International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.1540

(12/2019)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

Internet protocol aspects – Quality of service and network performance

Internet protocol data communication service – IP packet transfer and availability performance parameters

Recommendation ITU-T Y.1540

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

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
IPTV over NGN	Y.1900–Y.1999
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
Enhancements to NGN	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Network control architectures and protocols	Y.2500–Y.2599
Packet-based Networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999
FUTURE NETWORKS	Y.3000–Y.3499
CLOUD COMPUTING	Y.3500–Y.3999
INTERNET OF THINGS AND SMART CITIES AND COMMUNITIES	
General	Y.4000–Y.4049
Definitions and terminologies	Y.4050–Y.4099
Requirements and use cases	Y.4100–Y.4249
Infrastructure, connectivity and networks	Y.4250–Y.4399
Frameworks, architectures and protocols	Y.4400–Y.4549
Services, applications, computation and data processing	Y.4550–Y.4699
Management, control and performance	Y.4700–Y.4799
Identification and security	Y.4800–Y.4899
Evaluation and assessment	Y.4900–Y.4999

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

Recommendation ITU-T Y.1540

Internet protocol data communication service – IP packet transfer and availability performance parameters

Summary

Recommendation ITU-T Y.1540 defines the parameters that may be used in specifying and assessing the performance of speed, accuracy, dependability and availability of IP packet transfer of regional and international Internet protocol (IP) data communication services. The defined parameters apply to an end-to-end, point-to-point IP service and to the network portions that provide, or contribute to the provision of, such a service in accordance with the normative references specified in clause 2. Connectionless transport is a distinguishing aspect of the IP service that is considered in this Recommendation.

Following over 20 years as an in-force Recommendation, the 2019 edition recognizes many changes in the design of IP services and in the protocols employed by end users. It introduces the new Annex A that defines IP-layer capacity parameters in ways that cater toward assessment, and provides requirements for methods of measurement of IP-layer capacity. This new annex is the result of years of study, and application of ITU-T Study Group 12 principles of accurately evaluating performance parameters and methods of measurement against a "ground truth" reference in laboratory and field measurements. Flow-related throughput parameters and methods of measurement (reliable delivery transport), remain for further study, and the text makes a clear distinction between this IP-layer capacity parameters. In the same way, parameters describing performance of a specific reliable transport layer protocol (TCP) remain for further study, and recognize that reliable transport protocols for the Internet are constantly changing and the subject of ongoing research.

Annex B provides a second, more capable search algorithm for the IP capacity method of measurement defined in Annex A.

History

Edition	Recommendation	Approval	Study Group	Unique ID[*]
1.0	ITU-T I.380	1999-02-26	13	11.1002/1000/4573
1.0	ITU-T Y.1540	1999-02-26	13	11.1002/1000/5302
2.0	ITU-T Y.1540	2002-12-14	13	11.1002/1000/6189
2.1	ITU-T Y.1540 (2002) Amd. 1	2003-08-01	13	11.1002/1000/6975
3.0	ITU-T Y.1540	2007-11-13	12	11.1002/1000/9270
3.1	ITU-T Y.1540 (2007) Amd.1	2009-03-19	12	11.1002/1000/9727
4.0	ITU-T Y.1540	2011-03-01	12	11.1002/1000/11079
4.1	ITU-T Y.1540 (2011) Amd.1	2016-01-21	12	11.1002/1000/12761
5.0	ITU-T Y.1540	2016-07-29	12	11.1002/1000/12975
6.0	ITU-T Y.1540	2019-12-05	12	11.1002/1000/13933
6.1	ITU-T Y.1540 (2019) Amd. 1	2020-02-06	12	11.1002/1000/14161

[*] To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

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

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

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

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

NOTE

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

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

INTELLECTUAL PROPERTY RIGHTS

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

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

Table of Contents

Page

Page

Recommendation ITU-T Y.1540

Internet protocol data communication service – IP packet transfer and availability performance parameters

1 Scope

This Recommendation defines the parameters that may be used in specifying and assessing the performance of speed, accuracy, dependability and availability of IP packet transfer of regional and international Internet Protocol (IP) data communication services. The defined parameters apply to the end-to-end, point-to-point IP service and to the network portions that provide, or contribute to the provision of, such a service in accordance with the normative references specified in clause 2. Connectionless transport is a distinguishing aspect of the IP service that is considered in this Recommendation.

For the purpose of this Recommendation, end-to-end IP service refers to the transfer of user-generated IP datagrams (referred to in this Recommendation as IP packets) between two end hosts as specified by their complete IP addresses. This differs from the boundaries implied by the phrase "end-to-end" in some other Recommendations. For example, [ITU-T P.10] defines end-to-end quality as related to the performance of a communication system, including all terminal equipment. For voice services, end-to-end is equivalent to mouth-to-ear quality.

NOTE 1 – This Recommendation defines parameters that can be used to characterize an IP service provided using Internet Protocol version 4 (IPv4) and Internet Protocol version 6 (IPv6); applicability or extension of this Recommendation to other protocols (e.g., resource reservation protocol (RSVP)) is for further study.

NOTE 2 – Recommendations for the performance of a point-to-multipoint IP service are currently under development.

The [ITU-T Y.1540] performance parameters and methods of measurement (where specified) are intended to be used in planning and offering both regional and international IP services. The intended users of this Recommendation include IP service providers, equipment manufacturers, end users (subscribers to IP services) and others who seek to perform assessments and/or monitor IP services. This Recommendation may be used by service providers in the planning, development and assessment of IP services that meet user performance needs, by equipment manufacturers as performance information that will affect equipment design, and by end users in evaluating IP service performance.

The scope of this Recommendation is summarized in Figure 1. The IP service performance parameters are defined on the basis of IP packet transfer reference events (IPREs) that may be observed at measurement points (MPs) associated with specified functional and jurisdictional boundaries. For comparability and completeness, IP service performance is considered in the context of the 3×3 performance matrix defined in [ITU-T I.350]. Three protocol-independent communication functions are identified in the matrix: access, user information transfer and disengagement. Each function is considered with respect to three general performance concerns (or "performance criteria"): speed, accuracy and dependability. An associated two-state model provides a basis for describing IP service availability.

NOTE 3 – In this Recommendation, the user information transfer function illustrated in Figure 1 refers to the attempted transfer of any IP packet, regardless of its type or content.

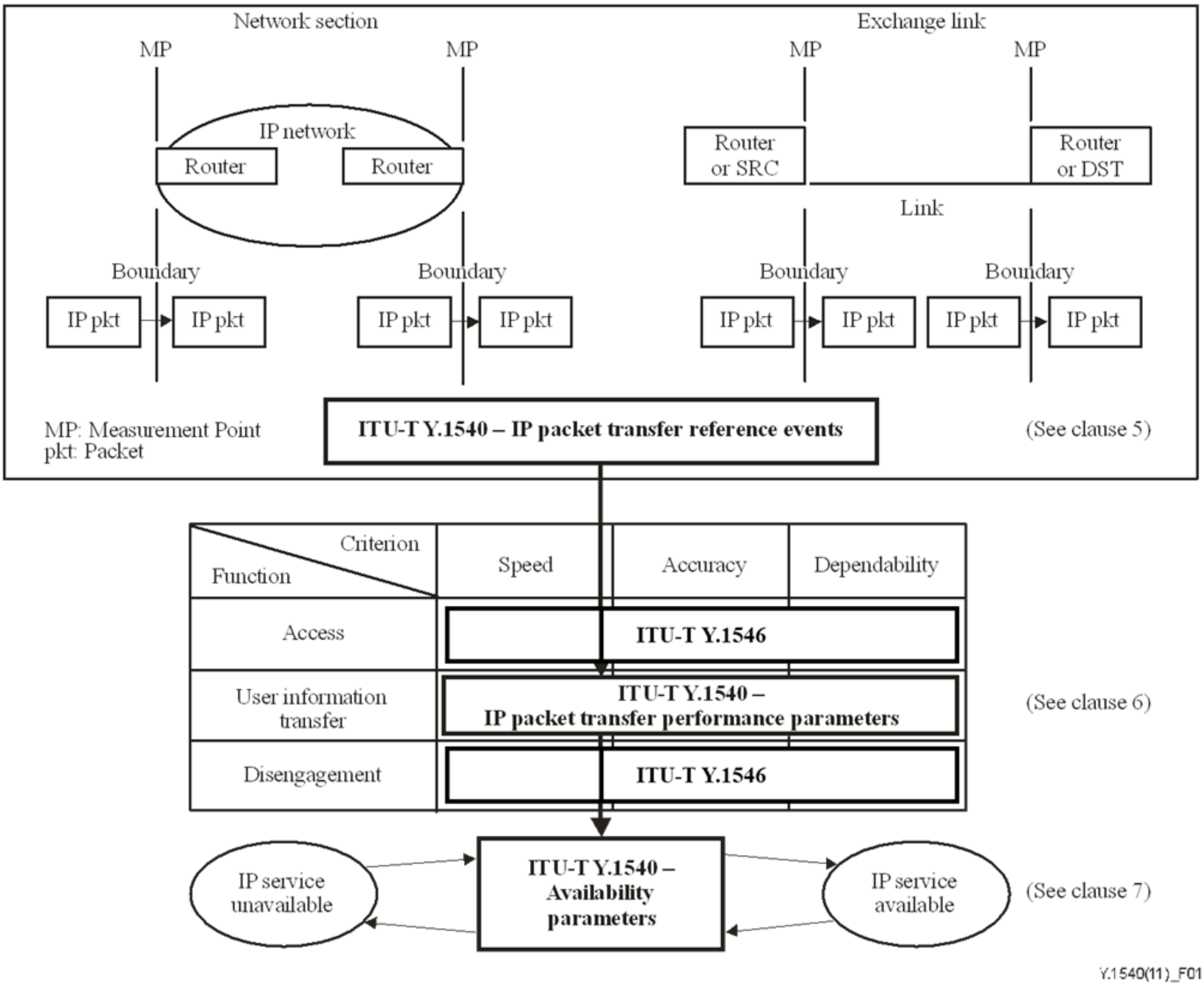


Figure 1 – Scope of this Recommendation

The performance parameters defined in this Recommendation describe the speed, accuracy, dependability and availability of IP packet transfer as provided by the IP data communication service. The end-to-end performance of regional and international IP services providing access and disengagement functions (e.g., domain name service) and higher-layer transport capabilities (e.g., transmission control protocol) have been addressed in separate Recommendations, such as [ITU-T Y.1546] Annex C for IP-based Service Availability.

This Recommendation is structured as follows: clause 1 specifies its scope; clause 2 specifies its normative references; clause 3 provides a list of abbreviations; clause 4 illustrates the layered model that creates the context for IP performance specification; clause 5 defines the model used for IP performance, including network sections and measurement points, reference events and outcomes; clause 6 uses this model to define IP packet transfer performance parameters; and clause 7 then defines IP service availability parameters.

Annex A defines IP capacity parameters in a way that caters towards assessment and provides requirements for methods of measurement, along with a plan to evaluate candidate methods of measurement in the laboratory and in (the field) production networks (new in the 2019 version). Appendix I describes IP packet routing considerations and their effects on performance. Appendix II provides secondary terminology for IP packet delay variation. Appendix III (Rate and throughput capacity related parameters) is deprecated in the 2019 edition. Appendix IV describes the estimation of IP service availability. Appendix V presents considerations for measuring the [ITU-T Y.1540] parameters. Appendix VI gives some background on IP service availability. Appendix VII offers background information on the stream repair parameters, and Appendix VIII adds information on capacity parameters (including a mapping to prior IETF metrics and items for further study). Appendix IX explains why TCP-based measurements fail to meet the normative requirements of clause 6.12.

The 2019 version adds many new appendices. The new appendix, Appendix X, provides substantial background material supporting the IP capacity parameters and methods of measurement, including a summary of laboratory and field test results comparing different methods of measurement. Appendix XI contains a survey of academic papers describing measurement campaigns making similar comparisons. Appendix XII contains the details of calculations required to make accurate bit rate measurements and comparisons at different layers of the protocol stack. Appendix XIII provides information on IP-based flow-related parameters and methods of measurement which is for further study, and indicates the complementary role with respect to IP capacity metrics and methods of measurement.

NOTE 4 – The ITU-T Y.1540 parameters may be augmented or modified based upon further study of the requirements of the IP applications (e.g., interactive, block, stream) to be supported.

NOTE 5 – The ITU-T Y.1540 speed, accuracy and dependability parameters are intended to characterize IP service in the available state.

NOTE 6 – The parameters defined in this Recommendation can apply to a single end-to-end IP service between two end hosts identified by their IP addresses. The parameters can also be applied to those IP packets from a given end-to-end IP service that are offered to a given network or exchange link (EL).

NOTE 7 – The ITU-T Y.1540 parameters are designed to characterize the performance of service provided by network elements between specified section boundaries. However, users of this Recommendation should be aware that network elements outside the specified boundaries can sometimes influence the measured performance of the elements between the boundaries. Examples are described in Appendix V.

NOTE 8 – The parameters defined in this Recommendation can also be applied to any subset of the IP packets offered to a given set of network equipment. Methods for aggregating performance over a set of network equipment or over an entire network are outside of the scope of this Recommendation.

NOTE 9 – This Recommendation does not provide the tools for explicit characterization of routing stability. However, the effects of route instability can be quantified using the loss, delay and severe loss block parameters defined in this Recommendation.

NOTE 10 – Specification of numerical performance objectives for some of the ITU-T Y.1540 performance parameters may be found in [ITU-T Y.1541].

NOTE 11 – The word "provisional", as used in this Recommendation, means that there is agreement on the stability of the value referenced, but that the value may change following further study, or on the basis of real network operational experience.

2 References

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

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

[ITU-T P.10] Recommendation ITU-T P.10/G.100 (2017), *Vocabulary for performance and quality of experience.*

[ITU-T Y.1541] Recommendation ITU-T Y.1541 (2011), *Network performance objectives for IP-based services.*

[ITU-T Y.1546] Recommendation ITU-T Y.1546 (2014), *Hand-over performance among multiple access networks.*

[ITU-T Y.1565] Recommendation ITU-T Y.1565 (2011), *Home network performance parameters.*

[IETF RFC 791] IETF RFC 791 (1981), *Internet Protocol*.
<http://www.ietf.org/rfc/rfc791.txt>

[IETF RFC 4737] IETF RFC 4737 (2006), *Packet Reordering Metrics.*
<http://www.ietf.org/rfc/rfc4737.txt>

[IETF RFC 5136] IETF RFC 5136 (2008), *Defining Network Capacity.*
<http://www.ietf.org/rfc/rfc5136.txt>

[IETF RFC 5481] IETF RFC 5481 (2009), *Packet Delay Variation Applicability Statement.*
<http://www.ietf.org/rfc/rfc5481.txt>

[IETF RFC 8200] IETF RFC 8200 (2017), *Internet Protocol, Version 6 (IPv6) Specification.*
<http://www.ietf.org/rfc/rfc8200.txt>

[IETF RFC 8337] IETF RFC 8337 (2018), *Model-Based Metrics for Bulk Transport Capacity.*
<http://www.ietf.org/rfc/rfc8337.txt>

3 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ARQ	Automatic Repeat-request
ATM	Asynchronous Transfer Mode
BTC	Bulk Transfer Capacity
DSCP	Differentiated Services Code Point
DST	Destination host
EL	Exchange Link
ER	Edge Router
FEC	Forward Error Correction
FTP	File Transfer Protocol
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
IPDR	Internet Protocol packet Duplicate Ratio
IPDV	Internet Protocol packet Delay Variation
IPER	Internet Protocol packet Error Ratio
IPIBR	Internet Protocol packet Impaired Block Ratio
IPIIR	Internet Protocol packet Impaired Interval Radio
IPLR	Internet Protocol packet Loss Ratio
IPOR	Octet-based IP packet Rate
IPPM	IP Performance Metrics
IPPR	Internet Protocol Packet Rate

IPRE	Internet Protocol packet transfer Reference Event
IPRR	Internet Protocol packet Reordered Ratio
IPSLB	Internet Protocol packet Severe Loss Block outcome
IPSLBR	Internet Protocol packet Severe Loss Block Ratio
IPTD	Internet Protocol packet Transfer Delay
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ISP	Internet Service Provider
LL	Lower Layers (protocols and technology supporting the Internet protocol layer)
M_{av}	The minimum number of packets recommended for assessing the availability state
MBM	Model-Based Metrics
MP	Measurement Point
MTBISO	Mean Time Between IP Service Outages
MTTISR	Mean Time To Internet protocol 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
PDH	Plesiochronous Digital Hierarchy
PDV	Packet Delay Variation
PIA	Percent Internet protocol service Availability
PIU	Percent Internet protocol service Unavailability
QoS	Quality of Service
R	Router
RIPR	Replicated Internet protocol Packet Ratio
RSVP	Resource reservation Protocol
RTCP	Real-Time Control Protocol
RTO	Retransmission Time Out
RTP	Real-time Transport Protocol
RTT	Round-Trip Time
SDH	Synchronous Digital Hierarchy
SPRT	Sequential Probability Ratio Test
SRC	Source host
STD	Standard
T_{av}	Minimum length of time of Internet protocol availability; minimum length of time of Internet protocol unavailability
TBF	Token Bucket Filter

TCP	Transmission Control Protocol
T_{max}	Maximum Internet protocol packet delay beyond which the packet is declared to be lost
ToS	Type of Service
T_s	Length of time defining the block in the severe loss block outcome
TTL	Time To Live
UDP	User Datagram Protocol

4 Layered model of performance for IP service

Figure 2 illustrates the layered nature of the performance of an IP service. The performance provided to IP service users depends on the performance of other layers:

- lower layers (LL) that provide (via "links") connection-oriented or connectionless transport supporting the IP layer. Links are terminated at points where IP packets are forwarded (i.e., "routers", "SRC" and "DST") and thus have no end-to-end significance. Links may involve different types of technologies, for example, asynchronous transfer mode (ATM), frame relay, synchronous digital hierarchy (SDH), plesiochronous digital hierarchy (PDH), ISDN and leased lines. There may be several layers of protocols and services below the IP layer, and these, in the end, make use of various types of physical media;
- the IP layer that provides connectionless transport of IP datagrams (i.e., IP packets). The IP layer has end-to-end significance for a given pair of source and destination IP addresses. Certain elements in the IP packet headers may be modified by networks, but the IP user data may not be modified at or below the IP layer;
- higher layers, supported by IP, that further enable end-to-end communications. Upper layers may include, for example, transmission control protocol (TCP), user datagram protocol (UDP), file transfer protocol (FTP), real-time transport protocol (RTP) and hypertext transfer protocol (HTTP). The higher layers will modify and may enhance the end-to-end performance provided at the IP layer.

NOTE 1 – Clause 5 defines an IP service performance model and more precisely defines key terms used in this layered model.

NOTE 2 – Performance interactions among these layers are for further study.

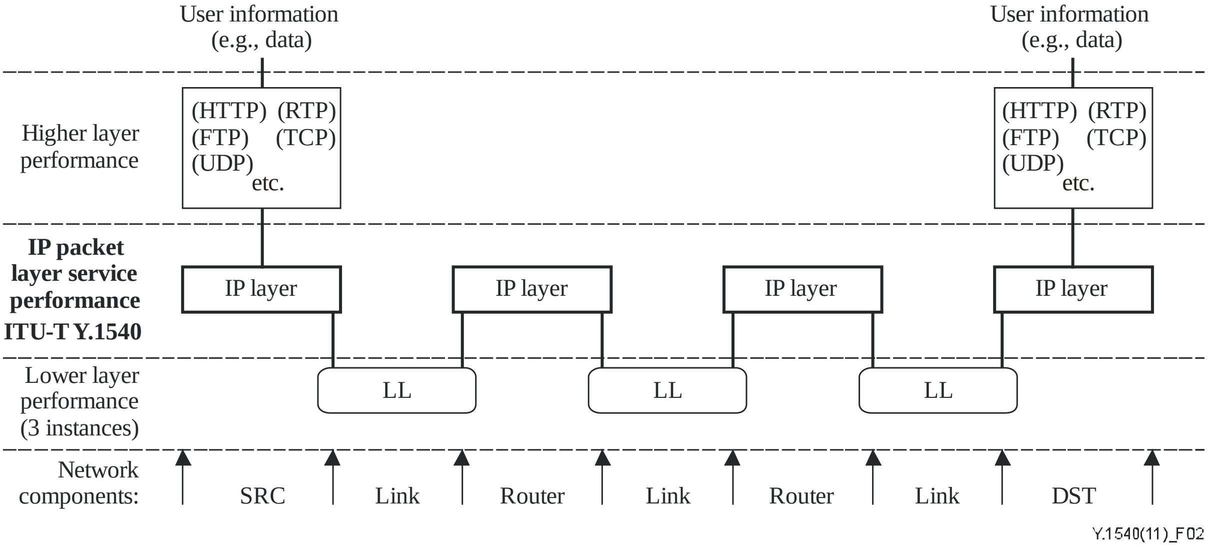


Figure 2 – Layered model of performance for IP service – Example

5 Generic IP service performance model

This clause defines a generic IP service performance model. The model is primarily composed of two types of sections: the exchange link and the network section (NS). These are defined in clause 5.2. They provide the building blocks with which any end-to-end IP service may be represented. Each of the performance parameters defined in this Recommendation can be applied to the unidirectional transfer of IP packets on a section or a concatenated set of sections.

Clause 5.4 specifies the set of IP packet transfer reference events that provide the basis for the performance parameter definition. These reference events are derived from and are consistent with relevant IP service and protocol definitions. Clause 5.5 then uses these reference events to enumerate the possible outcomes when a packet is delivered into a section.

NOTE – Incorporation of all or part of the ITU-T Y.1540 performance model and reference events into [b-ITU-T I.353] is for further study.

5.1 Network components

5.1.1 Host

A host is a computer that communicates using the Internet protocols. It implements routing functions (i.e., it operates at the IP layer) and may implement additional functions including higher layer protocols (e.g., TCP in a source or destination host (DST)) and lower layer protocols (e.g., ATM).

5.1.2 Router

A router is a host that enables communication between other hosts by forwarding IP packets based on the content of their IP destination address field.

5.1.3 Source host (SRC)

A source host is a host and a complete IP address where end-to-end IP packets originate. In general, a host may have more than one IP address; however, a source host (SRC) is a unique association with a single IP address. Source hosts also originate higher layer protocols (e.g., TCP) when such protocols are implemented.

5.1.4 Destination host (DST)

A destination host is a host and a complete IP address where end-to-end IP packets are terminated. In general, a host may have more than one IP address; however, a destination host is a unique association with a single IP address. Destination hosts also terminate higher layer protocols (e.g., TCP) when such protocols are implemented.

5.1.5 Link

A link is a point-to-point (physical or virtual) connection used for transporting IP packets between a pair of hosts. It does not include any parts of the hosts or any other hosts; it operates below the IP layer. For example, a link could be a leased line or it could be implemented as a logical connection over an Ethernet, a frame relay network, an ATM network or any other network technology that functions below the IP layer.

Figure 3 illustrates the network components relevant to an IP service between an SRC and a DST. Links, which could be dial-up connections, leased lines, rings or networks are illustrated as lines between hosts. Routers are illustrated as circles and both SRC and DST are illustrated as triangles.

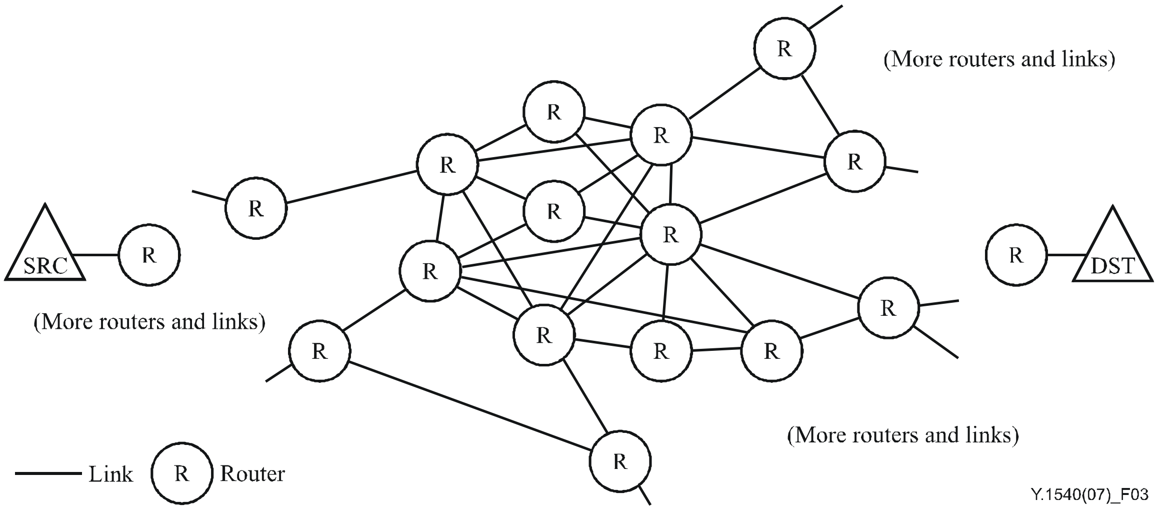


Figure 3 – IP network components

5.2 Exchange links and network sections

5.2.1 Exchange link (EL)

This is the link connecting:

1) a source or destination host to its adjacent host (e.g., router) possibly in another jurisdiction, sometimes referred to as an access link, ingress link or egress link; or

2) a router in one network section with a router in another network section.

Note that the responsibility for an exchange link, its capacity and its performance, is typically shared between the connected parties.

NOTE – "Exchange link" is roughly equivalent to the term "exchange" as defined in [b-IETF RFC 2330].

5.2.2 Network section (NS)

A network section is a set of hosts together with all of their interconnecting links that together provide a part of the IP service between an SRC and a DST, and are under a single (or collaborative) jurisdictional responsibility. Some network sections consist of a single host with no interconnecting links. Source NS and destination NS are particular cases of network sections. Pairs of network sections are connected by exchange links.

NOTE – "Network section" is roughly equivalent to the term "cloud" as defined in [b-IETF RFC 2330].

Any set of hosts interconnected by links could be considered a network section. However, for the (future) purpose of IP performance allocation, it will be relevant to focus on the set of hosts and links under a single (or collaborative) jurisdictional responsibility (such as an Internet service provider (ISP) or a network service provider (NSP)). These hosts typically have the same network identifier in their IP addresses. Typically, they have their own rules for internal routing. Global processes and local policies dictate the routing choices to destinations outside of this network section (to other NS via exchange links). These network sections are typically bounded by routers that implement the IP exterior gateway protocols.

5.2.3 Source NS

The source NS is the NS that includes the SRC within its jurisdictional responsibility. In some cases, the SRC is the only host within the source NS.

5.2.4 Destination NS

The destination NS is the NS that includes the DST within its jurisdictional responsibility. In some cases, the DST is the only host within the destination NS.

Figure 4 illustrates the network connectivity relevant to IP service between an SRC and a DST. At the edges of each NS, gateway routers receive and send packets across exchange links.

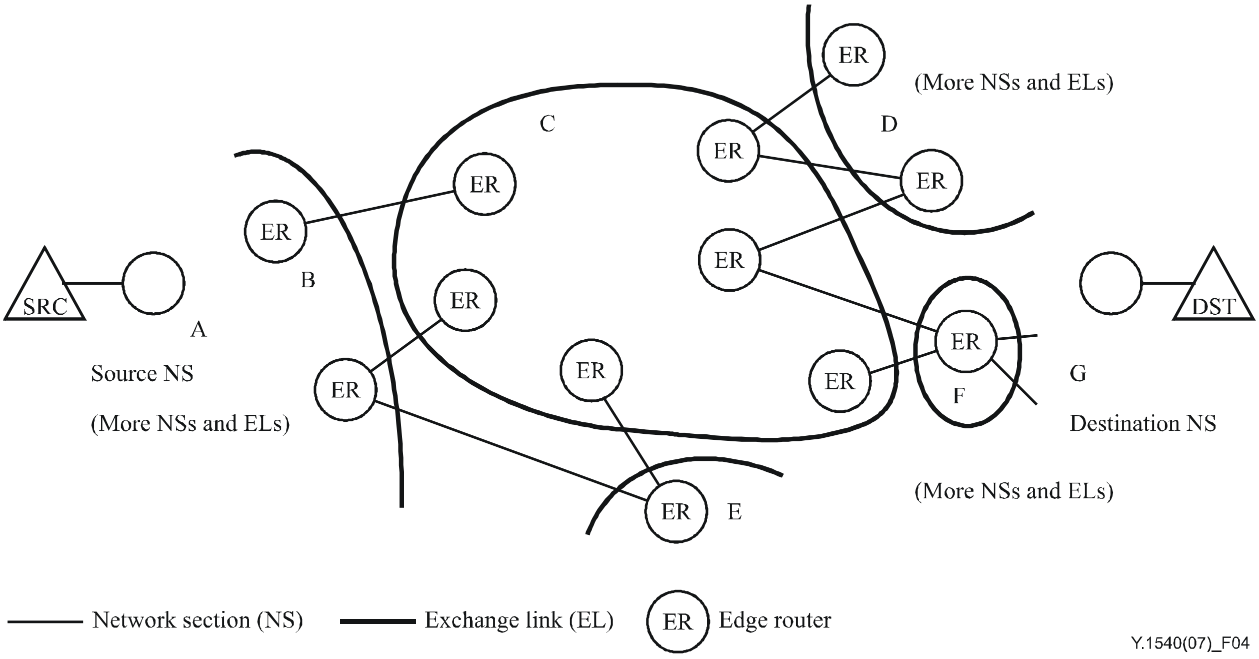


Figure 4 – IP network connectivity

5.3 Measurement points and measurable sections

5.3.1 Measurement point (MP)

The measurement point is the boundary between a host and an adjacent link at which performance reference events can be observed and measured. Consistent with [b-ITU-T I.353], the standard Internet protocols can be observed at IP measurement points (MPs). [b-ITU-T I.353] provides more information about MPs, for digital services.

NOTE – The exact location of the IP service MP within the IP protocol stack is for further study.

A section or a combination of sections is measurable if it is bounded by a set of MPs. In this Recommendation, the following sections are measurable.

5.3.2 Basic section

Basic sections are either an EL, an NS, an SRC or a DST. Basic sections are delimited by MPs.

The performance of any EL or NS is measurable relative to any given unidirectional end-to-end IP service. The *ingress MPs* are the set of MPs crossed by packets from that service as they go into that basic section. The *egress MPs* are the set of MPs crossed by packets from that service as they leave that basic section.

5.3.3 End-to-end IP network

The set of Els and NSs that provide the transport of IP packets transmitted from the SRC to DST. The MPs that bind the end-to-end IP network are the MPs at the SRC and the DST.

The end-to-end IP network performance is measurable relative to any given unidirectional end-to-end IP service. The *ingress MPs* are the MPs crossed by packets from that service as they go into the end-to-end network at the SRC. The *egress MPs* are the MPs crossed by packets from that service as they leave the end-to-end network at the DST.

