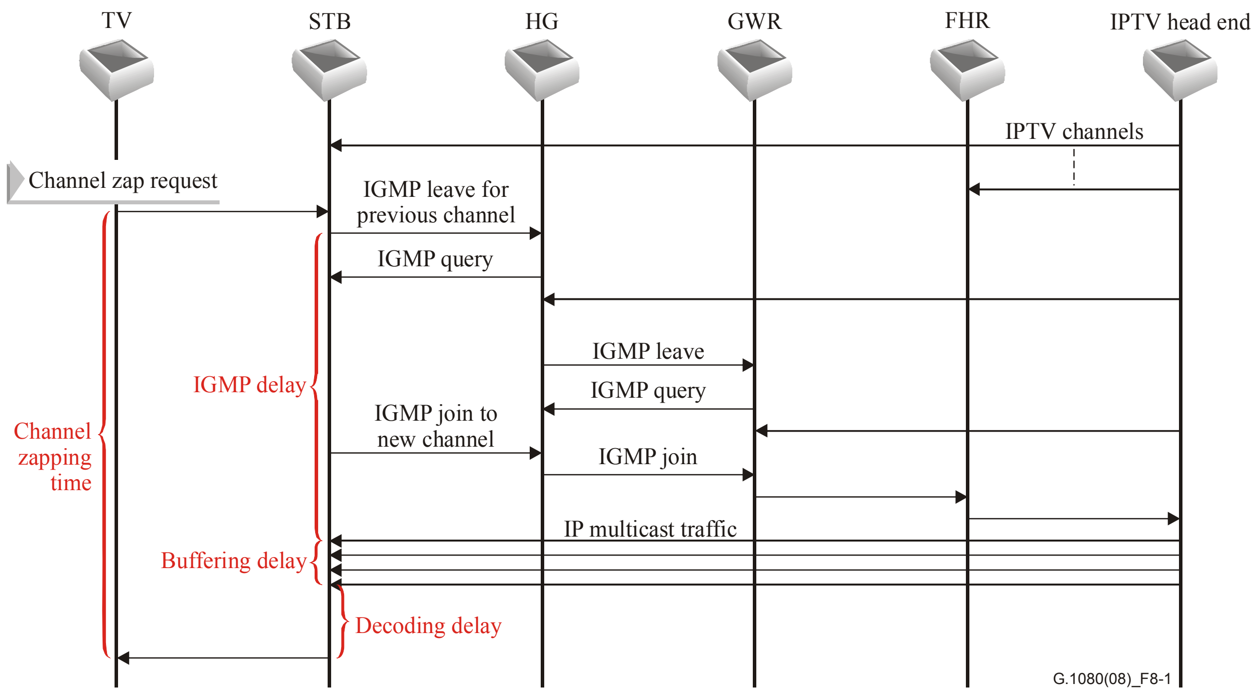
– the channel up/down button is pressed in the remote control;

– the channel up/down button is pressed in the STB front panel;

– a channel is selected in the IPG application menu;

– the STB/TV is powered on and tuned to the initial channel assigned by the IPG.

As a QoE parameter, channel zapping time can be described by three components: IGMP delay, buffering delay, and decoding delay, as shown in Figure 8-1 (note that the timings are not necessarily to scale).



FHR First Hop Router
GWR Gateway Router

Figure 8-1 – Components that contribute to channel zapping time

8.1.1 Classification of channel zapping time

8.1.1.1 IGMP delay

A `channel zap request` is triggered by a channel change which is mapped by the STB to a multicast group address carried in the IGMP message. The IGMP message, which includes a `Join message`, is sent to the homegate (HG). The HG, playing an IGMP proxy role, will process the IGMP message and send an `IGMP request` to the Gateway Router (GWR). After the IGMP message is sent by the GWR, the corresponding channel data should be delivered to the end point at some point. The time to get the content data after sending the first IGMP message is called the IGMP delay.

8.1.1.2 Buffering delay

As the STB receives IPTV multicast traffic, it stacks the packets in a buffer. Buffering delay is the time between the arrival of the first multicast traffic in the buffer and when the STB has sufficient data for playing to the screen.

8.1.1.3 Decoding delay

After the STB starts to receive and buffer a multicast stream, the decoding delay processes buffer data and render them to the TV screen. This type of delay includes both the codec decoding delay, which programs specific information frames in order to decide the target channel, and I-frame acquisition delay, which is used to reduce the bandwidth required for digital video transmission.

8.1.2 Requirements for channel zapping time

One of the key elements involved in validating quality of experience (QoE) in IPTV service is how quickly users can change TV channels, which is often referred to as channel zapping time. However, the explicit relation between channel zapping time and the user perceived quality expressed as a mean opinion score (MOS) is still under study [b-Kooij].

8.2 QoE requirements related to VoD trick mode

Video on demand (VoD) trick mode provides VCR-like features in VoD services. When a subscriber desires a video content through STB, the subscriber accesses the video content from the EPG which supports the contents search engine to help access the content information. To guarantee VCR-like flexibility, this mode provides the trick ability to handle pause, play, rewind, fast forward, and stop entries for these control features.

8.2.1 Trick latency

Each control function (video selection, play, pause, rewind, FFW, stop) has its own delay. QoE metrics for VoD transaction quality are expressed by the following indicators:

– Video selection process delay: Time period from the moment the subject is selected to the moment the content is displayed.

– Play delay: Time period from the moment the Play entry was selected to the moment the content is displayed.

– Stop delay: Time period from the moment the Stop play video entry was selected to the moment the content is stopped playing as indicated by video content display.

– Rewind delay: Time period from the moment the Rewind video entry was selected to the moment the rewind action is executed as indicated on the display device.

– Pause delay: Time period from the moment the Pause video entry was selected to the moment the pause action is executed as indicated on the display device.

– FFW delay: Time period from the moment the Fast Forward video entry was selected to the moment the FFW action is executed as indicated on the display device.

8.2.2 Requirements for VoD trick features

From a QoE perspective, trick feature latency is one of the most important issues to guarantee satisfaction of subscribers. As each trick feature latency directly affects QoE, the latency is required to be sufficiently low to meet user's requirement for QoE related to VoD trick features.

9 QoE for other IPTV services

9.1 QoE requirements for metadata

Figure 9-1 shows the components of metadata.

