International Telecommunication Union

ITU-T

TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU

Y.1531

(11/2007)

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

Internet protocol aspects – Quality of service and network performance

---

SIP-based call processing performance

ITU-T Recommendation Y.1531

ITU-T Y-SERIES RECOMMENDATIONS

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

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

ITU-T Recommendation Y.1531

SIP-based call processing performance

Summary

ITU-T Recommendation Y.1531 defines three performance parameters that may be used in specifying, measuring, and comparing the speed, accuracy, and dependability of call set-up processing in networks that employ the session initiation protocol (SIP), with other protocols, in establishing and terminating media sessions ("calls") between users. The parameters may also be used in call processing performance apportionment or accumulation. This Recommendation does not specify numerical performance values.

Source

ITU-T Recommendation Y.1531 was approved on 13 November 2007 by ITU-T Study Group 12 (2005-2008) under the ITU-T Recommendation A.8 procedure.

FOREWORD

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

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

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

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

NOTE

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

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

INTELLECTUAL PROPERTY RIGHTS

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

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

CONTENTS

ITU-T Recommendation Y.1531

SIP-based call processing performance

1 Scope

This Recommendation defines three call processing performance parameters – *call set-up delay*, *call misrouting probability* and *call set-up failure probability* – on the basis of session initiation protocol (SIP) message transfers that may be observed at calling and called user-network interfaces (UNIs). These parameters are intended to be used in describing the performance of IP-based networks in which SIP [IETF RFC 3261] is used to establish and terminate media sessions ("calls") between users. In this context, SIP is regarded not as a purely application layer protocol, but as part of a comprehensive signalling system with associated network transport resource and admission control functions and lower layer protocols. Standardized technologies that use SIP in this way include the IP multimedia subsystem (IMS) and IPCablecom 2.

As noted in [ITU-T I.350], end user concerns about the performance of a function fall in three general categories: speed, accuracy and dependability. These correspond, respectively, with the three general outcomes a discrete function can encounter: successful performance, incorrect performance and non-performance. The parameters call set-up delay, call misrouting probability, and call set-up failure probability address, respectively, these user concerns and corresponding performance outcomes. They thus provide a logically complete (although very basic) characterization of SIP-based call processing performance. The parameters are intended to be used in specifying and comparing user requirements with service provider offerings, and in measuring achieved performance levels. They may also be used in apportioning or accumulating performance values among concatenated network sections delimited by SIP-based interfaces. [ITU-T G.1040] provides methods for modelling the network contribution to transaction time, and is applicable to aspects of SIP call processing.

This Recommendation is focused on one particular SIP function (or "method"): the INVITE function. It does not address other SIP functions or other performance issues, some of which may also be of importance to network providers and users. These functions and issues may be considered in later work.[1]

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

---

[1] Among the SIP functions not addressed are querying for capabilities; registration; call redirection, forwarding, and queuing; multicast; forking; session description; offer/answer exchanges; retransmissions; and the security functions, e.g., authorization and authentication. Among the performance issues not addressed are apportionment and accumulation, handling of emergency services, call processing interactions with priority access and restoral, and the impact of call processing on service availability.

[ITU-T G.1040]	ITU-T Recommendation G.1040 (2006), *Network contribution to transaction time*.
[ITU-T Y.1540]	ITU-T Recommendation Y.1540 (2002), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.
[IETF RFC 3261]	IETF RFC 3261 (2002), *SIP: Session Initiation Protocol*.

3 Definitions

This clause is intentionally left blank.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACK	Acknowledgement
B2BUA	Back-to-Back User Agent
CMP	Call Misrouting Probability
CMTS	Cable Modem Termination System
CSD	Call Set-up Delay
CSFP	Call Set-up Failure Probability
DNS	Domain Name System
DSLAM	Digital Subscriber Line Access Multiplexer
IAD	Integrated Access Device
IMS	IP Multimedia Subsystem
IP	Internet Protocol
IPS	Interruptions per Second
LS	Location Server
MTA	Message Transfer Agent
NGN	Next Generation Network
PS	Proxy Server
PSTN	Public Switched Telephone Network
SIP	Session Initiation Protocol
UAC	User Agent Client
UAS	User Agent Server
UNI	User-Network Interface

5 Conventions

None.

