International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Q.3961

(09/2020)

SERIES Q: SWITCHING AND SIGNALLING, AND ASSOCIATED MEASUREMENTS AND TESTS

Testing specifications – Testing specifications for next generation networks

Parameters for bottleneck evaluation of the web-browsing service

Recommendation ITU-T Q.3961

ITU-T Q-SERIES RECOMMENDATIONS

SWITCHING AND SIGNALLING, AND ASSOCIATED MEASUREMENTS AND TESTS

SIGNALLING IN THE INTERNATIONAL MANUAL SERVICE	Q.1–Q.3
INTERNATIONAL AUTOMATIC AND SEMI-AUTOMATIC WORKING	Q.4–Q.59
FUNCTIONS AND INFORMATION FLOWS FOR SERVICES IN THE ISDN	Q.60–Q.99
CLAUSES APPLICABLE TO ITU-T STANDARD SYSTEMS	Q.100–Q.119
SPECIFICATIONS OF SIGNALLING SYSTEMS No. 4, 5, 6, R1 AND R2	Q.120–Q.499
DIGITAL EXCHANGES	Q.500–Q.599
INTERWORKING OF SIGNALLING SYSTEMS	Q.600–Q.699
SPECIFICATIONS OF SIGNALLING SYSTEM No. 7	Q.700–Q.799
Q3 INTERFACE	Q.800–Q.849
DIGITAL SUBSCRIBER SIGNALLING SYSTEM No. 1	Q.850–Q.999
PUBLIC LAND MOBILE NETWORK	Q.1000–Q.1099
INTERWORKING WITH SATELLITE MOBILE SYSTEMS	Q.1100–Q.1199
INTELLIGENT NETWORK	Q.1200–Q.1699
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR IMT-2000	Q.1700–Q.1799
SPECIFICATIONS OF SIGNALLING RELATED TO BEARER INDEPENDENT CALL CONTROL (BICC)	Q.1900–Q.1999
BROADBAND ISDN	Q.2000–Q.2999
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR THE NGN	Q.3000–Q.3709
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR SDN	Q.3710–Q.3899
TESTING SPECIFICATIONS	Q.3900–Q.4099
Testing specifications for next generation networks	**Q.3900–Q.3999**
Testing specifications for SIP-IMS	Q.4000–Q.4039
Testing specifications for Cloud computing	Q.4040–Q.4059
Testing specifications for IMT-2020 and IoT	Q.4060–Q.4099
PROTOCOLS AND SIGNALLING FOR P2P COMMUNICATIONS	Q.4100–Q.4139
SIGNALLING REQUIREMENTS AND PROTOCOLS FOR IMT-2020	Q.5000–Q.5049
COMBATING COUNTERFEITING AND STOLEN ICT DEVICES	Q.5050–Q.5069

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

Recommendation ITU-T Q.3961

Parameters for bottleneck evaluation of the web-browsing service

Summary

If the quality of service (QoS) of the web-browsing service drops, Internet service providers (ISPs) and Internet content providers (ICPs) would hope to be able to immediately find the reasons, fix the faults and improve the web-browsing service. To achieve the above aims, Recommendation ITU-T Q.3961 specifies parameters for bottleneck evaluation of the web-browsing service. These parameters can be divided into four groups: parameters in the application layer, parameters in the transportation layer, parameters in the network layer and the characteristic parameters. The relationship between these parameters is also introduced in this Recommendation.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Q.3961	2020-09-29	11	11.1002/1000/14417

Keywords

Parameter, QoS, web-browsing service.

* 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

Recommendation ITU-T Q.3961

Parameters for bottleneck evaluation of the web-browsing service

1 Scope

This Recommendation defines parameters for bottleneck evaluation of the web-browsing service, including parameters in the network layer, in the transportation layer and in the application layer, and the characteristic parameters. The relationship between these parameters is also introduced.

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 Q.3916] Recommendation ITU-T Q.3916 (2019), *Signalling requirements and architecture for the Internet service quality monitoring system*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

3.1.1 quality of service [b-ITU-T E.800]: Totality of characteristics of a telecommunication service that bear on its ability to satisfy stated and implied needs of the user of the service.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 web-browsing service: A service that enables a user to display text, images and other information from web servers.