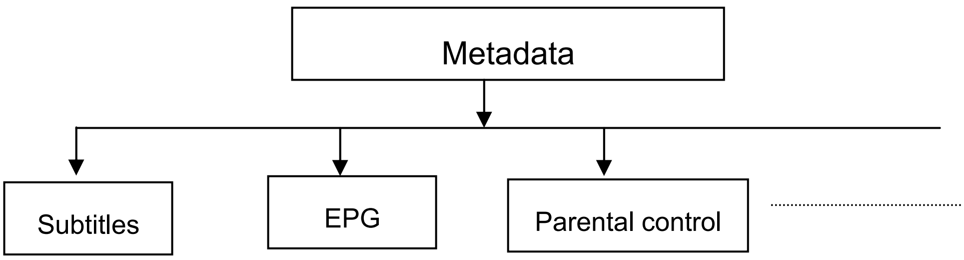


Figure 9-1 – Components of metadata

1) Availability

High availability is recommended to be ensured when transmitting the metadata over the network.

2) Data size

Metadata is recommended to be transported in such a way that the size of the transported data is sufficiently small, relative to such factors as the number of total services, the number of contents, and network bandwidth.

3) Correctness

The service provider should ensure that the metadata tagged to a particular content are correct.

An example to illustrate the importance of metadata is the correctness of "rating" of content. The correct rating on content is directly related to what the customer expects. An incorrect parental rating e.g., a "family" rating for an adult movie can have serious implications for the customer experience and business of the service provider.

9.1.1 QoE requirements for EPG

The following items are recommended to be considered as part of the definition of QoE for IPTV services.

1) User-friendliness

The EPG user interface is recommended to be designed for ease of use.

2) Response time to display EPG page

The response time – the interval of time elapsed between the moment the EPG button of the remote control is pushed and the moment the EPG page is displayed – is recommended to be sufficiently short.

9.2 QoE requirements for browser

If a browser, such as those for BML or HTML, is used to provide the user with an interactive content from the service provider, the following points are recommended to be taken into account.

1) Characteristics of the television set

The IPTV QoE requirements on browsers are recommended to take into account the fact that the behavioural patterns and expectations of television users typically differ from those of PC users.

Moreover, the differences in the capacities of typical TVs (and STBs), on one hand, and PCs, on the other, should be taken into account. For example, as the CPU performance of a television set is usually inferior to that of PC, the contents designed for PC-use do not necessarily work in the TV environment, making it necessary to set up QoE measures taking into account the difference in CPU performance between PC and television. It should be stressed that the browser in an IPTV service may not have the same capacity as the browser in a PC.

2) TV-like display

Some features of TV-like display are recommended to be considered necessary for browser QoE, for such that are commonly imposed by content providers. Examples are:

– overlay function;

– consistency of displayed pictures across terminals.

3) Character size

The character size is recommended to be sufficiently large.

4) Navigation

The navigation function is recommended to be considered for increasing the level of convenience and operability.

5) Cookie

The use of Cookies is recommended to be done with care because of the possible limitation on the non-volatile memory capacity of the terminal. The number, the size and the expiration date of cookies may need to be clearly specified.

9.3 QoE requirements for content navigation

Content navigation is defined as functions for content discovery and selection. So, content navigation is provided by various methods such as direct channel selection, EPG, recommendation. QoE requirements are described in the following clauses depending on the navigation methods.

9.3.1 Contents navigation by direct channel selection, using the up/down button

Ease of contents selection is recommended to be considered especially under numerous contents conditions. In this regard, the time required to select contents and subjective evaluations of ease of use (e.g., MOS) are recommended to be considered.

9.3.2 Contents navigation by EPG/ECG

EPG/ECG is one of the most useful ways of contents navigation. The time required to discover and select contents, and subjective evaluations of ease of use (e.g., MOS) are recommended to be considered.

9.3.3 Contents navigation by contents recommendation

Effective contents recommendation is useful for users. For example, IPTV service provider could recommend contents to the user according to his/her preferences. Contents recommendations from his/her friends are also effective.

For contents recommendation, recommendation accuracy and the security of personal information are recommended to be considered. If the recommended contents include many preferred contents, the QoE of recommendation would be high. Obviously, the security of personal information affects QoE, and communication functions such as the obtention of 3rd party metadata are recommended to be considered.

10 Accessibility requirements

The purpose of this clause is to compile specific performance requirements for IPTV services related to accessibility. This is for further study, but will encompass the following areas:

- Audio quality (including the provision of clean audio).
- Video quality (including sufficient frame rate and resolution for sign language, lip reading, etc. (see [b-ITU-T HSup1]).
- Audio/video synchronization.

11 Security considerations

Security aspects have not been addressed in this Recommendation.

Appendix I

Network QoS parameters affecting QoE

(This appendix does not form an integral part of this Recommendation)

In general, four main network segments: the content acquisition, encoding and play out; the core network; the access network; and the home network constitute an IPTV service network.

The core network is an IP network that is usually well engineered to handle different classes of traffic. Well-engineered networks still require the ability to manage traffic belonging to different applications. Packets belonging to real-time applications, such as IPTV services, should be transmitted before those that belong to non-real-time applications, such as email and file transfer. This differentiation is usually achieved by employing IP differentiated service and its related traffic conditioning and per-hop behaviour (PHB) mechanisms. The IP network may also implement a subset of the IP performance classes as those defined in [ITU-T Y.1541].

The access network could be based on a range of technologies including Ethernet, WiMax, WiFi, etc. The capacity of the access network is the limiting factor for the decision on how many channels are extended to the end user.

The home network includes a number of consumer electronic products that may be interconnected wirelessly using, for example, WiFi products or via a wired network such as an Ethernet.

Appendix II

Non-inclusive list of audio and video codecs

(This appendix does not form an integral part of this Recommendation)

II.1 Video codecs

This clause provides a non-inclusive list of video codecs for television applications.

The following video codecs are used for television applications:

- H.262 (a.k.a. MPEG-2 Video);
- H.264 (a.k.a. MPEG-4 AVC or MPEG-4 Part 10);
- SMPTE 421M (a.k.a. VC-1, previously known as VC-9, the standardized version of Windows Media™ 9);
- AVS.

II.2 Audio codecs

This clause provides a non-inclusive list of audio codecs for television applications.

Most video service offerings (e.g., those using MPEG transport streams or similar) are capable of supporting more than one audio codec along with a single or sometimes multiple video encoding schemes depending on the head-end equipment and set-top box.

Example audio formats used for television applications are:

- MPEG audio layer II (also known as Musicam, used in DVB systems, and MPEG-1 audio layer 2);
- Dolby digital used in ATSC systems (formerly known as AC-3);
- NICAM 728 (European digital format for PAL);
- Advanced audio coding – AAC (either MPEG-2 AAC or MPEG-4 AAC ([b-ISO/IEC 14496-3], Subpart 4));
- MP3 (MPEG-1 Audio layer 3) used particularly for music content.