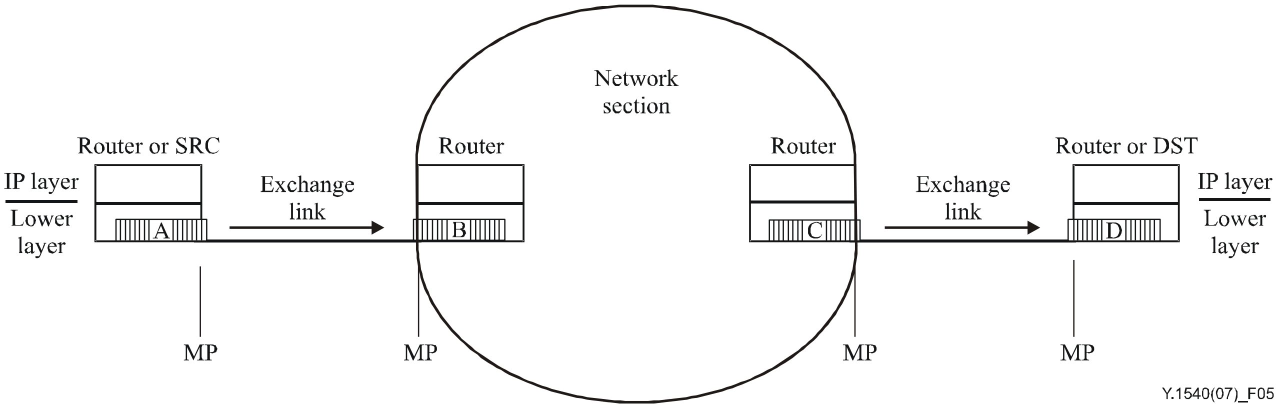
5.3.4 Network section ensemble (NSE)

A network section ensemble (NSE) refers to any connected subset of NSs together with all of the Els that interconnect them. The term "NSE" can be used to refer to a single NS, two NSs, or any number of NSs and their connecting Els. Pairs of distinct NSEs are connected by exchange links. The term "NSE" can also be used to represent the entire end-to-end IP network. NSEs are delimited by MPs.

The performance of any given NSE is measurable relative to any given unidirectional end-to-end IP service. The *ingress MPs* are the set of MPs crossed by packets from that service as they go into that NSE. The *egress MPs* are the set of MPs crossed by packets from that service as they leave that NSE.

5.4 IP packet transfer reference events (IPREs)

In the context of this Recommendation, the following definitions apply on a specified end-to-end IP service. The defined terms are illustrated in Figure 5.



NOTE 1 – IP exit events for packets A and C.
NOTE 2 – IP entry events for packets B and D.

Figure 5 – Example IP packet transfer reference events

An IP packet transfer event occurs when:

- – an IP packet crosses an MP; and
- – standard IP procedures applied to the packet verify that the header checksum is valid; and
- – the source and destination address fields within the IP packet header represent the IP addresses of the expected SRC and DST.

NOTE – The IP packet header contains information including the type of service (ToS) or differentiated services code point (DSCP). How such information may affect packet transfer performance is for further study.

IP packet transfer reference events are defined without regard to packet fragmentation. They occur for every IP packet crossing any MP regardless of the value contained in the "more-fragments flag".

Four types of IP packet transfer events are defined:

5.4.1 IP packet entry event into a host

An IP packet transfer entry event into a host occurs when an IP packet crosses an MP entering a host (NS router or DST) from the attached EL.

5.4.2 IP packet exit event from a host

An IP packet transfer exit event from a host occurs when an IP packet crosses an MP exiting a host (NS router or SRC) into the attached EL.

5.4.3 IP packet ingress event into a basic section or NSE

An IP packet transfer ingress into a basic section or NSE event occurs when an IP packet crosses an ingress MP into a basic section or an NSE.

5.4.4 IP packet egress event from a basic section or NSE

An IP packet transfer egress event from a basic section or NSE occurs when an IP packet crosses an egress MP out of a basic section or an NSE.

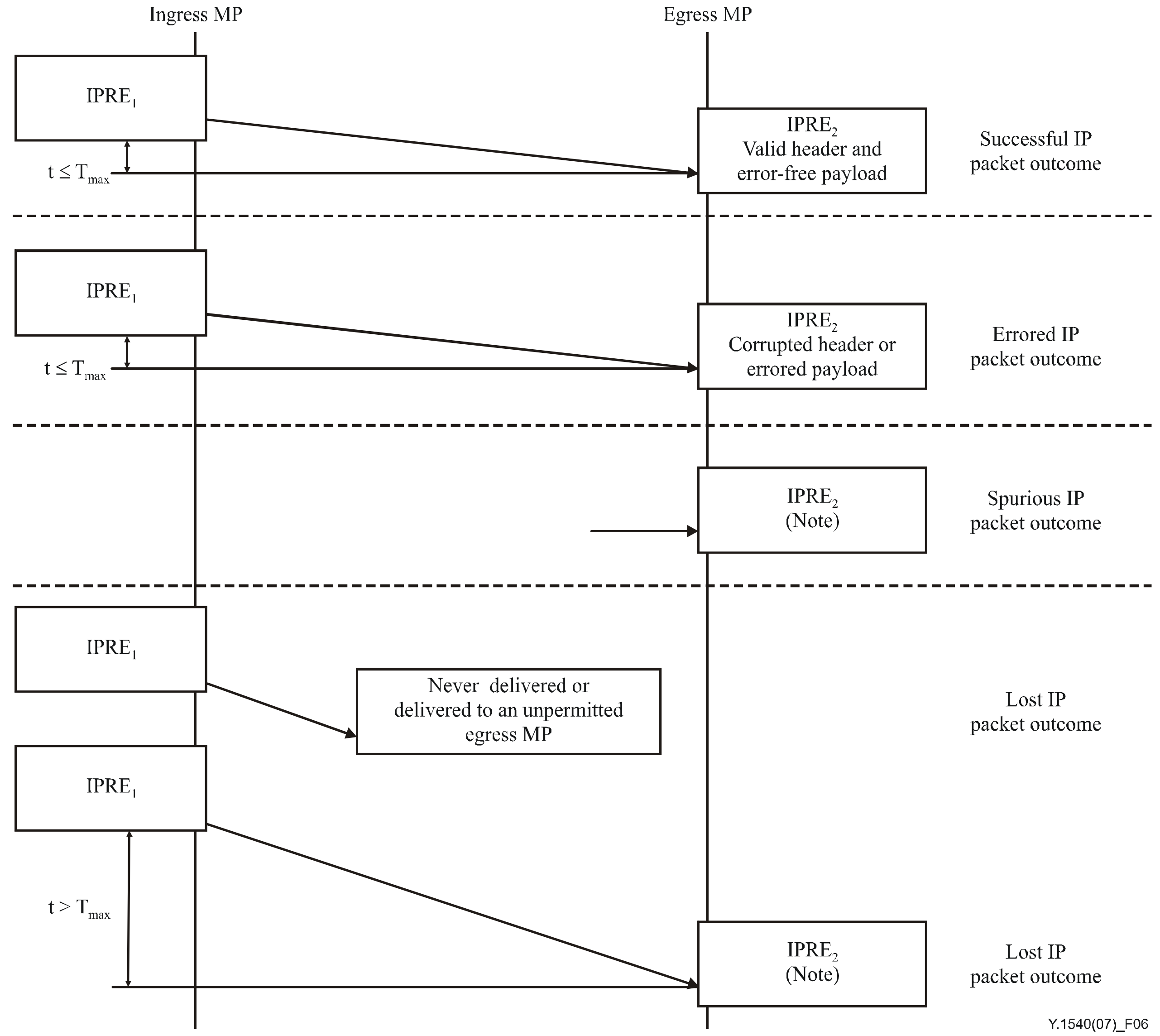
NOTE 1 – IP packet entry and exit events always represent, respectively, entry into and exit from a host. IP packet ingress events and egress events always represent ingress into and egress from a section or an NSE. To illustrate this point, note that an ingress into an EL creates an exit event from the preceding host, while an ingress into an NS is an entry event because, by definition, NSs always have hosts at their edges.

NOTE 2 – For practical measurement purposes, IP packet transfer reference events need not be observed within the IP protocol stack of the host. Instead, the time of occurrence of these reference events can be approximated by observing the IP packets crossing an associated physical interface. This physical interface should, however, be as near as possible to the desired MP. In cases where reference events are monitored at a physical interface, the time of occurrence of an exit event from a host is approximated by the observation of the first bit of the IP packet coming from the host or test equipment. The time of occurrence of an entry event into a host is approximated by the observation of the last bit of the IP packet going to the host or test equipment.

5.5 IP packet transfer outcomes

By considering IP packet transfer reference events, a number of possible IP transfer outcomes may be defined for any packet attempting to cross a basic section or an NSE. A transmitted IP packet is either *successfully transferred, errored* or *lost*. A delivered IP packet for which no corresponding IP packet was offered is said to be *spurious*. Figure 6 illustrates the IP packet transfer outcomes.

The definitions of IP packet transfer outcomes are based on the concepts of *permissible ingress MP, permissible egress MP* and *corresponding packets*.



NOTE – Outcome occurs independent of IP packet contents.

Figure 6 – IP packet transfer outcomes

5.5.1 Global routing information and permissible output links

In theory, in a connected IP network, a packet can be delivered to any router, NS or NSE, and still arrive at its destination. However, global routing information defines a restricted set of destination addresses that each network (autonomous system) is willing and able to serve on behalf of each of its adjoining NS. It is reasonable to assume that (in the worst case) an NS will completely discard any packets with destination addresses for which that NS has announced an inability (or an unwillingness) to serve. Therefore, all IP packets (and fragments of packets) leaving a basic section should only be forwarded to other basic sections as *permitted* by the available global routing information.

For performance purposes, the transport of an IP packet by an NSE will be considered successful only when that NSE forwards the entire packet contents to other basic sections as permitted by the currently available global routing information. If the destination address corresponds to a host attached directly to this NSE, the only permitted output and the only successful IP transport is a forwarding to the destination host.

NOTE 1 – IP procedures include the updating of global routing information. An NS that was permissible may no longer be permissible following an update of the routing information shared between NSs. Alternatively, an NS that was not previously permissible may have become permissible after an update of the global routing information.

NOTE 2 – Routing information can be supplemented by information about the relative suitability of each of the permitted output links. The performance implications of that additional information are for further study.

At a given time, and relative to a given end-to-end IP service and a basic section or NSE:

– an ingress MP is a *permissible ingress MP* if the crossing of this MP into this basic section or NSE is permitted by the global routing information;

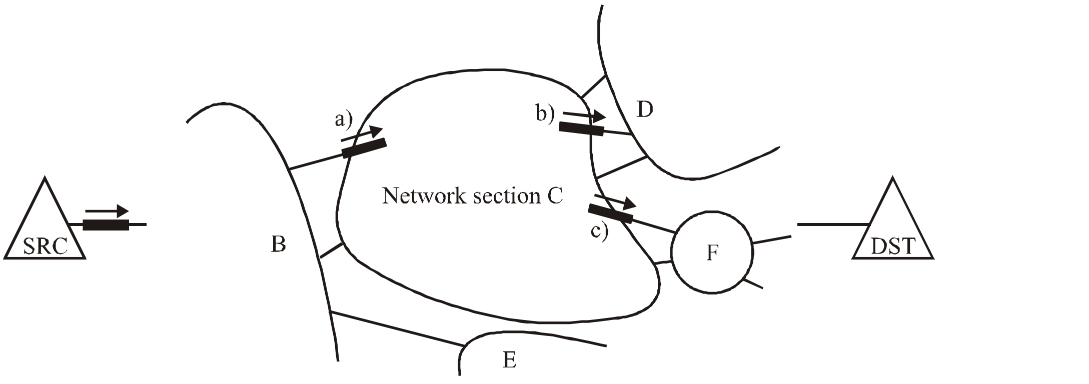
– an egress MP is a *permissible egress MP* if the crossing of this MP leads into another basic section that is permitted by the global routing information.

5.5.2 Corresponding events

Performance analysis makes it necessary to associate the packets crossing one MP with the packets that crossed a different MP. Connectionless routing means a packet may leave a basic section on any one of (possibly) several permissible egress MPs. Packet fragmentation means that a packet going into a basic section may leave in fragments, possibly into several different other basic sections. Finally, connectionless IP routing may even send a packet or a fragment back into a basic section it has already traversed (possibly due to the updating of routing tables).

An IP egress event is said to *correspond* to an earlier ingress event if they were created by the "same" IP packet. This concept applies whether the packet at the egress MP is the whole packet or just a fragment of the original. Figure 7 illustrates a case where a packet goes into NS C from NS B and is fragmented into two parts in NS C. One of the fragments is sent to NS D and the other to NS F. Both of these egress events *correspond* to the single ingress event. To avoid confusion resulting from packets re-entering the NSE, this concept of *correspondence* also requires that this be the first time (since its ingress) this particular content has departed from the NSE.

The practical determination of whether IP reference events are corresponding is usually *ad hoc* and will often rely on consideration of the IP addresses, the global routing information, the IP packet identification field, other header information and the IP packet contents.



An IP packet from SRC to DST enters NS C, creates an ingress event, is fragmented, and creates two corresponding egress events, b) and c).

Figure 7 – Corresponding events when fragmentation occurs

5.5.3 Notes about the definitions of successful, errored, lost and spurious packet outcomes

Each of the following definitions of individual packet outcomes is based on observing IP reference events at IP measurement points. By selecting the appropriate IP measurement points, each definition can be used to evaluate the performance of a particular EL, a particular NS, a particular NSE, and they can be applied to the performance of end-to-end services.

These outcomes are defined without restriction to a particular packet type (ToS, DSCP, protocol, etc.). IP performance will differ by packet type.

In each definition, the possibility of packet fragmentation is accounted for by including the possibility that a single IP reference event could result in several subsequent events. Note that if any fragment is lost, the whole original packet is considered lost. If no fragments are lost, but some are errored, the entire original packet is considered errored. For the delivery of the original packet to be considered successful, each fragment must be successfully delivered to one of the permissible output Els.

5.5.4 Successful IP packet transfer outcome

A successful packet transfer outcome occurs when a single IP packet reference event at a permissible ingress MP_0 results in one (or more) corresponding reference event(s) at one (or more) egress MP_i, all within a specified time T_{max} of the original ingress event and:

1) all egress MP_i where the corresponding reference events occur are permissible; and

2) the complete contents of the original packet observed at MP_0 are included in the delivered packet(s); and

3) the binary contents of the delivered IP packet information field(s) conform exactly with that of the original packet; and

4) the header field(s) of the delivered packet(s) is (are) valid.

NOTE – The value of T_{max} is recommended to be set at 3 seconds for general use. Some global end-to-end paths may require a larger value of T_{max} to ensure that packets with long transfer times have adequate opportunity to arrive. The value of 3 seconds has been used in practice.

5.5.5 Errored IP packet outcome

An errored packet outcome occurs when a single IP packet reference event at a permissible ingress MP_0 results in one (or more) corresponding reference event(s) at one (or more) egress MP_i, all within T_{max} time of the original reference event and:

1) all egress MP_i where the corresponding reference events occur are permissible; and

2) the complete contents of the original packet observed at MP_0 are included in the delivered packet(s); and

3) either:

   – the binary contents of the delivered IP packet information field(s) do not conform exactly with that of the original packet; or

   – one or more of the header field(s) of the delivered packet(s) is (are) corrupted.

NOTE – Most packets with errored headers that are not detected by the header checksum at the IP layer will be discarded or redirected by other IP layer procedures (e.g., based on corruption in the address or ToS/DSCP fields). The result is that no reference event is created for the higher layer protocols expecting to receive this packet. Because there is no IP reference event, these packet transfer attempts will be classified as lost packet outcomes. Errored headers that do not result in discarding or misdirecting will be classified as errored packet outcomes.

5.5.6 Lost IP packet outcome

A lost packet outcome occurs when there is a single IP packet reference event at a permissible ingress MP_1, and when some or all of the contents corresponding to that ingress packet do not result in an IP packet reference event at a permissible egress MP_n within the time T_{max}.

A lost packet outcome may in fact be one or more *misdirected packet* outcomes (which were not observed), as defined below.

A misdirected packet occurs when a single IP packet reference event at a permissible ingress MP_0 results in one (or more) corresponding reference event(s) at one (or more) egress MP_i, all within a specified T_{max} time of the original reference event and:

1) the complete contents of the original packet observed at MP_0 are included in the delivered packet(s); but

2) one or more of the egress MP_i where the corresponding reference events occur is (are) not permissible egress MP(s).

5.5.7 Spurious IP packet outcome

A spurious IP packet outcome occurs for a basic section, an NSE, on an end-to-end IP service when a single IP packet creates an egress event for which there was no corresponding ingress event.

5.5.8 Secondary IP packet outcomes

The following outcomes are based on the fundamental outcomes described above.

5.5.8.1 In-order and reordered IP packet outcomes

The definition of these IP packet outcomes requires some background discussion.

In-order packet delivery is a property of successful packet transfer attempts, where the sending packet order is preserved on arrival at the destination host (or measurement point). Arrival order is determined by the position relative to other packets of interest, though the extent to which a given packet has been reordered may be quantified in the units of position, time and payload byte distances. A reordered packet performance parameter is relevant for most applications, especially when assessing network support for real-time media streams, owing to their finite ability to restore order or when the performance implies a lack of that capability. Packets usually contain some unique identifier applied at the SRC, sometimes assumed to be a sequence number, so this number or other information (such as time stamps from the MP_0) is the reference for the original order at the source. The evaluation of arrival order also requires the ability to determine which specific packet is the "next expected" packet, and this is greatly simplified where sequence numbers are consecutive increasing integers.

An in-order packet outcome occurs when a single IP packet reference event at a permissible egress measurement point results in the following:

– The packet has a sequence number greater than or equal to the next expected packet value. The next expected value increases to reflect the arrival of this packet, setting a new value of expectation.

A reordered or out-of-order packet outcome occurs when a single IP packet reference event at a permissible egress measurement point results in the following:

– The packet has a sequence number lower than the next expected packet value and therefore the packet is reordered. The next expected value does not change due to arrival of this packet.

5.5.8.2 IP packet severe loss block outcome

An IP packet severe loss block outcome (IPSLB) occurs for a block of packets observed during time interval T_s at ingress MP_0 when the ratio of lost packets at egress MP_i to total packets in the block exceeds s1.

The value of time interval T_s is provisionally set at 10 seconds. The value of threshold s1 is provisionally set at 0.2. Evaluation of successive blocks (time intervals) should be non-overlapping.

NOTE – These values are intended to identify IP path changes due to routing updates, which cause significant degradation to most user applications. The values may change following further study and experience. Lower values of s1 would capture additional network events that may affect the operation of connectivity-sensitive applications. Also, significant degradation to video and audio applications may be well correlated with the IPSLB outcome when using T_s block lengths of approximately 1 second, and use of this value may be important in the future.

The minimum number of packets that should be used in evaluating the severe loss block outcome is M_{lb}, and these packets should be spread throughout a T_s interval. The value of M_{lb} is for further study.

5.5.8.3 Duplicate IP packet outcome

A duplicate packet transfer outcome is a subset of successful packet outcomes, and occurs when a single IP packet reference event at a permissible ingress MP_0 results in two or more corresponding reference event(s) on at least one permissible egress $MP_{i,}$ and the binary information fields of all the output packets are identical to the original packet. The egress reference event at MP_i for a duplicate packet occurs subsequently to at least one other corresponding egress reference event for the original packet (usually also at MP_i).

Note that in point-to-point communication, there is only one permissible egress MP_i where the destination host is directly attached to the NSE. In point-to-multipoint communication, there may be many permissible egress MP_i for the various destinations.

5.5.8.4 Replicated IP packet outcome

A replicated packet transfer outcome occurs when a single IP packet reference event at a permissible ingress MP_0 results in two or more corresponding reference event(s) on at least one permissible egress MP_i, and the binary information fields of all the output packets are identical to the original packet. The egress reference event at MP_i for a replicated packet is the first for the original packet and occurs prior to at least one other egress reference event for a duplicate packet (usually also at MP_i).

5.5.9 Stream-repair IP packet outcomes

The following outcomes are based on the fundamental outcomes, with additional analysis based on a model of stream repair systems. Appendix VII gives more background on this topic and the impairment mitigation techniques (above IP-layer) that are addressed.

5.5.9.1 Simple model of application-layer stream repair techniques

Appendix VII also defines a simple model, described below. Each stream of application-layer packets is modelled as containing two categories of packets:

- intervals or blocks of information packets;
- the maximum number of repairable packets associated with the information block.

The challenge to the repair technique designer is to choose the information block size in combination with the (maximum) repair capability that will be sufficient to compensate for a high percentage of packet network impairments (loss, excessive delay and corruption), while working within the overall packet transfer capacity limits of the system and delivering sufficient quality in the application stream.

The new performance parameters should aid these decisions.

5.5.9.2 Impaired packet outcome and IP packet impaired interval outcome

An *IP packet impaired interval outcome* occurs for a set of packets observed during time interval T_I at ingress MP_0 when the number of impaired packet outcomes at egress MP_i exceeds x. Note that the time interval T_I includes both information and overhead or repair packets (if embedded in the ingress stream).

Impaired packet outcomes are the sum of the following outcomes:

- lost packet outcomes, using a T_{max} associated with T_I and the nominal transfer time, and possibly equal to the minimum packet transfer delay for the population of interest plus (a multiple of) T_I. This would include packets that are subject to excessive queuing, as well as those that never arrive;
- errored packet outcomes.

Note that one distinguishing factor between this outcome and other packet loss/block metrics is the combination of exceptionally delayed packets (beyond a delay variation threshold) with packets that never arrive (and are truly lost during transfer) in a single category: Impaired packets.

There are no provisional values set for time interval T_I and threshold x. Instead, the analysis may involve a range of values for interval T_I and threshold x. The length of the IP packet payload should also be specified, as this influences the serialization time and therefore the time interval occupied by a block of packets.

5.5.9.3 IP packet impaired block outcome

An *IP packet impaired block outcome* occurs for a set of packets of block size b, observed at ingress MP_0 when the number of impaired packet outcomes at egress MP_i in the block exceeds x. There are no provisional values set for the block size b and the repair threshold x.

6 IP packet transfer performance parameters

This clause defines a set of IP packet information transfer performance parameters using the IP packet transfer outcomes defined in clause 5.5. All of the parameters may be estimated on the basis of observations made at the MP that bound the basic section or the NSE under test.

NOTE – Definitions of additional IP packet transfer performance parameters (e.g., severely errored IP packet block ratio) are for further study.

6.1 Packet qualifications

This clause defines key terminology for qualifying the applicability of performance parameters to sets of packets.

6.1.1 Populations of interest

Most of the performance parameters are defined over sets of packets called *populations of interest*. For the *end-to-end case,* the population of interest is usually the total set of packets being sent from the SRC to the DST. The measurement points in the end-to-end case are the MP at the SRC and DST.

For a basic section or NSE and relative to a particular SRC and DST pair, the population of interest at a particular permissible ingress MP is that set of packets being sent from the SRC to the DST that is routed into the basic section or NSE across that specific MP. This is called the *specific-ingress case*.

The total population of interest for a basic section or NSE relative to a particular SRC and DST pair is the total set of packets from the SRC to the DST that is delivered into the section or NSE across any of its permissible ingress MPs. This is called the *ingress-independent case*.

Each of these IP performance parameters are defined without reference to a particular packet type (ToS, DSCP, protocol, etc.) Performance will differ by packet type and any statement about measured performance should include information about which packet type or types were included in the population.

6.1.2 Packet flow

A packet flow is the set of packets associated with a given connection or connectionless stream having the same source host address (SRC), destination host address (DST), class of service and session identification (e.g., port numbers from a higher-layer protocol). Other documents may use the terms microflow or subflow when referring to packet streams with this degree of classification. A packet flow is the most common example of a population of interest.

IPv6 packets have an additional field for the source host to label sequences of packets which should receive some special treatment in IPv6 routers. This field is called the flow label and, in combination with the source address, uniquely defines a packet flow.

6.2 IP packet transfer delay

IP packet transfer delay (IPTD) is defined for all successful and errored packet outcomes across a basic section or an NSE. IPTD is the time, ($t_2 – t_1$) between the occurrence of two corresponding IP packet reference events, ingress event $IPRE_1$ at time t_1 and egress event $IPRE_2$ at time t_2, where ($t_2 > t_1$) and ($t_2 – t_1$) $\leq T_{max}$. If the packet is fragmented within the NSE, t_2 is the time of the final corresponding egress event. The end-to-end IP packet transfer delay is the one-way delay between the MP at the SRC and DST as illustrated in Figure 8.

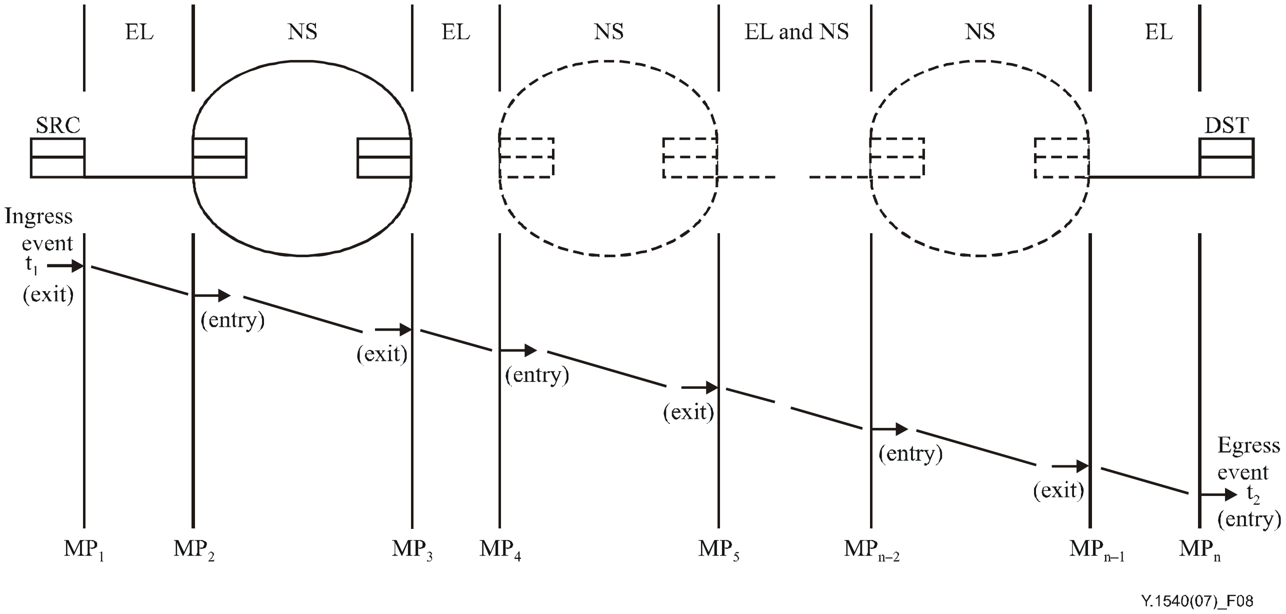


Figure 8 – IP packet transfer delay events (illustrated for the end-to-end transfer of a single IP packet)

6.2.1 Mean IP packet transfer delay

Mean IP packet transfer delay is the arithmetic average of IP packet transfer delays for a population of interest.

6.2.2 Minimum IP packet transfer delay

Minimum IP packet transfer delay is the smallest value of IP packet transfer delay among all IP packet transfer delays of a population of interest. This includes propagation delay and queuing delays common to all packets. Therefore, this parameter may not represent the theoretical minimum delay of the path between MPs.

6.2.3 Median IP packet transfer delay

The median IP packet transfer delay is the 50th percentile of the frequency distribution of IP packet transfer delays from a population of interest. The median is the middle value once the transfer delays have been rank-ordered. To obtain this middle value when the population contains an even number of values, then the mean of the two central values is used.

6.2.4 End-to-end 2-point IP packet delay variation

The variations in IP packet transfer delay are also important. Streaming applications might use information about the total range of IP delay variation to avoid buffer underflow and overflow. Extreme variations in IP delay will cause TCP retransmission timer thresholds to grow and may also cause packet retransmissions to be delayed or cause packets to be retransmitted unnecessarily.

End-to-end 2-point IP packet delay variation (PDV) is defined based on the observations of corresponding IP packet arrivals at the ingress and egress MPs (e.g., MP_{DST}, MP_{SRC}). These observations characterize the variability in the pattern of IP packet arrival events at the egress MP and the pattern of corresponding events at the ingress MP with respect to a reference delay.

The 2-point PDV (v_k) for an IP packet k between the SRC and DST is the difference between the absolute IP packet transfer delay (x_k) of packet k and a defined reference IP packet transfer delay, $d_{1,2}$, between those same MPs (see Figure 9): $v_k = x_k – d_{1,2}$.

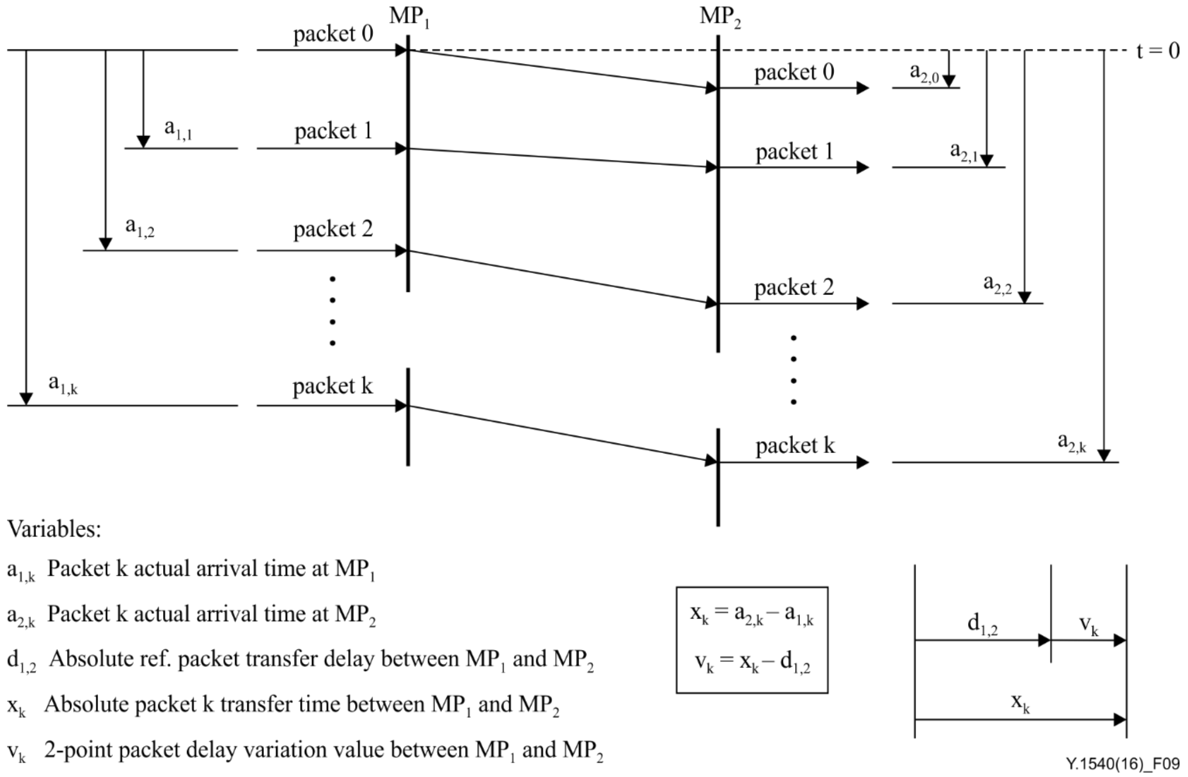

MP1
MP2
packet 0
t = 0
packet 0
a2,0
a1,1
packet 1
packet 1
a2,1
a1,2
packet 2
packet 2
a2,2
a1,k
packet k
packet k
a2,k
Variables:
a1,k Packet k actual arrival time at MP1
a2,k Packet k actual arrival time at MP2
d1,2 Absolute ref. packet transfer delay between MP1 and MP2
xk Absolute packet k transfer time between MP1 and MP2
vk 2-point packet delay variation value between MP1 and MP2
xk = a2,k – a1,k
vk = xk – d1,2
d1,2
vk
xk
Y.1540(16)_F09

Figure 9 – 2-point IP packet delay variation

The reference IP packet transfer delay, $d_{1,2}$, between the SRC and DST is the absolute IP packet transfer delay experienced by a selected IP packet between those two MPs.