3.2.2 HTML object: The HTML object tag represents an external resource embedded in the HTML file, which can be an image, a webpage or a plugin application.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACK Acknowledgement

BRAS Broadband Remote Access Server

DNS Domain Name Server

HTML Hypertext Markup Language

HTTP Hypertext Transfer Protocol

ICP Internet Content Provider

IP Internet Protocol

ISP	Internet Service Provider
PC	Personal Computer
QoS	Quality of Service
RTT	Round Trip Time
SYN	Synchronization
TCP	Transmission Control Protocol
URL	Uniform Resource Locator
XHTML	extensible Hypertext Markup Language

5 Conventions

The keywords "**is required to**" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this Recommendation is to be claimed.

The keywords "**is prohibited from**" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this Recommendation is to be claimed.

The keywords "**is recommended**" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

The keywords "**can optionally**" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with the specification.

In the body of this document and its annexes, the words shall, shall not, should, and may sometimes appear, in which case they are to be interpreted, respectively, as is required to, is prohibited from, is recommended, and can optionally. The appearance of such phrases or keywords in an appendix or in material explicitly marked as informative are to be interpreted as having no normative intent.

6 Parameters for bottleneck evaluation of the web-browsing service

If the QoS of the web-browsing service drops, many parameters are useful to diagnose reasons or faults, namely the bottleneck of the web-browsing service. To clearly describe those parameters, these are classified into four groups as shown in clauses 6.1 to 6.4. Except the characteristic parameters group, the groups are mainly derived from the Transmission Control Protocol /Internet Protocol (TCP/IP) network model. This means that the parameters are calculated according to protocols in the respective layer of the TCP/IP network model. The characteristic parameters represent influential factors of user terminals or web servers.

6.1 Parameters in the application layer

a) Webpage response time

Webpage response time refers to the interval from the start time of retrieving a webpage to the time a response is obtained at the terminal from a web server. The start time for retrieving a webpage could be defined as the time when the enter key is pressed after the uniform resource locator (URL) of the webpage is input. Anything of the webpage being displayed on a browser, usually the title of a webpage, could be regarded as getting a response from the web server. As a result, the webpage response time could be calculated as follows: if t1 is the time when the domain name server (DNS) request packet for the URL of the webpage is sent, and t2 is the time the webpage title is displayed, the webpage response time equals t2 minus t1.

b) Webpage display time

Webpage display time refers to the interval from the start time of retrieving a webpage to the time that the webpage is displayed fully on the browser. The start time for retrieving a webpage could be defined as the time when the enter key is pressed after the URL of the webpage is input. As a result, the webpage display time is calculated as follows: if that t1 is the time when the DNS request packet for the URL of the webpage is sent, and t2 is the time when the last data packet of the webpage is received, the webpage display time is t2 minus t1.

c) DNS resolution time

DNS resolution time refers to the interval in which the URL of the webpage is resolved by the DNS server. Apparently, it is calculated as follows: if t1 is the time when the DNS request packet for the URL of the webpage is sent, and t2 is the time when the DNS response packet is received, the DNS resolution time is t2 minus t1.

d) HTTP redirection time

It is known that hypertext transfer protocol (HTTP) redirection occurs when a new URL is used for the webpage whose old URL is retained. It is calculated as follows: if t1 is the time when the HTTP request packet of the old URL is sent, and t2 is the time when the HTTP request packet of the new URL is sent, the HTTP redirection time is t2 minus t1. Note that not all web-browsing services include the HTTP redirection time.

e) First packet time

First packet time refers to the interval from the time the first HTTP GET packet is sent, to the time the first data packet of the hypertext markup language (HTML) file is received. Apparently, it is calculated as follows: if t1 is the time when the first HTTP GET packet is sent, and t2 is the time when the first data packet of the HTML file is received, the first packet time is t2 minus t1. In a way, the first packet time reflects performance and load of web servers.

f) HTML object distribution

In addition to an HTML file, a webpage consists of many HTML objects, such as images and videos, that are embedded in the HTML file with a unique URL. These HTML objects are downloaded from the web servers separately. On the one hand, the size of these HTML objects varies. On the other hand, their servers could be different, which results in HTML objects to be in different locations, different ISPs, different network routes, etc.

HTML object distribution aims to summarize all HTML objects in the webpage in terms of URL, size, location, etc. In the end, the above information could be presented in the form of a table, as shown in Table 6-1. URL denotes the URL for an HTML object. IP address is the address of a web server. Note that URL and IP address can determine a unique row, much like the primary key of the relational database model. Size denotes how big an HTML object is. Starting time is the time to send an HTTP request for it. Duration denotes the interval from starting time to finishing time for getting it.