Appendix III

Supplementary information related to the selected provisional performance parameters

(This appendix does not form an integral part of this Recommendation)

III.1 Introduction

In general, the provisional parameters contained in the various tables in clause 6 were developed taking into consideration industry best practices (e.g., CableLabs specifications, encoder vendor guidelines), performance of competitive systems (e.g., cable, satellite benchmarks), telco deployment experiences, and the state of encoding technologies (e.g., H.262, H.264, SMPTE 421M, AVS commercial offerings) at the time of publication of this Recommendation. In the tables in clause 6, the minimum bit rate for each codec is shown as the target value for achieving sufficient quality. Not all of the codecs may have achieved these targets as of the time of publication of this Recommendation.

III.2 Standard definition TV (SDTV): General minimum objectives

Table 6-1 contains provisional video application layer performance requirements at the MPEG elementary stream level, prior to IP encapsulation for broadcast SDTV (480i / 576i).

In respect to Table 6-1, the following assumptions were used:

Source material:

- NTSC or PAL/SECAM.
- 4:3 aspect ratio.
- Source could enter the head-end in analogue or digital form.

Maximum viewable resolution:

- Horizontal x Vertical: 720 pixels x 480 lines (NTSC) [ITU-R BT.601-6] or 720 pixels x 576 lines (PAL).
- Lower resolutions (e.g., ¾ Horizontal or ½ Horizontal – so called ½ D1) could be used to ensure encoding quality is maintained for complex materials.

Frame rate:

- 29.97 fps (NTSC) or 25 fps (PAL/SECAM).
- 23.97/24 fps may also be used for film-based materials (with 3:2 pull-down for NTSC conversion to 30 fps).
- Two interlaced fields per frame.

Notes on SDTV video bit rate

The bit rates achieved by a particular video encoder undergo continuous improvement over time, particularly when first introduced. As is the case with [ITU-T H.262] since its commercialization in the mid-1990s, improvements have typically followed McCann's law, that states that encoder bit rate improves approximately 15% per year preserving the quality [b-McCannLaw]. In most cases, encoder improvements are done within the scope of the existing standards and therefore do not require upgrades to the decoders.

The H.262 bit rate examples shown in Table 6-1 are nearing the end of the improvement cycle and one may run at lower bit rates (particularly with proprietary pre-processing). A minimum bit rate value must be determined in order to satisfy all the requirements defined in each IPTV service context. Note that services provided over such access networks that are capable of higher bandwidth (e.g., optical fibre, digital cable and satellite) will use higher bit rates and often VBR encoding particularly for broadcast materials with highly complex image content, such as sports.

H.264, SMPTE 421M, and AVS codecs are newer (broadcast systems became commercially available in 2005 for SDTV and 2006 for HD) and are similarly expected to improve over time, although perhaps not aggressively as suggested by McCann's law of 15% per year. The minimum bit rate examples shown in Table 6-1 represent the state of commercially available encoders at present. Table 6-1 assumes similar quality / bit-rate performance of H.264, SMPTE 421M, and AVS.

Table 6-2 contains provisional audio application layer performance requirements for standard definition audio sources.

In respect to Table 6-2, the following assumptions were used:

Source material:

- NTSC or PAL/SECAM.
- Sources could include more than one stereo audio track to support multiple languages or multichannel audio for surround sound effects. Unless indicated in Table 6-2, recommended minimum bit rates are for one stereo pair only.

Audio channels:

- Many broadcasters are also using 5.1 (up to 6 channels) surround audio for prime-time series and special events, particularly concerts and sporting events.

Audio sample rate:

- 48 kHz sample rate for Dolby digital (AC-3).
- 16 kHz to 24 kHz for MPEG-2 audio layer III (MP3).
- 32 kHz to 44.1 kHz for MPEG-1 audio layer III (MP3).
- 32 kHz, 44.1 kHz or 48 kHz for MPEG-1 audio layer II as per [b-ETSI TR 101 154].
- 48 kHz for AAC.

In general, the audio codecs chosen should align with industry standards in the geography of deployment to ensure maximum compatibility with consumer receivers. There is a general trend to global support of Dolby Digital 5.1, particularly in North America (e.g., ATSC) and this is also an option for DVB-based systems. Bit rates should be aligned with original source material quality, and transcoding between formats should be avoided if possible. An MP3 target is provided to support music services.

Table 6-3 contains provisional audio-video synchronization requirements. These requirements were based on the guidelines provided by ATSC for SDTV program materials. Audio-Video synchronization requirements must be determined to satisfy all the requirements of each IPTV service context. Note that the asymmetry in the requirement is due to the unnaturalness of audio leading video since light travels faster than sound.

[ITU-R BT.1359-1] provides a thorough discussion of relative timing of audio and video, and it indicates the detectability threshold and the acceptability threshold for errors in sound/vision timing in television material for NTSC and PAL systems based on subjective evaluations undertaken in Japan, Switzerland and Australia. The threshold of detectability is about 45 ms (audio leading video) to 125 ms (audio lagging behind video). The threshold of acceptability is about 90 ms (audio leading video) to 185 ms (audio lagging behind video) on the average. Other SDOs such as EBU have different requirements on audio-video synchronization.

Inconsistent loudness levels between channels can negatively impact QoE. It is recommended that equipment be used in the service provider head-end to ensure similar loudness levels across the range of channels provided to the user. Another audio quality issue beyond the scope of this Recommendation is the dynamic range compression for links between the STB and TV.

III.3 Standard definition (SD) TV: VoD and premium content objectives

Video on demand (VoD) and other premium content such as pay per view in standard definition format will have similar application layer performance factors as regular broadcast materials. However, subscriber expectation may be higher because of additional fees paid to access the content and comparison to alternative delivery options. In the case of VoD, users may compare to VoD materials delivered over digital cable systems or even contained in DVDs.

In North America, VoD application layer parameters are defined by CableLabs [b-CLVoD2]. Since a great deal of existing VoD content is aligned with the parameters used by cable providers, and consumers will compare the quality levels, it is recommended that telco-based video service providers consider adopting these as the minimum guidelines. The current guidelines are limited to H.262 encoding. Recommendations for H.264, SMPTE 421M, or AVS encoded VoD materials assume a 1.5x improvement in bit rate as an example, aligned with the state of commercial deployments of these encoders.