6 SIP-based call processing function and outcomes

It is customary to describe network performance for a specified communication function by first, formally defining the *function* in terms of *reference events* observable at the network *interfaces* of interest; second, delineating a set of relevant possible *outcomes* of an individual "trial performance" of the function; and finally, defining one or more *parameters* to characterize the performance attributes of each outcome.[2] This clause defines the SIP-based call processing function and outcomes. The SIP-based call processing performance parameters are defined in clause 7.

6.1 Definition of the call set-up function

The interfaces of primary interest in this Recommendation are the user-network interfaces (UNIs) that separate the SIP-based customer equipment from the IP network (Figure 1). Such interfaces include, for example, the physical interface between a SIP user agent (e.g., SIP phone, residential gateway, IAD, or MTA) on customer premises and the physical link that connects it to an associated network access concentrator (e.g., DSLAM or CMTS). Other interfaces are shown in Figure 1 to provide transaction details. SIP is assumed to be used at all of the illustrated interfaces.[3]

The *reference events* of interest in describing call processing performance in SIP-based networks are transfers of SIP messages across the relevant interfaces in accordance with the standardized SIP protocols. A SIP message is either a *request* from a client to a server or a *response* from a server to a client. Equipment implementing SIP (e.g., SIP phones, access gateways) can function as both servers and clients. The six SIP request messages defined in [IETF RFC 3261] are REGISTER, OPTIONS, INVITE, ACK, BYE and CANCEL. Two of these request messages are involved in a normal (successful) call set-up attempt: INVITE and ACK. The standardized SIP response messages are divided into six categories, distinguished by *status codes*: 1xx (provisional), 2xx (success), 3xx (redirection), 4xx (client error), 5xx (server error), and 6xx (global failure). One of these response messages is involved in a normal (successful) call set-up attempt: 200 OK. The exact times of occurrence of reference events (e.g., first bit versus last bit of message transfer) are specified using the conventions defined in [ITU-T Y.1540].

The communication *function* of interest in this context is *call set-up*.[4] Referring to Figure 1, the call set-up function begins when the calling user issues an INVITE request (event A) and ends (successfully) when the called user receives the corresponding ACK to its final 200 OK (event F). The ACK is considered to be part of the call set-up function for the reasons explained in clause 6.2 below.

6.2 Possible outcomes of a call set-up attempt

Three possible outcomes of a call set-up attempt are distinguished, by observation of reference events, to provide a basis for the call set-up performance parameter definitions.[5] *Successful call set-up* is defined to occur when the reference events (**A, B, C, D, E, F**) occur, as illustrated in Figure 1, within a specified maximum call set-up time, T_m (to be specified). *Call misrouting* is defined to occur when event **D** occurs, but events **B** and **C** do not occur, within T_m. *Call set-up*

---

[2] In statistical terms, the set of outcomes is a *sample space* and the parameters are *random variables* defined on it.

[3] The events depicted in Figure 1 include a DNS query by the caller's proxy server and a location service query by the callee's proxy server. These exchanges are typical but not necessarily present in every call. Other message exchanges may occur at interfaces between the UNIs during a call set-up attempt. UNI is one of the set of called user interfaces to which the call can properly be routed.

[4] In this Recommendation, the term call set-up is used in preference to INVITE to emphasize that the defined function includes the calling user's ACK to the called user's 200 OK. The ACK is described as a separate transaction in [IETF RFC 3261].

[5] Correspondence among the SIP messages used in defining these reference events and associated outcomes can be verified by comparison of header fields. The outcome definitions assume that "forking" of INVITE messages does not occur.

failure is defined to occur when one of the following event sequences is observed during the call set-up attempt:

1) Both events **B** and **D** do not occur.
2) Event **C** occurs, but event **D** does not.
3) Event **E** occurs, but event **F** does not.