Table 6-1 – Illustration of HTML object distribution

URL	IP Address	ISP	Size	Starting time	Duration

g) HTML object download time

HTML object download time refers to the interval from the HTTP request of the first object to the time all objects have been obtained by a browser. Note that objects can be retrieved in parallel. Thus, HTML object download time does not equal the sum of download time of all objects. HTML object download time is calculated with the first HTML object and the last HTML object. If t1 is the time

of sending an HTTP request of the first HTML object is t1, and t2 is the time of finishing receiving the last HTML object, the HTML object download time is t2 minus t1.

6.2 Parameters in the transport layer

a) TCP establishment time

It is known that HTTP data are carried by the TCP protocol, and that TCP transmission begins with the establishment of a TCP connection. The TCP establishment time refers to the interval for the establishment of TCP connection before downloading HTML file of the webpage. To simplify its calculation, the first two TCP packets in the three-way handshake are considered. If t1 is the time of sending a synchronization (SYN) packet, and t2 is the time of receiving SYN ACK packet, the TCP establishment time is t2 minus t1.

6.3 Parameters in the network layer

a) Peak download rate

Peak download rate refers to the maximum download rate from all web servers related to the webpage when getting the webpage. Note that the download rate from web servers may overlap each other. Thus, the peak download is a sum value. It is calculated by total throughput per second. Besides, it is recommended that the download rate for each web server should be recorded.

b) Bandwidth

Bandwidth refers to the transmission capacity of the network, which is often chosen by customers when they subscribe the Internet connection line from an ISP.

c) RTT

Round trip time (RTT) refers to the round-trip time between a terminal and a web server. It can be calculated with the Ping request and Ping reply, which is usually used to evaluate the QoS of network. It can also be calculated with the three-way handshake of the TCP protocol. Note that it is required to record RTTs of all web servers of the webpage.

d) Packet loss rate

Packet loss rate is a ubiquitous measure to evaluate the QoS of a network, and is used here to examine the end-to-end route between terminals and web servers. It is usually calculated with the Ping request and Ping reply, so it can be obtained with RTT at the same time. Likewise, it is required to record packet loss rates of all web servers of the webpage.

6.4 Characteristic parameters

a) HTML file size

HTML file size refers to the size of the pure HTML file, which excludes HTML objects embedded in it. So, the size of an HTML file is usually small, for example in the order of KB.

b) Webpage size

Webpage size refers to the total size of a webpage, including the HTML file and all embedded HTML objects, namely all files to display the webpage.

c) Optimal bandwidth

Optimal bandwidth refers to the ideal maximum download rate to display a webpage as soon as possible. First, all files belonging to the webpage, namely the HTML file and HTML objects, are downloaded to the terminal. Next, the download rate is the maximum download rate achieved by the terminal supposing there is no constraint on the bandwidth available for the terminal. Therefore, the optimal bandwidth is calculated as follows: assuming there is no limitation on the available bandwidth for the terminal, the optimal bandwidth is the maximum throughput per second when the webpage is downloaded to the terminal.

6.5 Summary of parameters

For clarity, Table 6-2 summarizes the parameters described in clauses 6.1 to 6.4.

Table 6-2 – Description of parameter

Category	Name	Description
Application layer	Webpage response time	An interval to get response from web servers
	Webpage display time	An interval to get the whole webpage
	DNS resolution time	An interval to get IP address corresponding to URL
	HTTP redirection time	An interval from sending HTTP request of old URL to sending HTTP request of new URL
	First packet time	An interval to get first packet of HTML file
	Object distribution	Detailed information of all objects
	Object download time	An interval to get all objects
Transport layer	TCP establishment time	An interval between SYN and SYN ACK
Network layer	Peak download rate	Highest download rate
	Bandwidth	Transmission capacity of the access network
	Packet loss rate	A traditional indicator in network layer
	RTT	A traditional indicator in network layer
Characteristics	HTML file size	Size of HTML file excluding objects
	Webpage size	Size of the whole webpage
	Optimal bandwidth	Maximum download rate in ideal condition

7 Relationship of parameters

7.1 Relationship discussion

Figure 7-1 shows the relationships among the parameters. Webpage display time at the top of the chart mainly reflects the QoS of the web-browsing service. A user will get impatient if it takes a long time for the webpage to be displayed. Webpage response time represents the time when a user gets a response from web servers. All its branch parameters, as shown in Figure 7-1, such as DNS resolution time and packet loss rate, facilitate analyzing bottlenecks in this stage. HTML object download time and all its branch parameters particularly concentrate on download of webpage contents and analysis on bottlenecks in this stage.

