International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.1541

(02/2006)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

Internet protocol aspects – Quality of service and network performance

---

Network performance objectives for IP-based services

ITU-T Recommendation Y.1541

ITU-T

International
Telecommunication
Union

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

GLOBAL INFORMATION INFRASTRUCTURE	
General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899
INTERNET PROTOCOL ASPECTS	
General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	Y.1200–Y.1299
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	**Y.1500–Y.1599**
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
NEXT GENERATION NETWORKS	
Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	Y.2100–Y.2199
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Numbering, naming and addressing	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Network control architectures and protocols	Y.2500–Y.2599
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899

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

ITU-T Recommendation Y.1541

Network performance objectives for IP-based services

Summary

This Recommendation defines classes of network Quality of Service (QoS) with objectives for Internet Protocol network performance parameters. Two of the classes contain provisional performance objectives. These classes are intended to be the basis for agreements among network providers, and between end users and their network providers.

Appendix I provides information about how ATM might support IP layer performance. Appendix II discusses alternatives for defining IP delay variation. Appendix III presents the Hypothetical Reference Paths against which the Y.1541 QoS objectives were tested for feasibility. Appendix IV gives example computations of packet delay variation. Appendix V discusses issues that must be considered whenever IP measurements are made. Appendix VI describes the relationship between this Recommendation and the IETF-defined mechanisms for managing QoS. Appendix VII gives estimates of speech transmission quality for the Hypothetical Reference Paths of Appendix III. Appendix VIII discusses digital television transport on IP Networks. Appendix IX estimates TCP file transfer performance on paths conforming to Y.1541 objectives.

Source

ITU-T Recommendation Y.1541 was approved on 22 February 2006 by ITU-T Study Group 12 (2005-2008) under the ITU-T Recommendation A.8 procedure.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. 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 received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

ITU-T Recommendation Y.1541

Network performance objectives for IP-based services

1 Introduction and Scope

1.1 Introduction

Customers require network performance levels that, when combined with their hosts, terminals, and other devices, satisfactorily support their applications. The adoption of IP-based network services has not changed this fact, except that networks must be constrained in terms of packet transfer performance parameters (as defined in ITU-T Rec. Y.1540).

Application performance requirements are well-understood, but several key contributors are often beyond the network service provider's control (e.g., home networks, LAN, application gateways, terminals, hosts, and other customer devices). We note that objectives on the performance of customer equipment are available, such as ITU-T Rec. P.1010 for VoIP terminals and gateways, and combining these objectives with specific network performance levels (as appendices of this Recommendation illustrate), a view of application performance can be directly related to network performance.

In response, service providers have agreed on network performance levels that they will work together to meet, and codified the numerical objectives in this Recommendation. Agreement on levels of network performance is highly beneficial, because it constrains a critical and often dominating factor in application performance.

The objectives are organized in sets called network Quality of Service (QoS) classes (in Table 1) that can be matched with well-designed customer equipment to satisfactorily support various applications (as indicated in Table 2). Classes with provisional objectives are found in Table 3. The number of classes has been deliberately kept small to simplify the engineering of paths traversing multiple operators' networks, so the objectives in each class must satisfy the needs of multiple applications. Readers of this Recommendation should plan for *at least* eight classes when considering protocol fields and values, since future expansion of the classes is possible.

The objective values result from analysis of key applications such as conversational telephony, multimedia conferencing, reliable data exchange using TCP, and digital television, in concert with network feasibility analysis. The appendices provide significant, detailed testimony as to how the objectives in the network QoS classes can be used to determine the end-to-end (application) quality provided.

The network QoS classes form an important link in the chain of developments required to assure end-to-end performance. They are part of the lexicon for QoS negotiation among users and networks, especially when signalling protocols communicate QoS requests on a dynamic basis.

Verification that the service meets network objectives is another key area of customer interest. This has been addressed here through recommended evaluation intervals, packet payload sizes, and other aspects useful to measurement designers. In addition, the UNI-UNI objectives are directly verifiable by users, in contrast with objectives that apply to non-user interfaces or utilize information unknown to customers, such as route distance.

1.2 Scope

This Recommendation specifies network (UNI-UNI) IP performance values for each of the performance parameters defined in ITU-T Rec. Y.1540. The specific performance values vary, depending on the network QoS class. This Recommendation defines eight network QoS classes, two of which are provisional. This Recommendation applies to international IP network paths

(UNI-UNI). The network QoS classes defined here are intended to be the basis of agreements between end-users and network service providers, and between service providers. The classes should continue to be used when static agreements give way to dynamic requests supported by QoS specification protocols.

The QoS classes defined here support an extremely wide range of applications, including the following: conversational telephony, multimedia conferencing, digital video, and interactive data transfer. Other applications may require new or revised classes, but any desire for new classes must be balanced with the requirement of feasible implementation, and the number of classes must be small for implementations to scale in global networks.

The QoS objectives are primarily applicable when access link speeds are at the T1 or E1 rate and higher. This limitation recognizes that IP packet serialization time is included in the definition of IP Packet Transfer Delay (IPTD), and that sub-T1 access rates can produce serialization times of over 100 ms for packets with 1500 octet payloads. Also, this Recommendation effectively requires the deployment of Network QoS mechanisms on access devices in order to achieve the IP Packet Delay Variation (IPDV) objective, especially when the access rate is low (e.g., T1 rate). Network designs may include lower access rates if:

1) Network planners understand the effect of additional serialization time on the User-Network Interface (UNI) to UNI objective for IPTD.
2) QoS mechanisms limit the access contribution to IPDV, and the UNI to UNI objective for IPDV is met. The current IPDV objective is necessary to achieve high quality application performance, as Appendices III and VII clearly show.

This Recommendation provides the network QoS classes needed to support user-oriented QoS Categories. Accordingly, this Recommendation is consistent with the general framework for defining quality of communication services in ITU-T Rec. G.1000, and with the end-user multimedia QoS categories needed to support user applications given in ITU-T Rec. G.1010.

NOTE – This Recommendation utilizes parameters defined in ITU-T Rec. Y.1540 that can be used to characterize IP service provided using IPv4; applicability or extension to other protocols (e.g., IPv6) is for further study.

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.

[1] ITU-T Recommendation G.114 (2003), *One-way transmission time.*

[2] ITU-T Recommendation G.109 (1999), *Definition of categories of speech transmission quality.*

[3] ITU-T Recommendation G.826 (2002), *End-to-end error performance parameters and objectives for international, constant bit-rate digital paths and connections.*

[4] ITU-T Recommendation G.1020 (2003), *Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks.*

[5] ITU-T Recommendation I.113 (1997), *Vocabulary of terms for broadband aspects of ISDN.*

[6] ITU-T Recommendation I.350 (1993), *General aspects of quality of service and network performance in digital networks, including ISDNs.*

[7] ITU-T Recommendation P.1010 (2004), *Fundamental voice transmission objectives for VoIP terminals and gateways*.

[8] ITU-T Recommendation Y.1540 (2002), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

[9] IETF RFC 791 (STD-5) (1981), *Internet Protocol, DARPA Internet Program Protocol Specification*.

[10] ITU-T Recommendation Y.1231 (2000), *IP Access Network Architecture*.

[11] ITU-T Recommendation E.651 (2000), *Reference connections for traffic engineering of IP access networks*.

[12] ITU-T Recommendation G.1000 (2001), *Communications Quality of Service: A framework and definitions*.

[13] ITU-T Recommendation G.1010 (2001), *End-user multimedia QoS categories*.

[14] ITU-T Recommendation Y.1221 (2002), *Traffic control and congestion control in IP-based networks*.

[15] ITU-T Recommendation G.107 (2005), *The E-model, a computational model for use in transmission planning*.

[16] ITU-T Recommendation G.108 (1999), *Application of the E-model: A planning guide*.

3 Abbreviations

This Recommendation uses the following abbreviations:

AF	Assured Forwarding
ATM	Asynchronous Transfer Mode
CBR	Constant Bit Rate
CDV	Cell Delay Variation
CER	Cell Error Ratio
CLR	Cell Loss Ratio
CMR	Cell Misinsertion Ratio
CS	Circuit Section
DS	Differentiated Services
DST	Destination host
E1	Digital Hierarchy Transmission at 2.048 Mbit/s
E3	Digital Hierarchy Transmission at 34 Mbit/s
EF	Expedited Forwarding
FEC/I	Forward Error Correction and Interleaving
FIFO	First-In, First-Out
FTP	File Transfer Protocol
GW	Gateway
HRE	Hypothetical Reference Endpoint
HRP	Hypothetical Reference Path

HTTP	HyperText Transfer Protocol
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPDV	IP packet Delay Variation
IPER	IP packet Error Ratio
IPLR	IP packet Loss Ratio
IPOT	Octet based IP packet Throughput
IPPT	IP Packet Throughput
IPRE	IP packet transfer Reference Event
IPRR	IP Packet Reordering Ratio
IPTD	IP Packet Transfer Delay
ISP	Internet Service Provider
ITU-T	International Telecommunication Union – Telecommunication Standardization Sector
LL	Lower Layers, protocols and technology supporting the IP layer
LAN	Local Area Network
M_{av}	The minimum number of packets recommended for assessing the availability state
MP	Measurement Point
MPLS	Multi-Protocol Label Switching
MTBISO	Mean Time between IP Service Outages
MTTISR	Mean Time to IP Service Restoral
N	The number of packets in a throughput probe of size N
NS	Network Section
NSE	Network Section Ensemble
NSP	Network Service Provider
OSPF	Open Shortest Path First
PDB	Per Domain Behaviour
PDH	Plesiochronous Digital Hierarchy
PHB	Per Hop Behaviour
PIA	Percent IP service Availability
PIU	Percent IP service Unavailability
pkt	IP datagram (IP packet)
QoS	Quality of Service
R	Router
RFC	Request for Comment
RSVP	Resource reSerVation Protocol
RTP	Real-Time Transport Protocol

SDH	Synchronous Digital Hierarchy
SPR	Spurious Packet Ratio
SRC	Source host
STD	Standard
T1	Digital Hierarchy Transmission at 1.544 Mbit/s
T3	Digital Hierarchy Transmission at 45 Mbit/s
T_{av}	Minimum length of time of IP availability; minimum length of time of IP unavailability
TBD	To Be Determined
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TE	Terminal Equipment
T_{max}	Maximum IP packet delay beyond which the packet is declared to be lost
ToS	Type of Service
TTL	Time To Live
UDP	User Datagram Protocol
UNI	User Network Interface
VoIP	Voice over Internet Protocol
VTC	Video Teleconference

4 Transfer capacity, capacity agreements, and the applicability of QoS classes

This clause addresses the topic of network transfer capacity (the effective bit rate delivered to a flow over a time interval), and its relationship to the packet transfer Quality of Service (QoS) parameters defined in ITU-T Rec. Y.1540, and objectives specified here.

Transfer Capacity is a fundamental QoS parameter having primary influence on the performance perceived by end-users. Many user applications have minimum capacity requirements; these requirements should be considered when entering into service agreements. ITU-T Rec. Y.1540 does not define a parameter for capacity, however, it does define the Packet Loss parameter. Lost bits or octets can be subtracted from the total sent in order to provisionally determine network capacity. An independent definition of capacity is for further study.

It is assumed that the user and network provider have agreed on the maximum access capacity that will be available to one or more packet flows in a specific QoS class (except the Unspecified class). A packet flow is the traffic associated with a given connection or connectionless stream having the same source host (SRC), destination host (DST), class of service, and session identification. Other documents may use the terms microflow or subflow when referring to traffic streams with this degree of classification. Initially, the agreeing parties may use whatever capacity specifications they consider appropriate, so long as they allow both network provider enforcement and user verification. For example, specifying the peak bit rate on an access link (including lower layer overhead) may be sufficient. The network provider agrees to transfer packets at the specified capacity in accordance with the agreed QoS class.

When the protocols and systems that support dynamic requests are available, the user will negotiate a traffic contract. Such a contract specifies one or several traffic parameters (such as those defined in ITU-T Rec. Y.1221 [14], or RSVP) and the QoS class, and applies to a specific flow.

The network performance objectives may no longer be applicable when there are packets submitted in excess of the capacity agreement or the negotiated traffic contract. If excess packets are observed, the network is allowed to discard a number of packets equal to the number of excess packets. Such discarded packets must not be included in the population of interest, which is the set of packets evaluated using the network performance parameters. In particular, discarded packets must not be counted as lost packets in assessing the network's IPLR performance. A discarded packet might be retransmitted, but then it must be considered as a new packet in assessing network performance.

It is a network privilege to define its response to flows with excess packets, possibly based on the number of excess packets observed. When a flow includes excess packets, no network performance commitments need be honoured. However, the network may offer modified network performance commitments.

5 Network performance objectives

This clause discusses objectives for the user information transfer performance of public IP services. These objectives are stated in terms of the IP layer performance parameters defined in ITU-T Rec. Y.1540. A summary of the objectives can be found in Table 1 together with its associated general notes. All values in Table 1 are stable.

NOTE – From a users' perspective, network QoS objectives contribute only part of the transmission performance (e.g., mouth-to-ear quality in voice over IP). Appendix VII provides pointers to the appropriate Recommendations in this area.

5.1 General discussion of QoS

The QoS class definitions in Table 1 present bounds on the network performance between user network interfaces (UNI). As long as the users (and individual networks) do not exceed the agreed capacity specification or traffic contract, and a path is available (as defined in ITU-T Rec. Y.1540), network providers should collaboratively support these UNI-to-UNI bounds for the lifetime of the flow.

The actual network QoS offered to a given flow will depend on the distance and complexity of the path traversed. It will often be better than the bounds included with the QoS class definitions in Table 1.