A fourth outcome category, *excluded trial,* is defined to omit from network performance assessment any call set-up attempt that fails as a result of incorrect performance or non-performance by a user. Such outcomes include:

1) the calling user's termination of a call set-up attempt with a CANCEL request message;
2) the called user's failure to respond to an INVITE message (i.e., event **B** occurs but event **C** does not occur within T_m);
3) the calling user's failure to ACK the 200 OK message from the called user (i.e., event **D** occurs but event **E** does not occur within T_m).

In defining these parameters, call set-up trials should also be excluded if:

1) the network issues a redirection (3xx) response to the calling user;
2) a client error, server error, or global error (4xx-6xx) final response occurs, and the error is attributable to a user;[6] or
3) any event sequence not defined above occurs.

Transmission of the final ACK message is considered to be part of the call set-up function in these definitions because a SIP call is not firmly established until the ACK is received and accepted at the called user. Per [IETF RFC 3261], if a final ACK message is not received and accepted at the user agent server (UAS) within a maximum time (T2) after receipt of the 200 OK message, the call (session) should be terminated. The final ACK can also play an essential part in an offer/answer exchange.

7 SIP-based call processing performance parameters

This clause defines the three SIP-based call processing parameters: call set-up delay, call set-up misrouting probability, and call set-up failure probability.

7.1 Call set-up delay

Call set-up delay (CSD) is defined in terms of the reference events and variables shown in Figure 1. It is the elapsed time between the calling user's issuance of a SIP INVITE message at the calling UNI and the called user's receipt of the corresponding SIP ACK message at the called UNI, excluding:

1) the *called user delay*, between the called user's receipt of the INVITE message and issuance of the corresponding 200 OK message; and
2) the *calling user delay*, between the calling user's receipt of the 200 OK message and issuance of the corresponding ACK.

Referring to Figure 1:

$$CSD = (t6 - t1) - (t3 - t2) - (t5 - t4)$$

[6] In measurement situations, it may be necessary to verify proper attribution of failure outcomes by examining SIP message headers. Client errors can occur inside the network in certain cases, e.g., when the network contains back-to-back user agents (B2BUAs).

CSD is defined only for call set-up attempts that result in successful call set-up, as defined in clause 6.2. CSD includes the transfer time of the final ACK but excludes the response times of the calling and called users. The provisional (1xx) SIP responses have no effect on the CSD definition.

CSD will normally include transfer and processing delays associated with internal network functions that occur between the SIP-based reference events illustrated in Figure 1. Examples are the functions that implement admission control and resource reservation in particular IP network sections. In specifying or reporting CSD values, it will be important to carefully define the call flows assumed.

The defined CSD parameter is a *singleton* metric as described, for example, in [b-IETF RFC 2330]. This Recommendation does not consider the definition of sample or statistical metrics.

7.2 Call misrouting probability

Call misrouting probability (CMP) is the ratio of total call set-up attempts that result in *call misrouting* (as defined in clause 6.2) to the total call set-up attempts in a population of interest (omitting *excluded trials).*

Call misrouting is essentially the case of a "wrong number". It occurs when the network responds to a valid call request by erroneously establishing a call to a destination not designated in the request. It may be caused, for example, by network operator administrative or maintenance errors. Call misrouting is distinguished from successful call set-up by the fact that the intended called user is not contacted and committed to the session during the call set-up attempt.

7.3 Call set-up failure probability

Call set-up failure probability (CSFP) is the ratio of total call set-up attempts that result in *call set-up failure* (as defined in clause 6.2) to the total call set-up attempts in a population of interest (omitting *excluded trials*).

One type of call set-up failure in the SIP signalling context is a non-response by the network, analogous to the "high and dry" outcome users experience under some failure conditions in the PSTN. In the PSTN, users may also experience "call blocking", identified by a 120 IPS reorder ("fast busy") signal. SIP-based networks may report similar outcomes with a 503 (service unavailable) message. Both types of call set-up failure will be captured by the absence of expected network responses to user messages, as defined in clause 6.2.