Table 6-4 contains provisional recommended video encoding bit rates for standard definition, VoD and other premium content.

In respect to Table 6-4, the following assumptions were used:

Source material:

- NTSC or PAL/SECAM.
- 4:3 aspect ratio.
- Encoding could be done offline using multipass systems for stored content such as VoD assets.

Minimum viewable resolution:

- Horizontal x Vertical: 1/2 D1 352 pixels x 480 lines (NTSC) [ITU-R BT.601-6] or 352 pixels x 576 lines (PAL/SECAM) is permitted to ensure encoding quality is maintained for complex materials.
- However, it is recommended that ¾ D1 resolution (528 x 480 / 528 x 576) be used where possible to align with the maximum specified for cable systems.
- Telco service providers could run VoD assets at full D1 resolutions but would likely not be able to reuse assets pre-encoded for cable deployments.

Frame rate:

- 29.97 fps (NTSC) or 25 fps (PAL/SECAM).
- 23.97 fps may also be used for film-based materials (with 3:2 pull-down for NTSC conversion to 30 fps).
- Two interlaced fields per frame.

Notes on SDTV video bit rate for VoD and premium content

- Actual bit rate requirements must be individually determined to satisfy all the requirements of each IPTV service context.
- Bit rates for H.262 are based on typically used values for VoD content and align with the majority of VoD assets available.
- H.264, SMPTE 421M, and AVS bit rates are extrapolated from H.262 using a 1.5x factor.
- Total video plus audio bit rate for most commonly available H.262 encoded VoD assets is 3.75 Mbit/s.

- The QoE of a VoD service may also be impacted by the quality of the implementation of trick mode features such as fast forward and rewind. The fast forward and rewind modes should be as smooth as possible and include intelligible audio during trick modes, if possible.

Table 6-5 contains provisionally recommended audio codecs and bit rates for VoD and premium content.

In respect to Table 6-5, a sampling rate of 48 kHz was assumed.

III.4 High definition TV (HDTV): Objectives

Table 6-6 contains provisional recommended minimum video application layer performance objectives for broadcast HDTV (720p/1080i).

In respect to Table 6-6, the following assumptions were used:

Source material:

- ATSC or DVB.
- 16:9 aspect ratio.
- Source enters the head-end in digital form.

Resolution and frame rate:

- 720p60 (e.g., SMPTE 296M) or 720p50 (DVB):
  - Horizontal x Vertical: 1280 pixels x 720 lines.
  - 50, 59.94, 60 progressive scan frames per second.
- 1080i60 (e.g., SMPTE 274M) or 1080i50 (DVB):
  - Horizontal x Vertical: 1920 pixels x 1080 lines.
  - 29.97 (59.94i), 30 (60i) interlaced frames per second, two fields per frame.

Notes on HDTV video bit rate

Bit rates for video encoding and corresponding decoders undergo continuous improvement over time particularly when first introduced. As with SDTV, improvements have typically followed McCann's law [b-McCannLaw].

The H.262 bit rates shown in Table 6-6 are nearing the end of the improvement cycle, although one may run at lower bit rates (particularly with proprietary pre-processing). Bit rate values must be determined to satisfy all the requirements of each IPTV service context. It should be noted that services over such access networks that are capable of higher bandwidths (e.g., optical fibre, digital cable and satellite) will use higher bit rates and often VBR encoding, particularly for complex broadcast materials such as sports.

H.264, SMPTE 421M, and AVS codecs are newer (broadcast systems commercially available in 2005) and are expected to improve significantly over time. The minimum bit rate examples shown in Table 6-6 represent the state of commercially available encoders at the time of publication of this Recommendation, but lower bit rates with satisfactory quality are expected, as encoder technology improves. H.264 main profile is listed in Table 6-6, but as high profile encoders and compatible STBs become available, service providers may choose to take advantage of superior features available for HDTV encoding in the high profile. Table 6-6 also assumes similar quality / bit rate performance of H.264, SMPTE 421M, and AVS.

Table 6-7 contains the provisional audio application layer performance requirements for high definition audio sources. In general, the provisional parameters contained in Table 6-7 were guided by industry best practices, performance of competitive systems (e.g., cable, satellite), telco deployment experiences, and the state encoding technologies available at the time of publication of this Recommendation.

In respect to Table 6-7, the following assumptions were used:

Source material:

- ATSC or DVB.
- Sources could include more than one audio track to support multiple languages.
- For HDTV materials, multichannel audio for surround sound effects should be provided, where possible.
- Audio channels:
  - Many broadcasters are also using 5.1 surround audio for prime-time series and special events, particularly concerts and sporting events.
- Audio sample rate:
  - 48 kHz sample rate for Dolby digital (AC-3)
  - 16 kHz to 48 kHz for MPEG-2 Audio layer III (MP3)
  - 32 kHz to 48 kHz for MPEG-1 Audio layer III (MP3)
  - 32 kHz, 44.1 kHz or 48 kHz for MPEG-1 Audio layer II as per [b-ETSI TR 101 154]
  - 44.1 kHz or 48 kHz for MPEG-1 (or -2) Audio layer II

Notes on HDTV audio bit rate

In general, the audio codecs chosen should align with industry standards in the geography of deployment, to ensure maximum compatibility with consumer receivers. There is a general trend to global support of Dolby Digital 5.1, particularly in North America (e.g., ATSC) and this is also an option for DVB-based systems. Bit rates should be aligned with original source material quality, and transcoding between formats should be avoided if possible. An MP3 target is included to support music services.

A/V synchronization requirements for HDTV materials is currently under study by the ATSC and other bodies. Until additional data is available, the guidelines presented in Table 6-3 for SDTV material may be followed for HDTV material as well.

Appendix IV

Impact of transmission impairments on quality

(This appendix does not form an integral part of this Recommendation)

IV.1 Introduction

Key criteria for network transmission include loss, latency and jitter (see Appendix I). In general, reasonable end-to-end delay and jitter values are not problematic due to STB de-jitter buffers, provided the de-jitter buffer size is provisioned to match network and video element performance. Video streams however are highly sensitive to information loss and the QoE impact is in turn correlated to a number of variables including:

- Highly dependent on type of data loss:
  - System information and header losses produce different impairments.
  - Lost data from I and P frames produce different impairments than B frame packet losses due to temporal error propagation.
- Dependent on codec used.
- Dependent on MPEG transport stream packetization used.
- Loss distance and loss profile.
- With high encoding bit rates, the stream is more vulnerable to packet loss impairments:
  - For the same packet loss ratio, impairments due to loss occur more frequently on a higher rate video stream (i.e., there are more visible errors per unit time) simply because there are more packets per second transmitted and each one has the same probability to be affected.
- Decoder concealment algorithms can mitigate perceptual impact of some losses.