Static QoS class agreements can be implemented by associating packet markings (e.g., Type of Service precedence bits or Diff-Serv Code Point) with a specific class.

Protocols to support dynamic QoS requests between users and network providers, and between network providers, are under study. When these protocols and supporting systems are implemented, users or networks may request and receive different QoS classes on a flow-by-flow basis. In this fashion, the distinct performance needs of different services and applications can be communicated, evaluated, and acknowledged (or rejected, or modified).

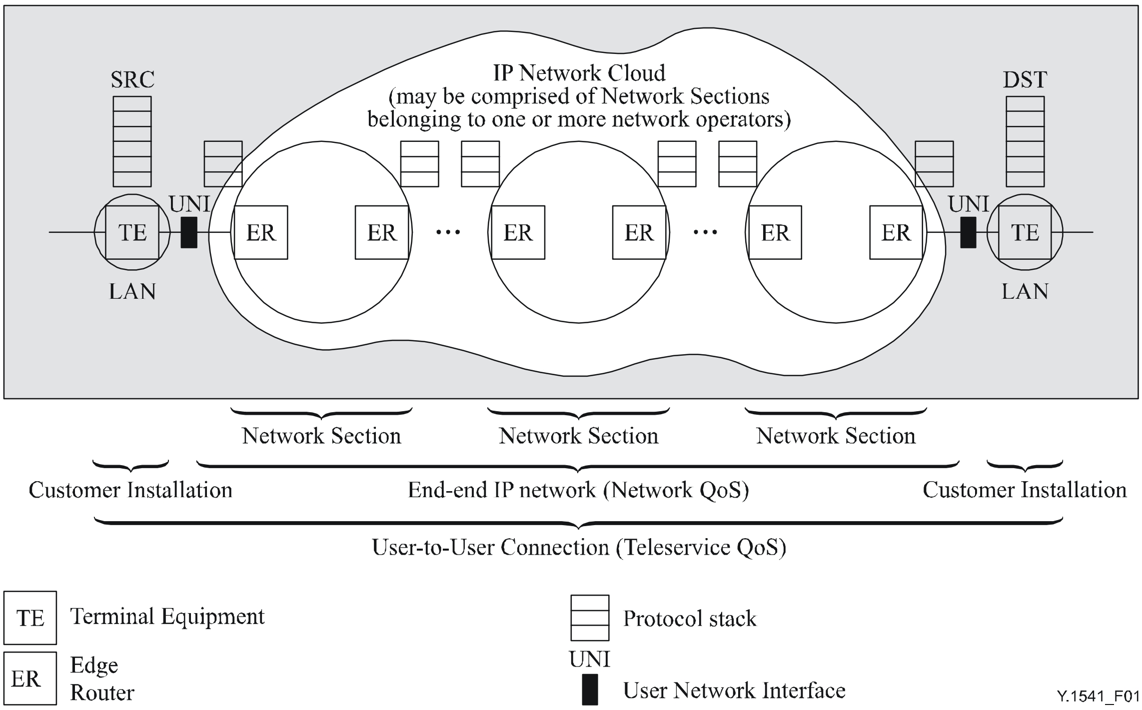
5.2 Reference path for UNI to UNI QoS

Each packet in a flow follows a specific path. Any flow (with one or more packets on a path) that satisfies the performance objectives of this clause can be considered fully compliant with the normative Recommendations of ITU-T Rec. Y.1541.

NOTE – The phrase "End-to-End" has a different meaning in Recommendations concerning user QoS classes, where end-to-end means, for example, from mouth to ear in voice quality Recommendations. Within the context of this Recommendation, end-to-end is to be understood as from UNI-to-UNI.

The UNI-to-UNI performance objectives are defined for the IP performance parameters corresponding to the IP packet transfer reference events (IPRE). The UNI-to-UNI IP performance objectives apply from User Network Interface-to-User Network Interface in Figure 1. The

UNI-to-UNI IP network path includes the set of Network Sections (NS) and inter-network links that provide the transport of IP packets transmitted from the UNI at the SRC side to the UNI at the DST side; the protocols below and including the IP layer (layer 1 to layer 3) may also be considered part of an IP network. Network Sections (NS) (defined in ITU-T Rec. Y.1540) are synonymous with operator domains, and may include IP Access Network Architectures as described in ITU-T Recs E.651 and Y.1231. The Reference Path in Figure 1 is an adaptation of the Y.1540 Performance Model.



NOTE – Customer Installation equipment (shaded area) is for illustrative purposes only.

Figure 1/Y.1541 – UNI-to-UNI reference path for network QoS objectives

The Customer Installation includes all Terminal Equipment (TE), such as a host and any router or LAN if present. There will be only one human User in some applications. It is important to note that specifications for TE and the User-to-User Connection are beyond the scope of this Recommendation. The edge routers that connect with terminal equipment may also be called Access Gateways.

Reference Paths have the following attributes:

1) IP clouds may support User-to-User connections, User-to-Host connections, and other endpoint variations.
2) Network Sections may be represented as clouds with Edge routers on their borders, and some number of interior routers with various roles.
3) The number of Network Sections in a given path may depend upon the Class of Service offered, along with the complexity and geographic span of each Network Section.
4) The scope of this Recommendation allows one or more Network Sections in a path.
5) The Network Sections supporting the packets in a flow may change during its life.
6) IP connectivity spans international boundaries, but does not follow circuit switched conventions (e.g., there may not be identifiable gateways at an international boundary if the same network section is used on both sides of the boundary).

5.3 Network QoS classes

This clause describes the currently defined network QoS classes. Each network QoS class creates a specific combination of bounds on the performance values. This clause includes guidance as to when each network QoS class might be used, but it does not mandate the use of any particular network QoS class in any particular context.

Table 1/Y.1541 – IP network QoS class definitions and network performance objectives

Network performance parameter	Nature of network performance objective	QoS Classes					
		Class 0	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5 Unspecified**
IPTD	Upper bound on the mean IPTD (Note 1)	100 ms	400 ms	100 ms	400 ms	1 s	U
IPDV	Upper bound on the $1 - 10^{-3}$ quantile of IPTD minus the minimum IPTD (Note 2)	50 ms (Note 3)	50 ms (Note 3)	U	U	U	U
IPLR	Upper bound on the packet loss probability	1×10^{-3} (Note 4)	1×10^{-3} (Note 4)	1×10^{-3}	1×10^{-3}	1×10^{-3}	U
IPER	Upper bound	1×10^{-4} (Note 5)					U

General Notes:

The objectives apply to public IP Networks. The objectives are believed to be achievable on common IP network implementations. The network providers' commitment to the user is to attempt to deliver packets in a way that achieves each of the applicable objectives. The vast majority of IP paths advertising conformance with ITU-T Rec. Y.1541 should meet those objectives. For some parameters, performance on shorter and/or less complex paths may be significantly better.

An evaluation interval of 1 minute is suggested for IPTD, IPDV, and IPLR and, in all cases, the interval must be recorded with the observed value. Any minute observed should meet these objectives.

Individual network providers may choose to offer performance commitments better than these objectives.

"U" means "unspecified" or "unbounded". When the performance relative to a particular parameter is identified as being "U" the ITU-T establishes no objective for this parameter and any default Y.1541 objective can be ignored. When the objective for a parameter is set to "U", performance with respect to that parameter may, at times, be arbitrarily poor.

NOTE 1 – Very long propagation times will prevent low end-to-end delay objectives from being met. In these and some other circumstances, the IPTD objectives in Classes 0 and 2 will not always be achievable. Every network provider will encounter these circumstances and the range of IPTD objectives in Table 1 provides achievable QoS classes as alternatives. The delay objectives of a class do not preclude a network provider from offering services with shorter delay commitments. According to the definition of IPTD in ITU-T Rec. Y.1540, packet insertion time is included in the IPTD objective. This Recommendation suggests a maximum packet information field of 1500 bytes for evaluating these objectives.

NOTE 2 – The definition of the IPDV objective (specified in ITU-T Rec. Y.1540) is the 2-point IP Packet Delay Variation. See ITU-T Rec. Y.1540 and Appendix II for more details on the nature of this objective. For planning purposes, the bound on the mean IPTD may be taken as an upper bound on the minimum IPTD and, therefore, the bound on the $1 - 10^{-3}$ quantile may be obtained by adding the mean IPTD and the IPDV value (e.g., 150 ms in Class 0).

Table 1/Y.1541 – IP network QoS class definitions and network performance objectives

NOTE 3 –This value is dependent on the capacity of inter-network links. Smaller variations are possible when all capacities are higher than primary rate (T1 or E1), or when competing packet information fields are smaller than 1500 bytes (see Appendix IV).
NOTE 4 – The Class 0 and 1 objectives for IPLR are partly based on studies showing that high quality voice applications and voice codecs will be essentially unaffected by a 10^{-3} IPLR.
NOTE 5 – This value ensures that packet loss is the dominant source of defects presented to upper layers, and is feasible with IP transport on ATM.

5.3.1 Nature of the network performance objectives

The objectives in Table 1 apply to public IP networks, between MPs that delimit the end-to-end IP network. The objectives are believed to be achievable on common implementations of IP Networks.

The left-hand part of Table 1 indicates the statistical nature of the performance objectives that appear in the subsequent rows.

The performance objectives for IP packet transfer delay are upper bounds on the underlying mean IPTD for the flow. Although many individual packets may have transfer delays that exceed this bound, the average IPTD for the lifetime of the flow (a statistical estimator of the mean) should normally be less than the applicable bound from Table 1.

The performance objectives for 2-point IP Packet Delay Variation (defined in ITU-T Rec. Y.1540) are based on an upper bound on the $1 - 10^{-3}$ quantile of the underlying IPTD distribution for the flow. The $1 - 10^{-3}$ quantile allows short evaluation intervals (e.g., a sample with 1000 packets is the minimum necessary to evaluate this bound). Also, this allows more flexibility in network designs where engineering of delay buildout buffers and router queue lengths must achieve an overall IPLR objective on the order of 10^{-3}. Use of lower quantile values will result in under-estimates of de-jitter buffer size, and the effective packet loss would exceed the overall IPLR objective (e.g., an upper quantile of $1 - 10^{-2}$ may have an overall packet loss of 1.1%, with IPLR = 10^{-3}). Other statistical techniques and definitions for IPDV are being studied as described in Appendix II, and Appendix IV discusses IPDV performance estimation.

The performance objectives for the IP packet loss ratios are upper bounds on the IP packet loss for the flow. Although individual packets will be lost, the underlying probability that any individual packet is lost during the flow should be less than the applicable bound from Table 1.

Objectives for less-prevalent packet transfer outcomes and their associated parameters are for further study, such as the Spurious Packet Ratio (SPR) defined in ITU-T Rec. Y.1540.

5.3.2 Evaluation intervals

The objectives in Table 1 cannot be assessed instantaneously. Evaluation intervals produce subsets of the packet population of interest (as defined in ITU-T Rec. Y.1540). Ideally, these intervals are:

- Sufficiently long to include enough packets of the desired flow, with respect to the ratios and quantiles specified.
- Sufficiently long to reflect a period of typical usage (flow lifetime), or user evaluation.
- Sufficiently short to ensure a balance of acceptable performance throughout each interval (intervals of poor performance should be identified, not obscured within a very long evaluation interval).
- Sufficiently short to address the practical aspects of measurement.

For evaluations associated with telephony, a minimum interval of the order of 10 to 20 seconds is needed with typical packet rates (50 to 100 packets per second), and intervals should have an upper limit on the order of minutes. A value of 1 minute is suggested and, in any case, the value used must be recorded with the observed value, along with any assumptions and confidence intervals. Any minute observed should meet the IPTD, IPDV, and IPLR objectives of Table 1. Minimally acceptable estimation methodologies are intended for future revisions of this Recommendation.

Methods to verify achievement of the objectives are for further study. Either continuous or non-continuous evaluation may be used. One possible method of measurement is given in RFC 3432, "*Network Performance Measurement with Periodic Streams*", where the requirement for random measurement start times and evaluation intervals of finite length result in a non-continuous evaluation.

5.3.3 Packet size for evaluation

Packet size influences the results for most performance parameters. A range of packet sizes may be appropriate since many flows have considerable size variation. However, evaluation is simplified with a single packet size when evaluating IPDV, or when the assessment targets flows that support constant bit rate sources and, therefore, a fixed information field size is recommended. Information fields of either 160 octets or 1500 octets are suggested, and the field size used must be recorded. Also, an information field of 1500 octets is recommended for performance estimation of IP parameters when using lower layer tests, such as bit error measurements.

5.3.4 Unspecified (Unbounded) performance

For some network QoS classes, the value for some performance parameters is designated "U". In these cases, the ITU-T sets no objectives regarding these parameters. Network operators may unilaterally elect to assure some minimum quality level for the unspecified parameters, but the ITU-T does not recommend any such minimum.

Users of these QoS classes should be aware that the performance of unspecified parameters can, at times, be arbitrarily poor. However, the general expectation is that mean IPTD will be no greater than 1 second.

NOTE – The word "unspecified" may have a different meaning in Recommendations concerning B-ISDN signalling.

5.3.5 Discussion of the IPTD objectives

Very long propagation times will prevent low UNI-to-UNI delay objectives from being met, e.g., in cases of very long geographical distances, or in cases where geostationary satellites are employed. In these and some other circumstances, the IPTD objectives in Classes 0 and 2 will not always be achievable. It should be noted that the delay objectives of a class do not preclude a network provider from offering services with shorter delay commitments. Any such commitment should be explicitly stated. See Appendix III for an example calculation of IPTD on a global route. Every network provider will encounter these circumstances (either as a single network, or when working in cooperation with other networks to provide the UNI-to-UNI path), and the range of IPTD objectives in Table 1 provides achievable network QoS classes as alternatives. Despite different routing and distance considerations, related classes (e.g., Classes 0 and 1) would typically be implemented using the same node mechanisms.