Positive values of 2-point IP packet delay variation (IPDV) correspond to IP packet transfer delays greater than those experienced by the reference IP packet; negative values of 2-point PDV correspond to IP packet transfer delays less than those experienced by the reference IP packet. The distribution of 2-point PDVs is identical to the distribution of absolute IP packet transfer delays displaced by a constant value equal to $d_{1,2}$.

6.2.4.1 Using minimum delay as the basis for delay variation

As illustrated in Figure 9, the delay variation of an individual packet is naturally defined as the difference between the actual delay experienced by that packet and a nominal or reference delay. The preferred reference (used in ITU-T Y.1541 IPDV objectives) is the minimum delay of the population of interest. This ensures that all variations will be reported as positive values, and simplifies reporting the range of variation (the maximum value of variation is equal to the range). Distributions of delay variation in IP networks often exhibit a bias toward the minimum (e.g., the minimum and the mode are equal). Many more useful capabilities of this form of delay variation – PDV, using the minimum delay as reference – are detailed in [IETF RFC 5481].

Use of the average delay as the delay variation reference is depreciated in this version of this Recommendation.

In previous versions of this Recommendation, there was an alternative to using the minimum packet delay as the nominal delay: to use the average delay of the population of interest as the nominal or reference delay. This has the effect of centring the distribution of delay variation values on zero (when the distribution is symmetrical), and produces both positive and negative variations. However, the average delay of the population may be distinctly different from the delay of any individual packet, creating an artificial reference for variation (e.g., when a bimodal distribution is present).

6.2.4.2 Quantile-based limits on IP packet delay variation

The preferred method (used in ITU-T Y.154] objectives) for summarizing the delay variation of a population of interest is to select upper and lower quantiles of the delay variation distribution and then measure the distance between those quantiles. For example, select the $1 – 10^{–3}$ quantile and the 0 quantile (or minimum), make measurements, and observe the difference between the delay variation values at these two quantiles. This example would help application designers determine the de-jitter buffer size for no more than 0.1% total buffer overflow.

An objective for IP packet delay variation could be established by choosing an upper bound for the difference between pre-specified quantiles of the delay variation distribution. For example, "The difference between the 99.9 quantile and the minimum of the packet delay variation should be no more than 50 ms."

6.2.4.3 Interval-based limits on IP packet delay variation

An alternative method for summarizing the IP packet delay variation experienced by a population of interest is to pre-specify a delay variation interval, e.g., 50 ms, and then observe the percentage of individual packet delay variations that fall inside and outside of that interval. If the 50 ms interval were used, application with fixed buffer sizes of at or near 50 ms would then know approximately how many packets would cause buffer over or under-flow.

NOTE – If this method is used for summarizing IP packet delay variation, the delay variant of individual packets should be calculated using the minimum delay as nominal in clause 6.2.4.1, instead of the definition of clause 6.2.4 using the first packet. Using the definition of clause 6.2.4, the pre-selected interval (e.g., the 50 ms) might occasionally be anchored on an unusually large or small value.

An objective for IP packet delay variation could be established by choosing a lower bound for the percentage of individual packet delay variations that fall within a pre-specified interval. For example, "≥99.9% of packet delay variations should be within the interval [0 ms, 50 ms]".

6.2.4.4 Secondary parameters for IP packet delay variation

One or more parameters that capture the effect of IP packet delay variations on different applications may be useful. It may be appropriate to differentiate the (typically small) packet-to-packet delay variations from the potentially larger discontinuities in delay that can result from a change in the IP routing. Appendix II gives several secondary definitions of delay variation and guidance on their use.

6.3 IP packet error ratio (IPER)

IP packet error ratio (IPER) is the ratio of total errored IP packet outcomes to the total of successful IP packet transfer outcomes plus errored IP packet outcomes in a population of interest.

6.4 IP packet loss ratio (IPLR)

IP packet loss ratio (IPLR) is the ratio of total lost IP packet outcomes to total transmitted IP packets in a population of interest.

NOTE – Metrics for describing one-way loss patterns may be found in [b-IETF RFC 3357]. Consecutive packet loss is of particular interest to certain non-elastic real-time applications, such as voice and video.

6.5 Spurious IP packet rate

Spurious IP packet rate at an egress MP is the total number of spurious IP packets observed at that egress MP during a specified time interval divided by the time interval duration (equivalently, the number of spurious IP packets per service-second)[1].

6.6 IP packet reordered ratio (IPRR)

An IP packet reordered ratio (IPRR) is the ratio of the total reordered packet outcomes to the total of successful IP packet transfer outcomes in a population of interest.

Figure 10 illustrates an out-of-order packet outcome for packet 2, and a hypothetical tolerance on arrival time with a playout buffer that can restore order.

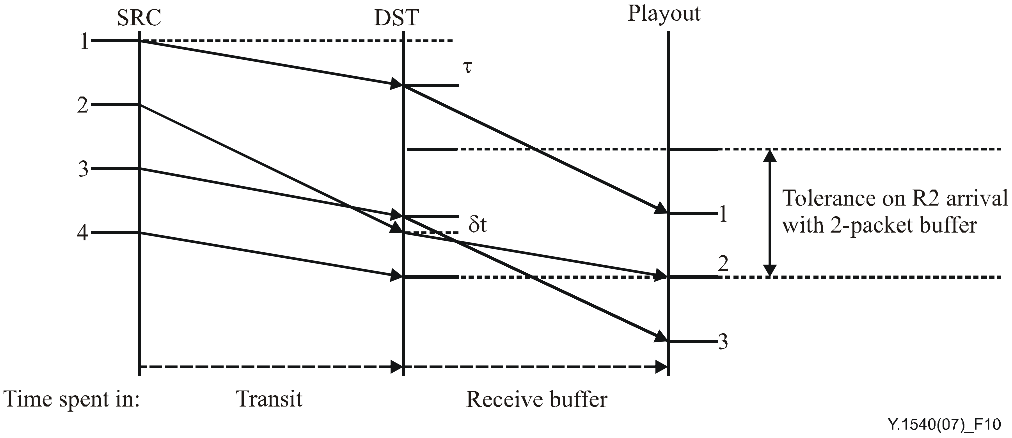


Figure 10 – Illustration of reordered arrival

If separate reordering events can be distinguished, then an event count may also be reported (along with the event criteria).

It is also possible to assert the degree to which a packet is reordered. Any packet whose sequence number causes the next expected value to increment by more than the standard increment indicates a discontinuity in the arrival order. From this point on, any (reordered) packets with a sequence number less than the next expected value can be quantified with a distance with respect to the discontinuity. The distance may be in units of position, time or the sum byte payloads of intervening packets. Referring to Figure 10 for an example, packet 2 can be said to be "late" by δt seconds, or 1 packet in terms of position.

[IETF RFC 4737] should be consulted for additional reordering parameters.

6.7 IP packet severe loss block ratio (IPSLBR)

An IP packet severe loss block ratio (IPSLBR) is the ratio of the IP packet severe loss block outcomes to total blocks in a population of interest.

NOTE – This parameter can identify multiple IP path changes due to routing updates, also known as route flapping, which causes significant degradation to most user applications.

[1] Since the mechanisms that cause spurious IP packets are expected to have little to do with the number of IP packets transmitted across the sections under test, this performance parameter is not expressed as a ratio, only as a rate.

6.8 IP packet duplicate ratio (IPDR)

IP packet duplicate ratio (IPDR) is the ratio of total duplicate IP packet outcomes to the total of successful IP packet transfer outcomes minus the duplicate IP packet outcomes in a population of interest.

6.9 Replicated IP packet ratio (RIPR)

The replicated IP packet ratio (RIPR) is the ratio of total replicated IP packet outcomes to the total of successful IP packet transfer outcomes minus the duplicate IP packet outcomes in a population of interest.

6.10 Stream repair parameters

Ideally, we would like to know the probability that a given packet interval (or information block, b) will contain more than x impaired packets.

$$P(b, x) = p, \text{ or } P(T_I, x) = p$$

Measurement of the impaired packet outcomes occurring in a *population of interest* should provide an empirical assessment of the probability during available time.

6.10.1 IP packet impaired interval ratio (IPIIR)

An IP packet impaired interval ratio is the ratio of the IP packet impaired interval outcomes to total non-overlapping intervals in a population of interest.

6.10.2 IP packet impaired block ratio (IPIBR)

An IP packet impaired block ratio (IPIBR) is the ratio of the IP packet impaired block outcomes to total non-overlapping blocks in a population of interest.

6.11 Capacity parameters

An end-to-end IP packet transfer service traverses an ordered sequence of basic sections from a source host to a destination host. The capacity parameters described below define properties for basic sections in terms of their ability to carry IP traffic, and corresponding properties for NSEs, also referred to as "paths". It is important to note that a basic section as well as a sequence of basic sections is associated with a direction. The direction is significant, as the properties of a sequence of sections in the forward direction need not be the same as in the reverse direction.

Note that, in contrast to the flow-related parameters defined in clause 6.12, the capacity-related parameters are not dependent on higher layer protocols on top of IP (e.g., TCP). Also, note that parameters for used capacity, utilization and available capacity are not externally measureable, and require support from monitoring systems with access to link utilization measurements, etc.

6.11.1 Section metrics

6.11.1.1 IP-layer bits transferred

For a given population of interest, the IP-layer bits transferred are defined as eight (8) times the number of octets in all IP packets generating successful IP packet transfer outcomes at an egress measurement point, from the first octet of the IP header to the last octet of the IP packet payload, inclusive.

Note that this definition is identical to the definition of IP-layer bits in [IETF RFC 5136]. Also, note that the definition of IP-layer bits is IP-version agnostic.

6.11.1.2 IP-layer section capacity

For a given population of interest, the IP-layer section capacity is:

$$C(t,\Delta t)=\frac{n_0(t,\Delta t)}{\Delta t}$$

where n_0 is the highest (total) number of IP-layer bits that can be transferred over a basic section generating successful IP packet transfer outcomes at the egress measurement point during a specified time interval [t, t + Δt].

Note that this is a conceptual definition, rather than a metric that can be measured repeatedly. The phrase "highest number" of bits is somewhat unclear, except in the case of repeated assessments. More likely, this refers to the total of bits in successful transfer outcomes during [t, t + Δt]. See Annex A for a normative definition of measureable IP-layer section capacity.

6.11.1.3 IP-layer used section capacity

For a given population of interest, the IP-layer used section capacity is:

$$U(t,\Delta t)=\frac{n(t,\Delta t)}{\Delta t}$$

where n is the actual number of IP-layer bits transferred over a basic section generating successful IP packet transfer outcomes at the egress measurement point during a specified time interval [t, t + Δt].

6.11.1.4 IP-layer section utilization

For a given population of interest, the IP-layer section utilization V(t, Δt) is defined as the ratio between the IP-layer used section capacity U(t, Δt) and the IP-layer section capacity C(t, Δt). That is:

$$V(t,\Delta t)=U(t,\Delta t)/C(t,\Delta t)$$

6.11.1.5 IP-layer available section capacity

For a given population of interest, the IP-layer available section capacity, A(t, Δt), is the unused portion of the IP-layer section capacity during a time interval [t, t + Δt]. This can be calculated as the difference between the IP-layer section capacity and the IP-layer used section capacity. That is,

$$A(t,\Delta t)=C(t,\Delta t)-U(t,\Delta t)$$

or, equivalently

$$A(t,\Delta t)=C(t,\Delta t)(1-V(t,\Delta t))$$

6.11.2 NSE metrics

6.11.2.1 IP-layer NSE capacity

The definition of IP-layer section capacity can be extended to an NSE, also referred to as a "path". For a given population of interest, the IP-layer NSE capacity C_{NSE}(t, Δt) during a specified time interval [t, t + Δt] is defined as the smallest IP-layer section capacity along that NSE. That is, the IP-layer NSE capacity is:

$$C_{NSE}(t,\Delta t)=\min_{i=1..n} C_i(t,\Delta t)$$

where C_i is the IP-layer section capacity of section number i (i=1..n) in the NSE.

6.11.2.2 IP-layer available NSE capacity

The definition of IP-layer available section capacity can be extended to an NSE, also referred to as a "path". For a given population of interest, the IP-layer available NSE capacity A_{NSE}(t, Δt) during a specified time interval [t, t + Δt] is defined as the smallest IP-layer available section capacity along that NSE. That is,

$$A_{NSE}(t,\Delta t)=\min_{i=1..n} A_i(t,\Delta t)$$

where A_i is the IP-layer available section capacity of the section number *i* (*i*=1..*n*) in the NSE. Note that the section number determining the IP-layer available NSE capacity may be different from the section number determining the IP-layer NSE capacity.

6.11.2.3 IP-layer tight section capacity

For a given population of interest, the IP-layer tight section is defined as the section in an NSE with the smallest IP-layer available section capacity. Note that if there are several sections fulfilling this condition the IP-layer tight section is not uniquely defined.

For a given population of interest, the IP-layer tight section capacity of an NSE is the IP-layer section capacity of the IP-layer tight section.

Note that the IP-layer available section capacity of the IP-layer tight section equals the IP-layer available NSE capacity. That is, the IP-layer tight section capacity is:

$$C_{TL}(t,\Delta t)=C_i(t,\Delta t) \text{ so that } A_i(t,\Delta t)=A_{NSE}(t,\Delta t)$$

Note that the IP-layer tight section does not necessarily have to be the same section as the section determining the IP-layer NSE capacity.

6.11.3 Variability

Each capacity metric P represents an average value over a time interval [*t*, *t* + Δ*t*]. For a set of consecutive observations $P_1..P_N$ for a given parameter *P* over an interval [*T*, *T* + *ΔT*], where *T* > *t*, the average, standard deviation and quantiles can be used to describe the variability.

6.11.3.1 Average

The average is calculated as:

$$a_P(T,\Delta T)=\frac{1}{n}\sum_{i=1..n}P_i(t,\Delta t)$$

6.11.3.2 Standard deviation

The standard deviation is calculated as:

$$s_P(T,\Delta T)=\sqrt{\sum_{i=1..n}\left(P_i(t,\Delta t)-a_P(T,\Delta T)\right)^2}$$

6.11.3.3 Quantiles

For a sorted list of *N* values $P_1..P_n$ the *k*th 100-quantile (i.e., *k*th percentile) is defined as:

$$P_I : I=\left\lceil N\frac{k}{100}\right\rceil$$

where P_I is the corresponding data value for the *k*th 100-quantile. (The symbol $\lceil\ \rceil$ means that if $N\frac{k}{100}$ is not an integer it should be rounded up to the next higher integer to get the list index *I*.)

The quantiles for minimum (*k* = 0), median (*k* = 50) and maximum (*k* = 100) are of special interest and should be reported. Other quantiles, such as *k* = 95 or *k* = 99, may also be used.

6.12 Flow-related parameters

It is useful to characterize performance in terms of flow or throughput-related parameters that evaluate the ability of IP networks or sections to carry quantities of IP packets. It should be noted that a parameter intended to characterize the throughput of an IP application would not be equal to the amount of resources (capacity) available to that application (as quantified in clause 6.11 and measured with the methods in Annex A). This is because the higher layer protocols employing closed-loop flow control over IP (e.g., TCP flow control) also influence the IP-layer throughput experienced.

In the present version of this Recommendation, it is recommended that all flow or throughput-related parameters should fulfil the following requirements:

1) A parameter characterizing the throughput offered to an IP service should relate the amount of IP packets successfully transported by an IP network or section to the amount of IP packets that were delivered into this network or section.
2) The throughput-related parameter should apply to an end-to-end IP network and to the IP transport across an EL, NS or NSE.

Some flow or throughput-related parameters attempt to characterize the throughput capacity of an IP network, i.e., its ability to sustain a given IP packet transfer rate. It is recommended that any such parameters and methods of measurement fulfil the following additional requirements:

1) The traffic pattern offered to the IP network or section should be described, since the ability of the IP network or section to successfully deliver these packets depends on this traffic pattern.
2) The rate at which traffic is offered should not exceed the capacity (in bits per second) of the link that connects the sections under test with the destination sections that are not under test.
3) In any individual statement about throughput performance, the type of IP packet considered should be declared (where type includes the IP version, the presence of extension headers, the transport layer protocol, the other protocol headers, and any other relevant information, such as the length of the packets used).
4) Some forms of flow control applied at the IP layer or above can result in measurement errors. For example, an IP layer throughput measurement which is subject to a higher-layer acknowledgement tracking, window size limitations, and/or flow control to mitigate congestion (e.g., TCP), requires assessment and reporting of the associated measurement error. The measurement error indicates possible unused IP layer capacity when comparing Internet service specifications and results from methods using flow control.

It is also recommended to follow the guidelines for throughput-related parameters and their measurement found in the IETF RFC 3148 framework for bulk transfer capacity (BTC) metrics.

Parameters related to flow and throughput are specified in Annex A. Appendix IX describes how measurements using TCP do not meet the requirements of this clause.

Note: Appendix III (Rate and throughput capacity related parameters) is deprecated in the 2019 edition.

7 IP service availability

IP service availability is applicable to end-to-end IP services, basic sections and NSE.

An availability function (defined in clause 7.1) serves to classify the total scheduled service time for an IP service into available and unavailable periods. On the basis of this classification, both percent IP availability and percent IP unavailability are defined in clause 7.2. Finally, a two-state model of IP service availability serves as the basis for defining related availability parameters in clause 7.2.

NOTE – Unless otherwise noted by an IP service provider, the scheduled service time for an IP service is assumed to be 24 hours a day, seven days a week.

7.1 IP service availability function

The basis for the IP service availability function is a threshold on the IPLR performance.

The IP service is available on an end-to-end basis if the IPLR for that end-to-end case is smaller than the threshold c_1 defined in Table 1.

Table 1 – IP service availability function

Outage criterion	Threshold
IPLR > c_1	$c_1 = 0.20$
NOTE – The value of 0.20 for c_1 is considered provisional and is identified as requiring further study. The previous provisional value for c_1 was 0.75. Values of 0.9 and 0.99 have also been suggested for c_1. However, at the time of approval of this Recommendation the majority of causes for unavailability appear to stem from failures where the loss ratio is essentially 100%, and many applications of IP networks are no longer operating when loss ratio is >0.20. When IP networks support multiple qualities of service, it may be appropriate to consider different values of c_1 for different services. In this case, c_1 values of between 0.03 and 0.2 (based on resilience of different speech coders) have been suggested for services offering ITU-T Y.1541 class 0 or class 1, and c_1 of 0.75 for ITU-T Y.1541 class 5. The threshold c_1 is only to be used for determining when the IP network resources are (temporarily) incapable of supporting a useful IP packet transfer service. The value c_1 should not be considered a statement about IPLR performance nor should it be considered an IPLR objective suitable for any IP application. Performance objectives established for IPLR should exclude all periods of service unavailability, i.e., all time intervals when the IPLR > c_1.	

Relative to a particular SRC and DST pair, *a basic section or an NSE is available for the ingress-independent case* if the IPLR for that pair is smaller than the threshold c_1, as measured across all permissible ingress MPs.

Relative to a particular SRC and DST pair, *a basic section or an NSE is available for the specific-ingress case* if the IPLR for that pair is smaller than the threshold c_1, as measured from a specific permissible ingress MP.

NOTE 1 – From an operations perspective, it will be possible to measure and/or monitor availability from a specific ingress MP and then use this information to create inferences about the ingress-independent availability.

NOTE 2 – The quantitative relationship between end-to-end IP service availability and the IP service availability of the basic section or NSE remains for further study.

If the outage criteria given by Table 1 is satisfied (i.e., IPLR exceeds its threshold), the IP service is in the unavailable state (experiences an outage). The IP service is in the available state (no outage) if the outage criteria is not satisfied. The minimum number of packets that should be used in evaluating the IP service availability function is M_{av} (the value of M_{av} is for further study. When tests of availability use end-user generated traffic, M_{av} of 60 packets has been suggested, disbursed within T_{av} at one packet per second). The minimum duration of an interval of time during which the IP service availability function is to be evaluated is T_{av}. T_{av} is provisionally defined to be one minute. Study has revealed that this value is consistent with practical limits on IP layer operations. The monitoring of lower layer performance and network element faults may be able to identify impending unavailability in a shorter time, and direct corrective action. Appendix VI gives the rationale for the current IP service availability function definition and values for T_{av} and c_1.

NOTE 3 – The outage criterion based on the IPLR is expected to satisfactorily characterize IP service availability. However, IP service availability might also take into account severely degraded performance for IPER and/or spurious IP packet rate. The inclusion of additional availability decision parameters and their associated thresholds remains for further study.

NOTE 4 – This unidirectional definition of availability is motivated by the fact that IP packets often traverse very different routes from the SRC to the DST than they traverse from the DST to the SRC. If, from an IP network user perspective, a bidirectional availability definition is needed, a bidirectional definition can be easily derived from this unidirectional definition, by summing the non-overlapping unavailable time of the reverse path.

It is intended that this definition of IP service availability be applicable to both end-user generated IP traffic (i.e., the normal flow of IP packets between the SRC and the DST) as well as to traffic generated by test sets and test methodologies. In either case, the source of the IP traffic should be documented when reporting availability findings. Such documentation should include the specific types of packets used in each direction of flow.

Traffic generated specifically to test the availability state should be limited so that it does not cause congestion. This congestion could affect other traffic and/or could significantly increase the probability that the outage criteria will be exceeded.

More information on the determination of the availability state can be found in Appendix IV.

7.2 IP service availability parameters

7.2.1 Percent IP service unavailability (PIU)

The percent IP service unavailability (PIU) is the percentage of total scheduled IP service time (the percentage of T_{av} intervals) that is (are) categorized as unavailable using the IP service availability function.

7.2.2 Percent IP service availability (PIA)

The percent IP service availability (PIA) is the percentage of total scheduled IP service time (the percentage of T_{av} intervals) that is (are) categorized as available using the IP service availability function; PIU and PIA are related as:

$$\text{PIU} = 100 - \text{PIA}$$

NOTE – Because the IPLR typically increases with increasing offered load from the SRC to the DST, the likelihood of exceeding the threshold c_1 increases with increasing offered load. Therefore, PIA values are likely to be smaller when the demand for capacity between the SRC and DST is higher.

Appendix IV provides information on sampling to determine the PIA and PIU.

Annex A

IP-based capacity and flow-related parameters and methods of measurement

(This annex forms an integral part of this Recommendation.)

A.1 Background

Standardization of measurement architectures characterizing "the Internet service" started in 2013. A consumer network access product supporting the Internet Protocol is a common way of connecting to the Internet. Traditionally, the majority of consumer applications communicate using TCP transport. TCP provides reliable datagram transmission. When trying to characterize an Internet service by standardized measurements, methods to characterize the TCP transport layer properties have been used for the judgement of Internet or link-layer properties, like access bandwidth. Concerns have been raised, that this approach may not work properly and alternate methods have been proposed to characterize consumer Internet access services based on measurements.

At the same time, there is a strong trend to replace TCP transport with UDP transport, payloads with open and encrypted portions, and application-layer retransmission and congestion-control. The use of Google QUIC and impending approval of IETF QUIC [b-QUIC] will rapidly change the transport landscape of the Internet, and consumers using popular browsers are/will be among the earliest adopters.

This annex specifies parameters and methods for access bandwidth measurement of maximum IP-layer capacity in the present text. In the future, flow-related (transport throughput) measurement parameters and method may be specified here after further study. These parameters and methods of measurement are fully applicable to Internet service subscribers (end users) and Internet service providers alike, with the caveat that the selected ingress and egress measurement points fully determine the applicability when comparing measurements with IP-layer service specifications.

ITU-T Study Group 12 (SG12) is experienced in developing measurement and modelling standards for telecommunication. Further, expertise on statistical methodology and toolsets allowing an objective comparison between proposed measurements and modelling methods are available. Using SG12 expertise and methods to decide on the applicability of competing proposals related to the characterisation of "consumer Internet service" is agreed as a good way to move from an exchange of views to a proof and comparison of concepts under reproducible testing conditions. [b-ITU-T P.800] provides the following requirements:

> It is important that the conditions simulated in the test are correctly specified and set up, and measured accurately before and after each experiment; … and that faithful records of the output of each test are kept.

The IETF has approved and published RFC 8337, "Model-Based Metrics for Bulk Transport Capacity", in March 2018 [IETF RFC 8337]. The model-based metrics (MBM) work was the result of many years considering the problem of transport capacity measurement, primarily in the IETF IP Performance Metrics (IPPM) working group. The specification carefully describes the many issues and difficulties with repeatability when testing with standard-compliant TCP (section 4), and solves these problems principally by designing a method and a set of diagnostic tests where the TCP's flow-control is disabled. The method involves the evaluation of a target transport performance in terms of transmission rate and round-trip time (RTT).

Until this Recommendation, there have been tools (e.g., iPerf) and limited mentions of UDP-based IP capacity measurements in published standards and reports. Today, the industry sees the delivery of both a measurement standard and a compliant tool as the ideal way to deliver new metrics and methods. This annex provides a sufficiently detailed specification for implementation that will produce statistically equivalent results (as described in [b-IETF RFC 6576] from IPPM work). A measurement tool will be provided separately which is compliant with this annex.

The test plan to evaluate relevant measurement methods is included in this annex. The test plan uses the SG12 principles of [b-ITU-T P.800] to establish a "ground truth" for comparing candidate methods of measurement. The testing was divided into phases: Laboratory tests were conducted according to the Phase 1 plan, where the implementation of "ground truth", the limits of test platform, and the candidate methods of measurement could be examined under controlled conditions. Phase 2 tests were conducted in the field, according to the test plan.

The summarized results of evaluation (using the test plan) are provided in Appendix X (informative), for both the Phase 1 Laboratory and Phase 2 field tests. Other appendices provide information collected to support the consensus developed here, and additional detail on calculations and conversion factors between measurements at different layers.

A.2 IP-layer access capacity parameters and methods of measurement (consumer Internet access)

A.2.1 Measurable IP capacity metric definition

Clause 6.11.1.2 defines an idealized parameter for IP-layer section capacity, which assesses the "highest (total) number of IP-layer bits" that can be transferred … during a specified time interval [*t*, *t* + Δ*t*]." This clause defines a measurable (more practical) version of the idealized definition.

For a given population of interest, the maximum IP-layer section capacity during time interval [*t*, *t* + Δ*t*] is:

$$Maximum\_C(t,\ \Delta t) = \frac{\max\limits_{[t,\Delta t]}\left(n_0(dt_n, dt_{n+1})\right)}{dt}$$

where:

time interval [*t*, *t* + Δ*t*] is composed of x equal sub-intervals, *dt in length*;

n_0 is the total number of IP-layer header and payload bits that can be transferred over a basic section generating successful IP packet transfer outcomes at the egress measurement point during a specified time interval, from [dt_1, dt_2] or other intervals *dt* in length, and

the maximum *C*(*t*, Δ*t*) corresponds to the maximum value of n_0 measured in any sub-interval [dt_n, dt_{n+1}] within time interval [*t*, *t* + Δ*t*], divided by the duration of the sub-interval.

Note that UDP transport shall be used when assessing the measureable IP capacity metric.

The method of measurement also needs a definition for its sending rate, supplied below.

A.2.1.1 IP packet sending bit rate (IPSBR)

For a given population of interest, the IP packet sending bit rate (IPSBR) generated by a sender at an ingress MP is 8 times the total number of octets transmitted in IP packet headers and payloads that result in an IP packet transfer reference event at that ingress MP during a specified time interval, divided by the time interval duration. Equivalently, the number of bits in the IP packet headers and payloads resulting in IP packet reference events per service-second.

A.2.2 Method of measurement

The overall steps in the method (procedure) are as follows:

- The sender shall arrange to send and receive the stream of IP packets using UDP transport-layer with key parameters defined, including:

  τype of packet, including header and payload lengths, headers and options present and any markings for special treatment in the network;

τhe starting/variable sending rate for packets during a specified time interval (e.g., the interval consistent with the HZ parameter, much less than *dt* which is the sub-interval for reporting during testing);

τhe length, duration and characteristics of a test preamble or priming phase (essential for certain types of networks, such as mobile networks);

τhe specific sending discipline, which includes allowed or intended burstiness (back-to-back sending).

- During a test, the sending rate shall be varied in accordance with a specified search algorithm, with:

    specified search goal, including one or more measured metrics and their corresponding operational thresholds, and tolerance above and below those thresholds;

    σpecified duration of trials, which are composed of the individual steps in a search algorithm;

    τhe set of measurements at sub-intervals, *dt*, which support the summary measurements conducted at the conclusion of each trial;

    τhe maximum duration of the search process (time limit).

    There shall be one mandatory search algorithm specified in this annex; other algorithms are mandatory to implement, or optional (as specified in the clause or annex containing the algorithm). Use of a mandatory or optional search algorithm shall be reported to the user with the results.

- There shall be storage of all measurements (results of trials) collected throughout the search, stored as a time series to allow analysis of the search process.
- There shall be defined post-processing of all measurements (results of trials) collected throughout the search, to determine the concluding summary value(s) of a particular search process. Examples include (for received rate, or here, measured capacity, other parameters include latency, loss and reordering):

    calculating an average of all measured values of received rate for all trials;

    calculating an average of all measured values of received rate where the search goal was met;

    calculating a maximum of all measured values of received rate where the search goal was met;

    calculating an average of all measured values of received rate for all trials where a specified result exclusion criteria has been met (e.g., removal of outliers, as determined by specified criteria);

    summarization of related parameters (see A.2.3 below).

- There shall be defined criteria to invalidate the results of a given search process, such as the detection of failure to achieve the required sending rate, or detection of competing traffic (but this traffic cannot be exhaustively detected in all circumstances).
- When tests are repeated to examine the consistency of the results or other reasons, the process to summarize the results should include corresponding post-test analysis to ensure data quality and to detect and exclude data artefacts (where possible). The post-test analysis methods shall be published with the results when used.
- Security: The test receiver (or server) shall arrange to accept testing requests only from the population of authorized users, and reject others.

- Capacity: The test receiver (or server) shall arrange to accept testing requests only when sufficient host and interface resources are available, and reject requests when this condition is not met.
- Reporting results: The measurement system shall report the maximum IP capacity, the trial IP packet loss ratio, and other metrics when available (see clause A.2.4). The measurement system may also report UDP capacity in terms of UDP payload bits delivered, because this is the capacity available to user applications after IP and UDP headers have been removed.