An error or sequence of errors in a video bit stream can cause effects ranging from no noticeable audio or video impact to the user, to complete loss of the video or audio signal, depending on what was lost and the robustness of the implementation. Figure IV.1 shows an example of the impact of a single lost IP packet (containing seven MPEG transport stream packets) on a video frame if the lost information is from a B or an I frame. As indicated, since the I frame is a key frame used in the compression of subsequent P and B frames, the I frame impairment propagates in time across 14 frames of video or almost a half second (assuming 33 ms per frame). If the lost packet impacted a B frame, the impairment impacted only that frame with a duration of 33 ms, because the codec did not employ inter-frame prediction coding from a B frame. Note that no loss concealment algorithms were running at the decoder in this example.

Single B-frame IP packet loss

Single I-frame IP packet loss

Figure IV.1 – Example of the impact of a single IP packet loss (B frame and I frame)

The following tables show the minimum thresholds required to achieve satisfactory service quality targets with respect to IP packet transport loss and jitter requirement for various video services. The associated assumptions are identified.

Network latency and jitter should be engineered to closely align with set-top box jitter buffer provisioning (wait time and buffer size) and overall network design, and therefore may vary from implementation to implementation. Network jitter must be less than the size of the de-jitter buffer. Delay variations beyond these limits will manifest themselves as a loss. Increasing buffering also negatively impacts such characteristics as channel change latency, so ideally the de-jitter buffers should be set as small as possible. Objectives outlined for jitter are based on the experience of the operators and STB buffering capabilities.

Packet loss objectives are stated in terms of loss period and loss distance, as defined in [b-IETF RFC 3357]. Essentially, loss distance is a measure of the spacing between consecutive network packet loss or error events; a loss period is the duration of a loss or error event (and how many packets are lost in that duration). The loss rates in the tables below are objectives designed to ensure satisfactory end user service level quality assuming no or minimal loss concealment. If the network infrastructure performance is below the required levels, service providers may make use of network level techniques (e.g., interleaving and FEC) and application layer mechanisms (e.g., loss concealment, application layer FEC, automatic repeat request (ARQ)) as outlined in Appendix II of [b-DSLF TR 126], Error Protection Mechanisms Overview, to achieve the required performance levels. In addition, the use of these techniques may provide an improved quality of experience over competing service offerings.

Ideally, the maximum loss period would correspond to one IP packet, since often random bit errors or minor amounts of congestion cause an isolated loss event with a loss period of one packet, and even such an isolated loss event can result in a very noticeable impairment, as shown in Figure IV.1. However, since a loss period which continues for tens of milliseconds may happen in some physical layer with loop impairment behaviour and FEC, like DSL, or in major amounts of congestion, we also need to consider it. This type of loss period will result in different numbers of packets being lost, depending on the bit rate of the video stream, as shown in the following tables. This maximum loss period objective is provisionally set until further studies allow better tuning of the maximum loss period allowed, considering various physical layers, protection mechanisms, and optimum settings.

Also, a loss event which continues for tens of seconds may happen in some situations like DSL modem resynchronization or IGP convergence, without some other protection mechanisms (e.g., MPLS fast reroute, fast IGP convergence). An IPTV system would not be expected to maintain normal service through such events, so they might be considered as service outage rather than quality defects.

The video application should be able to operate normally in the presence of normal operational defects. One such normal operational consideration is the operation of protection switching mechanisms in the network. SONET/SDH protection switching mechanisms may result in a potential packet loss duration in the order of, for example, 50 ms. For some other protection mechanisms (e.g., MPLS fast reroute, fast IGP convergence) the potential packet loss duration can be longer, for example, in the order of 250 ms. Service providers are encouraged to add mechanisms to minimize or eliminate the visible effects of such protection mechanisms, as these events cascade to a large number of subscribers.

Considering some other protection mechanisms, the potential packet loss duration can be longer. For example, a complete reconvergence of the IP (IGP) routing table would imply potential packet loss bursts in the order of 30 s. An IPTV system would not be expected to maintain normal service through such an event. Such events can be considered a service outage rather than a service quality defect.

All impairments are specified as end-to-end objectives (from video origin to the video output of a set-top box to the television, including any loss correction mechanisms that may be applied at the network or application layers):

- Loss distance of error events should be limited to at most the required number of loss events in a certain time period. An error event is defined as a loss or corruption of a group of a small number of IP packets, each containing up to seven MPEG transport stream packets.
- Set-top box decoders should employ error concealment techniques to minimize the impact of lost or corrupted video packets.
- Appendix II of [b-DSLF TR 126] provides additional details on access BER, FEC and ARQ mechanisms.

The goal is to minimize visible artefacts to as few as possible, using a combination of network performance requirements, loss recovery mechanisms (e.g., FEC, interleaver) and loss mitigation mechanisms (e.g., decoder loss concealment).

IV.2 Standard definition video

IV.2.1 H.262 codec example

The following example shows the minimal expected broadcast TV transport layer performance required to provide acceptable performance when H.262 codecs are used:

Assumptions for Table IV.1 are as follows:

- H.262 codec;
- MPEG transport stream;
- seven 188-byte packets per IP packet;
- no or minimal loss concealment (tolerable loss rates may be higher depending on degree and quality of STB loss concealment);
- metrics are end-to-end, from head-end encoder output to after any application layer protection mechanisms at the customer premises;
- metrics are for the IP flows containing video streams only; IP streams for other applications may have different performance requirements.

Table IV.1 – Minimum level of transport layer performance required to provide satisfactory QoE for H.262 encoded SDTV services

Transport stream bit rate (Mbit/s)	Jitter	Maximum duration of a single error event	Corresponding loss period in IP packets	Loss distance	Corresponding average IP video stream packet loss rate
3.0	<50 ms	<= 16 ms	<6 IP packets	<=1 error event per hour	<= 5.85E-06
3.75	<50 ms	<= 16 ms	<7 IP packets	<=1 error event per hour	<= 5.46E-06
5.0	<50 ms	<= 16 ms	<9 IP packets	<=1 error event per hour	<= 5.26E-06

Note that all the values contained in Table IV.1 are provisional, and they must be determined by taking into consideration the requirements defined by each IPTV service context.