According to the definition of IPTD in ITU-T Rec. Y.1540, packet insertion time is included in the IPTD objectives. This Recommendation suggests a maximum packet information field of 1500 bytes for evaluating the objectives.

5.3.6 Guidance on class usage

Table 2 gives some guidance for the applicability and engineering of the network QoS Classes.

Table 2/Y.1541 – Guidance for IP QoS classes

QoS class	Applications (examples)	Node mechanisms	Network techniques
0	Real-time, jitter sensitive, high interaction (VoIP, VTC)	Separate queue with preferential servicing, traffic grooming	Constrained routing and distance
1	Real-time, jitter sensitive, interactive (VoIP, VTC).		Less constrained routing and distances
2	Transaction data, highly interactive (Signalling)	Separate queue, drop priority	Constrained routing and distance
3	Transaction data, interactive		Less constrained routing and distances
4	Low loss only (short transactions, bulk data, video streaming)	Long queue, drop priority	Any route/path
5	Traditional applications of default IP networks	Separate queue (lowest priority)	Any route/path
NOTE – Any example application listed in Table 2 could also be used in Class 5 with unspecified performance objectives, as long as the users are willing to accept the level of performance prevalent during their session.			

Traffic policing and/or shaping may also be applied in network nodes.

Discussion of Broadcast Quality Television transport on IP may be found in Appendix VIII.

5.3.7 Provisional QoS Classes

This clause presents a set of Provisional QoS Classes. The distinction between these classes (see Table 3) and those in Table 1, is that the values of all objectives are provisional and they need not be met by networks until they are revised (up or down) based on real operational experience.

Table 3/Y.1541 – Provisional IP network QoS class definitions and network performance objectives

Network performance parameter	Nature of network performance objective	QoS Classes	
		Class 6	Class 7
IPTD	Upper bound on the mean IPTD	100 ms	400 ms
IPDV	Upper bound on the $1 - 10^{-5}$ quantile of IPTD minus the minimum IPTD (Note 1)	50 ms	
IPLR	Upper bound on the packet loss ratio	1×10^{-5}	
IPER	Upper bound	1×10^{-6}	
IPRR	Upper bound	1×10^{-6}	

Table 3/Y.1541 – Provisional IP network QoS class definitions and network performance objectives

General Notes: Evaluation intervals for these classes should be 1 minute or longer. Evaluations should use 1500 byte payloads. An evaluation interval of 1 minute is suggested for IPTD, IPDV, and IPLR, and any minute observed should meet these objectives. One rationale for IP Packet Loss Ratio (IPLR) objective was to minimize the effect of loss on TCP capacity, even when TCP parameters and the operating system have been tuned, and the Large Windows option has been utilized. Appendix IX provides background information on this and other support rationales. The value for IPLR is not sufficient to support all the quality levels envisioned by the community of digital video users, and Forward Error Correction and Interleaving (FEC/I) is likely to be required. Appendix VIII supplies background on the quality expectations of video transport users, and the FEC/I needed to supply even lower loss ratios. The objective for IP Packet Error Ratio (IPER) was set so as to contribute insignificantly to the overall packet loss. The IP Packet Reordering Ratio (IPRR) has been defined as supplementary terminology in Appendix VII/Y.1540. Reordered packets may appear as lost to a TCP sender, depending on the distance from their original positions. Therefore, the IPRR was set so as to contribute insignificantly to the overall packet loss. The value for IPDV is under study, and contributions are invited to examine the rationale and feasibility of other (lower) values. NOTE 1 – The definition of the IPDV objective (specified in ITU-T Rec. Y.1540) is the 2-point IP Packet Delay Variation. See ITU-T Rec. Y.1540 and Appendix II for more details on the nature of this objective. For planning purposes, the bound on the mean IPTD may be taken as an upper bound on the minimum IPTD, and therefore the bound on the $1 - 10^{-5}$ quantile may be obtained by adding the mean IPTD and the IPDV value (e.g., 150 ms in Class 6).

These classes are intended to support the performance requirements of high bit rate user applications that have more stringent loss/error requirements than those supported by Classes 0 through 4 in Table 1.

6 Availability objectives

This clause will include information about availability objectives based on the availability parameter defined in ITU-T Rec. Y.1540. The objectives require more study, since fundamental network design options are rapidly changing.

7 Achievement of the performance objectives

Further study is required to determine how to achieve these performance objectives when multiple network providers are involved. There are promising standards development activities that are intended to complete other aspects needed for UNI-UNI QoS assurance.

Clause 8 gives the relationships for concatenating the performance levels of two or more Network Sections to determine whether the UNI-UNI objectives are met.

8 Concatenating network sections and their QoS values

8.1 Introduction

This clause addresses the estimation of the UNI-UNI performance of a path, knowing the performance of sub-sections. The purpose is to provide standard relationships to compose these UNI-UNI estimates.

These relationships produce reasonably accurate estimates of the UNI-UNI performance. Errors in the estimation process are believed to be in balance with potential errors of the individual values themselves. When the values come from recent measurements or modelling activities, they can be subject to considerable error if conditions are not stationary, or the principal assumption of independence between network sections does not hold.

These relationships are intended to support accumulation of impairments facilitated by QoS signalling protocol(s). They must not be used to support allocation of UNI-UNI values.

8.2 Composing UNI-UNI values

8.2.1 Mean transfer delay

For the Mean IP packet Transfer Delay (IPTD) performance parameter, the UNI-UNI performance is the sum of the means contributed by Network Sections.

The units of IPTD values are seconds, with resolution of at least 1 microsecond. If lesser resolution is available in a value, the unused digits shall be set to zero.

8.2.2 Loss ratio

For the IP packet Loss Ratio (IPLR) performance parameter, the UNI-UNI performance may be estimated by inverting the probability of successful packet transfer across *n* Network Sections, as follows:

$$\mathrm{IPLR}_{\mathrm{UNI\text{-}UNI}} = 1 - \{ (1 - \mathrm{IPLR}_{\mathrm{NS1}}) \times (1 - \mathrm{IPLR}_{\mathrm{NS2}}) \times (1 - \mathrm{IPLR}_{\mathrm{NS3}}) \times \ldots \times (1 - \mathrm{IPLR}_{\mathrm{NSn}}) \}$$

This relationship does not have limits on the parameter values, so it is preferred over other approximations, such as the simple sum of loss ratios. All measurements will use the same value of T_{max} (the waiting time to declare a packet lost).

The units of IPLR values are lost packets per total packets sent, with resolution of at least 10^{-9}. If lesser resolution is available in a value, the unused digits shall be set to zero.

8.2.3 Error packet ratio

For the IP Packet Error Ratio (IPER) performance parameter, the UNI-UNI performance may be estimated by inverting the probability of error-free packet transfer across *n* Network Sections, as follows:

$$\mathrm{IPER}_{\mathrm{UNI\text{-}UNI}} = 1 - \{ (1 - \mathrm{IPER}_{\mathrm{NS1}}) \times (1 - \mathrm{IPER}_{\mathrm{NS2}}) \times (1 - \mathrm{IPER}_{\mathrm{NS3}}) \times \ldots \times (1 - \mathrm{IPER}_{\mathrm{NSn}}) \}$$

This relationship does not have limits on the parameter values, so it is preferred over other approximations, such as the simple sum of packet error ratios.

The units of IPER values are errored packets per total packets sent, with resolution of at least 10^{-9}. If lesser resolution is available in a value, the unused digits shall be set to zero.

8.2.4 Provisional relationship for delay variation

The relationship for estimating the UNI-UNI Delay Variation (IPDV) performance from the Network Section values must recognize their sub-additive nature and is difficult to estimate accurately without considerable information about the individual delay distributions. If, for example, characterizations of independent delay distributions are known or measured, they may be convolved to estimate the combined distribution. This detailed information will seldom be shared among operators, and may not be available in the form of a continuous distribution. As a result, the UNI-UNI IPDV estimation may have accuracy limitations. Since study continues in this area, the estimation relationship given below has been specified on a provisional basis, and this clause may change in the future based on new findings or real operational experience.

The provisional relationship for combining IPDV values is given below.

The problem under consideration can be stated as follows: estimate the quantile t of the UNI-UNI delay T as defined by the condition:

$$\Pr(T < t) = p$$

Step 1

Measure the mean and variance for the delay for each of n Network Sections. Estimate the mean and variance of the UNI-UNI delay by summing the means and variances of the component distributions.

$$\mu = \sum_{k=1}^{n} \mu_k$$

$$\sigma^2 = \sum_{k=1}^{n} \sigma_k^2$$

Step 2

Measure the quantiles for each delay component at the probability of interest, $p = 0.999$. Estimate the corresponding skewness and third moment using the formula shown below, where $x_{0.999} = 3.090$ is the value satisfying $\Phi(x_{0.999}) = 0.999$ where Φ denotes the standard normal (mean 0, variance 1) distribution function.

$$\gamma_k = 6 \cdot \frac{x_p - \frac{t_k - \mu_k}{\sigma_k}}{1 - x_p^2}$$

$$\omega_k = \gamma_k \cdot \sigma_k^3$$

Assuming independence of the delay distributions, the third moment of the UNI-UNI delay is just the sum of the Network Section third moments.

$$\omega = \omega_1 + \omega_2 + \omega_3 + \ldots = \sum_{k=1}^{n} \omega_k$$

The UNI-UNI skewness is computed by dividing by σ^3 as shown below.

$$\gamma = {}^{\omega}\!/_{\sigma^3}$$

Step 3

The estimate of the 99.9-th percentile ($p = 0.999$) of UNI-UNI delay t is as follows.

$$t = \mu + \sigma \cdot \left\{ x_p - \frac{\gamma}{6}\left(1 - x_p{}^2\right) \right\}$$

where $x_p = x_{0.999} = 3.090$.

As stated earlier, the nature of the IPDV objective is the upper bound on the $1 - 10^{-3}$ quantile of IPTD minus the minimum IPTD (i.e., the distribution of IPDV is normalized to the minimum IPTD). The units of IPDV values are seconds, with resolution of at least 1 microsecond. If lesser resolution is available in a value, the unused digits shall be set to zero.

8.3 Impairment accumulation procedures

There are two principal ways in which the relationships above may be applied to estimate the UNI-UNI performance levels. Both are acceptable.

When the values from all network sections in the path are available in one place for computation, then they should be used in the relationships above as individual values. In a signalling protocol, the individual values would be collected from the source to the destination and communicated to the entity responsible for computation and action on the result.

The values may also be accumulated each time a new value is available. In this case, the relationships above are used to combine the cumulative estimate with the value from the current network (or router, if that is the basis of combination). The calculated estimate becomes the new cumulative value, and would be communicated further along the path to the destination.

9 Security

This Recommendation does not specify a protocol, and there are limited areas where security issues may arise. All are associated with verification of the performance objectives with measurement system implementations.

Measurement systems that assess the performance of networks to determine compliance with numerical objectives defined in this Recommendation must limit the measurement traffic to appropriate levels to avoid abuse (e.g., Denial of Service Attack). Parties participating in measurement activities, including Administrations or Operators of networks that carry the traffic, should agree in advance on acceptable traffic levels.

Systems that monitor user traffic for the purpose of measurement must maintain the confidentiality of user information.

Systems that attempt to make measurements may employ techniques (e.g., cryptographic hash) to determine if additional traffic has been inserted by an attacker appearing to be part of the population of interest.

Appendix I

ATM network QoS support of IP QoS

This appendix presents an analysis of mapping IP performance parameters on top of the ATM QoS Class objectives as specified in ITU-T Rec. I.356. The purpose of this analysis is to estimate IP level performance obtained when ATM is used as the underlying transport. Because there are no routers considered in this analysis, the IP performance numbers shown here are the best that can be expected. In scenarios where intermediate routers exist, the IP performance will be worse.

Table I.1/Y.1541 – IP Packet Loss Ratio (IPLR) values corresponding to ATM QoS service classes 1 and 2 (IP packet size 40 bytes; all errored packets are assumed lost)

ATM QoS class	Delivered ATM CER	Delivered ATM CLR	Resulting IPLR
1	4.00 E-06	3.00 E-07	4.30 E-06
2		1.00 E-05	1.40 E-05

Table I.2/Y.1541 – IP Packet Transfer Delay (IPTD) values for a flow over a national portion and an end-to-end flow

Network portion	IPTD resulting from ATM QoS class 1 (no delay from IP routers)
National Portion	~27.4 ms
End-to-End	400 ms

Note that Class 0 and Class 2 mean IPTD cannot be met on the 27 500 km reference connection of I.356.

The value of the Cell Error Ratio (CER) in the ATM classes is 4×10^{-6}. If IP packets are long (1500 bytes) and errored cells cause errored IP packets, the value of IP packet error ratio will be about 10^{-4}.

Cell Misinsertion Ratio (CMR) is currently specified as 1/day. The implications of CMR on SPR requires more study.

Appendix II

IP delay variation parameter definition considerations

This appendix discusses considerations for the definition of IPDV and the use of alternate statistical methods for the IPDV objective.

In order to provide guidance to designers of jitter buffers in edge equipment, the parameter(s) need to capture the effects of the following on IPDV:

- routine congestion in the network (high frequency IPTD variations);
- TCP windowing behaviour (low frequency IPTD variations);
- periodic and aperiodic variations in average network loading (low frequency IPTD variations);
- routing update effects on IPTD (instantaneous (and possibly large) changes in IPTD).