Obviously, webpage display time is chiefly the sum of webpage response time and objects download time. Therefore, if the webpage display time is too large, webpage response time and objects download time will be checked according to their benchmarks.

As HTTP redirection rarely exists, webpage response time is chiefly the sum of DNS resolution time, TCP establishment time and first packet time. Therefore, if the webpage response time is abnormal, these parameters are first checked according to their benchmarks in order to find the accurate reason. For instance, DNS resolution time is related to DNS server and the corresponding network connection while TCP establishment time and first packet time are related to web servers and the corresponding network connection. The QoS of a network is traditionally measured by packet loss rate and RTT. The other reason for abnormal webpage response time is related to the HTML file size, which may take a long time to download.

Object download time is made up of the download time of all objects, but it is not their sum of these times as several objects can be retrieved in parallel. Hence, attention should be given not only to the

download time of each HTML object, but also to the starting time. Both of these are included in the HTML object distribution parameter. The download time of each HTML object helps to locate abnormal objects, such as those that are too big in size or too slow in speed. The starting times of all objects are used to restore the sequence in which a browser arranges them. An appropriate sequence promotes fast download of objects. From the perspective of the network layer, object download time is simply determined by webpage size and peak download rate that may be limited by bandwidth. It is worth noting that downloading objects includes DNS resolution, TCP establishment, HTTP request, etc.