IV.2.2 H.264 or SMPTE 421M codec example

The following example shows the minimal expected broadcast TV transport layer performance required to provide acceptable performance when H.264 or SMPTE 421M codecs are used:

Assumptions for Table IV.2 are as follows:

- H.264 or SMPTE 421M codec;
- MPEG transport stream with seven 188-byte packets per IP packet;
- no or minimal loss concealment (tolerable loss rates may be higher depending on degree and quality of STB loss concealment);
- metrics are end-to-end, from head-end encoder output to after any application layer protection mechanisms at the customer premises;
- metrics are for the IP flows containing video streams only; IP streams for other applications may have different performance requirements.

Table IV.2 – Minimum level of transport layer performance required to provide satisfactory QoE for H.264 or SMPTE 421M encoded SDTV services

Transport stream bit rate (Mbit/s)	Jitter	Maximum duration of a single error event	Corresponding loss period in IP packets	Loss distance	Corresponding average IP video stream packet loss rate
1.75	<50 ms	<= 16 ms	<4 IP packets	<=1 error event per hour	<= 6.68E-06
2.0	<50 ms	<= 16 ms	<5 IP packets	<=1 error event per hour	<= 7.31E-06
2.5	<50 ms	<= 16 ms	<5 IP packets	<=1 error event per hour	<= 5.85E-06
3.0	<50 ms	<= 16 ms	<6 IP packets	<=1 error event per hour	<= 5.85E-06

Note that all the values contained in Table IV.2 are provisional, and they must be determined by taking into consideration the requirements defined by each IPTV service context and compression coding scheme.

IV.2.3 VoD and premium content applications

The requirements for network performance of broadcast SDTV applications listed above in clauses IV.2.1 and IV.2.2 may also be applied to VoD and premium content services.

IV.3 High definition TV

The following examples show the minimal expected high definition TV transport layer performance required to provide acceptable performance.

It is commonly agreed upon that ideally HDTV services meet criterion such as one visible impairment event per 12 hours or better, for example. In the remainder of this clause, it is proposed that a value of four hours be used as the minimum loss distance for HDTV services, assuming that not all errors will result in a visible impairment, because:

- loss of B-frame information is sometimes below threshold of notability;
- error concealment will be used with HDTV decoders.

IV.3.1 H.262 codec example

Table IV.3 below shows the minimum level of transport layer performance, with respect to loss period and loss distance, required for H.262 HDTV.

Assumptions for Table IV.3 are as follows:

- H.262 codec;
- MPEG transport stream with seven 188-byte packets per IP packet;
- STB has some level of loss concealment;
- metrics are end-to-end, from head-end encoder output to after any application layer protection mechanisms at the customer premises;
- metrics are for the IP flows containing video streams only; IP streams for other applications may have different performance requirements.

Table IV.3 – Minimum level of transport layer performance required to provide satisfactory QoE for H.262 encoded HDTV services

Transport stream bit rate (Mbit/s)	Jitter	Maximum duration of a single error event	Corresponding loss period in IP packets	Loss distance	Corresponding average IP video stream packet loss rate
15.0	<50 ms	<= 16 ms	<24 IP packets	<=1 error event per 4 hours	<= 1.17E-06
17	<50 ms	<= 16 ms	<27 IP packets	<=1 error event per 4 hours	<= 1.16E-06
18.1	<50 ms	<= 16 ms	<29 IP packets	<=1 error event per 4 hours	<= 1.17E-06

Note that all the values contained in Table IV.3 are provisional, and they must be determined by taking into consideration the requirements defined by each IPTV service context.

IV.3.2 H.264, SMPTE 421M or AVS codec example

Table IV.4 below shows the minimum level of transport layer performance, with respect to loss period and loss distance, required for H.264, SMPTE 421M or AVS codecs.

Assumptions for Table IV.4 are as follows:

- H.264, SMPTE 421M, or AVS codec;
- MPEG transport stream with seven 188-byte packets per IP packet;
- STB has some level of loss concealment;
- metrics are end-to-end, from head-end encoder output to after any application layer protection mechanisms at the customer premises;
- metrics are for the IP flows containing video streams only; IP streams for other applications may have different performance requirements.

Table IV.4 – Minimum level of transport layer performance required to provide satisfactory QoE for H.264, SMPTE 421M, or AVS encoded HDTV services

Transport stream bit rate (Mbit/s)	Jitter	Maximum duration of a single error event	Corresponding loss period in IP packets	Loss distance	Corresponding average IP video stream packet loss rate
8	<50 ms	<= 16 ms	<14 IP packets	<=1 error event per 4 hours	<= 1.28E-06
10	<50 ms	<= 16 ms	<17 IP packets	<=1 error event per 4 hours	<= 1.24E-06
12	<50 ms	<= 16 ms	<20 IP packets	<=1 error event per 4 hours	<= 1.22E-06

Note that all the values contained in Table IV.4 are provisional, and they must be determined by taking into consideration the requirements defined by each IPTV service context and compression coding scheme.

Some cases may require special error control techniques to achieve the target, for example PLR in the range of 10^{-6} for video services. Appendix II of [b-DSLF TR 126] provides additional informative details on network BER, FEC performance and mitigation options.

IV.4 Plots of packet loss ratio vs bits sent

Figure IV.2 shows packet loss ratios as a function of bit rate and time between uncorrected loss events for isolated packet loss events. Curves are plotted for three examples, i.e., a loss distance of one hour between packet loss events, a loss distance of two hours between packet loss events, and a loss distance of four hours between packet loss events. Points are plotted on these curves to reflect required levels of performance as per Tables IV.1 through IV.4.

The figure assumes that each IP packet carries 7 MPEG transport stream packets. The plots implicitly assume that error statistics are stationary and time invariant.