The current definition of IP Delay Variation is:

$$\mathrm{IPDV} = \mathrm{IPTD}_{\mathrm{upper}} - \mathrm{IPTD}_{\mathrm{min}}$$

where:

$\mathrm{IPTD}_{\mathrm{upper}}$ is the $1 - 10^{-3}$ quantile of IPTD in the evaluation interval

$\mathrm{IPTD}_{\mathrm{min}}$ is the minimum IPTD in the evaluation interval

The definition of IPDV is based on the reference events given in 6.2.2/Y.1540. Here, the nominal delay is based on the packet with the minimum one-way delay (as an alternative to the first packet, or the average of the population as the nominal delay).

The specification of the $1 - 10^{-3}$ quantile (equivalent to the 99.9th percentile) is influenced by the size of the packet sample in a 1 minute measurement interval and the IPLR objective $\leq 10^{-3}$, resulting in overall loss ratio objective of about 10^{-3}. Smaller quantiles would add more losses, as shown below.

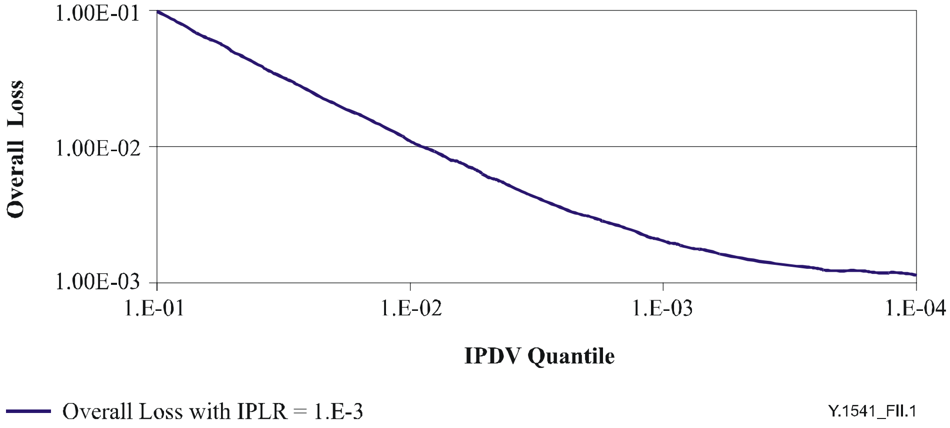


Figure II.1/Y.1541 – Effect of different IPDV quantiles on overall loss when IPLR = 0.001

An example alternate definition of IP Delay Variation is given here. IP Delay Variation may be defined as the maximum IPTD minus the minimum IPTD during a given short measurement interval.

$$\mathrm{IPDV} = \mathrm{IPTD}_{max} - \mathrm{IPTD}_{min}$$

where:

IPTD_{max} is the maximum IPTD recorded during a measurement interval

IPTD_{min} is the minimum IPTD recorded during a measurement interval

Several values of IPDV are measured over a large time interval, comprising several short measurement intervals. The 95th percentile of these IPDV values is expected to meet a desired objective. This is a simple and fairly accurate method for calculating IPDV in real-time. The actual value of the measurement interval is for further study. The measurement interval influences the ability of the metric to capture low and high frequency variations in the IP packet delay behaviour.

Appendix III

Example hypothetical reference paths for validating the IP performance objectives

This appendix presents the hypothetical reference paths considered in validating the feasibility of the end-to-end performance objectives presented in clause 5. These hypothetical reference paths (HRP) are examples only. The material in this appendix is not normative and does not recommend or advocate any particular path architectures.

Each packet in a flow follows a specific path. Any flow (with one or more packets on a path) that satisfies the performance objectives of clause 5 can be considered fully compliant with the normative recommendations in the main body of the Recommendation.

The end-to-end performance objectives are defined for the IP performance parameters corresponding to the IP packet transfer reference events (IPREs). The end-to-end IP network includes the set of Network Sections (NS) and inter-network links that provide the transport of IP packets transmitted from SRC to DST; the protocols below and including the IP layer (layer 1 to layer 3) within the SRC and DST may also be considered part of an IP network.

NOTE – For information concerning the effects on end-to-end quality as perceived by the user of the delay figures given by the presented hypothetical reference paths refer to Appendix VII.

III.1 Number IP nodes in the HRP

HRPs have similar attributes to the reference path of clause 5.

Network Sections are defined (in ITU-T Rec. Y.1540) as sets of hosts together with all of their interconnecting links that together provide a part of the IP service between a SRC and a DST, and are under a single (or collaborative) jurisdictional responsibility. Network Sections are synonymous with operator domains. Network Sections (NS) may be represented as clouds with Edge Routers on their borders, and some number of interior routers with various roles. In this case, HRPs are equivalent to the "path digest" of RFC 2330.

Each NS may be composed of IP Nodes performing Access, Distribution, and Core Roles, as illustrated in Figure III.1.

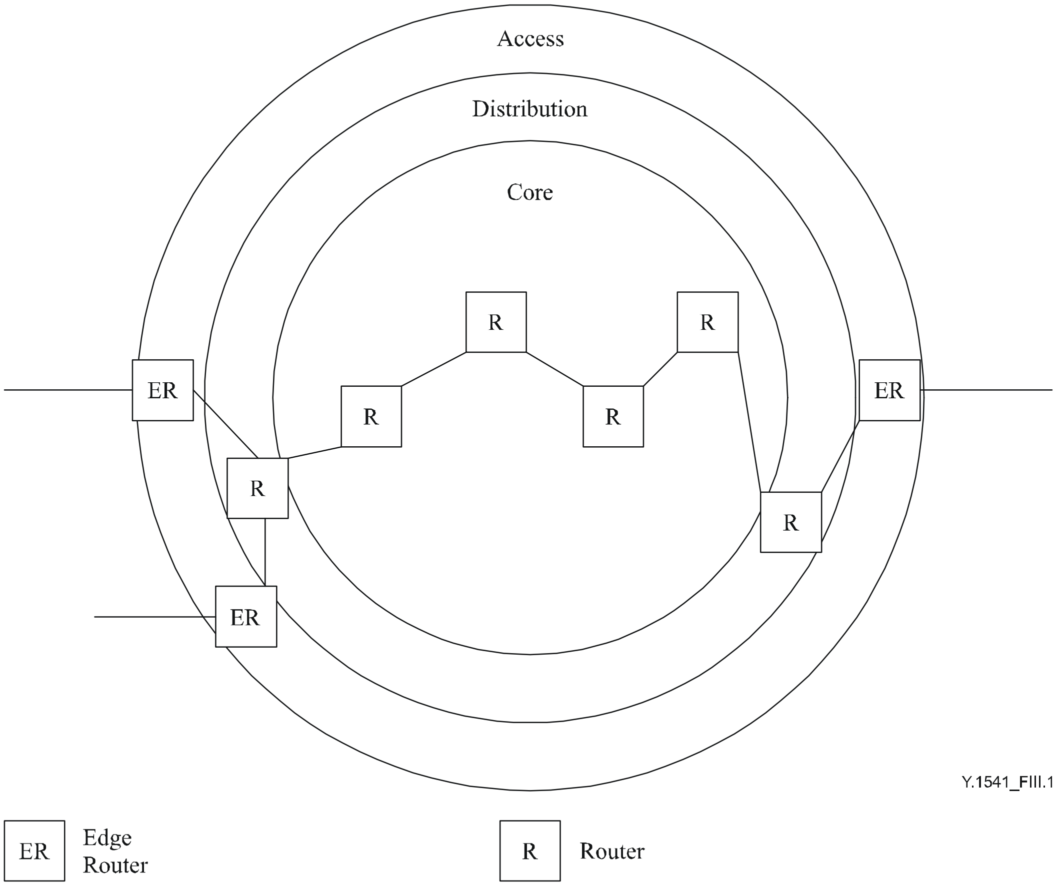


Figure III.1/Y.1541 – Role of IP nodes in a network section

Note that one or more routers are needed to complete each role, and the Core path illustrated has four routers in tandem. A path through a NS could encounter as few as three routers, or as many as eight in this example.

Router contribution to various parameters may vary according to their role. Edge Routers generally perform one of two roles, as Access Gateway Routers or Internetworking Gateway Routers.

Table III.1/Y.1541 – Examples of typical delay contribution by router role

Role	Average total delay (sum of queuing and processing)	Delay variation
Access gateway	10 ms	16 ms
Internetworking gateway	3 ms	3 ms
Distribution	3 ms	3 ms
Core	2 ms	3 ms

NOTE – Internetworking gateways typically have performance characteristics different from access gateways.

Route length calculation

If the distance-based component is proportional to the actual terrestrial distance, plus a proportional allowance for a typical physical-route-to-actual-distance ratio. The route length calculation used here is based on ITU-T Rec. G.826, and only for the long distances considered here. If D_{km} is the air-route distance between the two MPs that bound the portion, then the route length calculation is:

- if $D_{km} > 1200$, $R_{km} = 1.25 \times D_{km}$

The above does not apply when the portion contains a satellite hop.

III.2 Example computations to support end-end Class 0 and Class 1 delay

Class X Network Delay Computation (X = 0 through 4)

This clause calculates the IPTD for any path portion supporting a QoS class X flow. When a flow portion does not contain a satellite hop, its computed IPTD is (using the delay for optical transport given in ITU-T Rec. G.114):

IPTD (in microseconds) $\leq (R_{km} \times 5) + (N_A \times D_A) + (N_D \times D_D) + (N_C \times D_C) + (N_I \times D_I)$

In this formula:

- R_{km} represents the route length assumption computed above.
- $(R_{km} \times 5)$ is an allowance for "distance" within the portion.
- N_A, N_D, N_C, and N_I represent the number of IP access gateway, distribution, core and internetwork gateway nodes respectively; consistent with the network section example in Figure III.1.
- D_A, D_D, D_C, and D_I represent the delay of IP access gateway, distribution, core and internetwork gateway nodes respectively; consistent with the values for Class X (e.g., Table III.1).

Maximum IPDV may be calculated similarly.

As an example of this calculation, consider the following HRP. This path contains two IP networks, and an internetworking point.

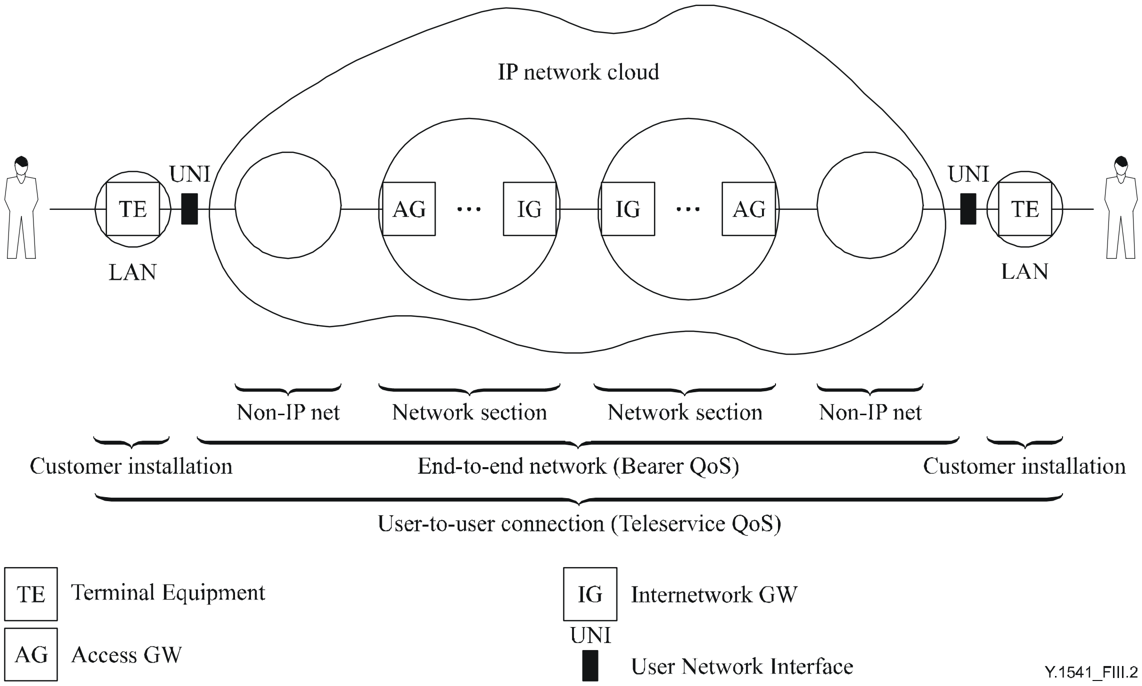


Figure III.2/Y.1541 – Hypothetical reference path for QoS class 0

Interior router configurations are not shown in the Hypothetical Reference Path (HRP) of Figure III.2. The number of Core and Distribution routers can be found in Table III.2.

Assumptions:

1) Distance used is approximately the span between Daytona Beach and Seattle (US Diagonal, longer than Lisbon to Moscow).
2) Access links are T1 capacity, others are larger than T1 (e.g., OC-3).
3) Largest Packet Size is 1500 bytes, and VoIP packet size is 200 bytes.
4) Non-IP networks are needed between the NI and Access GW.