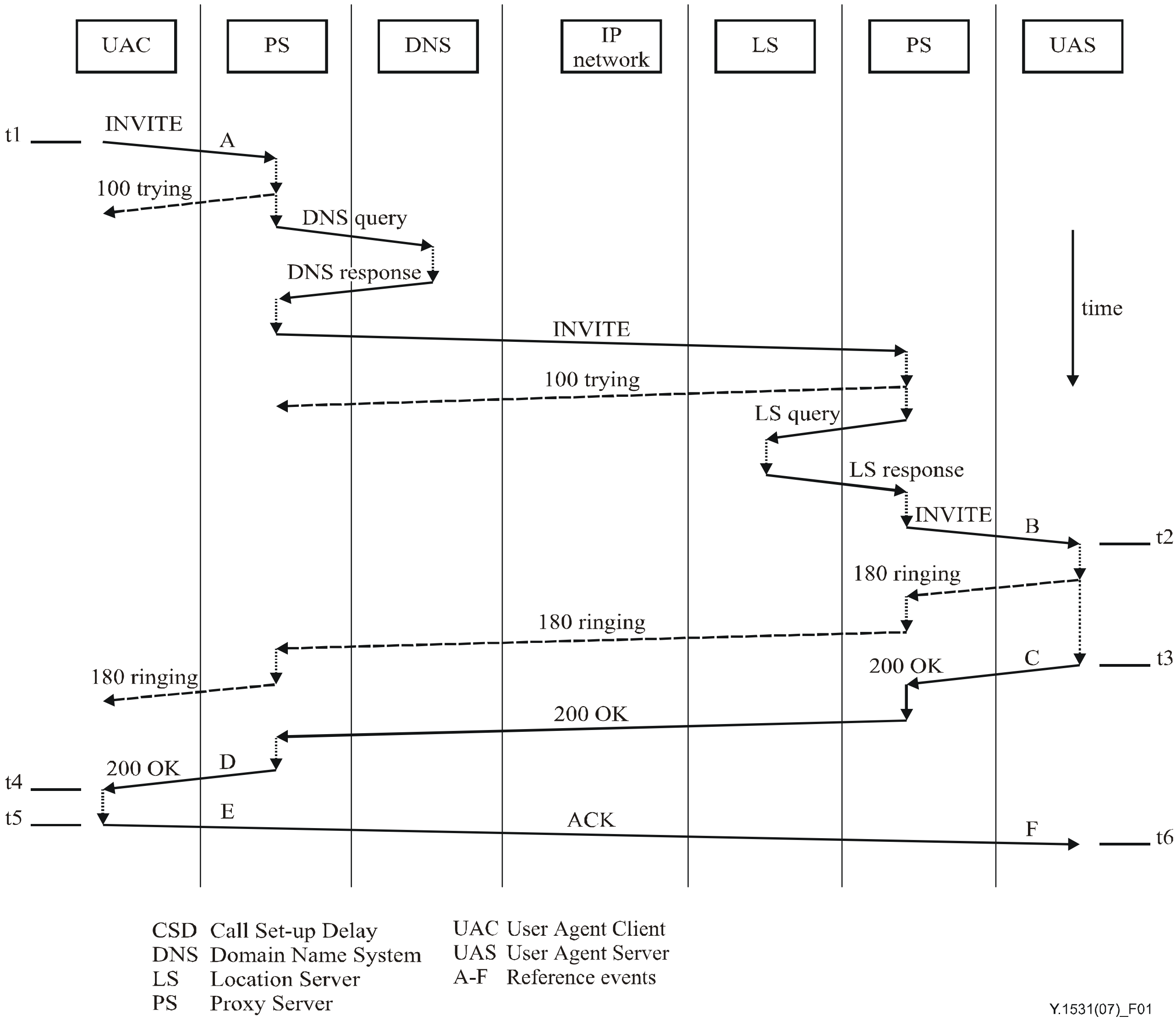

Figure 1 – Reference events used in defining the SIP-based call processing parameters

8 Security considerations

[IETF RFC 3261] describes security needs, services and mechanisms associated in general with the use of SIP. The SIP-based call processing performance parameter definitions presented in this Recommendation do not raise new security issues. Security issues will need to be considered in the future development and use of SIP-based call processing performance measurement methods.

Bibliography

[b-IETF RFC 2330] IETF RFC 2330 (1998), *Framework for IP Performance Metrics*.

SERIES OF ITU-T RECOMMENDATIONS

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

Printed in Switzerland
Geneva, 2008