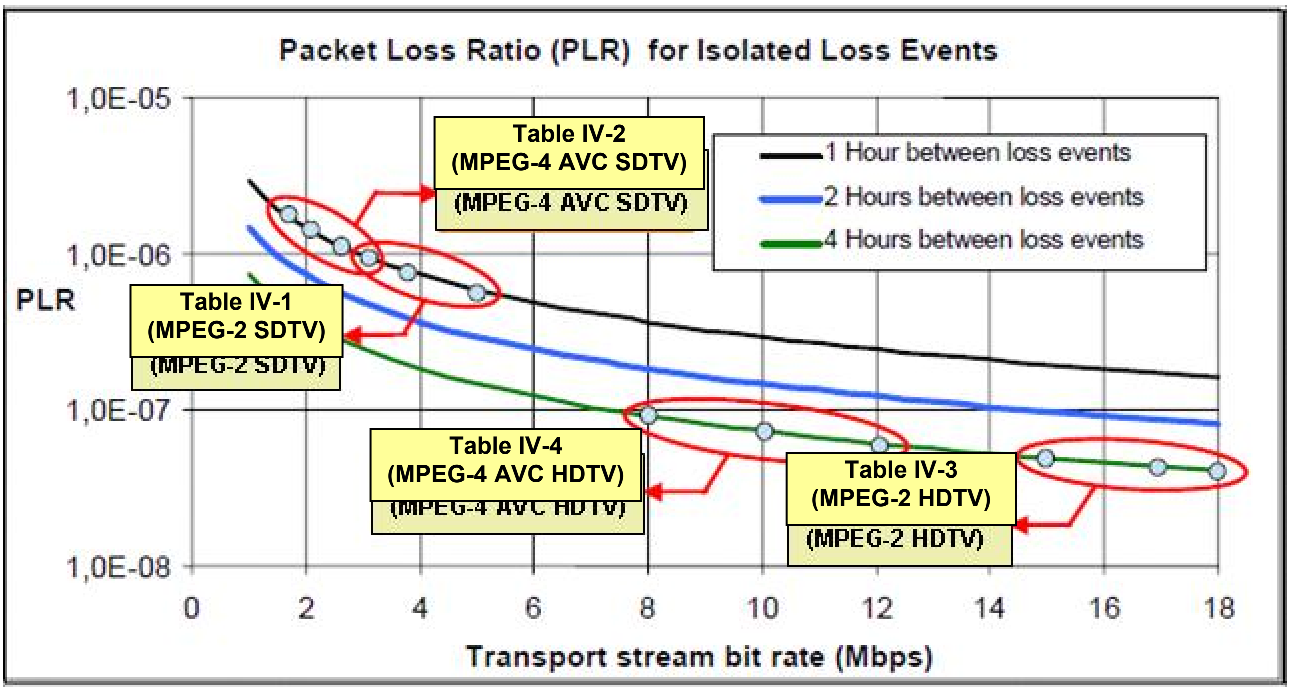


Figure IV.2 – PLR required to meet average time between loss events of 1, 2 and 4 hours assuming isolated lost packets

As described in clauses IV.3 and IV.4, there is a need to consider a loss period which continues for tens of milliseconds. Figures IV.3 and IV.4 show packet loss ratios as a function of bit rate and time between uncorrected loss events of two examples, which are 8 ms and 16 ms, respectively. The "ripple effect" in the charts is the result of rounding to an integer number of lost/corrupted IP packets. For example, 8 ms of lost video data in an MPEG transport stream at a bit rate of 3 Mbit/s:

Total MPEG packets per second	=	3 Mbit/s / 8 bits per byte / 188 bytes per MPEG transport stream packet
	=	1994.7 MPEG transport stream packets per second
Total IP packets per second	=	1994.7 / 7 MPEG transport stream packets per IP packet
	=	285 IP packets per second
A loss of 8 ms corresponds to	=	285 IP packets per second * 0.008 seconds
	=	2.28 IP packets lost.

Because an entire IP packet is lost if a part of a packet is lost, this is rounded to the next integer, i.e., 3 IP packets. Additionally, since the lost bytes are not necessarily aligned to IP packet boundaries, this would be further rounded to 4 IP packets.

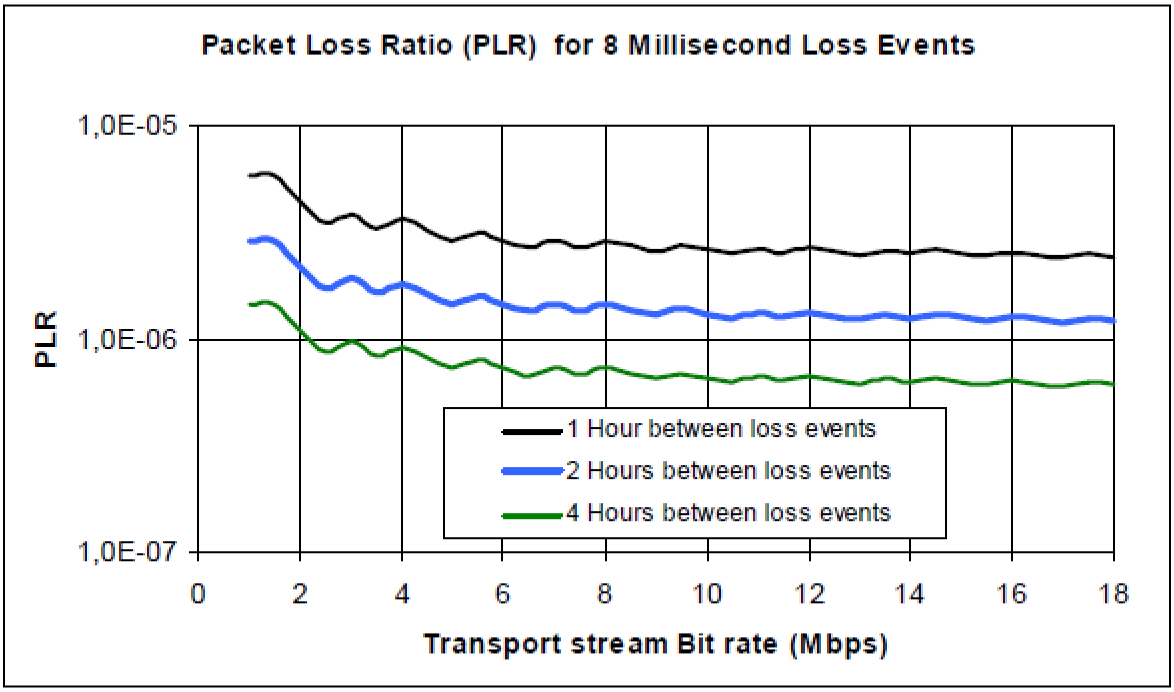


Figure IV.3 – PLR required to meet average time between loss events of 1, 2 and 4 hours – Loss event loses 8 milliseconds of contiguous data