Table III.2/Y.1541 – Analysis of example class 0 path

Element	Unit	IPTD/ Unit	Ave IPTD	IPDV/ Unit	Max IPDV
Distance	4070 km				
Route	5087.5 km		25		
Insertion Time	200 bytes (1500 bytes)		1 (8)		
Non IP Net 1			15		0
IP Net 1					
Access, N_A	1	10	10	16	16
Distribution, N_D	1	3	3	3	3
Core, N_C	2	2	4	3	6
Internetwork GW, N_I	1	3	3	3	3
IP Net 2					
Access, N_A	1	10	10	16	16
Distribution, N_D	1	3	3	3	3
Core, N_C	4	2	8	3	12
Internetwork GW, N_I	1	3	3	3	3
Non IP Net 2			15		0
Total, ms			**100**		**62**

Table III.2 gives the HRP configuration in terms of number and type of routers, distance, and contribution of all HRP components to delay (IPTD) and delay variation (IPDV). Note that the calculation of Maximum IPDV here is very pessimistic (assuming worst case addition of each node), and is therefore greater than the specification of IPDV in the body of this Recommendation.

III.3 Example end-end class 1 delay computation

Class 1 is available to support longer path lengths and more complex network paths. Using the same assumptions as described in Table III.2, but with a 12 000 km distance, the mean IPTD will be 150 ms, and an R-value of approximately 83 is possible.

In a second example, we add a transit IP Network Section, for a total of 3 NS.

Table III.3/Y.1541 – Example calculation for class 1 path

Element	Unit	IPTD/ Unit	Ave IPDT	IPDV/ Unit	Max IPDV
Distance	km				
Route	27 500 km		138		
Insertion Time	200 bytes (1500 bytes)		1 (8)		
Non IP Net 1			15		0
IP Net 1					
Access, N_A	1	10	10	16	16
Distribution, N_D	1	3	3	3	3
Core, N_C	2	2	4	3	6
Internetwork GW, N_I	1	3	3	3	3
IP Net 2					
Distribution, N_D	2	3	6	3	6
Core, N_C	4	2	8	3	12
Internetwork GW, N_I	2	3	6	3	6
IP Net 3					
Access, N_A	1	10	10	16	16
Distribution, N_D	1	3	3	3	3
Core, N_C	4	2	8	3	12
Internetwork GW, N_I	1	3	3	3	3
Non IP Net 2			15		0
Total, ms			**233**		**86**

Table III.3 gives the HRP configuration in terms of number and type of routers, distance, and contribution of all HRP components to delay (IPTD) and delay variation (IPDV).

III.4 Example computations to support end-end class 4 delay

Following the form of the calculation above, we can expand the number of NS having delay contributions given in Table III.1, or we can expand the contributions as follows:

Table III.4/Y.1541 – Class 4 delay contribution by router role

Role	Average total delay (sum of queuing and processing)
Access Gateway	200 ms
Internetworking Gateway	64 ms
Distribution	64 ms
Core	3 ms

Here, with a route length of 27 500 km, the average 1-way delay would be 884 ms (using the HRP with node configuration as described in Table III.2).

III.5 Loading within the HRP

The fraction of each transmission link occupied by active packets is one of the factors to be considered in the HRPs. The load levels at which the network will continuously operate is another factor.

III.6 Geostationary satellites within the HRP

The use of geostationary satellites was considered during the study of the HRPs. A single geostationary satellite can be used within the HRPs and still achieve end-to-end objectives on the assumption that it replaces significant terrestrial distance, multiple IP nodes, and/or transit network sections.

The use of low- and medium-Earth orbit satellites was not considered in connection with these HRPs.

When a path contains a satellite hop, this portion will require an IPTD of 320 ms, to account for low earth station viewing angle, low rate TDMA systems, or both. In the case of a satellite possessing on-board processing capabilities, 330 ms of IPTD is needed to account for on-board processing and packet queuing delays.

It is expected that most HRPs which include a geostationary satellite will achieve IPTD below 400 ms. However, in some cases, the value of 400 ms may be exceeded. For very long paths to remote areas, network providers may need to make additional bilateral agreements to improve the probability of achieving the 400 ms objective.

Appendix IV

Example calculations of IP packet delay variation

This appendix provides material to facilitate the calculation of the IP packet delay variation (IPDV) for those IP QoS classes where a rather strict value for the IPDV is specified, i.e., IP QoS class 0 and class 1.

For the calculations here it is assumed that a network operator provides a choice of different IP QoS classes also including QoS classes for which no IPDV objectives are specified. This mix of properties motivates the notion of "delay variation-sensitive" flows (e.g., QoS class 0 and class 1) and "delay variation-insensitive" flows (e.g., QoS classes 2, 3, 4, and 5). It is further assumed that an operator providing such a mix of QoS classes, makes a reasonable effort to separate the variation-sensitive from the variation-insensitive flows. Key elements in such an effort consist of a packet scheduling strategy and additional traffic control measures. For the calculations in this appendix, it is assumed that packets of variation-sensitive flows are scheduled with non-pre-emptive priority over packets from variation-insensitive flows, and that the scheduling within each of these two categories is FIFO.

NOTE – This simple assumption only serves the purpose to arrive at a 'calculable' model. Other packet scheduling strategies (such as Weighted Fair Queuing) or traffic control measures, are not excluded. It is further assumed that the performance of other approaches is either better, or not much worse than, the performance of the approach used for these calculations.

IV.1 Contributors to IP packet delay variation

The following factors are taken into account as the most significant contributors to IP packet delay variation (IPDV) for the variation-sensitive flows:

- Variable delay because the processing delay for the packet's forwarding decision (routing look-up) is not a single fixed value but may vary from packet-to-packet.
- Variable delay because the packet has to wait behind other variation-sensitive packets which arrived earlier.
- Variable delay because the packet has to wait for the service completion of a variation-insensitive packet which arrived earlier and is already in service.

IV.2 Models and calculation procedures to establish an upper bound to the IPDV

IV.2.1 Delay variation due to routing look-up

For an arriving packet, the router needs to establish the outgoing port to which the packet is to be forwarded, based on the IP address. The time required for this forwarding decision may vary from packet-to-packet.

High performance routers may cache recently used IP addresses to speed-up this process for subsequent packets. Then, all packets of a flow, except the first one, are expected to experience a short look-up delay and very small variation between them. Though, strictly, the longer delay of the first packet contributes to the IPDV, the exceptional delay of the first packet is disregarded in these calculations because it is a 'one off' event and its effect will vanish in flows with a relative long duration (e.g., a VoIP flow).

It is expected that the packet-to-packet variation in the routing look-up delay is not more than a few tens of microseconds in each router. For the calculations, the variability is assumed to be less than 30 μs per router.

Because there is little information available about the distribution of this delay component, the aggregated variability over several routers in tandem is set to the sum of the individual variabilities, i.e., statistical effects are not taken into account for this IPDV component.

IV.2.2 Delay variation due to variation-sensitive packets

A variation-sensitive packet will have to wait for other variation-sensitive packets to be serviced which arrived earlier (FIFO discipline). Each variation-sensitive flow is modelled as a continuous flow of packets with negligible 1-point IP packet delay variation, comparable to the concept of 'negligible CDV' used for an CBR stream of ATM cells (see ITU-T Rec. E.736).

For the calculations, it is further assumed that all variation-sensitive packets have a fixed size of 1500 bytes. This allows the well-known M/D/1 queuing model (see ITU-T Rec. E.736) to be applied for the calculation of this component in the packet delay variation. The fixed service time is determined by the assumed fixed packet size (1500 bytes) and the router's output link rate, e.g., 80.13 μs on an STM-1 link.

For the aggregation of this delay component over several routers in tandem, the convolution of the relevant delay distributions is to be used, taking into account different output link rates when applicable. The lower quantile is assumed to be zero, the higher $(1 - 10^{-3})$ quantile can be approximated accurately using large deviations theory, in particular the Bahadur-Rao estimate as worked out in [IFIP].

Figure IV.1 illustrates the result of such calculations. It shows the $(1 - 10^{-3})$ delay variation quantile for the aggregated delay component due to interference from variation-sensitive traffic, for different load levels of variation-sensitive traffic and for different numbers of router hops in tandem.

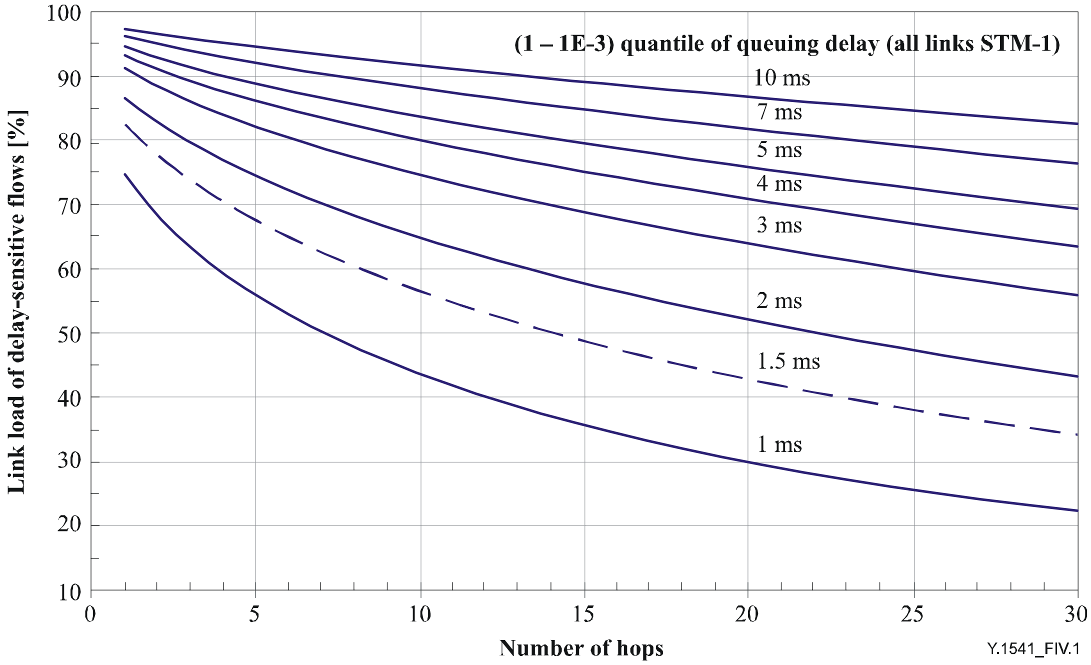


Figure IV.1/Y.1541 – The (1 – 10^{-3}) quantile of the aggregated queuing delay component due to variation-sensitive traffic for different levels of the variation-sensitive traffic and for different numbers of router hops in tandem

Figure IV.1 assumes that all links in the network are STM-1 and all links show the same load level for variation-sensitive traffic. If one or more links have a higher capacity than STM-1, the resulting end-to-end delay will be lower; if some links have a lower capacity, the resulting end-to-end delay will be higher. These effects can be calculated (see IV.2.4) but cannot easily be reflected in Figure IV.1.

Finally, it is assumed that in a network which supports both variation-sensitive and variation-insensitive traffic, the load of variation-sensitive traffic on a link is not more than 50% of the link to reflect the observed trend towards 'more data than voice'. Then, from Figure IV.1 it can be derived that this delay component contributes no more than about 2.48 ms to the IPDV on the path, even if the patch crosses a very high number of 25 STM-1 router hops.

IV.2.3 Delay variation due to a variation-insensitive packet

An arriving variation-sensitive packet does not pre-empt the servicing of a variation-insensitive packet which arrived earlier. Consequently, the variation-sensitive packet may experience a queuing component in each router bounded by the time it takes to serve a variation-insensitive packet.

For the calculation, it is assumed that each variation-sensitive packet experiences a random delay due to a variation-insensitive packet which is uniformly distributed between zero and the service time of maximum sized (1500 byte) variation-insensitive packets on the relevant output link rate. On an STM-1 output link this corresponds to a uniformly distributed delay between 0 and 80.13 μs in each router.

For the aggregation of this delay component over several routers in tandem, the convolution of the relevant delay distributions may be used, taking into account different output link rates when applicable. The lower quantile is assumed to be zero, the higher (1 – 10^{-3}) quantile can be calculated exactly. In most cases a good approximation is achieved by using an approximation by a normal (Gaussian) distribution or the worst case, whichever yields the smallest value. The (1 – 10^{-3}) quantile is found at ($\mu + 3.72 \cdot \sigma$).

IV.2.4 Aggregated delay variation for variation-sensitive packets

An upper bound to the IPDV on a HRP is found by adding the values calculated for each of the three components in IV.2.1 to IV.2.3.

NOTE – The resultant calculated value is expected to be higher than the value experienced in a real network. The following factors are noted:

- The addition of three quantile values yields a higher value than the actual delay quantile.
- The actual size of variation-sensitive packets (such as VoIP packets) is expected to be much smaller than the assumed size of 1500 byte. In addition, the load with variation-sensitive traffic on most links is expected to be smaller than the assumed value of 50%. Therefore, the actual queuing delay due to interference with variation-sensitive traffic is expected to be smaller than calculated.
- The actual distribution of variation-insensitive packets (e.g., TCP acks) also contains packets which are (much) smaller than the assumed size of 1500 byte. In addition, the total load (variation-sensitive plus variation-insensitive traffic) on most links is expected to be usually smaller than the assumed value of 100%. Therefore, the actual queuing delay due to interference with variation-insensitive traffic is expected to be smaller than calculated.

IV.3 Calculation examples

The following shows three examples for the calculation of the IPDV induced on a user-to-user HRP (see Figure II.1).

- An example where all links are relatively high speed (STM-1 or higher).
- An example where the links between customer and network and the links between network sections have a lower speed (E3 or T3).
- An example where the links between customer and network are low speed (e.g., 1.544 Mbit/s, T1).

IV.3.1 Example with STM-1 links

In this example, all links are assumed to be STM-1. The HRP between the network interfaces of the IP network cloud (see Figure III.2) consists of 12 router hops. Thus, the contributing factors to the IPDV on this path can be calculated as follows.