A.2.2.1 Mandatory search algorithm

The measurement system meets the requirements of clause A.2.2, and adds the following capabilities to support the search algorithm:

1. The tester should make a recommendation on maximum test packet size, and allow for some unexpected overhead to avoid fragmentation.
2. There is a table of transmit rates (IPSBR), the number of packets sent during each interval and packet sizes. The table has ascending values for offered load rates, between minimum and maximum supported load rates, inclusive.
3. The receiver of the offered load shall measure the following metrics: received rate, loss, reordering, delay variation (as per this Recommendation), and round-trip delay [Y.1565].
4. The receiver of the offered load shall periodically send a status feedback message to the sender with the results of the measured metrics.
5. Based on the results contained in the status feedback message, the sender shall adjust its offered load according to the flowchart in Figure A.1. In the flowchart, "one step" is a change in rate accomplished by using a new value (in the row above or below the current row for the sending) in the table of offered load rates.

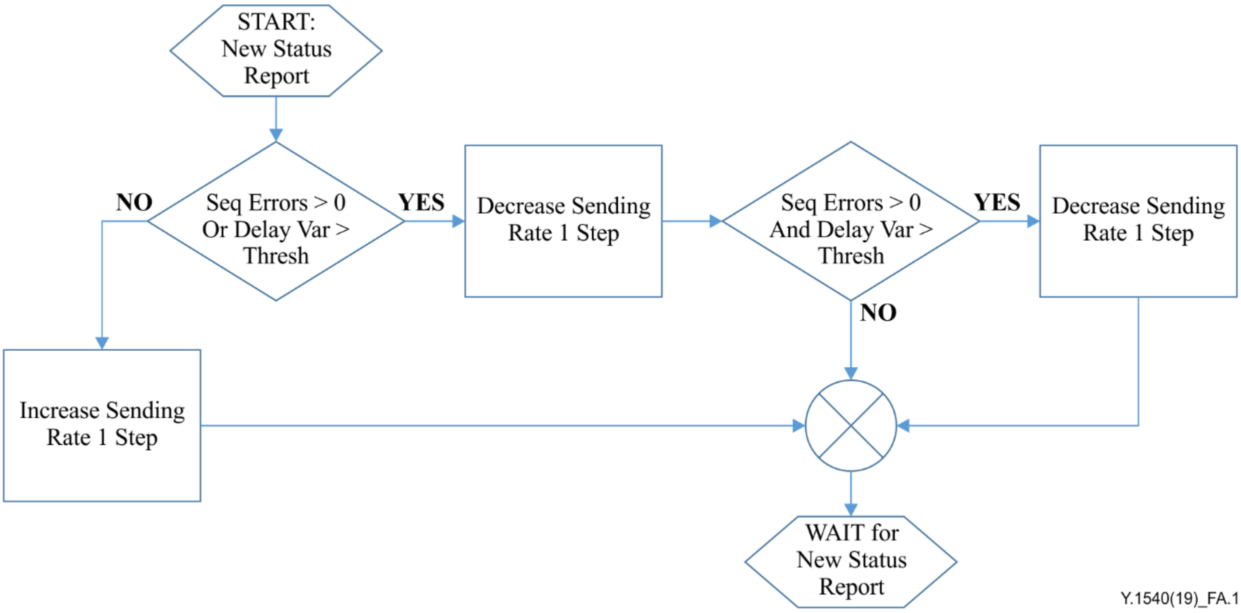


Figure A.1 – Flowchart for offered load adjustment as part of a search algorithm

NOTE – The algorithmic decisions could be performed by one of the measurement system's two participating hosts, which would make implementation at the other host much less complex and independent of the algorithm version. This also allows algorithm updates at a more conveniently accessible host.

In the future, alternative mandatory or optional search algorithms will be specified in separate annexes.

A.2.3 Trial IP packet loss ratio (IPLR) metric

For a given population of interest, the trial IP packet loss ratio (IPLR) is the ratio of total lost IP packet outcomes to the total of lost plus successful IP packet outcomes during each trial (or sub-interval, *dt*).

Note that the trial IPLR varies slightly from the IPLR definition in the body text, because it is not possible to use a long waiting time (T_{max}) to distinguish successful packet outcomes from lost packet outcomes when evaluating results of each trial (or sub-interval, *dt*) and immediately passing results to the search algorithm. Unusually long delayed or reordered packets would be accounted for in measurements of subsequent trials, in all parameters including IPDV.

A.2.4 Related parameters and methods

Clearly, trial IPLR, IPTD (evaluated as round-trip transfer time, or RT latency, in accordance with the parameter in [Y.1565]), and IPDV are of interest, and shall be reported along with the concluding summary of the received rate (measured capacity).

Additional measured parameters shall be available:

- the actual duration of the search process (intended to be Δt)
- the total number of lost packets observed during the search process
- The range and variation of results over repeated search processes.

A.3 Flow-related throughput parameters and methods of measurement (reliable delivery transport)

This clause, which is for further study (FFS), will specify metrics and methods of metrics for flow-related throughput, in accordance with clause 6.12 of this Recommendation. Note that Appendix X of this Recommendation explains why standard TCP with closed-loop flow-control does not meet these requirements. A candidate metric and method is described in Appendix XIII of this Recommendation, and this candidate meets the requirements of clause 6.12.

A.3.1 Parameter definition

FFS; see Appendix XIII.

A.3.2 Method of measurement

FFS.

A.3.2 Related parameters and methods

FFS.

A.4 Plan to qualify and compare access measurement methods

Recommendation ITU-T P.800 and other Recommendations of that series, although primarily aimed at voice, provide general guidance on how to set up, perform and evaluate measurement campaigns aiming to compare models based on measurement input with true conditions. To understand the capabilities and the limits of tests, tools and results designed to characterize Internet access types, a similar process is useful, as described below.

A set of various transmission conditions should be arranged for laboratory testing. The transmission conditions that are close to real operational conditions should first be defined. The conditions should be stable and verifiable during the tests, using laboratory measurement equipment. Then each method designed to characterize various types of Internet access can be tested against a set of various conditions. Measurement errors can be characterized. From the performance of each parameter and method, operational areas (conditions) can be identified where a measurement method has merits and where it is less reliable (and sources of measurement error can be quantified).

The testing conditions are listed below.

A.4.1 Measurement points

[b-IETF RFC 7398] defines a reference path and measurement points for commonly used performance metrics. Other similar measurement projects may also be able to use the extensions described here for measurement point location. The purpose of [b-IETF RFC 7398] is to create an efficient way to describe the location of the measurement point(s) used to conduct a particular measurement, especially pointing-out when a measurement includes managed and non-managed (private network) sub-paths.

Note that the measurement path, as delimited by the [b-IETF RFC 7398] measurement points, determines the applicability of subscription parameters, such as typical data rates offered and whether the subscription parameters inform the selection of parameters such as the [IETF RFC 8337] MBM target_data_rate. Also, parameters such as the frame size used in UDP tests, when considering the layer at which test traffic is inserted, to avoid fragmentation.

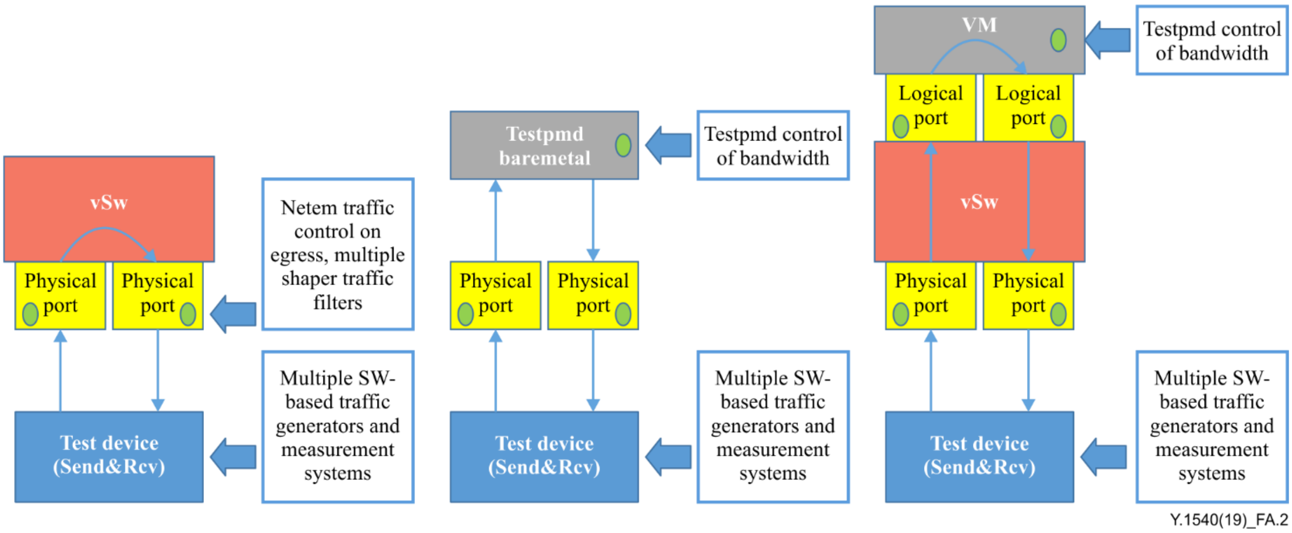
A.4.2 Testing conditions

All configurations and all background traffic behaviour should be as close as possible to operational networking conditions. It is suggested to define a baseline test and then to vary testing parameters. This is the Phase 1 testing.

A.4.2.1 Phase 1 laboratory testing conditions

Following on from the BEREC Requirements, [b-BEREC] the first phase test will be with Shapers (and policers up to 10 Mbps, [b-Google-Police]), and the tester must be aware of the layer where shapers, policers, passive observers, senders, receivers and end-of test results meter packets. In addition, network impairments such as delay will be implemented and tested.

The figure below illustrates two different utilities providing flexible traffic/rate control on a general purpose computing platform. The different traffic control utilities can be applied in three ways.



VM
Testpmd control of bandwidth
Logical port
Logical port
Testpmd baremetal
Testpmd control of bandwidth
vSw
vSw
Netem traffic control on egress, multiple shaper traffic filters
Physical port
Physical port
Physical port
Physical port
Physical port
Physical port
Multiple SW-based traffic generators and measurement systems
Multiple SW-based traffic generators and measurement systems
Multiple SW-based traffic generators and measurement systems
Test device (Send&Rcv)
Test device (Send&Rcv)
Test device (Send&Rcv)
Y.1540(19)_FA.2

Figure A.2 – Three alternative test paths on a general purpose computing platform

In Figure A.2, the test device is connected to a general purpose host via physical links at 10 Gbps. The test device is also a general purpose host, but it is completely isolated from the host performing traffic control, allowing each node to dedicate resources to their unique roles in the test environment. Different implementations of candidate measurement methods can be installed and tested through the traffic control node.

There are three alternatives to implement the traffic shaper function. The first on the left employs the Linux kernel-equipped emulator, netem, which can emulate delay and assist with the traffic control after configuration for the correct NIC and physical interfaces. The vSwitch simply switches frames between its two ports. This configuration is commonly referred to as "phy2phy".

The second alternative (middle) assumes that the Intel DPDK testpmd utility will be installed and configured to perform frame forwarding between the correct physical interfaces, along with control of bandwidth passing through the forwarding path.

The last alternative (right) uses testpmd installed in a virtual machine, or VM, and employs a vSwitch with configuration to connect the physical interfaces to the correct ports on the VM. Again, testpmd (or other utility running in the VM) controls the bandwidth of the path between the logical ports.

All three of these configurations are possible on the OPNFV VSPERF project pod, assigned for testing, development of, and evaluations using the VSPERF tool [b-Pod12].

Different conditions may then apply for single tests. In each test, only a single property is changed against the baseline test, all other configuration properties remain unchanged.

– The shaper/policer bandwidth are set to the rates as suggested by BEREC, up into the Gbps range, and will be tested.
– RTTs are set to the following values: 5, 10, 20, 40 ms.
– Random packet loss ratio is set to the following values: 0, 10**-4. 10**-5.
– The burst tolerance of the shaper may be set to 0 or up to the maximum the shaper allows (5kbit).
– The methods of rate limiting are shaping and policing (but not both simultaneously).
– The queue overflow strategy is tail drop.
– The tests will be executed without background traffic. An additional test may be performed with background (competing) traffic. The results and the results and the average background load will be published with the results.
– All test and background traffic will be best effort.
– The maximum measurement interval for an individual measurement is 30 seconds.
– For each individual measurement, tolerance of configured bandwidth as compared to a bandwidth measured during a test must be within 5% to accept a bandwidth as correctly measured by the candidate measurement system.
– The measurement software as used for the test must be available under an open source license. Commercial products may defer this requirement until standardization of a method begins. The test systems must be calibrated, and Q17 should know the limits of any system submitted for evaluation. Also, the details of the development environment and required operating system must be provided.
– Results will be published as an appendix to this Recommendation.
– The ETH frame sizes correspond to a layer 2 frame size of 64 bytes and in addition max ETH MTU size of 1512 (and 1516 including ETH CRC) bytes.
– It is desirable to test using the IPv6 address family, in addition to IPv4.

Standard laboratory test equipment may be used to calibrate network properties [b-TST 009] like maximum IP layer bandwidth, minimum and maximum round-trip time, buffer depth and so on prior to every single measurement tool per network test condition. In this way, network performance benchmark information against which to compare measurement tool results and related evaluations can be captured.

The benchmark for correct shaper operation according to configuration is UDP CBR transmission, and ultimately for correct assessment of the channel conditions (consistent with [b-PAM-12]). If the UDP capacity and shaper configuration differ, a third arbiter of correctness is packet capture of the UDP stream. Identification and discussion of any errors encountered is essential in benchmarking.

Preliminary laboratory tests for calibration and measurement method evaluation have been completed. Taking guidance from the existing text of Annex A of this Recommendation and the BEREC requirements for verifying measurement tools, AT&T tested one of the three alternative configurations using a general purpose compute platform that will meet the requirements (phy2phy).

Key points to take away from this testing:

- The bidirectional throughput was 213.85 Mbps, fairly close to the configured value of 100 Mbps x2. The excess may be an artefact of the burst size allowed.
- The binary search with loss verification was configured to allow a fairly large tolerance in Mbps to accept the result. This tolerance should be reconsidered when testing sub-Gbps rates.

A.4.2.2 Phase 2 testing conditions

To advance the testing program, this plan will employ the UDP benchmark on in-service networks (with validated service parameter specifications) for further comparisons of service specifications and methods, such as iPerf 2 TCP methods and UDP-based methods. This is similar to [b-PAM-12], and other references, as described below.

Test of production in-service network equipment configurations MAY also be conducted in the laboratory environment, where possible and on a volunteer basis.

Phase 2 tests will be conducted in a similar manner to the 2012 work by Goga and Teixeira [b-PAM-12]. From available knowledge of the measurement systems in widespread use today, none are using the method that has been shown in the past to produce the most accurate estimate of IP capacity – UDP-based measurements [b-PAM-12]. Unfortunately, [b-PAM-12] was conducted when typical access capacities were <50 Mbps, and now (over 5 years later), this capacity range has even been exceeded on mobile Internet access. Key points to take away from [b-PAM-12] are:

- iperf was used to emulate flooding-based tools, because it allows us to set the number of parallel connections, the duration or the size of the transfer among other parameters.
- "Benchmark. The first line of Table. 1 shows the UDP capacity obtained when flooding the link with iperf in UDP….. The UDP capacity represents the maximum achievable IP rate for each line."
- Spruce, pathload with large probes, and parallel TCP are the most accurate tools to estimate available bandwidth (a.k.a. residual capacity), but there are regions where parallel TCP tests fail to produce useful results, namely if unidentified parallel traffic is present.

The general drawbacks of existing measurement systems in use during the 2012 time frame are analyzed in [MortonPQS].

Phase 2 tests should be conducted for each basic access type.

Basic access types:

Wireline: DSL consumer access, broadband cable access, fiber access, others.

Wireless: WiFi hot spot, UMTS, LTE wireless access, others.

The results and test paths used in Phase 1 are defined as a foundation for Phase 2. Phase 2 will cover several types of Internet access, e.g., an access which has typical properties for a national telecommunications market. The same holds for all other parameters, e.g., a typical RTT for popular content should be assumed, background traffic should be typical and so on. Thus, the networking conditions should be close to a real subscriber environment.

Different conditions may then apply for single tests. In each test, only a single property is changed against the baseline test, all other configuration properties remain unchanged.

– Background traffic may be added on the access (with or without Diffserv features, as may be expected in a particular market).

Standard laboratory test equipment may be used in addition, to investigate the measured network properties.

Packet sizes should be provided when known, but there is potential size variability and this variation may be the result of network conditions requiring retransmission (primarily for TCP, other tools may use fixed or variable sizes).

Annex B

Additional search algorithm for IP-based capacity parameters and methods of measurement

(This annex forms an integral part of this Recommendation.)

B.1 Search algorithm

This measurement system meets requirements of clause A.2.2, and adds the following capabilities to support an alternate, and mandatory-to-implement, search algorithm, referred to as the Annex B Search Algorithm (which is test protocol agnostic):

1. The tester should make a recommendation on maximum test packet size, and allow for some unexpected overhead to avoid fragmentation.
2. A table of transmit rates, which are the number of packets sent during each time interval (corresponding to bits per second and a specified protocol layer) and packet sizes. The table has ascending values for offered load rates, between the minimum and maximum supported load rates, inclusive.
3. The receiver of the offered load shall measure the following metrics: received rate, loss, reordering, delay variation (as per this Recommendation), and round-trip delay [ITU-T Y.1565].
4. The receiver of the offered load shall periodically send a status feedback message to the sender with the results of the measured metrics.
5. Based on the results contained in the status feedback message, the sender shall adjust its offered load according to the flowchart in Figure B.1. In the flowchart, "one step" is a change in rate accomplished by using a new value (of the row above or below the current row for the sending rate and packet sizes) in the table of offered load rates.

The flow chart in Figure B.1 uses many variable names and, in some cases, configurable thresholds that determine the flowchart decisions. There are three main paths through the flowchart: when feedback indicates measured impairments are absent, or when impairments are first measured and some congestion may be present but sending rate change is deferred, or when measured impairments are confirmed by repeated measurement feedback.

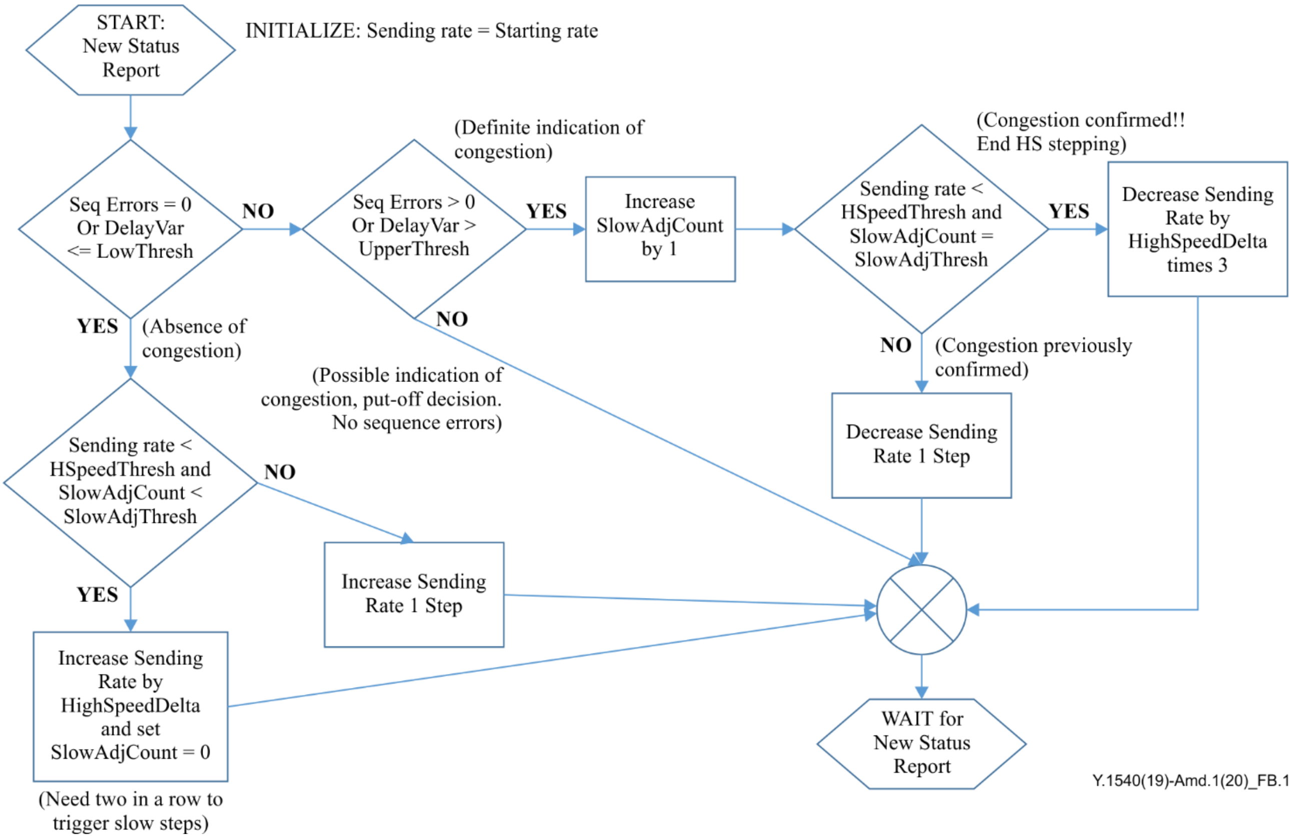

START: New Status Report
INITIALIZE: Sending rate = Starting rate
Seq Errors = 0 Or DelayVar <= LowThresh
NO
YES
(Absence of congestion)
(Definite indication of congestion)
Seq Errors > 0 Or DelayVar > UpperThresh
YES
NO
Increase SlowAdjCount by 1
Sending rate < HSpeedThresh and SlowAdjCount = SlowAdjThresh
YES
NO
(Congestion confirmed!! End HS stepping)
Decrease Sending Rate by HighSpeedDelta times 3
(Congestion previously confirmed)
Decrease Sending Rate 1 Step
(Possible indication of congestion, put-off decision. No sequence errors)
Sending rate < HSpeedThresh and SlowAdjCount < SlowAdjThresh
NO
YES
Increase Sending Rate 1 Step
Increase Sending Rate by HighSpeedDelta and set SlowAdjCount = 0
(Need two in a row to trigger slow steps)
WAIT for New Status Report
Y.1540(19)-Amd.1(20)_FB.1

Figure B.1 – Flowchart for offered load adjustment, Type B Search Algorithm

NOTE – The algorithmic decisions could be performed by one end of the measurement system's participating hosts, which would make implementation at the other host much less complex and independent of the algorithm version.

The variables and thresholds used in Figure B.1 are explained in Table B.1.

Table B.1 – Flowchart variables, descriptions, ranges and default values

Category/ Variable name	Description	Unit	Range	Default value
Sending Rate	The current sending rate (equivalent to a row of the table), initialized at minimum Sending Rate in the Table of Sending Rates	Kbps	500 ≤ # ≤ 10 000 000 (10 Gbps)	See starting rate
Start sending rate	Initial value of sending rate	Kbps	NA	500 Kbps
Seq Errors	Count of any of Loss or Reordering impairments measured (events where received packet sequence number did not increase by one)	number	NA	0 (no sequence errors)
DelayVar	Range of round trip time, RTT (or 1-way packet delay variation, above minimum delay when DelayVar 1-way measurements are reliable)	ms	NA	NA

Table B.1 – Flowchart variables, descriptions, ranges and default values

Category/ Variable name	Description	Unit	Range	Default value
LowThresh	Low threshold on the range of round trip time variation, RTT (Range is values above minimum RTT)	ms	$5 \leq$ ms ≤ 250	30 ms default
UpperThresh	High threshold on the range of round trip time variation, RTT (Range is values above minimum RTT)	ms	$5 \leq$ ms ≤ 250	90 ms default
HighSpeedDelta	The number of rows to move in a single adjustment when initially increasing offered load (to ramp-up quickly)	Number of rows	≥ 2	10 table rows (10 Mbps currently)
SlowAdjCount	Number of consecutive status reports indicating loss and/or delay variation above UpperThreshold	Count of occurrences	NA	See SlowAdjThresh
SlowAdjThresh	Threshold on SlowAdjCount used to infer congestion. Use values > 1 to avoid misinterpreting transient loss	Count of occurrences	> 1	2
HSpeedThresh	Threshold for transition between low and high sending rate step sizes (such as 1 Mbps and 100 Mbps). May result in use of jumbo frames if permitted.	Gbit/s		1 Gbit/s

Table B.2 gives the default input factors for Annex A method, for use with Annex B.

Table B.2 – Measurement variables, ranges, and default values

Category/ Variable Name	Parameter	Unit	Range	Default value
Max IP-layer Capacity				
	Number of parallel connections	#	$1 \leq \# \leq 10$	1 connection
	Duration of preamble to testing	s	$0 \leq s \leq 5$	~2 s
Δt	Duration of the test (either downlink or uplink) with search algorithm in use, which serves as the maximum duration of the search process.	s	$5 \leq s \leq 60$	10 s
Δt	Duration of the fixed rate test (either downlink or uplink)	s	$5 \leq s \leq 60$	10 s

Table B.2 – Measurement variables, ranges, and default values

Category/ Variable Name	Parameter	Unit	Range	Default value
dt	Duration of intermediate reporting intervals	s	0.1 ≤ s ≤ 10	1 s
	Timeout value	s	5 ≤ s ≤ 30	5 s
	Type of Test packet including header and payload lengths, headers and options present and any markings for special treatment in the network	NA	IPv4 or IPv6 UDP DSCP	No default UDP 00 = Best Effort
	Reference size of UDP Payload	KB	Minimum 1 kbyte, Maximum at 1472 bytes (Max 9000 with Jumbo Frames)	No default, recommend largest value that avoids fragmentation.
	Period of status feedback message (Receiver of offered load returns messages to the sender with the results of the measured metrics)	s	0.005 ≤ s ≤ 0.250	0.050 s
Supporting Metrics	These are metrics measured on the same stream as IP Capacity			
IPLR	Y.1540, RFC 7680			
Tmax	Maximum Waiting time for packets to arrive	s	0.05 ≤ s ≤ 3	1 s
Sampled RTT	Y.1545, RFC 2681: RTT uses feedback status messages from receiver.			
Tmax	Maximum Waiting time for packets to arrive	s	0.05 ≤ s ≤ 3	3 s
	Resolution of Timestamps	ms	0.001 ≤ ms ≤ 1	Suggested for fixed access: 0.001 (based on current implementation)
Supporting Metric: IPDV	Y.1540, RFC 3393, RFC 5481(PDV)			

Table B.2 – Measurement variables, ranges, and default values

Category/ Variable Name	Parameter	Unit	Range	Default value
Tmax	Maximum Waiting time for packets to arrive	s	0.05 ≤ s ≤ 3	1 s

Table B.2 – Measurement variables, ranges, and default values

Category/ Variable Name	Parameter	Unit	Range	Default value
	Resolution of Timestamps	ms	0.001 ≤ ms ≤ 1	Suggested for fixed access: .001 (based on current implementation)

Parallel connections introduce complexity as well as the advantage of reaching higher rates.

Possible benefits include:

- Parallel systems may be used to produce the aggregate rate needed with parallel connections.
- Parallel connections may be used as a way to saturate the path under test with a single pair of test hosts.
- Additional information could be derived for diagnostic purposes, or to validate the testing process. For example, comparing the data rates on each connection could be informative, where very different data rates might reveal abnormal operation.

The current view is that each connection would have its own feedback channel, calculation of measurements and flowchart, and a report of the aggregate results over all connections.

Appendix I

IP packet routing considerations

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

This appendix describes IP packet routing considerations relevant to the characterization of IP service performance.

IP packet routing is determined by each network operator's policies and configurations for routing protocols, and choices of the protocols themselves. For example, operators configure a parameter for the "cost" of traversing each link in their network, and the routing algorithm computes the lowest-cost route to the destination based on its knowledge of the current state of network topology. Clearly, the path a packet takes from source to destination greatly influences the transfer delay it will experience (from both transport and queuing), as well as exposure to other impairments such as loss, errors, duplication and reordering.

Another way in which routing protocols influence packet transfer performance is in their automated response to changes in network topology, such as link or router failures, or maintenance action to take a network element out of service. When the network topology changes due to failure, a recovery process restores the affected connectivity over the remaining network topology, if possible. This process is called "rerouting" or "re-convergence", and typically contains the following steps (each requiring time to execute):

1) failure/event detection

2) path computation

3) advertisement

4) forwarding table update.

Again, options for timers configured by the operator determine the duration of the rerouting process to a great extent. Operators also have the option to set waiting times between executions of the routing algorithm, which conserves processing resources but may lengthen the response to a failure in some cases.

Sub-IP networking technologies, such as SONET rings and MPLS-TE fast reroute, enable sub-second restoration from link or router failures.

Appendix II

Secondary terminology for IP packet delay variation

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

II.1 Introduction

This Recommendation specifies a single primary/normative definition that assesses the variation in a set of delays with respect to a reference delay. This appendix provides two informative/secondary definitions in the clauses that follow (based on IETF's inter-packet delay variation, and a modification of 1-point cell delay variation). This appendix also gives guidance on when each parameter is most appropriate, and relates the results of observations with the different parameters. Additional comparisons between different forms of delay variation are detailed in [IETF RFC 5481].

There are two additional approaches to quantifying delay variation:

1) a parameter based on [b-IETF RFC 3393] that ascertains the inter-packet delay variation;

2) a parameter similar to the 1-point cell delay variation described in [b-ITU-T I.356], which assesses the packet arrival spacing at a single interface with respect to an ideal arrival interval.

Note that [b-ITU-T I.356] included two different variation definitions, both 2-point and 1-point.

The [ITU-T Y.1541] IP performance objectives for PDV are in terms of the normative 2-point packet delay variation parameter in this Recommendation.

II.2 Definition of inter-packet delay variation

[b-IETF RFC 3393] defines delay variation as follows:

– A definition of the IPDV can be given for packets inside a stream of packets.

– The IPDV of a pair of packets within a stream of packets is defined for a selected pair of packets in the stream going from measurement point MP1 to measurement point MP2.

– The IPDV is the difference between the 1-way delay of the selected packets.

A selection function unambiguously determines the pair of packets used in each calculation of the delay variation metric. Only packets that arrive successfully are used in IPDV calculations.

The first selection function defined is for adjacent packets in the stream. The 1-way delay of the current packet has the 1-way delay of the previous packet subtracted from it to determine the current packet's IPDV. If either of the packets in the pair (or both) is lost, then the IPDV is undefined.

Another important example is the selection function that produces an equivalent delay variation assessment to the 2-point PDV parameter defined in clause 6.2.4. The pair of packets always includes the current packet and the packet with the minimum 1-way delay in the stream. The 2-point PDV for all arriving packets is calculated by subtracting the minimum delay from their 1-way delay values (the reference delay is the minimum delay).