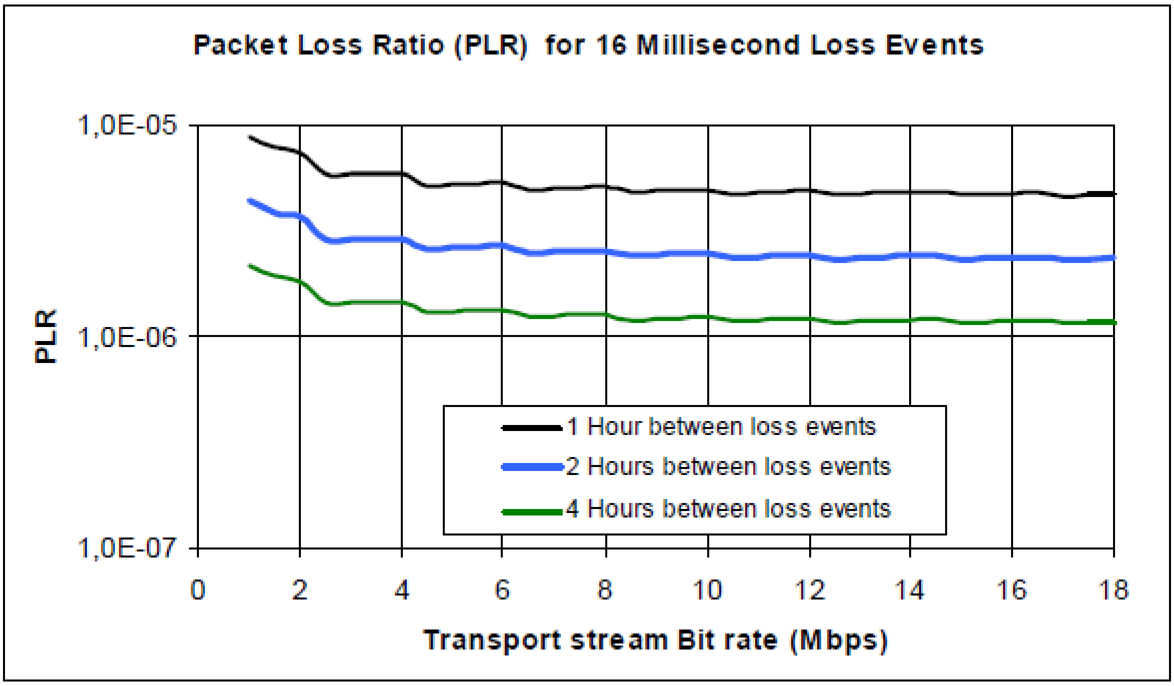


Figure IV.4 – PLR required to meet average time between loss events of 1, 2 and 4 hours – Loss event loses 16 milliseconds of contiguous data

IV.5 Severe error limits for SDTV and HDTV services

In addition to average packet loss rates impacting picture/audio quality and availability metrics, it may also be advantageous to define a second set of limits on severe impairments. These limits would apply to quality degradations that fall between the impairments generated by the packet loss limits specified above and total service outage (i.e., black screen) metrics specified by the dependability metrics. These gross impairments could include video frame drops, frame repetitions (freeze frames), or short duration loss of intelligible audio or video or control (e.g., due to protection switching). Metrics are TBD based on industry input and could be specified by frequency of the error event per time unit e.g., a maximum of one severe error per day and the duration of the impairment. This subject is for further study.

Bibliography

[b-ITU-T HSup1] ITU-T Series H Supplement 1 (1999), *Application profile − Sign language and lip-reading real-time conversation using low bit-rate video communication*.

[b-ATIS-0800004] ATIS-0800004 (2006), *A Framework for QoS Metrics and Measurements Supporting IPTV Services*.
<https://www.atis.org/docstore/product.aspx?id=22624>

[b-AVS] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. GB/T 20090.2-2006, *Information Technology, Advanced Coding of Audio and Video*, Part 2: Video.
<http://220.181.176.160/stdlinfo/servlet/com.sac.sacQuery.GjbzcxListServlet>

[b-DSLF TR 126] DSL Forum TR-126 (2006), *Triple-play Services Quality of Experience (QoE) Requirements*.
<http://www.broadband-forum.org/technical/download/TR-126.pdf>

[b-ETSI TR 101 154] ETSI TR 101 154 V1.8.1 (2007), *Digital Video Broadcasting (DVB); Implementation guidelines for the use of MPEG-2 Systems, Video and Audio in satellite, cable and terrestrial broadcasting applications*.
<http://pda.etsi.org/pda/home.asp?wki_id=3GDCON4OqcouoxxEh@PO>

[b-IETF RFC 3357] IETF RFC 3357 (2002), *One-way Loss Pattern Sample Metrics*.
<http://www.ietf.org/rfc/rfc3357.txt?number=3357>

[b-IETF RFC 3393] IETF RFC 3393 (2002), *IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)*.
<http://www.ietf.org/rfc/rfc3393.txt?number=3393>

[b-ISO-10646] ISO/IEC 10646 (2003), *Information technology – Universal Multiple-Octet Coded Character Set (UCS)*.
<http://www.iso.org/iso/search.htm?qt=10646&sort=rel&type=simple&published=on>

[b-ISO/IEC 14496-3] ISO/IEC 14496-3 (2005), *Information technology – Coding of audio-visual objects – Part 3: Audio*.
<http://www.iso.org/iso/search.htm?qt=14496&sort=rel&type=simple&published=on>

[b-NTIA264] Wolf, S. (1990), *Features for automated quality assessment of digitally transmitted video*, NTIA Report 264, June.
<http://www.its.bldrdoc.gov/pub/ntia-rpt/90-264/index.php>

[b-SMPTE 421M] SMPTE 421M (2006), *Television – VC-1 Compressed Video Bitstream Format and Decoding Process*.
<https://store.smpte.org/>

[b-CLVoD2] Cable Labs Specification MD-SP-VOD-CEP2.0-I02-070105 (2007), *Metadata 2.0 Specifications – Content Encoding Profiles 2.0 Specification*.
<http://www.cablelabs.com/specifications/MD-SP-VOD-CEP2.0-I02-070105.pdf>

[b-Möller] Sebastian Möller & Alexander Raake (2006), *A taxonomy of quality prediction models recommended by the ITU-T*, ITU Workshop on End-to-End QoS/QoE, June.
<http://www.itu.int/ITU-T/worksem/qos/200606/presentations/s1p1-moeller.pdf>

[b-McCannLaw] Ken McCann, *Review of DTT HD Capacity Issues – An Independent Report from ZetaCast Ltd, Commissioned by Ofcom*, v3.0 (31 October 2007). <http://www.ofcom.org.uk/consult/condocs/dttfuture/report.pdf>

[b-Kooij] Kooij, R., Ahmed, K. and Brunnstrom K. (2006), *Perceived quality of channel zapping*, fifth IASTED International Conference Spain, pp. 155-158.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	**Transmission systems and media, digital systems and networks**
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2009