- Router look-up delay variation (see IV.2.1): 12 × 30 µs = 0.36 ms.
- Queuing delay variation due to variation-sensitive traffic (see Figure IV.1 for 50% load and 12 hops STM-1): ≈ 1.36 ms.
- Queuing delay variation due to variation-insensitive traffic (see IV.2.3): ≈ 9.01 × 80.13 µs = 0.72 ms.

Thus, the IPDV on this high link rate path can be expected to be smaller than **2.44 ms**.

IV.3.2 Example with E3 interconnecting links

In this example, all links are assumed to be STM-1 except the user-network links and the link between network sections which are assumed to be E3 (34 Mbit/s). The HRP between the network interfaces of the IP network cloud (see Figure III.2) consists of 12 router hops, of which 2 hops have the lower E3 bit rate. Thus, the contributing factors to the IPDV on this path can be calculated as follows.

- Router look-up delay variation (see IV.2.1): 12 × 30 µs = 0.36 ms.
- Queuing delay variation due to variation-sensitive traffic (for 50% load and 10 hops STM-1 plus 2 hops E3): ≈ 2.92 ms.
- Queuing delay variation due to variation-insensitive traffic (for 10 hops STM-1 plus 2 hops E3): ≈ 1.19 ms.

Thus, the IPDV on this mixed link rate path can be expected to be smaller than **4.47 ms**.

IV.3.3 Example with low rate access links

In this example, all links are assumed to be STM-1 except the user-network links which are assumed to be about 1.5 Mbit/s T1. The HRP between the network interfaces of the IP network cloud (see Figure III.2) consists of 12 router hops, of which 1 hop has the lower bit rate. In this case, the access link contribution is treated separately. The contributing factors to the IPDV on the high rate part of this path can be calculated as follows.

- Router look-up delay variation (see IV.2.1): 12 × 30 µs = 0.36 ms.
- Queuing delay variation due to variation-sensitive traffic (for 50% load and 11 hops STM-1): ≈ 1.29 ms.
- Queuing delay variation due to variation-insensitive traffic (for 11 hops STM-1): ≈ 8.364 × 80.13 µs = 0.67 ms.

Thus, the IPDV on this high link core path can be expected to be smaller than **2.32 ms**.

On the access links, the delay contribution due to interference with a variation-insensitive packet may be as much as 15.6 ms when two 1500 byte packets are served ahead of a variation-sensitive packet (one of these packets may be part of the delay sensitive flow). The contribution to the IPDV due to interference with other variation-sensitive flows highly depends on the number of these flows and on the actual packet sizes used.

Note that the number of variation-sensitive flows, and the related packet size on the low rate access link, is determined by applications selected by the end-users. Without some influence, the network operator will find himself in a difficult position to commit to a stringent value for the IPDV network performance objective in the presence of a low rate access link.

If the delay-sensitive traffic has constant packet size (each containing 20 ms of G.711 coded voice, consistent with Appendix III), and occupies no more than 50% of the access link, then delay can be estimated as follows. There may be up to 9 voice flows of 50 packet/s, each 160 byte payload plus 40 byte RTP, UDP and IP headers (each total 80 kbit/s).

- Queuing delay variation due to variation-sensitive traffic (for 46.9% load and 1 hop T1), using the M/D/1 queuing model shows that the delay contribution, due to those relatively small variation-sensitive packets on the access link, is 5.12 ms.
- Queuing delay variation due to variation-insensitive traffic (for 1 hop T1): 7.81 ms.

The contribution to the delay variation on the access link thus aggregates to 12.93 ms thus totalling to 15.25 ms. The access link contribution thus dominates the IPDV in this case.

IV.3.4 Example summary and conclusions

The calculation examples show that a network operator who makes a modest effort to support both variation-sensitive and variation-insensitive traffic can commit to rather stringent values for the IPDV on a long HRP where all links have a reasonably high rate (e.g., a mix of STM-1 and E3/T3 or higher). Committing to an IPDV value in the order of 10 ms leaves ample room for additional lower rate (E3/T3) links or for an additional network section.

If a low rate link (1.5 Mbit/s T1, or E1) is present, committing to any low IPDV value becomes difficult. The network operator has little or no control over the actual number of variation-sensitive flows and the actual packet size of the variation-sensitive packets. Therefore, the IPDV commitments made by the network in this case will be dominated by the access link, and will need to be considerably larger than 10 ms, as shown in Table 1. On the access link, the end-user has control over the number and type of flows designated for a delay sensitive class, and therefore over the resulting IPDV. Under the assumption that the access link is only modestly loaded (<50%) with variation-sensitive traffic and that the dominant size of those packets will be small compared to the

1500 byte maximum size, an additional allowance of **20 ms** for one low rate access link may be sufficient.

Appendix V

Material relevant to IP performance measurement methods

This appendix, which is for further study, will describe important issues to be considered as IP performance measurement methods are developed. It will describe the effects of conditions external to the sections under test, including traffic considerations, on measured performance.

The following conditions should be specified and controlled during IP performance measurements:

1) The exact sections being measured:
   - SRC and DST for end-to-end measurements;
   - MP bounding an NSE being measured;

     NOTE – It is not necessary to measure between all MP pairs or all SRC and DST pairs in order to characterize performance.
2) Measurement time:
   - how long samples were collected;
   - when the measurement occurred.
3) Exact traffic characteristics:
   - rate at which the SRC is offering traffic;
   - SRC traffic pattern;
   - competing traffic at the SRC and DST;
   - IP packet size.
4) Type of measurement:
   - in-service or out-of-service;
   - active or passive.
5) Summaries of the measured data:
   - means, worst-case, empirical quantiles;
   - summarizing period:
     - short period (e.g., one minute);
     - long period (e.g., one hour, one day, one week, one month).

Appendix VI

Applicability of the Y.1221 transfer capabilities and IETF differentiated services to IP QoS classes

This appendix addresses the applicability of the transfer capabilities defined in ITU-T Rec. Y.1221 in support of the Y.1541 IP QoS classes. It also specifies the relationship between Y.1221 transfer capabilities and IETF Differentiated Services Per Hop Behaviours consistent with what is specified in ITU-T Rec. Y.1221.

ITU-T Rec. Y.1221 defines three transfer capabilities (TC) called Dedicated Bandwidth (DBW), Statistical Bandwidth (SBW), and Best-effort (BE). Each of the service models specified as part of the definitions of the Y.1221 transfer capabilities currently specify a set of network performance parameters consistent with those specified in Table 1. Transfer capabilities defined in ITU-T Rec. Y.1221 can be used to meet the performance objectives of the six QoS classes defined in ITU-T Rec. Y.1541.

QoS classes 0 and 1 in Table 1 define bounds on both IP packet delay and delay variation, and on IP packet loss ratio. The transfer capability of Y.1221 that allows a traffic contract to specify bounds on IP Packet Delay/Delay variation and IP packet loss is the Dedicated Bandwidth transfer capability. QoS classes 2, 3 and 4 in Table 1 define bounds on IP packet loss ratio but not on IP packet delay variation. The transfer capability of Y.1221 that allows a traffic contract to specify bounds on both IP packet loss and delay is Under Study. QoS class 5 in Table 1 does not define bounds on IP packet loss ratio or IP packet delay/delay variation. The transfer capability of ITU-T Rec. Y.1221 that does not offer any QoS commitment is the Best-effort transfer capability. Table VI.1 specifies the mapping between Y.1541 QoS classes and Y.1221 transfer capabilities.

ITU-T Rec. Y.1221 provides a mapping between the three transfer capabilities it defines and the IETF Differentiated Services Per Hop behaviours that should be used in networks that use the DiffServ architecture. Table VI.1 specifies the mapping between Y.1221 transfer capabilities and IETF DiffServ Per Hop behaviours.

Table VI.1/Y.1541 – Association of Y.1541 QoS classes with Y.1221 transfer capabilities and differentiated services PHBs

Y.1221 transfer capabilities	Associated DiffServ PHBs	IP QoS class	Remarks
Best-effort (BE)	Default	Unspecified QoS class 5	A legacy IP service, when operated on a lightly loaded network may achieve a good level of IP QoS.
Delay-sensitive Statistical Bandwidth (DSBW)	AF	QoS classes 2, 3, 4	The IPLR objective only applies to the IP packets in the higher priority levels of each AF class. The IPTD applies to all packets.
Dedicated Bandwidth (DBW)	EF	QoS classes 0 and 1	

Appendix VII

Effects of network QoS on end-to-end speech transmission performance as perceived by the user

This appendix gives calculations of end-to-end speech quality using the objectives of Y.1541 network QoS Class 0 and Class 1 as a starting point. These objectives constrain key contributors to application performance that are often dominant in the calculations. When combined with the performance of well-designed customer equipment, it is believed that the objectives provided by this Recommendation do allow for the achievement of a high end-to-end speech transmission performance as perceived by the users. However, the material provided by the G.100-series of Recommendations should also be taken into account.

ITU-T Recs G.107, G.108, G.109, G.113 and G.114 are the key documents required to derive an estimation of the mouth-to-ear speech quality which can be achieved with the values of the relevant network QoS class.

ITU-T Rec. G.114 provides end-to-end limits and allocations for mean one-way delay, independent of other transmission impairments. The need to consider the combined effects of all impairments on overall transmission quality is addressed by ITU-T Rec. G.107, the so-called E-model as the common ITU-T Transmission Rating Model, which is the recommended ITU-T method for end-to-end speech transmission planning. ITU-T Rec. G.108 gives detailed examples on how to use the model to assess the transmission performance of connections involving various impairments, including delay; and ITU-T Rec. G.109 maps transmission rating predictions of the model into categories of speech transmission quality. Thus, while ITU-T Rec. G.114 provides useful information regarding mean one-way delay as a parameter by itself, ITU-T Rec. G.107 (and its companion ITU-T Recs G.108 and G.109) should be used to assess the effects of delay in conjunction with other impairments (e.g., distortions due to speech processing).

Furthermore, ITU-T Rec. G.101 (the transmission plan) and related Recommendations are currently undergoing a basic revision.

VII.1 Example VoIP Calculations with Y.1541 Class 0 network performance

As an example, a telephony Hypothetical Reference Endpoint (HRE) for speech media may be as shown below. Information flows from the Talker down through the protocol stack on the left, across the HRP, and up the protocol stack on the right to the Listener (only one sending direction is shown).

Talker		**Listener**
G.711 coder		G.711 decoder, Appendix I/G.711 Packet Loss Concealment
RTP 20 ms payload size		60 ms Jitter Buffer
UDP		UDP
IP		IP
	(lower layers)	

Figure VII.1/Y.1541 – Example VoIP hypothetical reference endpoint

Using the Hypothetical Reference Endpoint in Figure VII.1, endpoint delay is as below. These calculations follow from the formulas given in ITU-T Rec. G.1020 for Overall delay.

Table VII.1/Y.1541 – Endpoint delay analysis

	Delay, ms	Notes
Packet Formation	40	2 times frame size plus 0 look-ahead
Jitter Buffer, ave.	30	centre of 60 ms buffer
Packet Loss Conceal.	10	one PLC "frame"
Total, ms	**80**	

The endpoint delay calculated in Table VII.1 is consistent with the objective for a P.1010 Category B terminal. If we combine this mean endpoint delay with the Class 0 network delay, the total average delay for the user-to-user path is 100 + 80 = 180 ms. The example Class 0 reference path in Appendix III indicates that this delay may be achieved over a distance of 4070 km.

A 50 ms Customer Installation (1-way send and receive) is possible with a packet formation time of 10 ms and a 50 ms de-jitter buffer.

Table VII.2/Y.1541 – Low delay endpoint delay analysis

	Delay, ms	Notes
Packet Formation	20	2 times frame size plus 0 look-ahead
De-Jitter Buffer, ave.	25	centre of 50 ms buffer
Packet Loss Conceal.	0	"Repeat Previous" requires no additional delay
Other Equipment	5	
Total, ms	**50**	

The endpoint delay calculated in Table VII.2 is consistent with the objective for a P.1010 Category A terminal. The Class 0 path IPTD and Customer Installation delays sum to a 1-way mouth-to-ear transmission time of 150 ms, satisfying the needs of most applications (as per ITU-T Rec. G.114).

It must be noted that a de-jitter buffer's contribution to mouth-ear delay is based on the average time packets spend in the buffer, not the peak buffer size. Packets that encounter the minimum transfer delay will wait the maximum time in the de-jitter buffer before being played out as a synchronous stream, while the reverse is true for packets with the maximum accommodated transfer delay (these packets spend the minimum time in the de-jitter buffer). In this way, the de-jitter buffer compensates for transfer delay variations and ensures that packets can be removed according to a synchronous play-out clock. ITU-T Rec. G.1020 gives a more detailed description of the de-jitter buffer and its contribution to overall delay.

VII.2 Example VoIP Calculations with Y.1541 Class 1 network performance

Using the same assumptions and the Hypothetical Reference Path Endpoint delays of Table VII.1, and the Class 1 example path from Appendix III, the total average delay for a 27 500 km user-to-user path is 233 + 80 = 313 ms.

VII.3 Speech quality calculations for Y.1541 hypothetical reference paths

It is possible to estimate the speech quality of IP Networks using the G.107 Transmission planning tool, also known as the E-model.

Appendix III gives assumptions and configuration details of calculations for Network (UNI-UNI). The example endpoint assumptions and delay calculations above include codec (G.711), packet size, packet loss concealment, de-jitter buffer size, etc. Alternate speech codecs with lower bit rates, alternate packet sizes, and other variations are possible.