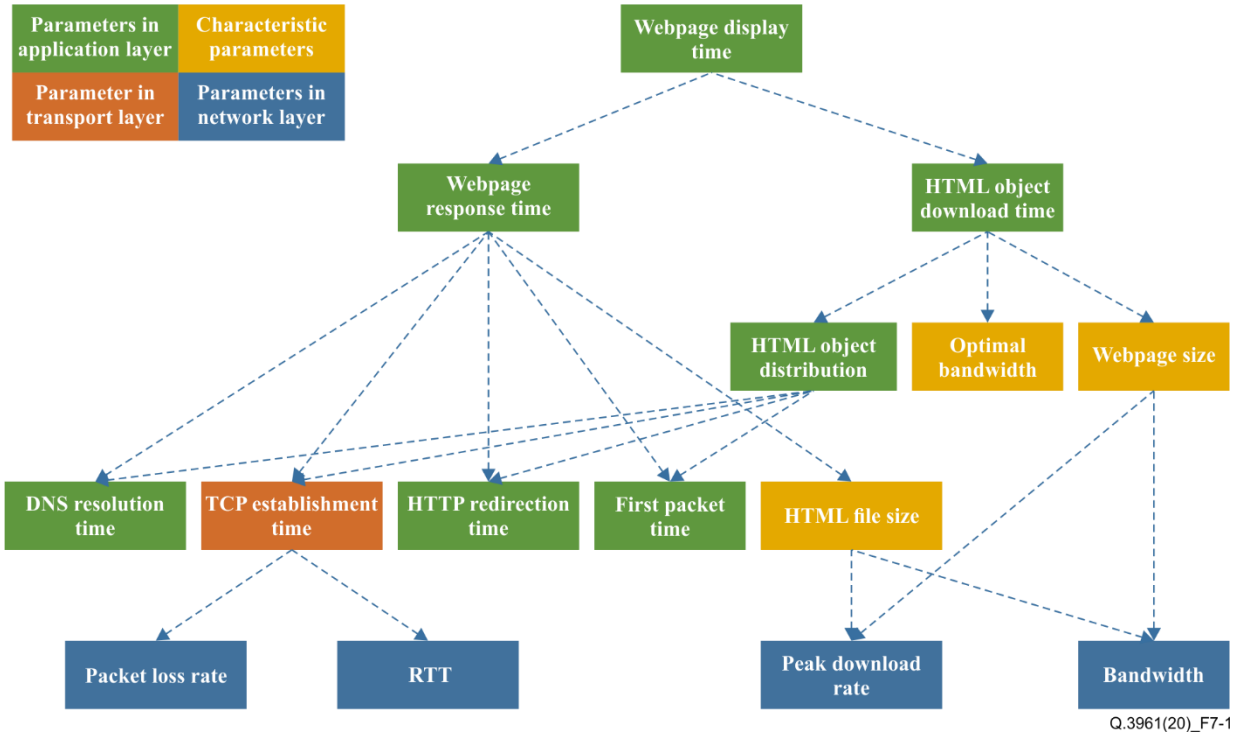

Figure 7-1 – Relationship illustration

7.2 Bottleneck analysis

Bottleneck analysis of the web-browsing service is generally performed to check service characteristics, cloud performance and network performance when the webpage display time is higher than a benchmark value. Service characteristics are reflected by characteristic parameters, cloud performance refers to performance of web servers, and network performance particularly refers to performance in the network layer like RTT.

Service characteristics like HTML layout and webpage size have an influence on HTML objects download time and are chiefly determined by design and development of a webpage. It will be thought to be a factor to the bottleneck if the webpage display time does not get significant improvements when the bandwidth is equal to or greater than optimal bandwidth parameter.

Cloud performance could become bottleneck in terms of response time, connection traffic, and outbound bandwidth of web servers, which are detailed as follows.

– Response time. Parameters of DNS resolution time, TCP establishment time, and HTTP redirection time primarily consist of RTT and response time of web servers. So, the response time can be approximately got by subtracting RTT from these parameters. The response time should be in a reasonable range if web servers are normal otherwise abnormal. Thus, bottleneck can be deduced from unreasonable response time.

– Connection traffic. When web servers are unable to establish connections for new requests, an error or redirection would occur, which is a signal of heavy connection traffic.

– Outbound bandwidth. There exists constraint of outbound bandwidth if peak download rate is less than end-to-end theoretical bandwidth that can be calculated with packet loss rate, RTT, window size of TCP, etc. Therefore, constraint of outbound bandwidth could be a factor of bottleneck.

At last, network performance could become bottleneck in terms of bandwidth, packet loss rate, RTT, and server location. If bandwidth is less than optimal bandwidth, it would be an obstacle to downloading objects as soon as possible. Packet loss rate and RTT negatively impact on download rate. If web servers are geographically far from users, the RTT and packet loss rate are likely to increase accordingly. Hence, the server location relative to users is an important factor.

8 Assessment tool

An assessment tool is needed to obtain all parameters mentioned in Table 6-2 in order to evaluate bottlenecks of the web-browsing service. First of all, it is required that the assessment tool be able to collect all parameters such as webpage response time, first packet time, TCP establishment time, etc. Second, it is recommended to have an interface to access fixed network, mobile network, or even routers (e.g., broadband remote access server (BRAS)) since QoS of web-browsing service at different locations in the network needs to be evaluated. Finally, it is better to embed core module of popular browsers (e.g., Internet Explorer, Firefox or Chrome) in the assessment tool in order to simulate the conditions of a real user. Especially, concurrent connections of TCP and sequence to download objects are different for popular browsers.

It is recommended to use the Internet service quality monitoring system described in [ITU-T Q.3916] as the assessment tool to evaluate web-browsing service. It not only fully supports simulation of the web-browsing service, but it also has capabilities of data storage, data analytics and data visualization, which facilitates diagnosing bottlenecks as described in clause 7.

Appendix I

Introduction to the process of browsing webpages

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

The web-browsing service enables a user to browse a webpage in the form of hypertext markup language (HTML) or extensible hypertext markup language (XHTML) with a popular browser such as Internet Explorer, Firefox and Chrome using a device such as a personal computer (PC), a smartphone or a tablet PC.

When the user tries to access a webpage, the domain name of the webpage is input to the browser. Next, a DNS request package is sent to a DNS server to ask for the IP address corresponding to the domain name, namely the IP address of the web server. Then, a TCP connection to carry HTTP data is established between the browser and the web server. The browser first downloads the main HTML file, which defines the structure of the webpage. Note that download of the main HTML file is conducted with only one TCP connection. The main HTML file is processed to construct a tree that shows relations between HTML objects (e.g., images) embedded in the webpage and to create their download sequence. Then, the objects are downloaded by sequence. However, several HTML objects are downloaded in parallel, the number of which differs depending on the restrictions or configurations of the browser. As soon as an HTML object is completely downloaded, the browser will render it and display it to the user. The interactions in browsing a webpage, which are shown in Figure I.1, can be mainly divided into four phases.