II.3 Definition of 1-point packet delay variation

The fundamental notion of a 1-point delay variation parameter is the comparison between the actual arrival pattern and the intended (usually periodic) arrival pattern. Some variations of this definition include a "skipping clock" adjustment (when cells or packets arrive late/behind their ideal arrival time), as in [b-ITU-T I.356]. The definition below does not implement the skipping clock feature, since there is no clear bias if the reference pattern is established arbitrarily.

The 1-point PDV (y_k) for packet k at an MP is the difference between the packet's reference arrival time (c_k) and actual arrival time (a_k) at the MP: $y_k = c_k - a_k$. The reference arrival time pattern (c_k) is defined as follows:

$$c_0 = a_0 = 0,$$

$$c_{k+1} = c_k + T$$

where T is ideal packet spacing.

Positive values of 1-point PDV ("early" packet arrivals) correspond to packet clumping; negative values of 1-point PDV ("late" packet arrivals) correspond to gaps in the packet stream.

II.4 Guidance on applying the different parameters

Guidance that serves the practical side of measurement is as follows:

– When synchronized clocks are not possible (or temporarily unavailable) in measurement devices:

1) 1-point packet delay variation (1-point PDV) is a possible substitute for 1-way delay range/histogram, applicable for measurements on packet streams with periodic sending times (once the reference arrival time is appropriately set).
2) IP performance metrics (IPPM) inter-packet delay variation is applicable to all traffic flow types.
3) When clock error is stable, the ITU-T Y.1540 2-point PDV can be calculated and used.

– When synchronized clocks are available in measurement devices:

1) The ITU-T Y.1540 PDV 1-way delay range/histogram calculation is useful for a range of assessment tasks, including assessment of de-jitter buffer size.
2) IPPM inter-packet delay variation adds a parameter with sensitivity to sequential/short-term variation and some immunity to route changes.

The inter-packet metric, IPDV, defined by the IETF IPPM working group (WG), is similar to the calculation of inter-arrival jitter measurement in real-time control protocol (RTCP) reports. RTP gives the calculation of inter-arrival jitter in clause 6.4 of [b-IETF RFC 3550], with a sample implementation in an appendix. Although there are some differences in method (RTCP inter-arrival jitter uses order of arrival, as opposed to sending sequence with IPDV), there should be a favourable comparison between a "smoothed jitter" computed using IPDV singletons and the RTCP reports of jitter in many circumstances (if many packets were reordered, the results would probably not agree). It would be valuable to have a parameter that can be related to measurements made by user's end points. The IPDV metric with adjacent packet pairs is also less susceptible to route changes during a measurement interval, where the effect would only be observed in measurement pairs spanning the route change.

A positive attribute of 1-point PDV is its simplicity. The capability of assessing periodic streams within a single network element is highly advantageous.

A point that must be made clear in all variation parameter specifications is the effect of packet length. Since insertion time is included in transfer delay (first-bit to last-bit), packets with varying size have an inherent delay variation. Network specifications and tests should use packets with a single size to simplify interpretation of the results (and the size must be recorded).

Appendix III

Rate and throughput capacity related parameters

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

This appendix is deprecated in the 2019 edition.

Appendix IV

Tests of IP service availability state and sampling estimation of IP service availability parameters

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

This appendix, which is for further study, describes tests for determining whether an IP service, a basic section or an NSE is in the available state or the unavailable state. In a future version, it will provide methods for sampling estimation of the IP service availability parameters.

IV.1 Minimal test of IP service availability state (for test methodologies and test sets)

Clause 7.1 requires that at least M_{av} packets be used to evaluate the availability state. Test methodologies and test sets should attempt at least M_{av} packets spread throughout a T_{av} interval of time. For end-user generated traffic, successive T_{av} intervals of time might be concatenated until the requirement of at least M_{av} ingress events is fulfilled. This is for further study.

The following describes the minimum amount of effort that is necessary to decide the availability state during a single T_{av} interval of time. Repeated applications of this test are necessary in order to determine the PIA and the PIU. This minimum test of IP service availability is applicable to test methodologies and test sets; some requirements for end-user generated traffic are presented in clause 7.1. Any other test of IP service availability that (statistically) performs at least as well as this test is an acceptable test of IP availability. This test of IP availability is applicable end-to-end or in the specific-ingress case for a basic section or an NSE.

– Step 1: Determine the SRC and the DST.

– Step 2: Position test sets or activate test scripts at the appropriate measurement points.

– Step 3: At a predetermined time, start sending M_{av} IP packets distributed over the time duration T_{av}.

– Step 4: If the number of lost packet outcomes is greater than $c_1 \times M_{av}$ then the IP service is unavailable over the T_{av} interval of time.

– Step 5: If the IP service (basic section or NSE) is not declared unavailable as per the results of step 4, then it is available over this T_{av} interval of time.

The minimal test provides an unknown level of confidence depending on the size of the sample, M_{av}, so the following test is preferred.

IV.2 Test of IP service availability state (using sequential probability ratio test)

This clause describes a non-parametric test, which makes no assumption of the underlying distribution on losses, relies on the sequential probability ratio test (SPRT) to determine whether the c_1 loss threshold has been exceeded with a predetermined level of error. SPRT also allows the tester to stop testing when a much lower loss ratio has been observed over a specified number of packets and time. The outcome may also be indeterminate, in which case further testing is warranted. SPRT was first applied in [b-Morton] to evaluate packet loss ratios and associated with target rates in Internet testing.

For the null hypothesis, H0, we set the probability of loss (or defects) equal to $c_1 = p_0 = 0.20$. We also set the loss ratio for the alternate hypothesis, H1, at $p_1 = 0.05$. Finally, the Type I and II errors are alpha = beta = 0.001.

SPRT equations [b-Montgomery], [b-Wald] follow:

$$X_A = -h_1 + sn \text{ (acceptance line)} \quad (1)$$

$$X_R = h_2 + sn \quad \text{(rejection line)} \tag{2}$$

where n increases linearly over all packets sent, and

$$h_1 = \left(\log \frac{1-\text{alpha}}{\text{beta}}\right) k^{-1} \tag{3}$$

$$h_2 = \left(\log \frac{1-\text{beta}}{\text{alpha}}\right) k^{-1} \tag{4}$$

$$k = \log \frac{p_1(1-p_0)}{p_0(1-p_1)} \tag{5}$$

$$s = \left(\log \frac{(1-p_0)}{(1-p_1)}\right) k^{-1} \tag{6}$$

for p_0 and p_1 as defined in the null and alternative hypotheses, above.

Using the equations above, calculate the minimum number of packets needed to accept H0 when x defects are observed, for example x=0 (no losses).

$$X_A = 0 = -h_1 + sn \tag{7}$$

$$n = \frac{h_1}{s} \tag{8}$$

With c_1 = p_0 = 0.20 used as the H0 level, p_0 =0.05 for alternative H1 and errors at 0.001, it is found that at least 41 packets are needed to prefer H1 (with zero loss), and observing 9 losses in these 41 packets would result in a preference for H0.

Figure IV.1 shows the results from the R tool [b-Rdev] operating with the [b-CVST] package installed using the values above.

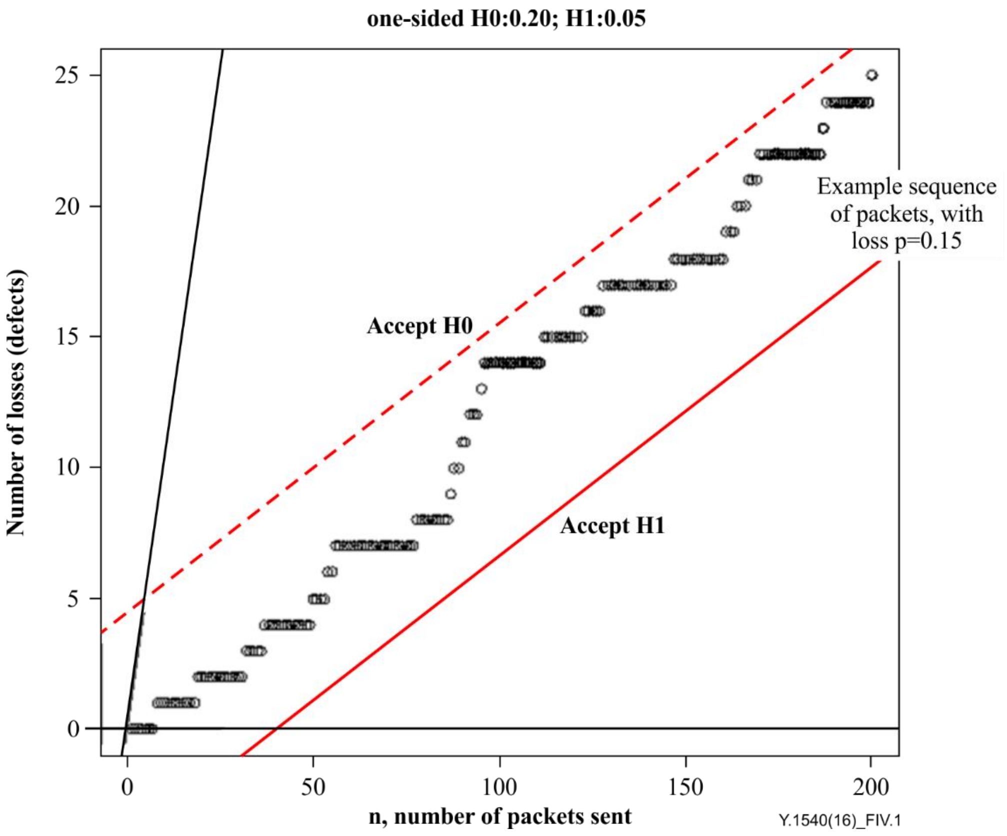


Figure IV.1 – Example of sequential probability ratio testing

Figure IV.1 illustrates that at least 41 packets are needed to prefer H1 (with zero loss), and observing 9 losses in these 41 packets would result in a preference for H0.

IV.3 Alternate test of statistical significance to determine IP service availability

[ITU-T Y.1540] determines the IP service to be available during a measurement interval if the IPLR for that interval is smaller than a threshold c_1. Since a packet is either successfully transmitted or lost, the packet loss can be modelled by a binomial distribution.

The null hypothesis H_0 is that the IP service is available during the measurement interval. H_0 is assumed to be true, if the average packet loss during the measurement interval is below or equal c_1 (the z-test deems the IP service to be available if the packet loss rate equals c_1). The one hypothesis H_1 is, that the IP service is unavailable during the measurement interval (packet loss > c_1 during the measurement interval). A z-test is proposed to decide whether H_0 or H_1 is supported through measurement. Following [b-C-298], take the confidence level to be 95% (meaning the significance level α=0.05).

The test consists of one sample compared against a threshold, c_1. The threshold mean $\mu_0 = c_1$ and its variance applicable for the test is $\sigma = c_1 * (1 – c_1)$.

The number of packets $n = \text{packets}_{\text{transmitted}} + \text{packets}_{\text{dropped}}$. The average packet loss ratio then is $x_{\text{mean}} = \text{packets}_{\text{dropped}} / n$.

The test statistic for a threshold text is $z_{\text{available}} = \text{sqrt}(n) * (x_{\text{mean}} – \mu_0) / \sigma$.

With the z-value for a confidence level of 95% and α=0.05 for a single sided test, H_1 (IP service is unavailable during the measurements) is accepted if $z_{\text{available}} > 1.645$.

With the z-value for a confidence level of 99.9% and α=0.001 for a single sided test, H_1 (IP service is unavailable during the measurements) is accepted if $z_{\text{available}} > 3.09$.

IV.4 Sampling estimation of IP service availability

Random samples of the availability state using the minimum test above may be sufficient for estimating PIA and PIU. In order to estimate the duration of contiguous time in an available or an unavailable state, sampling must be much more frequent. [b-ITU-T X.137] provides procedures for ITU-T X.25/ITU-T X.75 networks that might also be suitable for IP services.

Appendix V

Material relevant to IP performance measurement methods

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

This appendix, which is for further study, will describe important issues to consider 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) 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 hour);
     – long period (e.g., one day, one week, one month).

Appendix VI

Background on IP service availability

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

VI.1 Introduction

This appendix gives the rationale for the current IP service availability function definition in clause 7. The purpose is to provide additional background information and aid the appreciation for this complex and important topic.

VI.2 Background

There are many ways to define availability, and many perspectives that translate into evaluation using a range of sensitivities and timescales. This Recommendation uses a simple, adequate definition (from a network operator's perspective) that specifies the minimum evaluation conditions. In order to understand why the IP service availability function is sufficient, an understanding of the causes of unavailability is needed.

Figure VI.1 shows a Venn diagram where the universe is all service time. The body of this Recommendation notes that IP service providers may identify maintenance intervals where service availability is not guaranteed. Thus, the service time universe is usually different from the universe of *all* time.

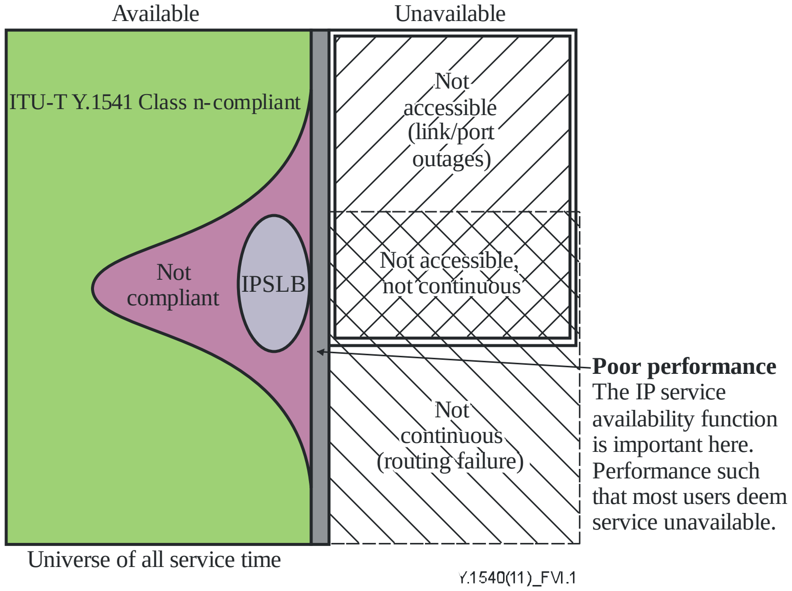

Available
Unavailable
ITU-T Y.1541 Class n-compliant
Not compliant
IPSLB
Not accessible (link/port outages)
Not accessible, not continuous
Not continuous (routing failure)
Poor performance
The IP service availability function is important here. Performance such that most users deem service unavailable.
Universe of all service time
Y.1540(11)_FVI.1

Figure VI.1 – Illustration of service time as a Venn diagram

We indicate that service time is divided in two main categories: available time (on the left) and unavailable time (on the right). Note that the relative sizes are not to scale, since available time is usually much larger than unavailable time.

VI.3 Definitions of the regions in Figure VI.1

Unavailable time is composed of the following regions:

– **Not accessible**: The service user is unable to communicate with the IP network because of failure in the access network transport or network elements. The access link itself or router interface failure are common causes. Packet loss ratio is typically 100%, and this failure will often take much longer than 1 minute to correct. Maintenance forces should be almost immediately alerted to the failure by fault management systems.

- **Not continuous**: The service user is unable to communicate with the desired destination because of a failure in IP network global routing information. The user may be able to communicate with some destinations, but not the desired destination. Packet loss ratio is typically 100% and this failure will often take much longer than 1 minute to correct.
- **Not accessible, not continuous**: The service user is unable to communicate while both of the above conditions exist simultaneously.
- **Poor performance**: The service user is unable to communicate reliably with the desired destination. The packet loss ratio is 20% or greater, and the user will deem the service unavailable for communicating with almost any form of IP network application. When congestion is the primary cause for this level of packet loss, end-to-end flow control should be activated to alleviate it (as provided in TCP).

Available time is composed of the following regions:

- **[ITU-T Y.1541] class n-compliant**: The service user is able to communicate with the desired destination and the packet transfer performance is compliant with the objectives of the agreed class. Evaluation of this state is usually conducted in 1-minute intervals. Note that any user application will have specific capacity needs; the ability to support a traffic contract (as defined in [b-ITU-T Y.1221]) must also be considered.
- **Not compliant**: The service user is able to communicate with the desired destination, but the packet transfer performance does not meet one or more of the objectives of the agreed class. Evaluation of this state is usually conducted in 1-minute intervals.
- **IP packet severe loss block (IPSLB)**: The service user is able to communicate with the desired destination, but the packet transfer performance does not meet one or more of the objectives of the agreed class. Specifically, the loss ratio is sufficient to determine that an IPSLB has occurred (provisionally defined as more than 20% loss in a 10-second interval).

VI.4 Summary

It is observed that the criteria of the IP service availability function are only important in the poor performance region, and that the unavailable time contributed by this region is small compared to the other causes of unavailability. Therefore, the evaluation of state based on loss alone, and the criteria provisionally agreed for state evaluation (1 minute, 20% loss), are deemed sufficient.

Appendix VII

Packet performance parameters for estimation and optimization of stream repair techniques

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

VII.1 Introduction

IP-layer performance parameters have many uses, with network monitoring and trouble identification being one class of use. The parameters are also used as the basis of service level agreements (SLA). Both the aforementioned uses describe packet transfer as a characterization of the network which provided the UNI-UNI transport.

There is a second perspective: IP-layer performance parameters also characterize networks in terms which can be relevant to the application designer. Although many of the parameters used in network monitoring are useful to application designers, there are likely to be unique parameters for each use case. Figure VII.1 illustrates the two different perspectives, or use cases for IP performance parameters.

Recommendation ITU-T Y.1540 defines performance and availability parameters for IP-based networks. It defines primary and secondary packet transfer outcomes and a range of packet performance parameters based on these outcomes, including the IP service availability function.

This version of Recommendation ITU-T Y.1540 builds on the fundamental definitions and concepts to standardize a new set of normative stream repair performance parameters. The objective of the new parameters is to provide information relevant to the design and configuration of higher-layer (application-layer) techniques to compensate for packet loss due to various causes (including errors and delay variation). Thus, the design and/or optimization and performance estimation of application-stream repair techniques should be simplified if these new metrics for packet performance assessment meet their goal.

This appendix begins with a short background on application-layer stream repair techniques. It then goes on to offer a very simple model intended to be applicable to many different repair techniques.

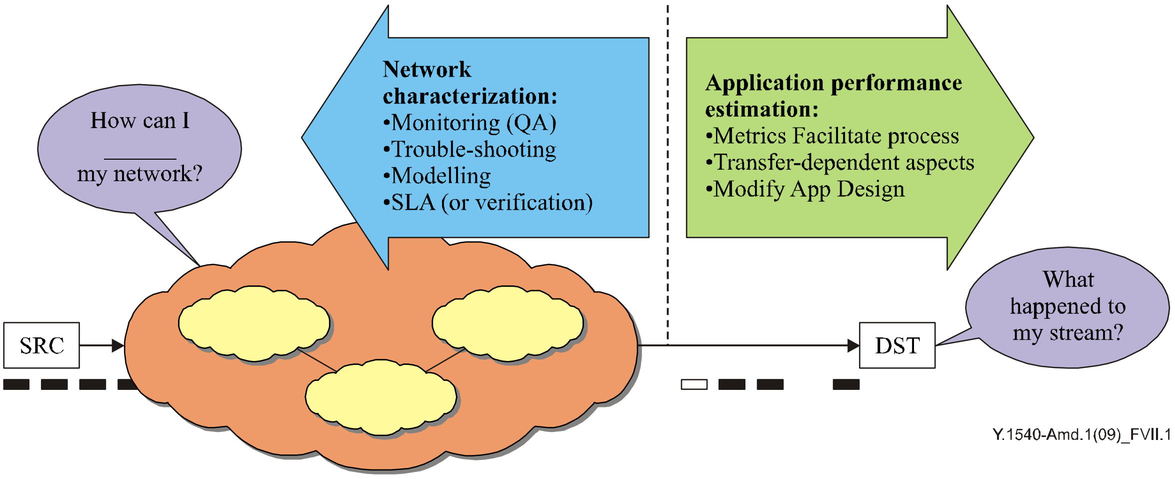

Network characterization:
•Monitoring (QA)
•Trouble-shooting
•Modelling
•SLA (or verification)
Application performance estimation:
•Metrics Facilitate process
•Transfer-dependent aspects
•Modify App Design
How can I ______ my network?
What happened to my stream?
SRC
DST
Y.1540-Amd.1(09)_FVII.1

Figure VII.1 – Two different use cases for IP performance parameters

The usual procedure is to introduce new metrics as informative appendices, so that potential users have the opportunity to evaluate them prior to their incorporation as normative parameters in the body of the Recommendation. These new metrics have followed the informative-first path to incorporation in Recommendation ITU-T Y.1540. In its studies, ITU-T has considered many contributions detailing experience with the stream repair performance parameters that serves as the foundation for their promotion to normative status.

VII.2 Short description of application-layer stream repair techniques

There are three main types of application-layer techniques to compensate for packet transport impairments. We focus on continuous real-time or near-real-time applications (audio, video) that are non-elastic – information delivery must take place according to a predetermined time schedule, and not the class of elastic data transfer applications usually served by TCP and its reliable octet stream transfer services.

Forward error correction (FEC): This is a technique where streams of packets are organized into blocks prior to transfer. There are calculations performed on each block, and overhead packets added to the stream which the receiver can use to reproduce some fraction of the packets in the block if they are lost, or successful but delayed, or corrupted in transport. Typical overhead represents 5% to 20% of the information block. In an *ideal* FEC scheme, the number of lost packets that can be corrected is *equal* to the number of overhead packets. The key aspects of this scheme are:

- the size of the information block, in packets and time;
- the amount of overhead packets relative to the information block, which approximately represents the corrective capability of the scheme.

Automatic repeat-request (ARQ): In this technique, there is a reverse communication channel available where the receiver, having detected that specific individual packets are lost, delayed or corrupted, can request retransmission (this is referred to as a selective ARQ). The lost packets are resent in time for them to take their place as the information is passed to higher layers for decoding and play-out. TCP has sometimes been modified to serve non-elastic streams in the role of ARQ. There is a waiting time for determining whether packets are simply delayed or lost, and this is similar to the information block used in FEC schemes. There may also be a limit on retransmitted packets which can accompany the primary stream in any time interval, and this is parallel to the overhead of FEC schemes. The ARQ technique can retransmit a number of lost packets in a block, equal to its limit on retransmission overhead. Note that the retransmitted packets will represent overhead on a subsequent block of information packets, but the concept still applies.

Thus, the ARQ and FEC techniques can both be described using the same basic variables of information block size and maximum repairable size.

Application-layer error concealment: This is a technique where decoders attempt to compensate for lost or corrupted information, using a variety of application-specific techniques, some of which have been standardized. The applicability of the simple model (derived below) to this class of techniques is for further study.

VII.3 Simple model of application-layer stream repair techniques

Each stream of application-layer packets is modelled as containing two categories of packets:

1) time intervals, T_I, or blocks, b, of information packets;

2) overhead packets, or the maximum repairable packets, x, associated with the information block.

The challenge to the repair technique designer is to choose the information block size in combination with the (maximum) amount of overhead packets that will be sufficient to compensate for a high percentage of packet network impairments (loss, excessive delay, and corruption), while working within the overall packet transfer capacity limits of the system and delivering sufficient quality in the application stream.

The new performance parameters (described in clause 6.10) should aid these decisions.

VII.4 Example of performance parameters to characterize stream repair variables

Figure VII.2 below gives an example of the stream repair parameter calculations, where b = 9 packets and x = 3 packets.

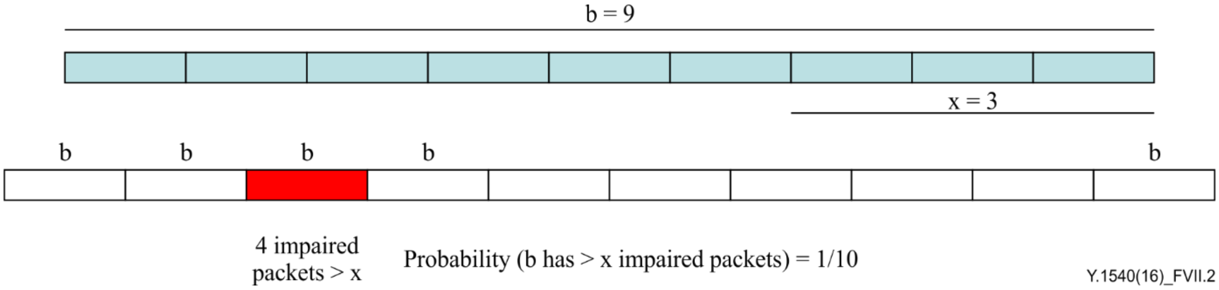


Figure VII.2 – Illustration of stream repair performance parameter

VII.5 Discussion of parameter measurement and usage

When attempting to estimate the performance of a repair system with unknown block alignment, the time intervals, T_I, or blocks, b, may be overlapping to allow assessment of different interval vs impairment alignments (sliding interval analysis). There is an issue with using a single fixed, non-overlapping interval for performance estimation and analysis, that the actual information block + overhead may experience worse performance owing to the difference in alignment.

There are two approaches to characterizing packet streams to determine the optimum combination of stream repair variables:

1) using (multiple) arbitrarily-established packet intervals (in terms of time or number of packets), as done above;

2) counting intervals of consecutive impaired packets and intervals of unimpaired packet transfers.

The approach of counting consecutive intervals appears to have flexibility not available with evaluation based on fixed intervals; it can determine the actual size of impaired/unimpaired intervals in a stream and does not suffer from the interval alignment issue. However, summary parameters describing impaired/unimpaired interval lengths are independent from the actual sequence in which they occurred. This sequence of changes between impaired intervals and unimpaired intervals may be important. Also, the consecutive interval counting approach requires some way to evaluate whether the x threshold has been crossed, as this is essential to the definition of an impaired outcome. If more than one value of x is to be evaluated, then multiple passes through stored data may be needed.

In either case, the results can be expressed as probability or cumulative distributions over the dependent and independent variables, as the example below shows (Figure VII.3).

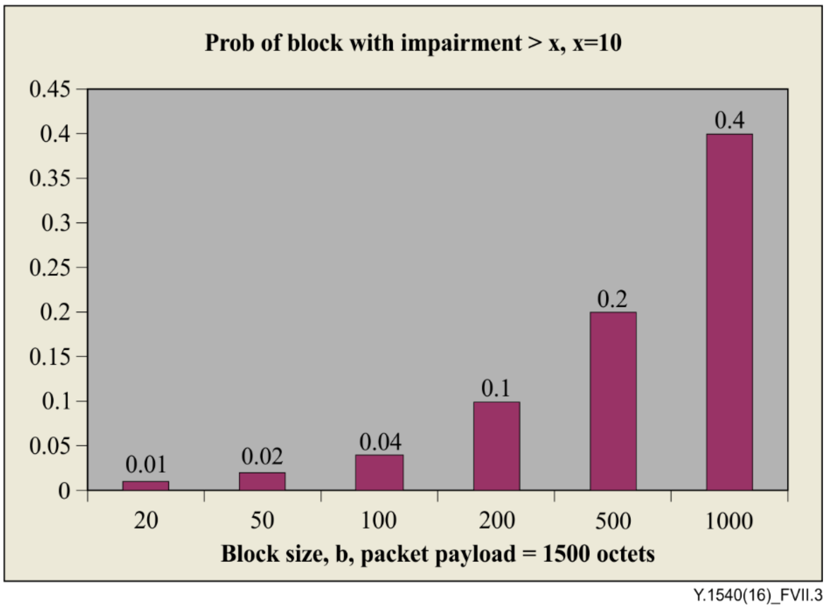


Figure VII.3 – Example plot of stream repair parameter results for a range of block sizes, where x is fixed, packet size is fixed

VII.6 Additional considerations

Although network characterization using the parameters defined above may be useful, the application repair system details should be known to begin to predict the quality delivered to users. FEC and ARQ techniques produce different packet loss patterns when operating beyond their ability to perform complete loss correction. The typical block sizes associated with each technique are different, with ARQ often characterized by larger block sizes.

FEC schemes organize the information block and overhead packets in different ways (sometimes called one-dimensional or two-dimensional forms) with less sophisticated schemes having more sensitivity between the exact pattern of losses and their ability to correct the losses. The performance margin between simple FEC schemes and the ideal performing scheme predicted by the parameters above should be known to the designer and taken into account.

Some applications may use chains of the various techniques described above. For example, a system might use FEC or ARQ in combination with application-layer error concealment. In another example, there could be FEC used in one part of the path, with ARQ or a different FEC used in another part of the path, and finally employing application-layer error concealment.

Finally, the short-term performance parameters defined above may be useful in troubleshooting by helping to identify the signatures of network problems, but this is for further study.

Appendix VIII

IP-layer capacity framework

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

VIII.1 Introduction

This appendix provides further information related to the capacity metrics defined in clause 6.11.

Knowing how much IP-layer capacity is available in real-time across an IP network (congested or not) is valuable information to the network operators and to the application users. This parameter can be used for network optimization, network monitoring, troubleshooting, server or gateway selection, load balancing, admission control, congestion control or to verify the service level agreement (SLA) of a guaranteed or business class service offering across a network provider.

The parameters and methods of measurement defined in normative Annex A supersede the list of academic projects and example tools previously listed in this appendix, and address several items listed for further study below.

VIII.2 Terminology and relation to IETF RFC 5136

The terms "available capacity" and "available bandwidth" are used interchangeably in the literature. [IETF RFC 5136] provides a discussion on terminology, mainly whether to use the word capacity or bandwidth for describing IP characteristics. [IETF RFC 5136] proposes to use the term capacity, and in order to harmonize with IETF, the term capacity is also used in Recommendation ITU-T Y.1540.

[IETF RFC 5136] defines capacity-related parameters similar to what is defined in clause 6.11. However, one major difference between the ITU-T and IETF definitions is that Recommendation ITU-T Y.1540 takes into account that network hosts may affect IP-layer capacity parameter values. This is not covered by [IETF RFC 5136], but it has been up for discussion in IETF. The ITU-T Y.1540 parameters are defined over basic sections which inherently take into account the capacity of both links and hosts in that section.

Table VIII.1 provides a mapping between the parameters that constitutes the definitions in clause 6.11 and the definitions in [IETF RFC 5136].

Table VIII.1 – Parameter mapping between ITU-T Y.1540 and IETF RFC 5136

ITU-T Y.1540 clause 6.11	IETF RFC 5136
IP-layer bits transferred	IP-layer Bits
IP-layer section capacity	IP-type-P Link Capacity
IP-layer used section capacity	IP-type-P Link Usage
IP-layer section utilization	IP-type-P Link Utilization
IP-layer available section capacity	IP-type-P Available Link Capacity
IP-layer NSE capacity	IP-type-P Path Capacity
IP-layer available NSE capacity	IP-type-P Available Path Capacity
IP-layer tight section capacity	Not defined

VIII.3 Items for further study

The definitions of capacity parameters in this Recommendation do not explicitly address multipoint paths; however, this is identified as an item for further study.

Discuss and identify methods of measurement that fulfil requirements from operators in terms of measurement accuracy, speed and overhead.

Is there a way of introducing a system for identification of the IP-layer tight link?