Figure VII.2 gives the reference connection for this analysis.

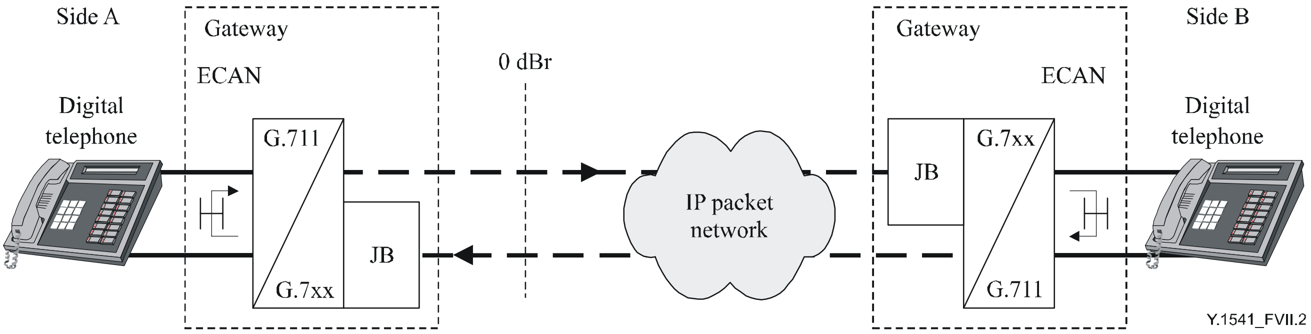


Figure VII.2/Y.1541 – Reference connection

Table VII.3 gives the E-model parameters used in the analysis.

Table VII.3/Y.1541 – E-model parameters

Parameters		**Model input values**		
Symbol	**Definition**	**G.107 default**	**Input values**	**Unit**
Nc	Electric Circuit Noise Referred to at the 0 dBr point	(–70)	–70.0	dBm0p
Pos	Room Noise (Send)	(35)	35.0	dB(A)
Por	Room Noise (Receive)	(35)	35.0	dB(A)
SLR	Send Loudness Rating	(8)	8.0	dB
RLR	Receive Loudness Rating	(2)	2.0	dB
Ds	D-factor (Send)	(3)	3.0	
LSTR	Listener's Sidetone Rating	(equ.)	18.0	dB
Nfor	Noise Floor	(–64)	–64.0	dBmp
STMR	Sidetone Masking Rating	(15)	15.0	dB
qdu	Quantizing Distortion Units	(1)	1.0	units
T	Mean One-Way Delay	(0)	**150.0**	ms
TELR	Talker Echo Loudness Rating	(65)	65.0	dB
WEPL	Weighted Echo Path Loss	(110)	110.0	dB
Ta	Absolute Delay from (S) to (R)	(0)	**150.0**	ms
Tr	Round-Trip Delay	(0)	**300.0**	ms
Ie	Equipment Impairment Factor	(0)	0.0	
Bpl	Packet Loss Robustness Factor	(1)	**4.8**	
Ppl	Random Packet Loss Probability	(0)	0.0	%
A	Expectation Factor	(0)	0.0	
Dr	D-factor (Receive)	(3)	3.0	

We have assumed the default values for all parameters, except T, Ta, and Tr. The mean absolute 1-way delay was calculated using 100 ms for network delay (UNI-UNI, conforming to the QoS Class 0 objective) and 50 ms for the end-terminal, including G.711 packetization and de-jitter buffer (100 + 50 = 150 ms = T = Ta = Tr/2). Here, R = 89.5.

Packet loss also influences speech quality. We include a column below where approximately 0.1% loss is combined with a Packet Loss Robustness Factor, Bpl = 4.8 when the packet loss concealment used with G.711 is Repeat 1, followed by silence. When using Appendix I/G.711 PLC, we take the Packet Loss Robustness Factor, Bpl = 25.1.

Appendix III also provides calculations showing longer mean network delays, and larger terminal delays. Table VII.4 summarizes the findings.

Table VII.4/Y.1541 – E-model results with Y.1541 hypothetical reference paths and end-terminals

Network, mean 1-way delay, ms	Terminal mean 1-way delay, ms	Total, mean 1-way delay, ms	Packet size, ms	Packet loss conceal.	R, no loss	R, with ~0.1% packet loss	Y.1541 QoS class
100	50	150	10	Rpt.1/Sil	89.5	87.6	0
100	80	180	20	G.711ApI	87.8	87.5	0
150	80	230	20	G.711ApI	81.9	81.5	1
233	80	313	20	G.711ApI	71.1	70.7	1

Appendix VIII

Effects of IP network performance on digital television transmission QoS

VIII.1 Introduction

This appendix details a part of the analysis behind the specification of Provisional Network QoS classes 6 and 7 in Table 3. The objective values were selected in order to support digital television transmission. The IP Packet Loss Ratio (IPLR) objective in classes 0 through 4 was insufficient to support this application, as stated in the previous version of this appendix.

VIII.2 Hypothetical Reference Endpoint (HRE) for high-bandwidth video signals

It is important to first establish a reference endpoint for video transport. The proposed endpoint is based on work done previously by the ATIS T1A1.3 committee as well as analysis of typical video transport endpoint models spanning both compressed and uncompressed video by the Video Services Forum. There may ultimately be a need to establish more than one HRE to allow point-to-point and point-to-multipoint transmission, but this analysis is restricted to the simpler case of the point-to-point HRE.

Sender		Receiver
Video (uncompressed SDI, multi- or single-compressed-MPEG-stream DVB-ASI, etc.), Multiple Audio Streams, Ancillary Data		Video (uncompressed SDI, multi- or single-compressed-MPEG-stream DVB-ASI, etc.), Multiple Audio Streams, Ancillary Data
Embedder		De-Embedder
Packetizer/Interleaver/FEC		FEC-1/De-inteleaver/De-packetizer
RTP		100 ms Jitter Buffer
UDP		UDP
IP		IP
	(Physical Layer)	

Figure VIII.1/Y.1541 – Hypothetical reference endpoint for digital television

The digital television transport uses an IP network where uncompressed video packets or MPEG-compressed video packets are encapsulated into either UDP/IP or RTP/UDP/IP. We assume that RTP/UDP/IP is the protocol used and that the following protocol overhead applies:

IP packet length = (7 × 188-Byte MPEG packets) + RTP/UDP/IP packet overhead

The following clauses describe three profiles of video services and give a rationale for the deployment of error correction mechanisms in IP networks to guarantee the appropriate level of quality and reliability.

VIII.3 Service profiles and end-to-end packet performance requirements

The technical requirements for this appendix will be limited to three service profiles: Contribution Services Profile, Primary Distribution Service Profile and Access Distribution Service Profile. These three profiles encompass the vast majority of the video industry's applications and needs. We also present the performance requirements for these profiles in terms of packet loss at three different viewer quality levels, or hit rates.

VIII.3.1 Contribution video services profile

Contribution services typically have the highest performance and can vary from uncompressed to mildly compressed video and audio signals. Contribution connections allow exchange of content by a network or its affiliates for further use, e.g., for bringing signals back from fixed, temporary, or remote locations to the studio for editing or immediate rebroadcast. In those scenarios, for long-haul applications, terrestrial fibre, microwave or satellite infrastructure endpoint connections can be utilized.

Contribution can also mean the outbound delivery of signals from the main network studio to network affiliates for re-broadcasting and typically employs satellite or long-haul terrestrial network services. Today, these outbound connections are provided by way of fixed or on-demand private leased lines (fibre), or in certain, less-extensive applications, ATM services offering DS-3, OC-3, or OC-12 bandwidth.

In addition to those real-time applications, sometimes IP services are used for non-real-time file exchange between video and audio servers and for monitoring and control of remote systems. As the same user may use their IP service for contribution video and file transfer, the contribution service profile also easily accommodates file transfer and remote control.

VIII.3.2 Primary distribution video service profile

Distribution means delivery of video and audio content either directly to the consumer or to cable head-ends for transmission through a cable television plant. In these applications, typically a lower signal quality (lower data rates) is needed, as little additional signal processing will be applied. Traditionally for these applications, terrestrial or satellite services are used. There are two types of distribution signals, primary and access. Primary distribution connections are feeds from the local affiliate to the cable head-end or to the television transmission tower, and ordinarily, these connections are comparable to, or slightly lower in quality than, contribution connections. Primary distribution may be provided by satellite, short-haul terrestrial microwave, or fibre optic connection. Access distribution involves the delivery of the content from the cable head-end to the final consumer over the cable television plant or through the air in the form of a broadcast emission from the television transmitter tower antenna. The VSF recommends that 40 Mbit/s represent the bit rate of this type of service.

VIII.3.3 Access distribution service profile

Access distribution service profile is defined as TV services currently being delivered by Cable and Satellite networks. Since the quality achieved by these networks is somewhat subjective, this contribution will characterize quality as an upper bound on video data errors (due to network) in a specific window of time.

VIII.3.4 Performance requirements for the service profiles

Quality of service for this application will be given in terms of actual number of errors (performance hits) in a specific time period. Table VIII.1 was constructed based on recommendations from active members of the Video Services Forum and represents expected error rates that service providers (e.g., DirecTV), as well as users (e.g., Fox Sports Network), would demand.

Table VIII.1/Y.1541 – Digital television loss/error ratio recommendations

Profile (Typical bit rate)	One performance hit per 10 days	One performance hit per day	10 performance hits per day
Contribution (270 Mbit/s)	4×10^{-11}	4×10^{-10}	4×10^{-9}
Primary Distrib. (40 Mbit/s)	3×10^{-10}	3×10^{-9}	3×10^{-8}
Access Distrib. (3 Mbit/s)	4×10^{-9}	4×10^{-8}	4×10^{-7}

This table assumes all lost packets cause a performance hit (possibly visible or audible impairment), and seven MPEG TS packets are encapsulated in a single IP packet. The required packet loss ratio is given at the intersection of a hit rate and profile. For example, Access Distribution allowing a quality level of 1 performance hit per day requires a packet loss ratio of 4×10^{-8}.

VIII.4 Forward Error Correction (FEC)/Interleaving to improve UNI-UNI performance

An IP network conforming to QoS Classes 6 or 7 is not capable of providing the packet loss rates required for the profiles above, and edge equipment is needed to correct for packet errors, packet losses and reordered packets. We assume the service uses FEC/Interleaving as defined by the Pro-MPEG Forum COP-3 recommendation (Code of Practice) and as reflected in Table VIII.2.

Table VIII.2/Y.1541 – FEC/Interleaving to achieve desired end-to-end hit rates

	Minimal Correction	**Moderate Correction**	**High Correction**
Minimum Network Performance			
Loss Distance (Packets)	100	50	50
Loss Period (Packets)	5	5	10
Applied FEC			
FEC L, D	5, 20	5, 10	10, 5
FEC Overhead (%)	5	10	20
Resulting Video Performance Quality?	High	High	High

Note that the specification of network performance above utilizes two new terms. Loss Distance (LD) and Loss Period (LP), defined in RFC 3357, are packet loss pattern parameters. LP defines the maximum number of consecutive packets that can be lost, while LD defines the minimum number of good packets that must arrive between lost packets for the algorithm to properly correct for losses. The LD and LP values describe the minimum network performance correctable by the corresponding FEC in the same column. The FEC is defined by Length (L) and Depth (D) algorithm parameters that define the robustness of the method.

Correction of network impairments is not free, as it consumes additional bandwidth. The overhead values in the table represent three levels of robustness, where 5% represents minimal correction, 10% represents moderate correction and 20% represents the highest amount of correction. Note that the more robust the algorithm we choose, the higher the overhead. It is the VSF's position that these three values encompass the majority of needs in the industry.

As an example, a 2 Mbit/s video service requiring minimal correction would be configured with (L, D) settings of (5, 20). This would generate an extra 100 kbit/s (5% of 2 Mbit/s) of network traffic for the FEC packets, resulting in a total data rate of 2.1 Mbit/s. Similarly, a 270 Mbit/s service requiring high correction would be configured with (L, D) values of (10, 5) which would generate an additional 54 Mbit/s of network traffic, resulting in an aggregate rate of 324 Mbit/s.

VIII.5 Laboratory assessment of Forward Error Correction (FEC)/Interleaving effectiveness

Laboratory test results with FEC/Interleave (5, 50) indicate that:

– UNI-UNI loss ratio of 10^{-4} improves to 1.5×10^{-8} (covers most of the Access Profile);

– UNI-UNI loss ratio of 10^{-5} improves to 2×10^{-10} (covers most Profiles).

It was concluded that an IP network with UNI-UNI IPLR and IPER conforming to Table 3, Class 6 or 7 will support the digital television application described above, providing that the appropriate FEC/Interleaving is applied.

VIII.6 Additional performance parameters

The Video Services Forum concluded that the values for IPTD and IPDV specified in Table 3, Classes 6 and 7 are sufficient for digital television transport.

Appendix IX

Effects of network QoS on end-to-end data transmission performance using TCP

IX.1 Introduction

This appendix details a part of the analysis behind the specification of Provisional Network QoS Classes 6 and 7 in Table 3. The objective values were selected in order to support applications using the reliable byte stream transfer services of the Transmission Control Protocol (TCP) [RFC793] at the largest possible data rate. The existing IP Packet Loss Ratio (IPLR) objective (in classes 0 through 4) supports TCP with the limitations of widely deployed legacy settings, or assumes that some bottleneck will be encountered beyond the UNI-UNI path.