– DNS resolution. The browser launches the DNS request to resolve the domain name of the webpage in order to get its IP address. This DNS request is also performed for every HTML object possessing an own domain name.
– TCP connection. A TCP connection is established between the browser and the web server with "three-handshake". On one TCP connection, HTTP version 1.1 now supports consecutive HTTP requests that the next HTTP request can start on the same TCP connection after the previous one is finished. On the other hand, most browsers support parallel TCP connections, which are for downloading HTML objects simultaneously in order to improve efficiency.
– Main HTML file download. It is the first HTML file downloaded by the browser. After getting it, the browser starts to download other HTML files and objects.
– Object download. Most browsers support multiple TCP connections to download objects simultaneously in order to increase the download rate.

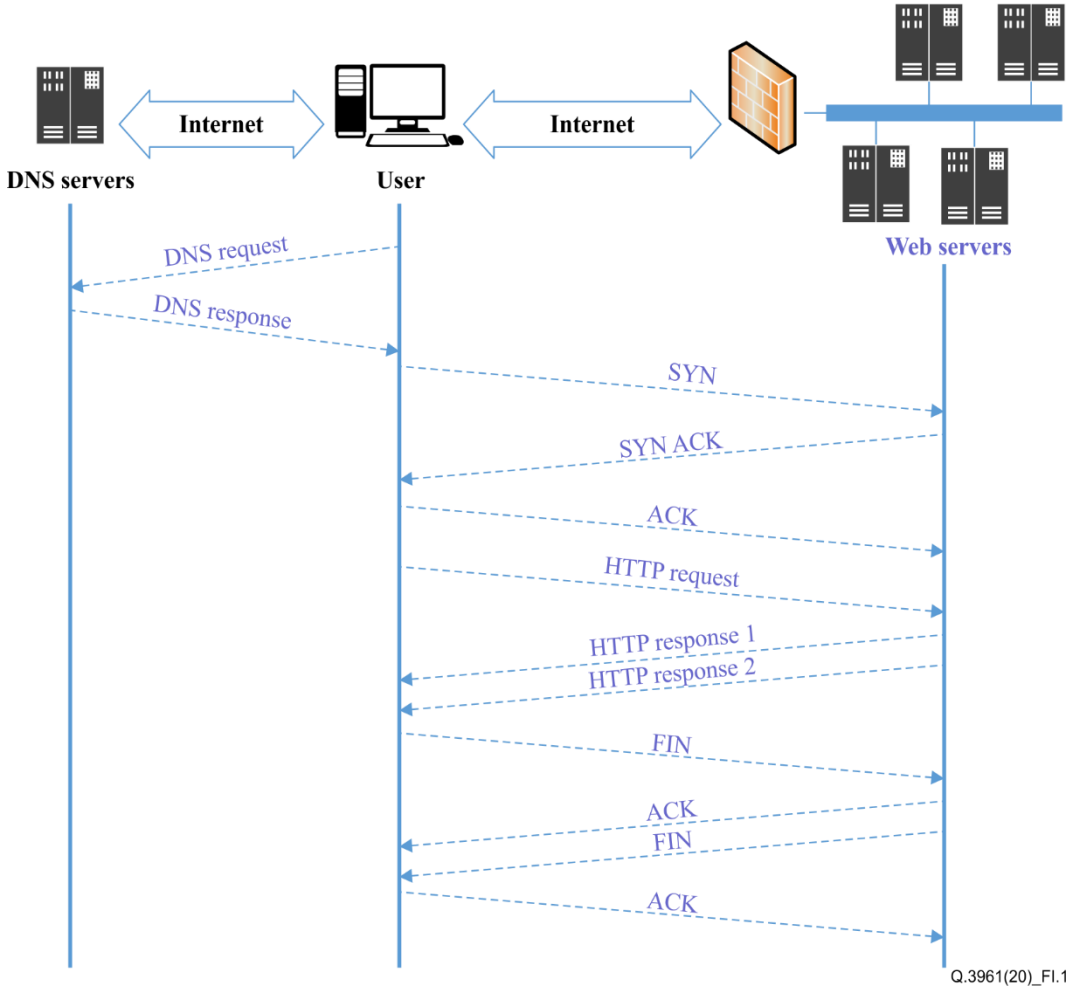

Figure I.1 – Process of browsing a webpage

Bibliography

[b-ITU-T E.800] Recommendation ITU-T E.800 (2008), *Definitions of terms related to quality of service*.

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