For future methods of measurement, policing functions cause packet loss, and this form of limitation may require a different method of assessment from methods that rely on packet dispersion.

Appendix IX

Explanation of TCP-based measurement inadequacy to meet normative requirements

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

IX.1 Introduction

Readers of this Recommendation may find it useful to understand the implications of the normative requirements in clause 6.12 when considering measurement methodologies, especially those based on available implementations of the TCP protocol. While TCP-based measurements are considered useful for informative surveys of user experience, they do not constitute the basis for standard metrics, methods of measurement or numerical objectives. Comparison of TCP protocol with the requirements of clause 6.12 in this appendix clarifies its status as a measurement method.

IX.2 Comparison with normative requirements

The requirements in clause 6.12 are organized in two numbered lists. The first requirement list is for all parameters, and the second list is for parameters that assess the ability to sustain a given IP packet transfer rate.

For the first list of requirements (all parameters):

1) Regarding the required accounting for packet delivery into the network and successful transfer: Some versions of TCP may make available the count of retransmitted segments during a connection (through a management interface), but retransmissions are based on the adaptive retransmission time out (RTO), not on whether the packets were actually lost, or are acknowledged after the time out expires, or whether an ACK was lost following successful delivery. TCP receivers do not distinguish whether the original or retransmitted (or both) packets arrive successfully. Furthermore, different TCP congestion control algorithms vary in their methods to achieve fairness to other flows and throughput, resulting in a larger number of lost packets when aggressive algorithms are used, or resulting in unnecessarily lower sending rates when packet losses are incorrectly interpreted as a signal of congestion (note the fixed mapping of packet loss interpreted as congestion in TCP flow control).

2) Regarding the required ability to measure partial paths: A TCP's congestion control is highly sensitive to round-trip time (RTT) in non-linear and sometimes unexpected ways. Thus, a TCP-based measurement on a partial path (EL or NS) will not typically predict the performance of a complete path, and TCP's dependence on RTT is one key reason.

For the list of requirements for assessment of sustained packet rate:

1) Regarding the required description of the traffic pattern offered to the network: TCP slow-start and congestion avoidance phases determine the sending pattern, and these patterns vary widely according to the conditions on the path, especially the presence of cross-traffic and characteristics of any bottlenecks encountered. Thus, the pattern is difficult or impossible to constrain or predict with TCP's flow control operating.

2) Regarding the requirement to limit traffic rate to less than the capacity of connecting links: TCP's flow control continues to test for available capacity, assuming that conditions may change. It is not practical to limit a TCP sender to an exact capacity using the parameters available, partly due to the variation of RTT during the life of a TCP connection. In other words, a TCP can always send traffic at a rate that exceeds connecting links.

All difficulties caused by TCP flow control are further exacerbated by operating multiple simultaneous TCP connections, each independently evaluating their connection on the same path.

In conclusion, the transport protocol is determined and implemented in user hosts, and outside the purview of IP-based packet transfer service providers. Standard assessments of the service provider performance should avoid the contribution of layers that are chosen by others and must meet the normative requirements of clause 6.12.

Appendix X

Summary of Laboratory (Phase 1) and Field (Phase 2) results: Annex A evaluation plan

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

X.1 Introduction

In their April 2018 interim meeting, and the ad-hoc meetings during the SG12 Plenary that followed in May 2018, Question 17/12 agreed on a plan to develop metrics (a.k.a. parameters) and methods of measurement for IP capacity assessment (along with packet latency and loss, other key performance metrics). The work began with the development of a plan for laboratory evaluation of some existing metrics and methods. A new Annex A documents the plan to qualify and compare access measurement metrics, methods, models and tools in a stable and repeatable laboratory environment. The work is guided by the current parameters in clause 6.11, and the requirements of clause 6.12.

As work proceeded, it became clear that a parallel effort to harmonize standardized IP performance across multiple SDOs would be needed to achieve both wide and prompt industry adoption. These efforts included the proposal of new work in TC ETSI STQ, and a call for collaboration in the IETF IP Performance Measurements working group (with 2 volunteers). Other SDOs (ITU-T SG 11, ETSI TC INT and BBF) have received multiple liaisons describing current progress.

Meetings in the autumn of 2018 produced the decision to divide the evaluation plan into two phases, and collected the first test results of Phase 1. The Phase 1 test plan was informed by the BEREC evaluation of Internet access measurement systems, available in [b-BEREC], where Requirement 127 described mandatory testing of speed accuracy using "traffic shaping software or hardware" at multiple speeds up to 500 Mbps. There were no other details, and the critical factor of latency was omitted. The new Annex A addresses this and other omissions of the BEREC plan. The Phase 2 evaluation will investigate the Phase 1 conclusions on access networks.

The contributions to Q17/12 have also included two surveys of academic research on Internet access performance measurement. The most recent survey is available, along with points from the November 2018 SG12 workshop, and several key points are summarized at the end of this contribution, the most relevant being that UDP-based tests are considered the benchmark for capacity.

This contribution summarizes tests and results from SG12-C275, TD627 and TD701 R2, through January 2019. A companion contribution follows up on discussion and questions at the Q17 interim meeting (Virtual, Jan 2019), and provides new laboratory results for multiple metrics, tools and methods of measurement.

X.2 Phase 1 laboratory test set-up

This clause describes two main alternatives to create a controlled and isolated test environment as a foundation for repeatable comparisons between test methods.

The figure below illustrates two different utilities providing flexible traffic/rate control on a general purpose computing platform. The different traffic control utilities can be applied in three ways.

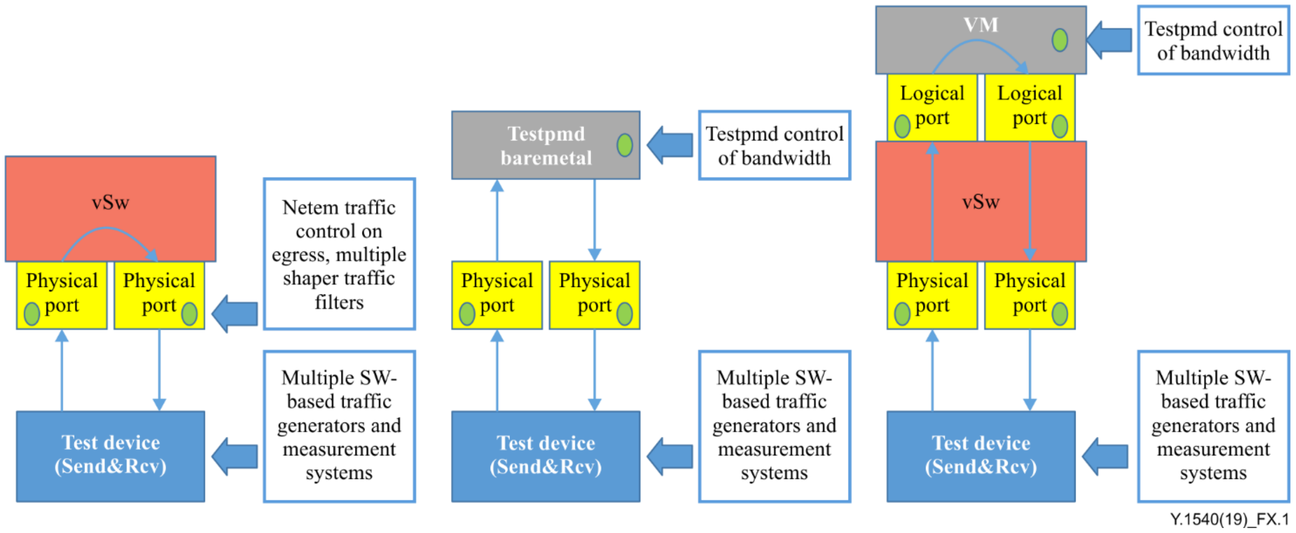

VM
Testpmd control of bandwidth
Logical port
Logical port
Testpmd baremetal
Testpmd control of bandwidth
vSw
Netem traffic control on egress, multiple shaper traffic filters
vSw
Physical port
Physical port
Physical port
Physical port
Physical port
Physical port
Test device (Send&Rcv)
Multiple SW-based traffic generators and measurement systems
Test device (Send&Rcv)
Multiple SW-based traffic generators and measurement systems
Test device (Send&Rcv)
Multiple SW-based traffic generators and measurement systems
Y.1540(19)_FX.1

Figure X.1 – Three possible test paths on a general purpose computing platform

In Figure X.1, the test device is connected to a general purpose host via physical links at 10 Gbps. The test device is also a general purpose host, but it is completely isolated from the host performing traffic control, allowing each node to dedicate resources to their unique roles in the test environment. Different implementations of candidate measurement methods can be installed and tested through the traffic control node.

There are three alternatives to implement the traffic shaper function. The first on the left employs the Linux kernel-equipped emulator, netem, which can emulate delay and assist with the traffic control after configuration for the correct NIC and physical interfaces. The vSwitch simply switches frames between its two ports. This configuration is commonly referred to as "phy2phy".

The phy2phy configuration was used for the tests and results described below. See Annex A for a description of the remaining configurations (using the Intel DPDK testpmd).

X.3 Detailed test set-up

The figure below illustrates the details of the "phy2phy" test set-up with two host nodes, measurement systems, network interfaces with 10 Gbps links, and configurations including a network namespace and an open virtual switch (OVS). The hosts are located in an OPNFV laboratory hosted by Intel [b-Pod12].

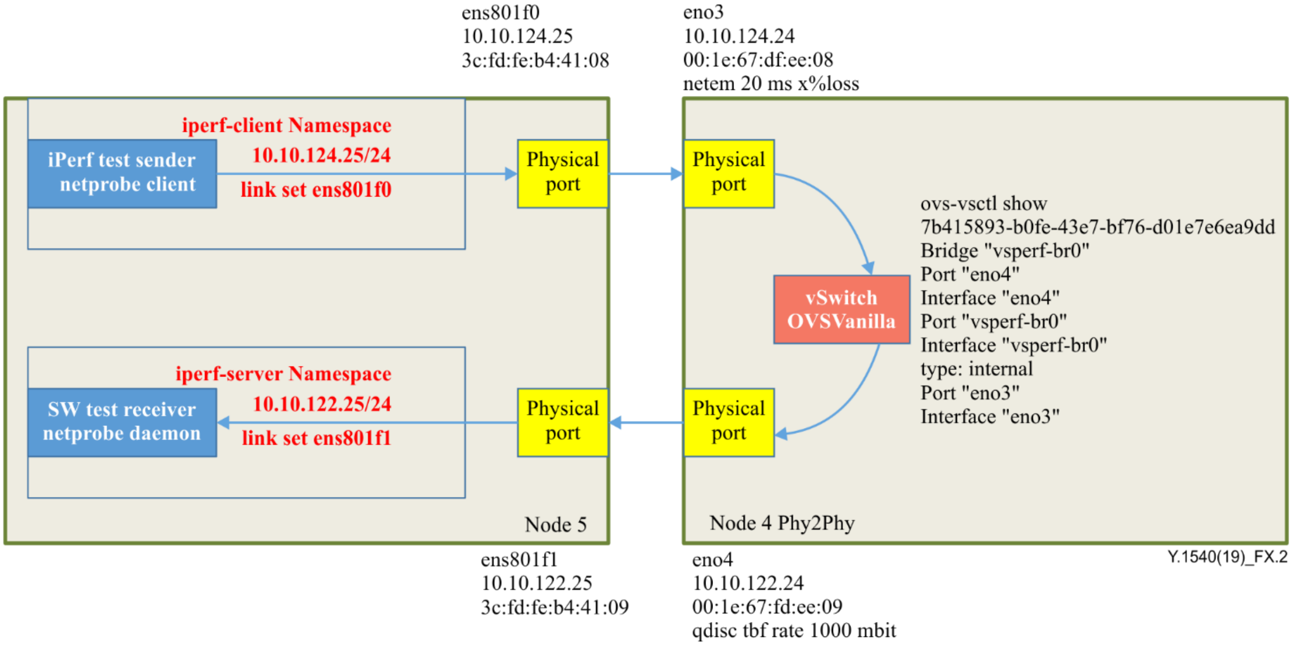


Figure X.2 – Test set-up with iPerf, calibrated network (DUT) and competing (Meas.) traffic

The network namespace is essential to force traffic to exit on the associated network interfaces and avoid internal routing through the kernel. The Netprobe tool can provide both competing UDP traffic and a sub-stream of the overall traffic with per-packet loss and latency measurements (both 1-way and round-trip in millisecond resolution).

X.4 Test tools

Multiple open-source measurement tools have been used in evaluations that lead to the present summary and decisions along the way. These are Cisco's T-rex, iPerf 2, iPerf 3, NetProbe, and a new tool yet to be named. Traffic shaper and policer functions are part of the typical Linux distribution, as is the netem network impairment emulator.

Early UDP tests with T-rex supported the Phase 1 evaluation, and deployed the binary search with loss verification (BSwLV) advanced search algorithm specified in [b-TST 009]. However, methodological choices (loss determined using un-synchronized Tx and Rx counters) and difficult activation of TCP testing encouraged the investigation of other tools.

Although iPerf 2 has been succeeded by the development of iPerf 3, current testing has determined that iPerf 2 is more predictable to configure, once a few tricks have been learned. iPerf 2 is used throughout this summary (except where indicated). MTU-sized packets are used throughout. iPerf 3 on Linux has some "features" which appear to need sorting out (some configurations encounter sending rate limits). There are currently parallel developments of iPerf3, and a sub-version number is necessary to distinguish different sources.

NetProbe has been used exclusively as a complementary measurement system (providing measurements of latency that iPerf 2/3 does not), and as a competing traffic generator (with measurement capability).

X.5 Calibration of reported results with iPerf 2

iPerf 2 reports rate measurements based on the transport payload bytes delivered (above the UDP or TCP layer). The token bucket filter (TBF) rate is specified in "bits in Layer 2 frames without ETH CRC", so the headers added to transport payloads are included in TBF calculations (the rate includes ETH, IP and transport header bits).

Received rate calculations "with OH" include the per packet overhead of octets in the following headers: ETH (14), IP(20), and UDP(8) or TCP(20). A typical UDP payload rate of 972 Mbps corrected for overhead (1.0286) is 999.799 Mbps.

A TCP frame size is variable, because iPerf 2 submits an 8kB block to the sender which results in five frames at MTU size and 892 bytes left for the remaining frame to complete the block. However, packet size varies from this model when observed in a packet trace. The simplest correction factor for overhead (OH) on TCP streams uses only the maximum segment size (MSS) in calculation:

54+1446 = 1500; 1500/1446 = 1.0373 * the measured TCP payload rate.

A high-end TCP payload rate with 3 connections (and with no complicating factors such as delay) of 956 Mbps is 991 Mbps when corrected for overhead. Most TCP payload rates measured with a 1 Gbps calibrated shaper rate are appreciably less than 956 Mbps (See figures X.4 and X.5).

There is also some inconsistency between the calculated TCP maximum receive window size (RWIN) reported by iPerf 2 and the measured value with 20 ms RTT. One test with 3 TCP connections yielded:

956 Mbps/3 connections = 318,666,666 bps per connection;

318,666,666 / 50 windows per sec = 6,373,333.32 bits in RWIN

or 796,666 bytes ~ 0.8 MBytes (not 0.08 MBytes as iPerf 2 reported).

X.6 Summary of test approach and results

The following figure summarizes part of the initial testing (through November 2018).

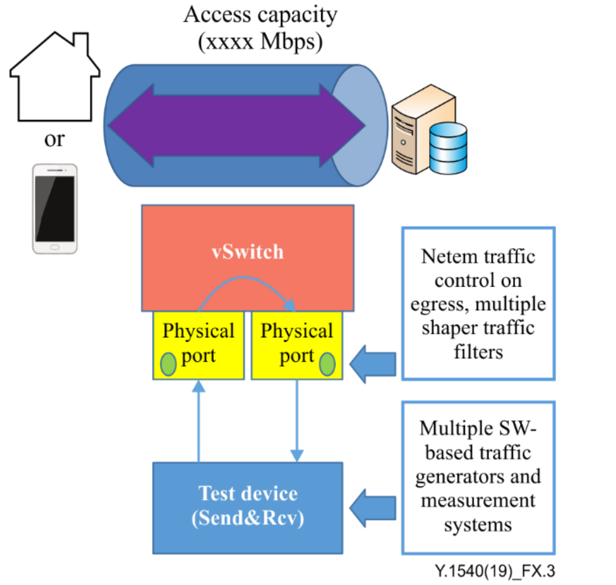
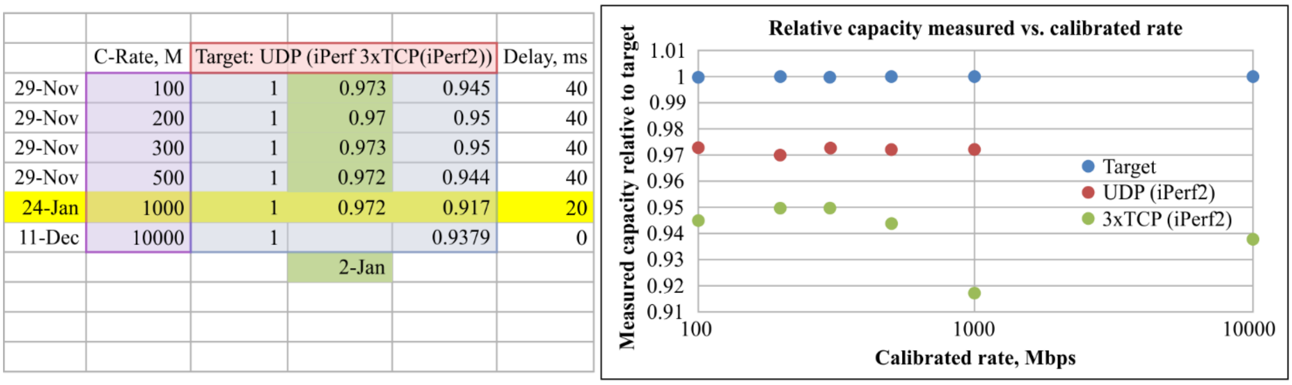
Evaluation phase 1: Benchmarking and IP access measurement crossover	
• Access moving to Gbps and low latency • Benchmark methods: UDP and new robust search alg. • Today's access test Methods: N x TCP conn. • Test using **Calibrated DUT** • Results: UDP found Cap Limits! Best perf using MTU • TCP methods have variable results and a tendency to underestimate Cap limits	 Access capacity (xxxx Mbps) or vSwitch Physical port Physical port Netem traffic control on egress, multiple shaper traffic filters Test device (Send&Rcv) Multiple SW-based traffic generators and measurement systems Y.1540(19)_FX.3

Figure X.3 – Initial testing summary

X.7 Summary of tests comparing measured capacity vs calibrated PHY rates

The test set-up and configuration (described in Figure X.2 above) uses a token bucket filter (TBF) in Node 4 eno4 with configurable target rate, and allows some burstiness, but also sets the maximum time any packet can stay in processing (typically 4 ms). 100 Mbps is one of the higher rates that [b-BEREC] intended to use for verifications in Req-127. Most of the results in this section were influenced by simulated delay (in netem).



	C-Rate, M	Target: UDP (iPerf 3xTCP(iPerf2))			Delay, ms
29-Nov	100	1	0.973	0.945	40
29-Nov	200	1	0.97	0.95	40
29-Nov	300	1	0.973	0.95	40
29-Nov	500	1	0.972	0.944	40
24-Jan	1000	1	0.972	0.917	20
11-Dec	10000	1		0.9379	0
			2-Jan		

Y.1540(19)_FX.4

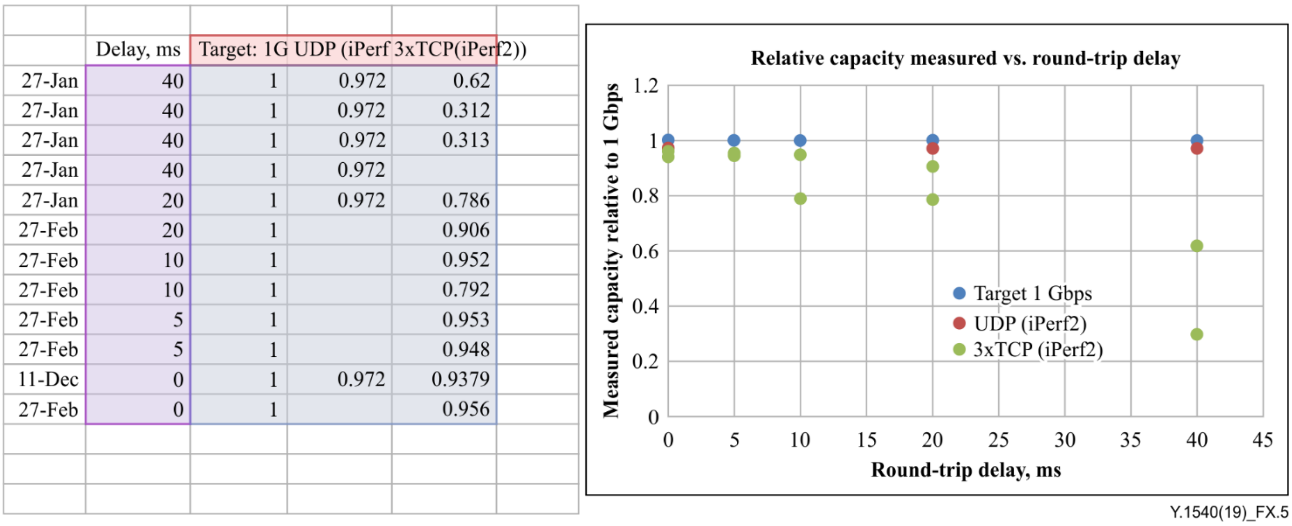
Figure X.4 – Summary of measured capacity vs calibrated rates

As Figure X.4 shows, iPerf 2 UDP-based methods operate approximately loss-free at the rate configured in the TBF, e.g., 972 Mbps in UDP payloads (which corrects to 1001 Mbps taking the headers into account). Measurements with 3 TCP connections are sensitive to the round-trip delay, particularly as the TBF rate increases to 1000 Mbps (and the header correction is not sufficient to compensate (although TCP start-up time is a factor, 956 Mbps is achieved after 2 seconds).

The measurements of 10 000 Mbps (10 Gbps) links alone (with no TBF or netem) represent a very practical source of rate limiting. Tests with 3 TCP connections and 5 TCP connections under these realistic conditions continued to underestimate the PHY rate after applying header correction factors. iPerf 2 and T-rex were unable to generate a single UDP stream above ~5 Gbps in this test set-up (this finding was investigated further: two iPerf 2 clients simultaneously generated 3.94 + 4.18 = 8.12 Gbps, and a third stream appears to be needed).

X.8 Summary of tests comparing measured capacity vs round-trip delay

Using a single target TBF rate of 1 Gbps, constant delay was emulated in the round-trip path (on Node 4 eno3) using netem.



	Delay, ms	Target: 1G UDP (iPerf 3xTCP(iPerf2))		
27-Jan	40	1	0.972	0.62
27-Jan	40	1	0.972	0.312
27-Jan	40	1	0.972	0.313
27-Jan	40	1	0.972	
27-Jan	20	1	0.972	0.786
27-Feb	20	1		0.906
27-Feb	10	1		0.952
27-Feb	10	1		0.792
27-Feb	5	1		0.953
27-Feb	5	1		0.948
11-Dec	0	1	0.972	0.9379
27-Feb	0	1		0.956

Y.1540(19)_FX.5

Figure X.5 – Summary of measured capacity vs round-trip delay @1 Gbps

Netprobe latency measurements (see the next section) confirm that when the TBF is limiting the rate (and some packets are dropped), the configured TBF latency adds as much as 4 ms to the netem latency (20 or 40 ms, above) for some packets. UDP capacity test results are immune to latency, but allow UDP latency measurements to be added conveniently. However, the TCP flow-control loop is sensitive to latency (specifically, the round-trip time), and its capacity measurements suffer with increasing delay (especially when the BW*Delay product exceeds the Max Receive window size configuration, see measurements at 40 ms in Figure X.5).

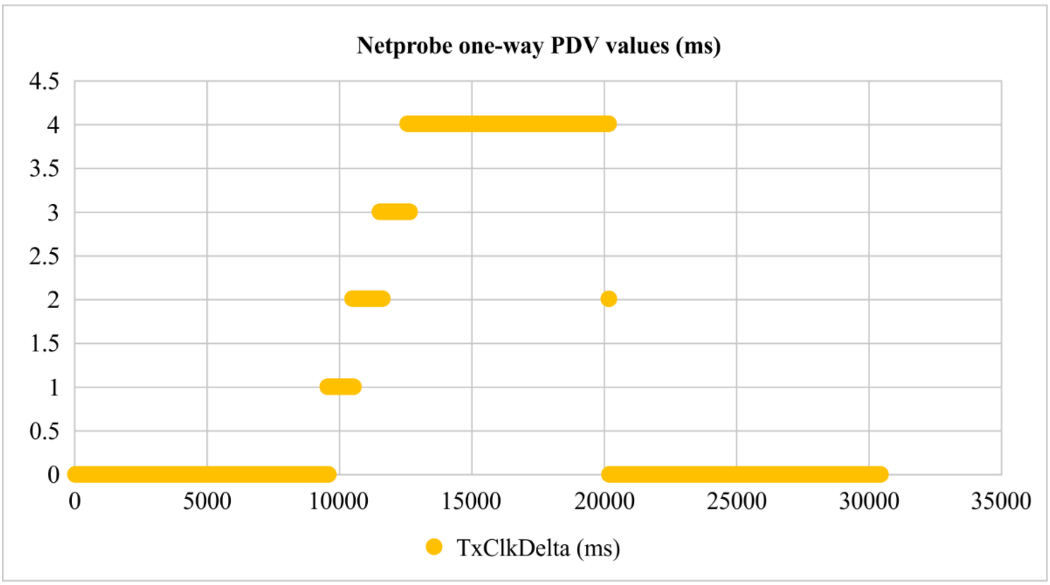
At the request of ETSI STQ members, tests of TCP at 5 ms and 10 ms were conducted, and all tests from 0 ms to 20 ms were repeated. The results show that the TCP-based measurement has variable results, and delay tends to increase the potential variability. [IETF RFC 8337] notes this issue for measurements with TCP flow control, in section 4.

X.9 Summary of tests with competing traffic

This clause summarizes three tests where UDP and TCP streams attempt to make IP capacity measurements in the presence of a competing traffic stream (a constant bit rate stream from NetProbe, which enables the measurement of packet delay and packet delay variation, or PDV).

Table X.1 – Competing 1 Mbps with 1000 Mbps token bucket Filter+Delay "phy2phy"

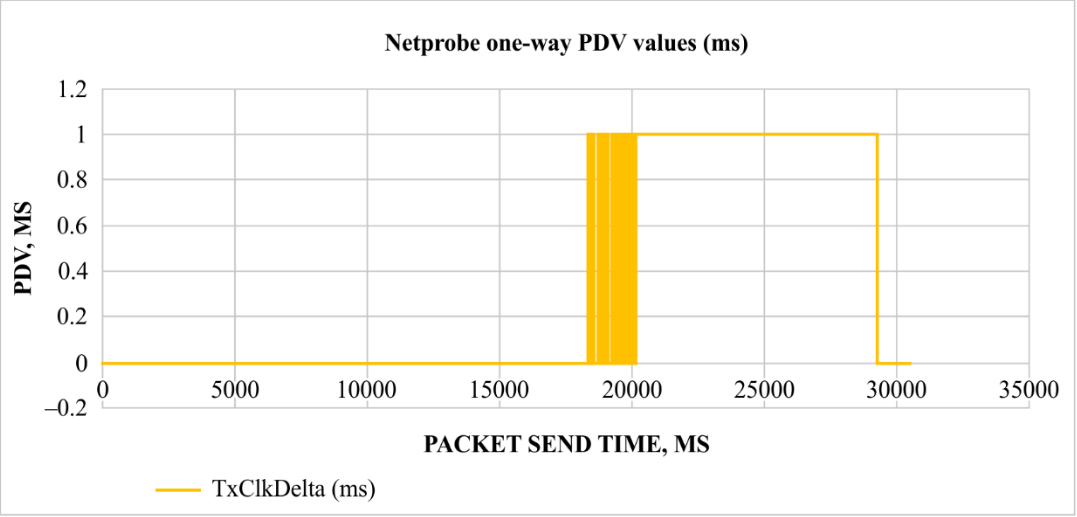
Notes	TBF rate	TBF burst, latency	Netem RT latency	Tool	Frame size, bytes	Rx rate, Mbps	Meas Lat	Loss count	Rate accuracy
27 Jan with netprobe sending 1,156,800 bps	1000 Mbps	5kb 4.0 ms	40 ms (eno3)	iPerf2; UDP uni-dir **–b 972000000**	1470 data	971 Mbps	Fig X.6	798 in most of test	0.971
27 Jan with netprobe sending 1,156,800 bps			40 ms (eno3)	iPerf2; UDP uni-dir **–b 971000000**	1470 data	971 Mbps	Fig X.7	245 (in 1st sec only)	0.971
27 Jan with netprobe sending 1,156,800 bps			20 ms (eno3)	iPerf2; TCP 3 connect uni-dir 12 sec	5@MTU + remain 892	**786 Mbps Ave** reached 955 M after 10 seconds	Fig X.8	X	0.786 ave 0.955 peak



Y.1540(19)_FX.6

Figure X.6 – NetProbe measurement of packet delay variation for 11 seconds UDP @972 Mbps

NetProbe measurements illustrate the delay encountered when the token bucket filter rate is exceeded (by about 1 Mbps) during the 11 second iPerf UDP test duration (during the 30 second NetProbe test). There were 3 packet losses on the NetProbe stream, sending 1 156 800 bps. The UDP rate assessment is reduced by the amount of the competing traffic in this test. The Figure X.6 latency (one-way PDV) measurements indicate that the TBF maximum latency was encountered a few seconds after the iPerf 2 stream began.

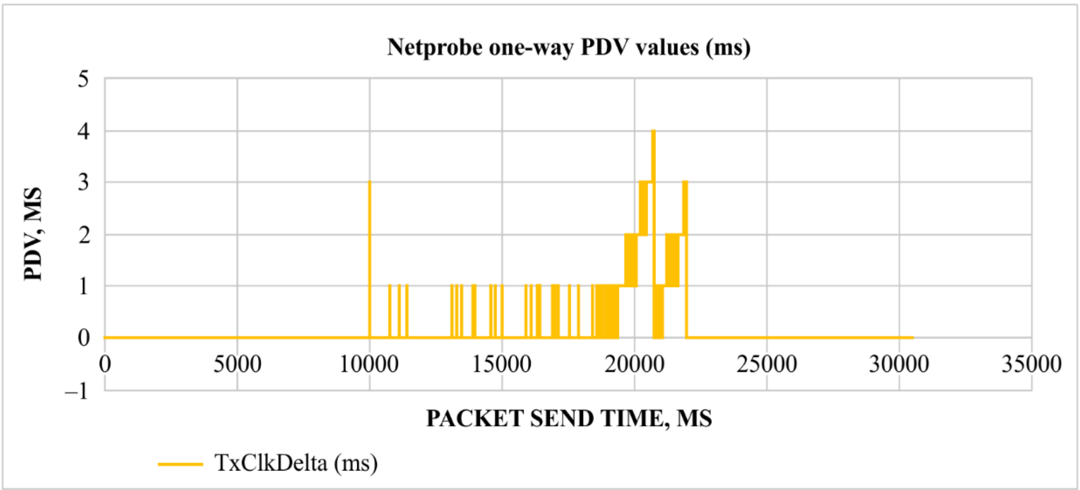


Y.1540(19)_FX.7

Figure X.7 – NetProbe measurement of packet delay variation for 11 seconds UDP @971 Mbps

NetProbe measurements illustrate the delay encountered when the token bucket filter rate is reached during the 11 second iPerf UDP test duration (during the 30 second NetProbe test). There were zero packet losses on the NetProbe stream, sending 1,156,800 bps.