There are two key factors that limit TCP transfer capacity:

1) The **congestion-aware flow-control mechanisms** infer that congestion has been encountered on the path when packet loss occurs. In response to loss, the flow-control cuts the sending window in half, and allows linear increase when a full window of packets has been transferred successfully. Thus, **packet loss can limit capacity**.
2) The **maximum window size** may be limited by the sender or receiver TCP settings, or by the operating system itself (limiting the amount of memory available to a specific application for buffering network data). This is the classic Delay Bandwidth Product, where the transmission rate is given as **one window of octets per round-trip time** (for acknowledgement).

Given that packet transfer time is usually dominated by propagation time, the goal of the analysis was to determine an objective for IPLR that provides very high TCP transfer capacity when other factors, such as window size or bottleneck bandwidth, do not encumber the process. A packet loss ratio of 10^{-5} was selected for classes 6 and 7, and the analysis below shows what capacities can be achieved.

IX.2 Model of TCP performance

The basis for this study is the model of TCP Reno [RFC2001] developed and verified by Padhye et al [Padhye98]. Their model can be approximated as follows:

$$B(p) \approx \min\left(\frac{W_{\max}}{RTT}, \frac{1}{RTT\sqrt{\frac{2bp}{3}} + T_0 \min\left(1, 3\sqrt{\frac{3bp}{8}}\right) p\left(1 + 32p^2\right)} \right)$$

where:

$B(p)$: approximate model of TCP throughput [packet/s]

W_{max}: maximum window buffer size of receiver [packets]

RTT: Round Trip Time [sec]

b: number of packets that are acknowledged by a received ACK

p: probability that a packet is lost

T_0: time-out for re-transmitting an unacknowledged (lost) packet [sec]

There are many combinations of TCP features, and the different combinations are sometimes named according to the meeting place where they were agreed (Vegas, Tahoe, and Reno). A discussion of TCP features is available in [Morton98], and many other references. For an even simpler TCP model with a single fitting parameter that is useful across versions, see [Mathis97].

IX.3 TCP Hypothetical Reference Endpoint (HRE)

Various appendices of this Recommendation specify Hypothetical Reference Endpoints (HRE) and pair them with Hypothetical Reference Paths to assess the user application quality levels that the network performance objectives can support. We define a TCP Hypothetical Reference Endpoint below.

Sending Application		**Receiving Application**
TCP Reno Max Window = 16 kbyte, 64 kbyte, or 256 kbyte T0 timeout = 1 s Large Windows option		TCP Reno Max Window = 16 kbyte, 64 kbyte, or 256 kbyte b = 1 ACK/2 packets Large Windows option
IP		IP
	(lower layers)	

Figure IX.1/Y.1541 – TCP hypothetical reference endpoint

We assume that the sending application supplies a continuous byte stream with no idle intervals, and that the receiving host contribution to RTT is insignificant. Note that the sending and receiving Max Window sizes will vary in the analysis that follows.

IX.4 Observations

Figure IX.2 shows the estimate of "Legacy" TCP Reno capacity vs. round-trip time (including host processing) and packet loss. The 3-dimensional surface is scribed with lines that correspond to Round Trip Times (RTT) of 20, 40, 100, 200, 400, 1000, 2000, and 4000 ms, intersecting with lines at Loss Ratios of 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6}. The height of the surface indicates the TCP capacity in bits/second, and the surface color changes when it crosses a labeled capacity level.

We note that none of the long-delay mitigations have been applied here, such as RFC 1323 Large Windows or RFC 2018 Selective Acknowledgements (SACK).

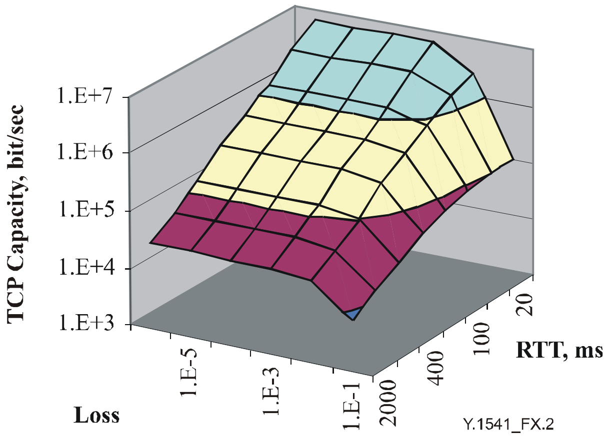


Figure IX.2/Y.1541 – TCP capacity with 16 kbyte Window ("Legacy")

An 8 kbyte or 16 kbyte window is the default setting for many legacy TCP implementations. Figure IX.2 shows that Packet Loss > 10^{-3} has an affect on capacity, but the window size limitation dominates the capacity vs. loss performance over a wide range of Round-Trip Times (RTT). Therefore, the IPLR objective <10^{-3} is sufficient under these circumstances, and network QoS classes 2, 3, and 4 will produce satisfactory capacity.

Although transfer capacities in the order of 10 Mbit/s are possible at very low RTT, packet transfer time also influences capacity for the "legacy" TCP sender-receiver pair.

Figure IX.3 shows the TCP Reno Capacity when the maximum window size is set to 64 kbytes. This is usually possible with simple tuning procedures, but the overwhelming majority of IP network users do not attempt tuning, or have no need. Users who want to realize the full potential of Broadband Access while reducing the transfer time for extremely large files (e.g., Linux distribution ISO-files contain 700 Mbyte CD-ROM images) may seek the benefits of tuning.

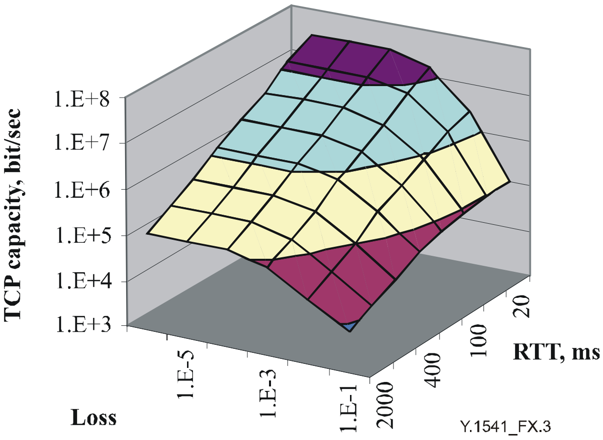


Figure IX.3/Y.1541 – TCP capacity with 64 kbyte Window

A 64 kbyte window is the maximum setting for standard TCP implementations that do not enable RFC 1323 Large Windows. Figure IX.3 shows that Packet Loss > 10^{-4} has an affect on capacity, but the window size limitation dominates the capacity from there on.

Figure IX.4 shows the TCP Reno Capacity when the maximum window size is set to 256 kbytes. This is possible with many operating systems, and the TCP Large Windows option must be available.

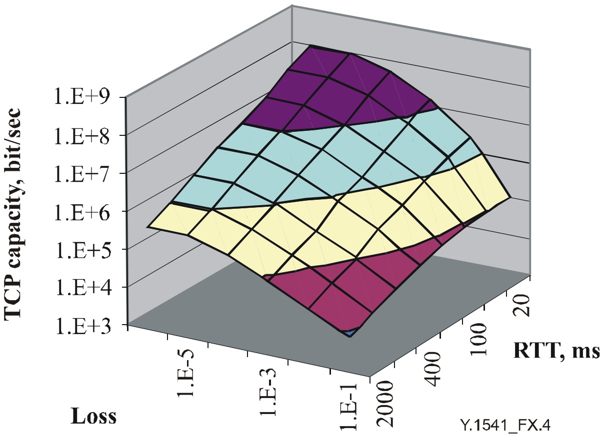


Figure IX.4/Y.1541 – TCP capacity with 256 kbyte Window (and RFC 1323)

Figure IX.4 shows that Packet Loss > 10^{-5} has an affect on capacity, but the window size limitation dominates the capacity vs. loss performance characteristic beyond that point. Therefore, these are circumstances where the new provisional classes (with IPLR objective < 10^{-5}) are needed to support maximum capacity.

Transfer capacities on the order of 100 Mbit/s are possible at very low RTT, and the Large Window option (RFC 1323) reduces the negative affect of RTT on capacity.

IX.5 Summary of TCP capacity estimates

Table IX.1 provides a numerical summary of Figures IX.2 through IX.4 at the values of delay and loss ratio appearing in the objectives.

Table IX.1/Y.1541 – Summary of TCP capacity estimates, bits/s

Window size	Packet Loss, p	IPTD = RTT/2 = 100 ms	IPTD = RTT/2 = 400 ms
16 kbytes	10^{-3}	640 000	160 000
	10^{-5}	640 000	160 000
64 kbytes	10^{-3}	**1 624 887**	**409 640**
	10^{-5}	2 560 000	640 000
256 kbytes	10^{-3}	**1 624 887**	**409 640**
	10^{-5}	10 240 000	2 560 000

Note that **Bold** values are limited by packet loss ratio, otherwise window size limits capacity. Packet Loss ratio of 10^{-5} does not limit capacity at any window size examined, clearly showing the benefits of the new network QoS classes.

BIBLIOGRAPHY

- ITU-T Recommendation J.241 (2005), *Quality of service ranking and measurement methods for digital video services delivered over broadband IP networks*.
- ITU-T Recommendation P.911 (1998), *Subjective audiovisual quality assessment methods for multimedia applications*.
- ETSI TIPHON TR 101 329 – *Part 2, Quality of Service (QoS) Classes*.
- IETF RFC 768 (STD-6) (1980), *User Datagram Protocol*.
- IETF RFC 792 (STD-5) (1981), *Internet Control Message Protocol*.
- IETF RFC 793 (STD-7) (1981), *Transmission Control Protocol − DARPA Internet program Protocol specification*.
- IETF RFC 919 (STD-5) (1984), *Broadcasting Internet datagrams*.
- IETF RFC 922 (STD-5) (1984), *Broadcasting Internet datagrams in the presence of subnets*.
- IETF RFC 950 (1985), *Internet Standard Subnetting Procedure (updates RFC 792)*.
- IETF RFC 959 (STD-9) (1985), *File Transfer Protocol (FTP)*.
- IETF RFC 1305 (1992), *Network Time Protocol (v3) − Specification, Implementation and Analysis*.
- IETF RFC 1786 (1995), *Representation of IP Routing Policies in a Routing Registry*.
- IETF RFC 1812 (1995), *Requirements for IP Version 4 Routers*.
- IETF RFC 1889 (1996), *RTP: A transport protocol for real-time applications*.
- IETF RFC 2018 (1996), *TCP Selective Acknowledgment Options*.
- IETF RFC 2330 (1998), *Framework for IP performance metrics*.
- IETF RFC 2474 (1998), *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*.
- IETF RFC 2475 (1998), *An Architecture for Differentiated Services*.
- IETF RFC 2597 (1999), *Assured Forwarding PHB Group*.
- IETF RFC 2598 (1999), *An Expedited Forwarding PHB*.
- IETF RFC 2679 (1999), *A One-way Delay Metric for IPPM*.
- IETF RFC 2680 (1999), *A One-way Packet Loss Metric for IPPM*.
- IETF RFC 2681 (1999), *A Round-trip Delay Metric for IPPM*.
- IETF RFC 2733 (1999), *An RTP payload format for generic forward error correction*.
- IETF RFC 3086 (2001), *Definition of Differentiated Services Per Domain Behaviours and Rules for their Specification*.
- IETF RFC 3357 (2002), *One-way loss pattern sample metrics*.
- IETF RFC 3432 (2002), *Network Performance Measurement with Periodic Streams*.

[RFC1323] IETF RFC 1323 (1992), *TCP Extensions for High Performance*.

[RFC2001] IETF RFC 2001 (1997), *TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms*.

– IFIP: MANDJES (Michel), VAN DER WAL (Kees), KOOIJ (Rob), BASTIAANSEN (Harrie): End-to-end delay models for interactive services on a large-scale IP network, *Proceedings* (edited by Guido H. Petit) *of the seventh IFIP workshop on Performance Modelling and Evaluation of ATM Networks: IFIP ATM'99*, Paper 42, Antwerp, Belgium, June 1999.

[Mathis97] MATHIS (M.), SEMKE (J.), MAHDAVI (J.), OTT (T.): The Macroscopic Behaviour of TCP Congestion Avoidance Algorithm, *Computer Communications Review*, Vol. 27, No. 3, July 1997, ISSN# 0146-4833. http://www.psc.edu/networking/papers/model_ccr97.ps

[Morton98] MORTON (A.C): Transmission Control Protocol Overview, T1A1.3/98-015; PROJECT#: T1A1-14; MEETING DATE: 03/16/98. ftp://ftp.t1.org/pub/t1a1/98-t1a1.3/8a130150.doc

[Padhye98] PADHYE (J.), FIROIU (V.), TOWSLEY (D.), AND KUROSE (J.): Modelling TCP Throughput: a Simple Model and its Empirical Validation, *SIGCOMM 1998*. ftp://gaia.cs.umass.edu/pub/Padhye-Firoiu98:TCP-throughput.ps.Z

– PADHYE (J.), FIROIU (V.), TOWSLEY (D.), KUROSE (J.): Modelling TCP Reno Performance: A Simple Model and its Empirical Validation, *IEANEP*, Vol. 8, No. 2, pp. 133-145, April 2000.

– Study Group 12 Delayed Contribution D15: The effect of Packet Losses on Speech Quality, C. Karlsson, *Telia AB*, Feb. 2001.

– Study Group 12 Delayed Contribution D22: A Framework for Setting Packet Loss Objectives for VoIP, J. Rosenbluth, *AT&T*, Oct. 2001.

– T1 Standard* T1.522-2000, Quality of Service for Business Multimedia Conferencing.

---

* T1 Standards are maintained since November 2003 by ATIS.

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, 2006