The UDP sending rate was reduced to accommodate competing traffic, indicating that the previous (Figure X.6) test correctly assessed the remaining capacity when competing traffic was present. The combined streams more closely match the TBF rate. Figure X.7 latency (one-way PDV) measurements indicate that the TBF max latency was not observed (and that latency can be a useful input to a rate searching algorithm [b-TST 009], in addition to packet loss).



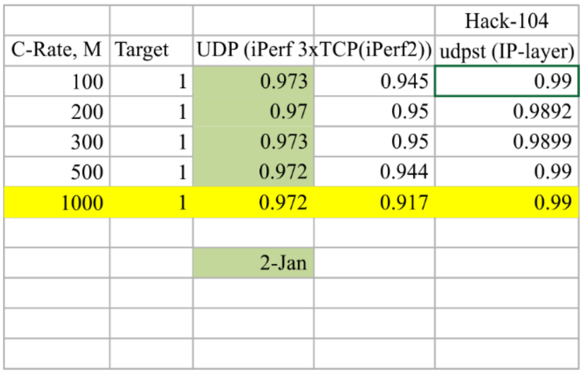
Y.1540(19)_FX.8

Figure X.8 – NetProbe measurement of PDV during 12 seconds of three TCP connections

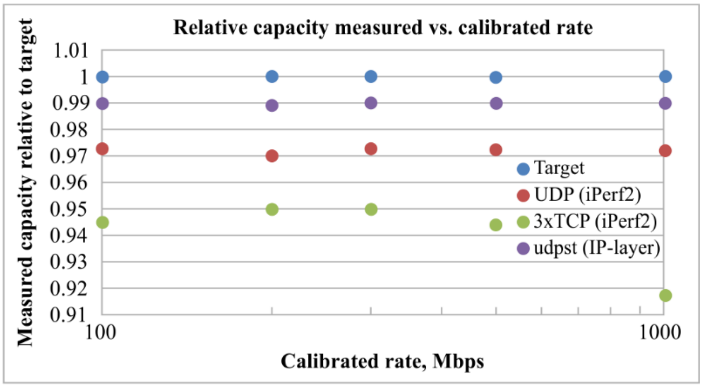
NetProbe measurements illustrate the delay encountered when the token bucket filter rate is reached during the 12 second iPerf TCP 3 connection test duration (during the 30 second NetProbe test). There were zero packet losses on the NetProbe stream, sending 1,156,800 bps. Near the end of the 12 second TCP test, the 3 connections appear to have synchronized and NetProbe PDV indicates a delay saw tooth at the TBF (possibly following the AIMD pattern), up to the Max TBF latency of 4 ms.

X.10 Tests with early implementation of a new UDP testing tool

One of the features that iPerf 2 and 3 UDP testing lacks is a search algorithm that can determine the IP capacity in an automated way. In previous testing, an exploratory iPerf 2 test would determine the maximum received packet rate for the test path (usually while significant packets were lost), and a second test at that maximum received packet rate would determine if it was possible to send at that rate with zero packet loss. The T-rex tool made use of the [b-TST 009] binary search with loss verification algorithm and was successful at finding the "ground truth" shaper rate, but it uses several trials. It was observed that a faster search algorithm that took account of loss and other impairments might help.



				Hack-104
C-Rate, M	Target	UDP (iPerf 3x	TCP(iPerf2))	udpst (IP-layer)
100	1	0.973	0.945	0.99
200	1	0.97	0.95	0.9892
300	1	0.973	0.95	0.9899
500	1	0.972	0.944	0.99
1000	1	0.972	0.917	0.99
		2-Jan		



Y.1540(19)_FX.9

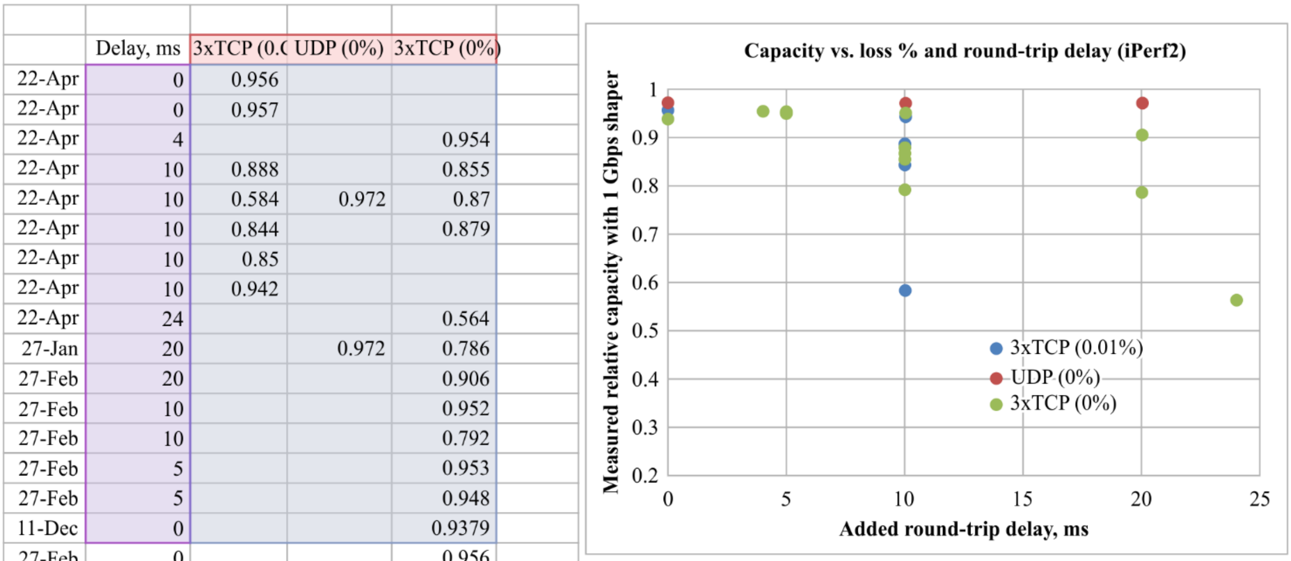
Figure X.9 – Tool "udpst ver1.4" comparison with iPerf measured capacity vs calibrated rates

The udpst is a prototype measurement tool authored by Len Ciavattone. udpst's search algorithm seeks the maximum IP capacity by adjusting its sending rate according to feedback status messages that include measurements loss and reordering, as well as information on delay variation. The udpst receiver sends status messages at regular intervals (default 50 ms). Either packet loss, reordering or excessive delay variation will trigger sending rate reductions, until the impairments are no longer present. The results are reported at the IP layer, including header bits. This means that the correction factor for an overhead only includes the ETH leading header (14 octets). An overhead correction factor of 1.0112 (1264/1250) applied to the typical 990 Mbps measurement yields 1.001 Gbps for the 1 Gbps calibrated shaper rate.

X.11 Tests of the effects of low-level packet loss

One of the Phase 1 laboratory evaluation variables was packet loss. The original loss ratios selected were 10^-4 and 10^-5, or 0.01% and 0.001% of total packets (as specified in `netem`). The loss distribution was not specified, so random was chosen.

Figure X.10 illustrates the test results, where new results with 0.01% loss or no loss were conducted at various delays (on 22 April), and plotted along with previous results for no loss conditions.



	Delay, ms	3xTCP (0.(	UDP (0%)	3xTCP (0%)
22-Apr	0	0.956		
22-Apr	0	0.957		
22-Apr	4			0.954
22-Apr	10	0.888		0.855
22-Apr	10	0.584	0.972	0.87
22-Apr	10	0.844		0.879
22-Apr	10	0.85		
22-Apr	10	0.942		
22-Apr	24			0.564
27-Jan	20		0.972	0.786
27-Feb	20			0.906
27-Feb	10			0.952
27-Feb	10			0.792
27-Feb	5			0.953
27-Feb	5			0.948
11-Dec	0			0.9379
27-Feb	0			0.956

Y.1540(19)_FX.10

Figure X.10 – Summary of measured capacity vs loss and round-trip delay @1 Gbps

First, note that 0.01% loss alone, with no added delay, had little effect on capacity measured with TCP at the transport layer. Of course, a higher loss ratio (1%) would cause a rather significant reduction in the measured rate (about a factor of 10 reduction). When 10 ms round-trip delay was added, the variation in TCP rates is substantial with or without loss. One test with loss and 10 ms delay yielded <600 Mbps, and this poor performance was similar to a recent measurement of 24 ms delay (possibly with a small amount of loss present). In one case with 10 ms delay, inserting the 0.01% loss appeared to increase the TCP rate, which is the sort of non-linear performance described in section 4 of [IETF RFC 8337].

When testing UDP with iPerf2, the emulated loss ratios are reflected in the results (with the expected variability among 1 second reports) and the final sum.

The conclusion is the TCP measurement variability is chiefly dominated by round-trip delay when the loss ratio is low (0.01%).

X.12 Examination of the maximum transmission limitations of the test tools and platform

Although calibrated rates/ground truth up to 1 Gbps were the main focus of the Phase 1 laboratory test plan, the upper limits of the test tool software in combination with the hardware hosting these tools is also useful to know. Tests to examine this limitation were conducted with iPerf2 for UDP and TCP transport at several times. The test set-up is simply the path to the OVS vSwitch over 10 Gbps links with no TBF traffic shaper or netem impairments.

The most conclusive testing of these limits was conducted during the IETF-104 Hackathon. Two tests with 3xTCP connections yielded 9385 Mbps and 9380 Mbps, and this compared favourably tests dating back to 11 December 2018. Tests with 3xUDP streams yielded 9308 Mbps, and there was 1.8% loss and about 1500 packets reordered in a 10 second test (each stream was attempting to generate 3330 Mbps, but did not achieve this rate).

The conclusion was that there is sufficient UDP and TCP packet generation capability and reception capability to conduct tests at 1 Gbps, and likely at somewhat higher rates considering the near-10 Gbps measurements above.

X.13 Examination of tests with early impairments on UDP streams

In several iPerf 2 UDP tests, the first second of the packet stream was reported to encounter packet reordering (typically less than 40 packets) and a small amount of packet loss, and these observations apply to cases with a 1 Gbps TBF shaper. To investigate the cause of the early impairments, the UDP stream was captured (using tshark) on the eno4 interface (just prior to the shaper operation).

The packet capture was further examined with the Wireshark tool, especially the timing of the early packets in the UDP stream. The early inter-packet arrival times were found to be irregular until 20 packets or more into the stream, where the nominal 12 microsecond spacing was finally achieved. There are no sequence numbers present, so the iPerf sender or the test path may have caused some reordering on the bursty portion of the stream, and the shaper may have discarded packets that exceeded its specifications.

X.14 Examination of TBF shaper parameters used in tests and comparison with policer filter

A series of tests was conducted to determine if the shaper parameters used in most testing had unfairly disadvantaged the TCP streams. Tests with 1 Gbps shaper rates concluded that when the maximum latency TBF parameter was reduced from 4 ms to 0.1 ms, the TCP measured capacity dropped significantly (TCP tests with 4 ms saw a maximum of 956 Mbps, but the performance with 0.1 ms latency were highly variable and in the range of 763 to 862 Mbps). On the other hand, increasing the maximum latency TBF parameter from 4 ms to 20 ms offered no benefit for TCP measured capacity. iPerf 2 UDP testing was insensitive to these changes.

When a policer filter was implemented instead of the shaper, again TCP measured capacity dropped significantly, despite attempts to improve performance with parameter tuning (TCP tests with shaper saw a maximum of 956 Mbps, but the highest performance with the policer was 2.02 Mbps). This is due to the very limited buffering available with the policer filter. iPerf 2 UDP testing was insensitive to use of the policer filter, owing to its constant bit rate stream.

X.15 Summary of Phase 1 laboratory testing

iPerf 2 TCP-based assessments tend to underestimate the calibrated capacity, and the estimates are:

- less accurate than UDP for high rates, such as 1 Gbps (the present state of the art Internet access service);
- more sensitive to round-trip delay than UDP, especially as it pertains to TCP measurement variability, due to TCP's flow-control;

- more sensitive to competing traffic, resulting in low average measurement due to a longer time needed to achieve equilibrium.

iPerf 2 UDP measurements confirm the status of "benchmark" for capacity assessment, aligned with calibrated rates (the experimental "ground-truth") and substantial research with lower-rate access technologies/speeds.

Latency can be a useful input to a UDP-based rate searching algorithm, in addition to packet loss, as speculated in [b-TST 009]. A prototype UDP test tool (udpst ver1.4) shows promise to fulfil this need to automatically search for a maximum rate with UDP.

X.16 Platform specifications

See: https://wiki.opnfv.org/display/pharos/Intel+POD12

X.17 Summary of Phase 2 field test results

The results summarized below (conducted in two test campaigns) used the following access types:

1. Fixed: DOCSIS 3.0 cable modem with "triple-play" capability and embedded WiFi and Wired GigE switch.
2. Mobile: LTE cellular phone with a Cat 12 modem (600 Mbps Downlink, 50 Mbps uplink).
3. Fixed: passive optical network (PON) "F", 1 Gbps service.
4. Fixed: PON "T", 1000 Mbps Service.
5. Fixed: VDSL, service, at various rates <100 Mbps.
6. Fixed: ADSL, 1.5 Mbps.
7. Mobile: LTE enabled router with ETH LAN to host (stationary).

The consensus on measurement results was that UDP is the preferred transport protocol for capacity assessment:

- UDP indicated more consistent results.
- UDP tools were able to measure loss, delay, delay variation and reordering.
- TCP registered lower rates than UDP tests and greater rate variability under various circumstances.
- TCP measurements on 1 Gbps PON exhibit a significant underestimation of capacity.
- Laboratory conclusions on UDP as the benchmark, and TCP as underestimating capacity were supported by the field measurements.
- Tests of LTE access indicated significant variability, as expected with any wireless network.

Appendix XI

A brief survey on Internet access related QoS and QoE research

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

XI.1 Introduction

Service providers, regulatory authorities (and consumer protection), as well as vendors and operators of measurement systems, are interested in Internet access speed (IAS) measurements for various reasons. Surveys characterizing service provider Internet access products are well established in many countries. In this area, standardization may consist of tracking existing deployments. Surveys reporting IAS estimates do not require a high accuracy and are often accepted by competing service providers, if the party performing the survey transparently provides a fair set of measurement campaign related parameters to the service providers to be surveyed.

There is interest in standardized exact IAS measurements too. Like with other exact speed or volume measurements on which obligations or penalties are based in commercial markets, standardized exact IAS measurements require a defined accuracy. Determination of a measurement accuracy requires determination of a measurement benchmark.

Research publications are expected to allow fellow researches reproduction of published results. This contribution examined a set of research publications, looking for results relevant for Internet access related QoS parameters, measurements and evaluation. The focus includes IAS measurements, but it is not limited to it. The research community mainly is interested in QoE measurement and Internet access related QoS parameters. Some of the approaches taken by QoE related research seem to indicate QoS parameters characterizing Internet access and components involved in an end-to-end communication more thoroughly, than access speed estimation. Also, the latter set of QoS parameters is investigated. The text extracts and references provided by this contribution helps to base discussion of evolving standards on arguments rather than opinions.

Note that whenever possible, studies were picked which are based on approaches validated by more or less large scale measurements using commercial accesses. References [2], [4], [6], [7], [8] and [10] fulfil this criterium.

XI.2 Key findings

This clause sums up Internet access and access usage related QoS, KPIs and measurement methods published by research.

A benchmark measurement of the maximum achievable access speed is obtained by flooding the access link with UDP traffic [1]. Also, a large scale research survey uses UDP to accurately measure access speed [2]. A major OTT provider includes UDP measurement facilities in an internal speedometer tool [3].

A conclusion related to parallel TCP based measurements is that they can be used to estimate Internet access bandwidth [1]. Some publications discuss particular factors impacting the accuracy of parallel TCP based measurements and examine some of these impairments in more detail [1], [2], [4]. Parallel TCP based measurements capture the effective available bandwidth in the presence of background traffic [1]. Note that the available bandwidth is identical with the Internet access speed only in the absence of background traffic. Parallel TCP based measurements fail to detect the presence of background traffic. Tools and methods to diagnose the presence of background traffic are available [1], [2].

UDP was found to provide a viable common transport protocol by [14]. This claim is based on tests in North and South America, Europe, Asia and Oceania (notably Africa misses). The authors found UDP traffic to be blocked on some ports or entirely in rare cases. UDP impairment was found to occur in enterprise networks and "networks in geographic regions with otherwise-challenged connectivity". A fallback to TCP transport is recommended in the cases where UDP transport fails. The Wi-Fi impact on application QoE may be characterized by passively collected commodity access point statistics as proposed by [10]. The authors report a generally reduced QoE for Wi-Fi TxPhymean below 15 Mbit/s.

Many applications apply encrypted transport. This leaves packet and IP flow information for network based QoE assessment. [11], [12] and [13] suggest network QoS parameters correlating to streaming QoE. The parameters suggested there may be useful to compare and characterize IP access too.

The spread of smartphones increases the interest of OTTs, service providers and research in the QoE resulting from mobile network QoS. The investigated research publications indicate or explicitly state that parallel TCP based measurements fail to provide useful input to this purpose.

The mobile network bandwidth fluctuation is correlated with the subscriber QoE under common wireless network conditions (like, e.g., subscriber mobility). Popular video streaming adapts video quality to deal with these conditions. In an environment with bandwidth fluctuations, subscriber QoE correlation to QoS KPIs requires capturing an additional parameter besides average available bandwidth, like average maximum throughput; see [6], [7], [8]. In general, user application latency or RTT measurements are mentioned as obligatory KPIs. The publications indicate, that deducing subscriber satisfaction with LTE Internet access from QoS measurements requires capturing KPIs related to the bandwidth-delay product (or bits in flight) or the bandwidth fluctuation, respectively [6], [7], [8].

Bandwidth remains cheap in relevant markets of a major OTT provider. User-perceived web performance is the primary metric for modern network services to this OTT. Web latency is the main impediment to improving user QoE in the relevant markets of the OTT [5]. TCP's design limits further improvement from the OTT's perspective. Some time after publication of [5], IETF started to work on QUIC, a new, UDP based transport protocol.

One publication concludes that within mobile LTE environments of many OTTs core markets, the implications their research for the end user seems straightforward: subscribers do not need a super high speed cellular contract, if their main use of the mobile access is streaming videos or web browsing ([7], note, that this statement reliably only holds for the location and point in time of the evaluation). Similar statements were made for fixed Internet access [5], also [9] seems to indicate this. These assessments are not globally valid. The statements indicate that subscriber satisfaction with an Internet access may no longer be related Internet access speed, if the latter provides a stable average throughput much higher than that required to produce a high subscriber QoE.

XI.3 Analysis of research publications related to QoS and QoE measurements

In the years 2010 to 2012, Internet access speed measurements were the focus of some research publications. Research community interest shifted to Internet consumer quality of experience afterwards. This does not mean, that Internet access bandwidth, its properties and QoS measurements were out of the researchers scope in the later years. Internet access QoS parameters and their measurement are still researched, if their measurement correlates with consumer QoE.

XI.3.1 Fixed Internet access speed measurement

Goga and Texeira [1] investigated tools by which to estimate Internet access speed. To eliminate undesired interferences, measurements were set up during off peak hours and with a fixed network. The aim of the work was to compare available bandwidth measurement tools. To do so, the accuracy and the measurement load of these tools are determined by measuring commercial ADSL and cable providers' Internet access with controlled cross traffic.

The authors benchmark the IAS by measuring the access UDP capacity obtained when flooding the link with UDP traffic. Afterwards, they estimate IAS using TCP based tools.

In the absence of cross traffic the following relative errors of estimation versus benchmark can be calculated from their results for off peak / no cross traffic measurements:

TCP based flooding ("Speedtest"): 4,06% best case, 5,28% average error, 6,8 % worst case

On average, access bandwidth is underestimated.

TCP based flooding fails to detect cross traffic. This is proven by a test with a single TCP flow being active while a TCP based speed test is performed. The latter estimates its available bandwidth, but not the IAS. Tools to detect the presence of background traffic are existent.

The authors detect that home gateway packet processing rates are limited, if the home gateways act as network address translators. Hence the selection of measurement packet size impacts IAS measurement accuracy.

Finally the authors inform about the load required to estimate IAS. TCP based speed tests are creating the highest load, whereas other tools estimate IAS by less than 10% of the load created by TCP based speed tests.

Also, Kanuparthy [2] uses UDP based measurements to measure IAS. He observes, that the presence of shapers at the bottleneck link decreases the accuracy of TCP speed test based IAS estimates. In the best case, such a test reports a rate between link speed and shaping rate.

Flach et al. [4] provide an algorithm determining the rate (and other properties) of policers used to limit Internet access bandwidth. IAS estimation is not directly discussed. On average, policed TCP flows experience a six times higher packet loss than those which are not policed. TCP based speed test IAS estimates without the suggested algorithm will likely be of poor accuracy.

The authors add, that policing was still deployed to some extent for commercial Internet access products in Asia and Africa by the time of publication.

A tool and measurement based test to investigate whether UDP is a viable basis as a general Internet transport protocol shows that UDP indeed is a viable common transport protocol [14]. This claim is based on tests in North and South America, Europe, Asia and Oceania (notably Africa misses). The authors found UDP traffic to be blocked on some ports or entirely in rare cases. A fallback to TCP transport is recommended in the cases where UDP transport fails.

The blocking of UDP occurred for between 2% and 4% of terrestrial access networks. UDP blocking was primarily linked to an access network. The authors of [14] found that UDP impairment is especially concentrated in enterprise networks and networks in geographical regions with otherwise challenged connectivity. Where UDP worked on these terrestrial access networks, no evidence of systematic impairment of traffic with UDP headers was found.

To trigger a fallback to TCP, a node does not need to measure or remember anything about its peers, but only about its connectivity to the Internet (also this recommendation given by the authors of [14] indicates that UDP impairments are based on particular access links rather than networks).

XI.3.2 Wireless LAN access speed measurement

Kanuparthy [2] earlier investigated IEEE 802.11 WLAN access performance. Da Hora, Van Doorselaer, Van Oost, and Teixeira published a model to estimate the impact of a Wi-Fi network on consumer QoE [10]. The authors of the latter model base their work on passively collected parameters of commodity Wi-Fi access points. A measurement campaign in provider networks results in the conclusion that the majority of the Wi-Fi networks perform well. Still more than 10% of the APs exposed an estimated MOS < 3 for at least 5% of all collected samples (or more).

The following Wi-Fi parameters were collected ([10], Table II):

Table X1.3.2-1 – Wi-Fi metrics measured on the access point (Table II of [10])

Metric	Description	Period
BUSY	% of time the medium is busy	2s
WiFi	% of time busy due to Wi-Fi traffic	2s
nonWiFi	% of time busy due to non Wi-Fi traffic	2s
TxPhy	PHY rate of last frame sent	1s
FDR	Frames sent / retransmitted to station	1s
RSSI	Received signal strength indicator	1s

The samples taken by the authors contain the mean, standard deviation (std), min, max, 25%-ile, and 75%-ile for each metric of Table X1.3.2-1.

Sample statistics are evaluated for a period T which depends on the application:

- T = 10s for audio and video experiments.
- T = 10s for web browsing.
- T = 120s for Youtube streaming (each video was played for two minutes during testing).

The researchers found the following application feature vector subsets to correlate consumer application QoE best with the subset of Wi-Fi features ([10], Table III):

- Video: TxPhy25%-ile , BUSY25%-ile, BUSYmax, RSSImean, RSSI75%-ile, WiFi25%-ile.
- Audio: TxPhymin, RSSIstd, WiFi25%-ile, WiFimax, nonWiFimax, FDRmean.
- YouTube: TxPhymean, BUSY75%-ile, RSSImean, RSSI25%-ile, WiFi25%-ile, nonWiFimin.
- Web: TxPhymax, BUSYstd, RSSImin, WiFimax, nonWiFimax, FDRmean.

The approach taken in [10] is based on machine learning. This contribution does not discuss, whether the results of [10] are applicable in general. It is assumed, that the mentioned QoS parameters are correlated to application QoS. The complete model and the per application parameter-fitting approach may be found in the referenced document. As a final notice and a rule of thumb quoted here, the authors of [10] expect a reduced QoE when TxPhymean is below 15 Mbit/s.

XI.3.3 Mobile network access QoE and speed measurement

Publications of Dimopoulos et al. [6] and Casas at al. ([7] and [8]) deal with measuring streamed video QoE for mobile terminals, [7] investigates additional applications. [7] concludes that the subscriber QoE "for certain applications is very sensitive to bandwidth fluctuations. Throughput fluctuations due to bandwidth variation are very common in cellular networks, but unfortunately, its QoE effect is not captured in today's network measurements, as only average throughput values are typically considered." Note that "today's network measurement" refers to TCP flooding based measurements.

Also, [6] shows, that the average TCP throughput is not correlated well with subscriber QoE, if throughput variability is present.

The publications correlate subscriber QoE with network QoS measurements by the following parameters:

- [6] concludes to use mean bandwidth delay product, max packet retransmissions, avg max bits in flight and cumulative sum of throughput min.

- [6] further suggests to "reduce the noise introduced by the start-up phase in the detection of resolution variations" by "removing the first 10 seconds of all video sessions from their dataset". The throughput within this start-up phase seems to deviate from the throughput and its fluctuations during the remaining time of the video streaming (average session duration was 180 seconds).
- [7] provides results of laboratory measurements with throughput fluctuations and [8] introduces two QoS metrics to capture them: maximum session downlink flow throughput and average session downlink flow throughput. These are classified as "most relevant parameters".
- [8] further evaluated more key performance indicators and their correlation to MOS prediction and their accuracy. Besides average and max throughput, the following KPIs were found to impact MOS prediction accuracy: average signal strength, session volume and session duration.
- The KPIs investigated in [8] are based on measurements within an area offering very good mobile network access. [7] shows laboratory results, indicating that additional KPIs like RTT and short-duration bandwidth outages (caused by handovers) can have a more or less significant impact on QoE.
- [7] further mentions network neutrality as a relevant topic. One of the applications investigated was bandwidth limited by one of the ISPs participating.

XI.3.4 QoE measurement of encrypted applications

Network based QoE judgment of encrypted applications requires a more thorough investigation of relevant QoS parameters as compared to non-encrypted application QoE sharing. The proposed parameters also characterize access properties. [11], [12] and [13] suggest parameters to this purpose. Their common aim is to recognise or optimise the following video streaming QoE parameters by observing network QoS parameters:

- initial start-up delay (start-up time);
- multi-chunk (segment) average bit rate;
- stream average bit rate switches (caused by varying video quality levels);
- re-buffering ratio.

The common approach taken is to measure the throughput of a video (and audio [12]) stream. The recommended duration of a sample is typical chunk or segment duration of the video (or audio, respectively) stream. These lie between 2 and 15 seconds, with a value of 4 seconds recorded for YouTube video streams [13].

The stream of video data rate in Mbit/s directed to the replaying terminal is recommended to be captured by 1 second samples, which are then evaluated by statistics related to the chunk duration [11], [13]. The average bit rate is calculated by 5 second simple moving average samples by [11]. [13] expands this to capture data of 10 consecutive second "windows" (resulting in 5 average values per window, each calculated from five 1 second samples).

[13] proposes the following QoS parameters and statistics to correlate transport layer properties (which may not be detectable by the network) with network layer parameters during a learning phase:

Table X1.3.4-1 – List of overall features considered from [13], Table I

	Network-layer	Transport-layer
10 s evaluation window	Byte counts Packet counts Throughput Idle Time	TCP Flag Counts (SYN, ACK, FIN, URG, PSH and RST) Out-of-order bytes/packets TCP Goodput ‡Retranmission ratio (Ratio of 0, 1, 2 and >2) In case of real time QoE assessment: Starting bytes-in-flight Ending bytes-in-flight
Packet Statistics: mean, min, max, median, standard deviation, skewness, kurtosis.	Packet inter-arrival times Bytes per packet	Retransmissions per packet Transport receive window RTT (only for upstream traffic) Bytes-in-flight

More detailed measurement definitions and instructions may be found in [13].

To correlate network and transport measurements, a laboratory learning under controlled networking conditions is suggested.

[13] characterizes the start-up phase only by statistics collected during the first 10 second evaluation window.

To characterize retransmission and re-buffering events, [13] evaluates and compares properties of the first five seconds of each window with the properties of the second five seconds of each evaluation window.

[13] found the following network parameters correlate with YouTube QoE:

- Initial start-up delay (median downloaded bytes, average packet inter-arrival times, median downstream throughput – packet statistics compared after 3.3, 6.6 and 10 seconds);
- Rr-buffering events (min downloaded bytes first half of the evaluation window, downstream throughput first half of the evaluation window);
- video quality (downstream throughput, upstream throughput).

[13] continues by adding a machine learning approach to develop and parameterize a QoE model. This contribution only extracts QoS measurements which correlate with application QoE. Interested readers may find the complete model in [13].

XI.4 General tendencies related to Internet access performance

Two reports also provide general statements to market requirements.

Flach et al. [5] mention that in an environment, where "bandwidth remains relatively cheap, Web latency is now the main impediment to improving user perceived performance." This is related to fixed network access, but the commodity of smartphones as end-devices is mentioned as one reason for this trend. Note that some of the authors work with a major OTT.

The authors of [9] measured QoE according to Recommendation ITU-T P.1203 with a 1920 × 1080 pixel display using a major OTT's content received via a commercial fixed network DSL access. A shaper was inserted between the measurement device and DSL router and downstream bandwidth was varied in 11 steps between 0.256 and 37.5 Mbit/s. MOS values of 4 and higher were scored for downstream bandwidths of 3.073 Mbit/s.

The authors of [7], a mobile network related document, conclude "that a downlink bandwidth of 4 Mbps is high enough to reach near optimal results in terms of overall quality and acceptability for [video steaming] when accessed by smartphones. The implications for the end-user are straightforward: ….an expensive LTE contract is not necessary to have a near optimal experience today" Note that the document was published in 2016 and the authors used HD720p videos to achieve their results.

All three statements indicate, that for a high enough Internet access bandwidth, the satisfaction of subscribers and content providers with the performance of an Internet access may be independent of the maximum Internet access speed.

References for Appendix XI

[1] "Speed Measurements of Residential Internet Access", Oana Goga and Renata Teixeira, PAM 2012, 2012.

[2] "End-to-end Inference of Internet Performance Problems", Partha Kanuparthy, PhD Thesis, Georgia Institute of Technology, 2012.

[3] "Diagnosing Path Inflation of Mobile Client Traffic", Kyriakos Zarifis, Tobias Flach, Srikanth Nori, David Choffnes, Ramesh Govindan, Ethan Katz-Bassett, Z. Morley Mao, and Matt Welsh, 2014.

[4] "An Internet-Wide Analysis of Traffic Policing", Tobias Flach, Pavlos Papageorgey, Andreas Terzis, Luis D. Pedrosa, Yuchung Chengy, Tayeb Karimy, Ethan Katz-Bassett, and Ramesh Govindan, IEEE SIGCOMM, 2016.

[5] "Reducing Web Latency: the Virtue of Gentle Aggression" Tobias Flach, Nandita Dukkipati, Andreas Terzis, Barath Raghavan, Neal Cardwell, Yuchung Cheng, Ankur Jain, Shuai Hao, Ethan Katz-Bassett, and Ramesh Govindan, IEEE SIGCOMM 2013.

[6] "Measuring Video QoE from Encrypted Traffic", Giorgos Dimopoulos, Ilias Leontiadis, Pere Barlet-Ros, and Konstantina Papagiannaki, IMC '16, 2016.

[7] "Next to You: Monitoring Quality of Experience in Cellular Networks from the End-devices", Pedro Casas, Michael Seufert, Florian Wamser, Bruno Gardlo, Andreas Sackl, and Raimund Schatz, IEEE Transactions on Network and Service Management Vol 13 issue 2, 2016.

[8] "Predicting QoE in Cellular Networks using Machine Learning and in-Smartphone Measurements", Pedro Casas, Alessandro D'Alconzo, Florian Wamser, Michael Seufert, Bruno Gardlo, Anika Schwind, Phuoc Tran-Gia, Raimund Schatz, QoMEX 2017.

[9] "Measuring YouTube QoE with ITU-T P.1203 under Constrained Bandwidth Conditions", Werner Robitza, Dhananjaya G. Kittur, Alexander M. Dethof, Steve Göring, Bernhard Feiten, Alexander Raake, QoMEX 2018.

[10] "Predicting the effect of home Wi-Fi quality on QoE: Extended Technical Report." Diego Da Hora, Karel Van Doorselaer, Koen Van Oost, Renata Teixeira. [Research Report] INRIA; Technicolor; TelecomParisTech. 2018. <hal-01676921>

[11] "QoE-based low-delay live streaming using throughput predictions." Konstantin Miller, Abdel-Karim Al-Tamimi, and Adam Wolisz. 2016. ACM Trans. Multimedia Comput. Commun. Appl. 13, 1, Article 4 (October 2016), 24 pages. DOI: http://dx.doi.org/10.1145/2990505.

[12] "eMIMIC: Estimating HTTP-based Video QoE Metrics from Encrypted Network Traffic", Tarun Mangla, Emir Halepovicy, Mostafa Ammar, Ellen Zegura. Georgia Institute of Technology yAT&T Labs – Research.

[13] "Real-time Video Quality of Experience Monitoring for HTTPS and QUIC", M. Hammad Mazhar, Zubair Shafiq, The University of Iowa.

[14] "copycat: Testing Differential Treatment of New Transport Protocols in the Wild", Korian Edeline, Mirja Kühlewind, Brian Trammell, Benoit Donnet, ANRW '17, Prague.

Appendix XII

Accurate bit rate measurements

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

XII.1 Introduction

Every traffic meter along a communication path is configured to capture or enforce a bit rate on a single communication layer (or does so by default). Awareness of the communication layer on which a bottleneck bit rate is enforced reduces the measurement error.

Intermediate devices along a laboratory or network measurement path may be configured to condition traffic, and in their absence a physical interface may enforce a bit rate. A bit rate measurement system might indicate a rate above or below the bit rate, which is enforced by the bottleneck, if both do not meter the bit rate at the same communication layer.

Benchmark testing and calibration especially require accurate knowledge of the layers, on which traffic is metered by these traffic conditioners or physical interfaces. Transferring bandwidth measurement results accurately from one layer to another requires knowledge of the measurement packet header sizes on different layers. Accurate and comprehensible presentation of measurement results requires indicating the layer at which a bit rate is enforced and accompanying information like the packet PDU size on that layer.

Note that errors caused by ignoring packet headers and the layer, at which a bit rate is enforced are one source of a bit rate measurement error, but not the only one. Measurements using closed loop congestion feedback based on packet drop, like TCP or QUIC, introduce additional sources of error, which depend on round-trip delay and packet loss.

This appendix also includes information on token bucket filters, which are a key component of traffic shapers and policers.

XII.2 Key findings

Interface and protocol performance is comprehensibly presented by a rate or a bandwidth, like some Mbit/s. For general purposes, like surveys, rough knowledge of a bandwidth sufficiently characterizes a measurement result. This changes if bandwidth measurements are intended to calibrate or benchmark the performance of a device or a communication path under test. To be able to compare bandwidth measurement results collected by different methods or collected along measurement paths with multiple traffic conditioning points, accurate results require additional information. The error introduced by a measurement system and a bit rate enforcement on different layers is one of them (but not necessarily the only one).

Consider a simple bandwidth measurement set-up as shown in Figure XII.2-1.

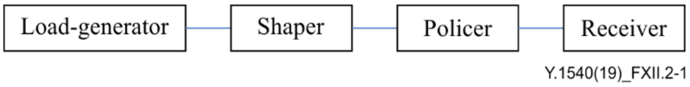


Figure XII.2-1 – A simple bandwidth measurement set-up

Note that there is no claim that the shown configuration is typical. Parts of it may be present in many networks.

The aim of the test may be to determine the maximum loss free bandwidth of the test set-up. The threshold against which to compare the measurement results is the bandwidth, by which the shaper and policer rates are configured (let us assume that both are configured to the same rate C_{Access}).

A measurement may reveal a maximum loss-free bandwidth of 0,95* C_{Access}. If the purpose of the test is to calibrate or benchmark the shaper and policer performance, this alone is however not a meaningful result (even if collected in a laboratory under controlled conditions).

Assume the load-generator and receiver to be the same device. By default this device generates and counts packet sizes and resulting bit rates like C_{Access} on layer 2. If the shaper and policer are configured to count and schedule or constrain bandwidth on layer 1, they may accurately forward traffic at bandwidth C_{Access}, if C_{Access} is measured at layer 1. All devices correctly work as designed and configured. The seemingly inaccuracy of 5% deviation from the expected result is caused by the layer on which the devices used in the test set-up meter traffic.

Now consider the test staff to introduce a small change, like diminishing the size of the measurement packets by 50%. The resulting bandwidth measurement now may be 0,93* C_{Access}. Assume that the bottleneck is not caused by a limited packet per second performance of any processor along the test set-up. Then this result indicates a layer or metering incompatibility, respectively, along the tested path. Smaller packets result in a smaller bit rate at layer 2, while the layer 1 bit rate remains constant. The production chain works correctly as configured.

If the test flow bandwidth at the conditioning device is unknown or cannot be controlled by configuring the sender or if the layer at which the conditioning device limits traffic is unknown, calibrating or benchmarking such a device is only possible with limited accuracy.

XII.3 Header size inflicted bandwidth measurement error estimation

The traffic is limited to a constant bit rate only at the communication layer on which the conditioning shaper. Policer or physical interface works. Shapers and policers should not be expected to work on any particular and well specified layer. In some cases, network provider devices allow to configure the layer on which the rate meters of shapers and policers work.

To accurately compare bandwidth measurements for calibration and benchmarking, the following information is required:

- communication layer and PDU size of the sender and receiver;
- all sizes of headers added or removed, if sender and receiver do not meter traffic on the same layer.
- The layer of the configuration parameters for which traffic conditioning devices to be calibrated or benchmarked are configured. If the bottleneck is a physical interface, most often the layer 1 interface bandwidth is known.
- The exact sizes of headers inserted or removed, if sender, receiver and traffic conditioning devices are not metering bandwidth at exactly the same layer with identical PDU sizes.
- The layer of a threshold against which a measurement result is compared and all header sizes which are added or removed, if the layer of the threshold is not identical with the layers of sender or receiver.

Measurement impairments result, if the threshold bit rate to which bandwidth measurement results of a sender and receiver are compared to, are not on the same communication layer. Further, the higher the number of active nodes, policy and interworking points along a path of production, the more likely is the presence of additional headers which may be absent at the sending and receiving device, e.g., tunnel headers between layer 2 and layer 3 or between layer 3 and layer 4.

If Ethernet headers are to be used in calculations, the presence and number of VLAN tags should be known.

In some cases, layer 1 requires escape sequences for correct framing. The presence of an escape code sequence may depend on payload bit patterns. In that case, measurement accuracy is limited.

In general, a bit rate C_x is limited to a (physical or) configured bandwidth at a layer x. At this layer x, and only at this layer x, the following equation holds for the bit rate C_x to be measured at layer x (generalized from ITU-T Y.1540):

$$C_x\ (t, \Delta t) = N_x(t + \Delta t)\ /\ \Delta t \tag{1}$$

where N_x is the total number of layer x bits that can be transferred over a basic section generating successful layer x packet transfer outcome at the egress measurement point during a specified time interval [t, t + Δt]. On a communication layer y which transports layer x packets, a packet header of constant length h_y bits needs to be added per layer x packet. If the only known or measured bandwidth is that of layer x, the layer y bandwidth consumed by the measured packet stream can only be determined if the number of packets p_x at layer x and the layer y header size h is known:

$$C_y\ (t, \Delta t) = [N_x\ (t + \Delta t)+p_x*h_y]\ /\ \Delta t = C_x\ (t, \Delta t)\ +\ p_x*h_y\ /\ \Delta t \tag{2}$$

Two changes impacting the measurement error are obvious:

- As expected, the layer y C_y bandwidth is bigger than the layer x C_x bandwidth.
- C_y depends on the packet number p_x on layer x, while C_x is independent of the layer x packet number p_x.

The number of packets p_x and the average packet size s_x are linked by equation (3):

$$s_x = C_x\ (t, \Delta t)\ /\ p_x \tag{3}$$

The layer x packet size s_x impacts the resulting bit rate C_y measured at layer y.

Standards, public specifications and product information may allow to determine header and maximum packet sizes on different communication layers for some markets. The layer at which a network or service-provider device enforces a configured rate should be considered as unknown.

The best accuracy improvement for a bandwidth measurement results if detailed information on measurement packet size and specified bottleneck header format information from as many different layers as possible is taken into account.

In the absence of detailed information, a conservative guess is to assume the minimum overhead size which is allowed to assume by either the measurement set-up used or by general information about an access, if no layer header format specification of the latter is available. This still allows to reduce the resulting measurement error.

The average sizes of the received measurement packets or the number of received measurement packets must be known to reduce the measurement error.

If layer headers and the average received measurement packet size or number are not exactly known, but maximum and minimum values can be determined, then a corrected measurement result corridor can be determined.

Formulas (1), (2) and (3) allow to calculate bandwidths at different communication layers, if communication layer headers and the average received measurement packet size or number or maximum and minimum values of these are known (in addition to the bandwidth measurement result).

XII.4 Example overhead calculation for a wireline IEEE 802.3 Ethernet interface

Please note that all calculations shown below are only valid for a bottleneck transmitting packets via a physical link conforming to Ethernet as specified by IEEE 802.3 series standards.

This SG12 project performed laboratory access bandwidth measurement tests (see Appendix X). The measurement tool "iPerf 2 reports rate measurements based on the transport payload bytes delivered (above the UDP or TCP layer). The token bucket filter rate [of a shaper configured to act as bottleneck] is specified in "bits on the wire", so the headers added to transport payloads are included in token bucket filter calculations (the rate includes ETH, IP and transport header bits).

The UDP payload of a first measurement is known to be 1470 Bytes. The Ethernet connection is a native one (no VLAN headers).

Lower layer header calculations:

- UDP header: 8 Byte;
- IPv4 header: 20 Byte;
- Ethernet.

Header

14 Byte (layer 2 without cyclic redundancy check, CRC);
18 Byte (layer 2 including cyclic redundancy check, CRC);
26 Byte (layer 1).

97.2 % of the configured rate were received without packet loss over various configured bottleneck rates. At first Ethernet layer 2 frame size was assumed to include CRC bytes. With a header of 46 Bytes per packet, a correction factor of 1.0313 resulted. The "bits on the wire" shaper bandwidth was calculated to be 1.00242 times the configured rate. The error is 0.24%, which is low.

A closer investigation reveals that the shaper rate excludes the CRC bytes. The header correction based on 42 Byte header size then is 1.0286. The "bits on the wire" shaper bandwidth was calculated to be 0.999799 times the configured rate. The error is -0.02%, which is another order lower. A 2 ppm measurement error indicates, that the laboratory shaper configuration worked pretty exact on Ethernet layer 2 (without CRC bytes).

Note, that also network provider equipment may meter Ethernet traffic on layer 2 without CRC. There is no default Ethernet bit rate metering layer yet, network provider hardware may meter Ethernet on layer 1 or on layer 2 including CRC bytes also.

XII.5 Description of token bucket filter functionality

Shapers and policers constrain traffic rates. The underlying rate control is often based on a token bucket filter. A token bucket filter generically works as follows:

- Configure a *Rate* in [bits/s].
- Configure a *Burst-Tolerance_Byte* in [Byte].

Optinally, systems often offer configuration of the *Burst-Tolerance_ms* in [ms]. Expect the following buffer to be assigned system-internally:

$$\textit{Burst-Tolerance\_Byte [byte]} = \textit{Rate} / 8 / \textit{Burst-Tolerance\_ms} * 1000 \qquad (4)$$

A token bucket policer will add a one bit token to the bucket every *1/Rate* seconds. If the bucket is filled by *Burst-Tolerance_Byte* tokens, additional tokens are dropped.

When a packet of *Packet-Length* bytes arrives, the filter checks whether tokens corresponding to *Packet-Length* are present in the bucket.

- If yes, the packet is forwarded and *Packet-Length* bytes are removed from the bucket.
- If no, the packet is dropped and the bucket remains unchanged (unless an additional shaper buffer is present; see the note below).

A token bucket shaper operates like a policer, but it operates an additional buffer of *Buffer_bytes*. If the policer drops the packet, it is stored at the buffer, as long as the latter does not overflow. Once a sufficient number of tokens is present in the token bucket filter, the first packet stored in the buffer is forwarded (given first-in-first-out scheduling is deployed).

If the shaper buffer has been completely consumed by arriving packets, any additional packet is dropped.

Note that shapers have an additional buffer that may be configurable in *Buffer_ms* [ms] or in a number of *Packets*, and there is typically an average packet size assumed when the system allocates bytes to the packet buffer. Again, the configured *Rate* is used to calculate the system internal buffer depth *Buffer_bytes* as shown in (1).

Appendix XIII

IP-based flow-related parameters and methods of measurement

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

XIII.1 Background

The IETF has approved and published RFC 8337, "Model-Based Metrics for Bulk Transport Capacity", in March 2018 [IETF RFC 8337]. The model-based metrics (MBM) work was the result of many years considering the problem of transport capacity measurement, primarily in the IETF IP Performance Metrics (IPPM) working group. The specification carefully describes the many issues and difficulties with repeatability when testing with standard-compliant TCP (section 4), and solves these problems principally by designing a method and a set of diagnostic tests where a TCP's flow-control is disabled. The method involves the evaluation of a target transport performance in terms of transmission rate and round-trip time (RTT).

Evaluation of MBM began before the RFC was published. In the paper, "Improved Internet speed tests can enhance QoS and QoE", Morton examined multiple measurement methods to evaluate the MBM model framework in the context of identifying the many accuracy issues with existing Internet measurement techniques and commencing the work to solve them [MortonPQS].

XIII.2 Why MBM meets the requirements of this Recommendation

Clause 6.12 provides a set of requirements that any flow-related measurement method must meet, as a reasonable participant in the context of the shared resources of the Internet. MBM meets these requirements, as described below (refer to section 3 of [IETF RFC 8337] for terms and definitions).

All flow or throughput-related parameters should fulfil the following numbered requirements:

1) A parameter characterizing the throughput offered to an IP service should relate the amount of IP packets successfully transported by an IP network or section to the amount of IP packets that were delivered into this network or section.

The number of packets sent ("delivered into this network or section") is completely controlled by the selection of the test stream. Further, "successful transport" of IP packets is directly measured in the MBM method (in order to derive the loss ratio and measure run length).

2) The throughput-related parameter should apply to an end-to-end IP network and to the IP transport across an EL, NS or NSE.

The MBM metrics and measurements are designed to be vantage-point independent, and therefore applicable to EL, NS, or NSE test paths (as meeting several other key requirements, see section 4.3 of [IETF RFC 8337].

Further requirements from this Recommendation are listed below, with analysis of MBM compliance:

Some flow or throughput-related parameters attempt to characterize the throughput capacity of an IP network, i.e., its ability to sustain a given IP packet transfer rate. It is recommended that any such parameters should fulfil the following additional requirements:

1) The traffic pattern offered to the IP network or section should be described, since the ability of the IP network or section to successfully deliver these packets depends on this traffic pattern.

The traffic pattern ("offered to the IP network or section") is constantly controlled by the selection of the test stream (conforming to the specified test parameters).

2) The rate at which traffic is offered should not exceed the capacity (in bits per second) of the link that connects the sections under test with the destination sections that are not under test.

The traffic pattern offered is selected and controlled under the constraints of the test stream and the target_rate MBM model parameter.

3) In any individual statement about throughput performance, the type of IP packet considered should be declared.

The IETF IP Performance Metrics (IPPM) Framework [b-IETF RFC 2330] (which guides all metric development and measurements conducted to their specifications) makes clear requirements for specifying the packet details; see section 13, Packets of Type P. Note that this section is currently being updated to include IPv6 requirements and other recent developments.

Note that Appendix IX describes how measurements using standard-compliant TCP do not meet the requirements of clause 6.12.

XIII.3 Role and status of the MBM method of measurement

The role of the MBM method is to determine whether a path or sub-path has sufficient performance to support a target rate of reliable byte stream transfer of a single transport-layer connection. This method is useful when evaluating whether a path or sub-path exhibits the transport rate required by a specific application, such as the sub-path between a content delivery server and the head-end of an access link. On the other hand, evaluation of IP capacity at Gigabit rates is not the role of MBM methods.

The remaining sections of this appendix describe topics which are for further study at this time. Additional laboratory and field evaluation is needed.

XIII.4 Test stream selection

[IETF RFC 8337] section 6 suggests several different test stream designs which can be selected as part of a targeted IP diagnostic suite (TIDS).

Section 6.1 describes stream design to **Mimic TCP Slow-Start** (which occurs at the beginning of every TCP connection). Key stream parameters are (noting that target values are determined by the service subscription and the path under test):

- burst size in packets (4, but smaller sizes may also be used);
- target_window_size;
- target_RTT;
- target_data_rate.

[IETF RFC 8337] says: "…at timescales longer than the target_RTT and when the burst size is equal to the target_window_size, the average rate is equal to the target_data_rate."

Section 6.2 describes the design of the **Constant Window Pseudo Constant Bit Rate (CBR) Stream**, along with the issue that a window of integer packet sizes combined with a fixed RTT may cause the delivered data rate to represent the target_data_rate exactly (operating slightly above or below the target_data_rate instead) for example when RTT or target_data_rate is small.

Again, the four key stream parameters are as listed for the stream to mimic TCP Slow-Start.

Variation from the target_data_rate may also occur when the RTT varies, due to the use of self-clocking in this stream (TCP ACK arrivals determine the sending rate, and they depend on RTT). Unexpected competing traffic can cause RTT variation (increase and decrease).

As [IETF RFC 8337] suggests, "Conventional paced measurement traffic may be more appropriate for these environments" where the target_data_rate cannot be matched or RTT/self-clocking present issues. **Pseudo-CBR traffic** may still contain bursts, but it is sent at the target_data_rate throughout the test.

The pseudo CBR traffic streams above are used in basic data rate evaluations, supporting the test in section 8.1.2 of [IETF RFC 8337], for example.

Note that section 6.3 describing Scanned Window Pseudo CBR stream generation increases the complexity to attain more realism in response to network conditions, and is left for further study at this time.

XIII.5 Measurement points

[b-IETF RFC 7398] defines a reference path and measurement points for commonly used performance metrics. Other similar measurement projects may also be able to use the extensions described here for measurement point location. The purpose of [b-IETF RFC 7398] is to create an efficient way to describe the location of the measurement point(s) used to conduct a particular measurement, especially pointing-out when a measurement includes managed and non-managed (private network) sub-paths.

Note that the measurement path, as delimited by the [b-IETF RFC 7398] measurement points, determines the applicability of subscription parameters, such as typical data rates offered and whether the subscription parameters inform the selection of MBM parameters such as target_data_rate.

The figure below indicates two scopes of non-overlapping measurements: access scope and distribution scope.

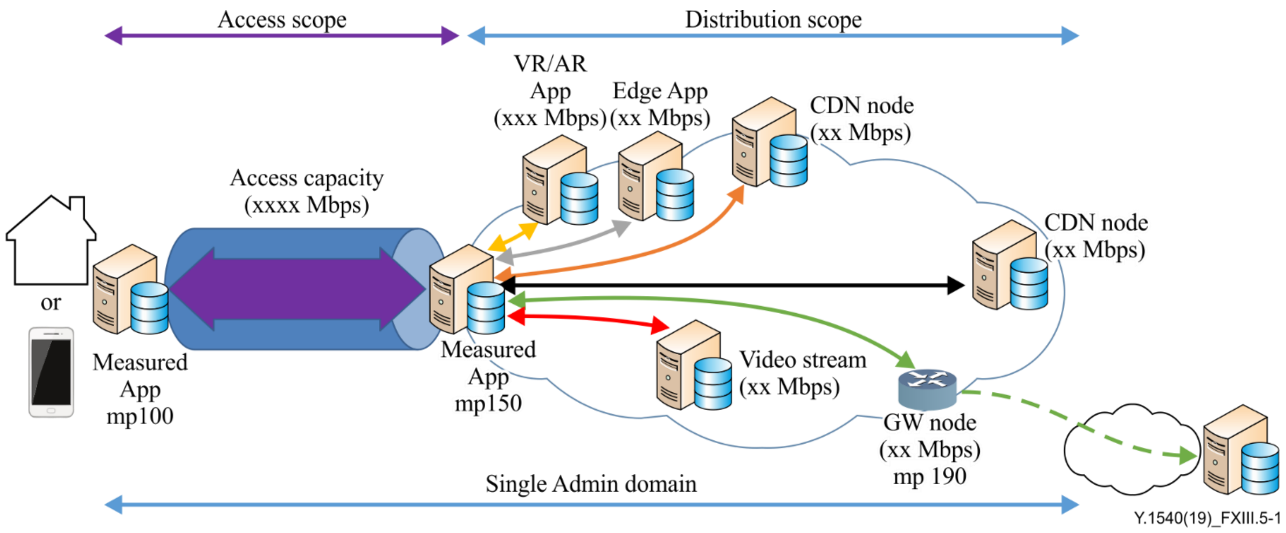


Figure XIII.5-1 – Separate measurements for access and distribution

In [b-IETF RFC 7398], the access scope exists between mp100 and mp150, and this is the intended scope of the UDP-based IP capacity metric and methods of Annex A.

On the other hand, the distribution scope exists between [b-IETF RFC 7398] mp150 and mp190, and this is the intended application of the MBM-based target_data_rate evaluation methods (between hosts at the edges or within the distribution scope) when they are refined further.

XIII.6 Target model parameter specification

See sections 5.1 and 5.2 of [IETF RFC 8337].

XIII.7 Setting acceptance criteria and interpreting the results

See sections 7.1 and 7.2 of [IETF RFC 8337].

XIII.8 Test methods

[b-IETF RFC 6673] and [MortonPQS] note the use of many repeated tests. A single test does not constitute an accurate evaluation of any subscription service that is expected to be available on demand, but may be sufficient for simple verification purposes when the results confirm expectations.

XIII.9 Example(s)

See section 9 of [IETF RFC 8337].

Bibliography

[b-ITU-T I.353] Recommendation ITU-T I.353 (1996), *Reference events for defining ISDN and B-ISDN performance parameters.*

[b-ITU-T I.356] Recommendation ITU-T I.356 (2000*), B-ISDN ATM layer cell transfer performance.*

[b-ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for objective and subjective assessment of quality*.

[b-ITU-T X.25] Recommendation ITU-T X.25 (1996), *Interface between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) for terminals operating in the packet mode and connected to public data networks by dedicated circuit.*

[b-ITU-T X.75] Recommendation ITU-T X.75 (1996), *Packet-switched signalling system between public networks providing data transmission services.*

[b-ITU-T X.137] Recommendation ITU-T X.137 (1997), *Availability performance values for public data networks when providing international packet-switched services.*

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

[b-IETF RFC 768] IETF RFC 768 (1980), *User Datagram Protocol*.
<http://www.ietf.org/rfc/rfc768.txt>

[b-IETF RFC 792] IETF RFC 792 (1981), *Internet Control Message Protocol*.
<http://www.ietf.org/rfc/rfc792.txt>

[b-IETF RFC 793] IETF RFC 793 (1981), *Transmission Control Protocol.*
<http://www.ietf.org/rfc/rfc793.txt>

[b-IETF RFC 919] IETF RFC 919 (1984), *Broadcasting Internet Datagrams.*
<http://www.ietf.org/rfc/rfc919.txt>

[b-IETF RFC 922] IETF RFC 922 (1984), *Broadcasting Internet datagrams in the presence of subnets*.
<http://www.ietf.org/rfc/rfc922.txt>

[b-IETF RFC 950] IETF RFC 950 (1985), *Internet Standard Subnetting Procedure.*
<http://www.ietf.org/rfc/rfc950.txt>

[b-IETF RFC 959] IETF RFC 959 (1985), *File Transfer Protocol*.
<http://www.ietf.org/rfc/rfc959.txt>

[b-IETF RFC 1305] IETF RFC 1305 (1992), *Network Time Protocol (Version 3) Specification, Implementation and Analysis.*
<http://www.ietf.org/rfc/rfc1305.txt>

[b-IETF RFC 1786] IETF RFC 1786 (1995), *Representation of IP Routing Policies in a Routing Registry (ripe-81++).*
<http://www.ietf.org/rfc/rfc1786.txt>

[b-IETF RFC 1812] IETF RFC 1812 (1995), *Requirements for IP Version 4 Routers.*
<http://www.ietf.org/rfc/rfc1812.txt>

[b-IETF RFC 2018] IETF RFC 2018 (1996), *TCP Selective Acknowledgment Options.*
<http://www.ietf.org/rfc/rfc2018.txt>

[b-IETF RFC 2330] IETF RFC 2330 (1998), *Framework for IP Performance Metrics.*
<http://www.ietf.org/rfc/rfc2330.txt>

[b-IETF RFC 3148] IETF RFC 3148 (2001), *A Framework for Defining Empirical Bulk Transfer Capacity Metrics.*
<http://www.ietf.org/rfc/rfc3148.txt>

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

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

[b-IETF RFC 3432] IETF RFC 3432 (2002), *Network performance measurement with periodic streams.*
<http://www.ietf.org/rfc/rfc3432.txt>

[b-IETF RFC 3550] IETF RFC 3550 (2003), *RTP: A Transport Protocol for Real-Time Applications.*
<http://www.ietf.org/rfc/rfc3550.txt>

[b-IETF RFC 6576] IETF RFC 6576 (2012), *IP Performance Metrics (IPPM) Standard Advancement Testing*.
<https://www.rfc-editor.org/rfc/rfc6576.txt>

[b-IETF RFC 6673] IETF RFC 6673 (2012), *Round-Trip Packet Loss Metrics.*
<https://www.rfc-editor.org/info/rfc6673>

[b-IETF RFC 7398] IETF RFC 7398 (2015), *A Reference Path and Measurement Points for Large-Scale Measurement of Broadband Performance*.
<https://www.rfc-editor.org/info/rfc7398>

[b-BEREC] BoR (18) 32: TENDER SPECIFICATIONS, No BEREC/2018/01/OT Net Neutrality Measurement Tool, March 2018.

[b-C-298] Kotanis, Irina (2015), *Proposals for E.802 Annex: minimum required of samples, statistical significance for benchmarking and quality trends evaluations and minimum required number of mobile agents, (with revisions)*, ASCOM, Switzerland.

[b-CVST] Krueger, T. and M. Braun (2012), *R package: Fast Cross – Validation via Sequential Testing, version 0.1*.

[b-Damjanovic] Damjanovic, Welzl et al. (2008), *Extending the TCP Steady-State Throughput Equation for Parallel TCP Flows*, University of Innsbruck, Budapest University of Technology.
<http://heim.ifi.uio.no/~michawe/research/publications/mulPadhye-TechnicalReport.pdf>

[b-Ekelin] Ekelin, S., Nilsson, M., Hartikainen, E., Johnsson, A., Mångs, J., Melander, B., Björkman, M. (2006), *Real-time measurement of end-to-end available bandwidth using kalman filtering*, IEEE/IFIP Network Operations and Management Symposium, Vancouver, Canada.

[b-Google-Police] "An Internet-Wide Analysis of Traffic Policing", Flach, Papageorge et al., University of Southern California and Google, 2016.

[b-Lautenschlaeger] Lautenschlaeger, W. (2014), A Deterministic TCP Bandwidth Sharing Model, Bell-Labs Alacatel-Lucent
<https://arxiv.org/abs/1404.4173>

[b-Montgomery] Montgomery, D. (1990), *Introduction to Statistical Quality Control – 2nd edition, ISBN 0-471-51988-X.*

[b-Morton] Morton, Al (2013), *Improved Internet speed tests can enhance QoS and QoE*, Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS 2013), Vienna, Austria.

[b-Mou] Mou, M. (2017), *Evaluating a TCP Model-Based Network Performance Measurement Method*, Masters Thesis at MIT, June 2017. <https://dspace.mit.edu/handle/1721.1/113177>.

[b-PAM-12] Oana Goga & Renata Teixeira (2012), *Speed Measurements of Residential Internet Access*, Passive and Active Measurements Conference, PAM-12. <https://people.mpi-sws.org/~ogoga/papers/PAM12-speed.pdf>

[b-Pod12] OPNFV Project, Intel POD12. <https://wiki.opnfv.org/display/pharos/Intel+POD12>

[b-Prasad] Prasad, R.S., Murray, M., Dovrolis, C., Claffy, K.C. (2003), *Bandwidth Estimation: Metrics, Measurement Techniques, and Tools, IEEE Network.*

[b-QUIC] "draft-ietf-quic-recovery-11", Iyengar and Swett. work in progress, IETF 2018. <https://datatracker.ietf.org/doc/draft-ietf-quic-recovery/>

[b-Rdev] R Development Core Team (2016), R*: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.r-project.org/>

[b-TST 009] ETSI GS NFV-TST 009 V3.1.1, (2018), *Network Functions Virtualisation (NFV) Release 3; Testing; Specification of Networking Benchmarks and Measurement Methods for NFVI.* <https://www.etsi.org/deliver/etsi_gs/NFV-TST/001_099/009/03.01.01_60/gs_NFV-TST009v030101p.pdf> <https://docbox.etsi.org/ISG/NFV/Open/Drafts/TST009_NFVI_Benchmarks>

[b-Wald] Wald, A. (1947), *Sequential Analysis*, Wiley.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
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	Environment and ICTs, climate change, e-waste, energy efficiency; 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, and associated measurements and tests
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, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2020