ETSI TR 103 733 V1.1.1 (2022-03)

Speech and multimedia Transmission Quality (STQ);
Best practices of testing the performance of
web content delivery

Reference
DTR/STQ-00227m

Keywords
measurement, network measurement, network performance, performance, QoS, service, testing, user

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Contents

List of figures

List of tables

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "will not", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

The present document focuses on testing and performance measuring of web content delivery. Retrieving web content is more than web browsing, many apps on smartphones use the same or similar techniques like browsing. This means that the delivered information is based on HTML with embedded resources in other formats like images, scripts or files and the transport is using HTTP/HTTPS mainly.

Web browsing or HTTP browsing as in [i.1] is a prominent example for delivery of web content but can be seen as representing a much wider range of popular smartphone applications.

Even when looking at HTTP browsing solely, it is one of the most used web applications in mobile networks. Consequently, testing HTTP browsing is a core metric for benchmarking Quality of Experience (QoE). This applies even more when HTTP browsing is seen to be representing other applications retrieving web content as well. The techniques applied e.g. by using a banking application, a newspaper app or accessing social media via app are the same or at least very similar to web browsing.

In principle, a browsing or web content delivery session is structured into accessing a start or landing page like index.html and further retrieval of content linked in this landing page. The initial content is usually very small and often just a background colour or a frame.

The main information of this initially accessed page are all the URLs to further locations where the additional text and format information, pictures and animations are stored. On those next level HTML locations again, further links can be inserted. The visual web page appears usually in a multi-step process along with the received information from the accessed links. This is a process that works at the same time sub-sequently (level by level) as well as in parallel, where multiple threads are started to download the content of the various URLs.

In this context, testing HTTP browsing with a (mobile) browser can be seen as an archetype test for all applications downloading HTML content from multiple sources to the end-user devices, which is especially typical for most smartphone apps. Consequently, testing HTTP browsing is even more a core metric for benchmarking and provides information about web content delivery performance.

When testing HTTP browsing, several factors need to be considered. This includes success ratio and download time, which depend not only on the mobile network and the RAN but heavily on the website structure and its used syntax, as well as the connections to the Content Delivery Network (CDN).

Websites popular at time of publication of the present document are highly dynamic, which means that content and advertisements change within short periods. Therefore, it is highly recommended to include multiple different websites in benchmarking campaigns to diversify and average the sites' individual behaviors. This is to avoid focusing on results caused by an individual web site structure and position of, and connection to, its content. Typically, five to ten different websites are used in one benchmarking campaign. They should be continuously observed and may be replaced in case the applied rules are no longer met.

1 Scope

The present document focuses on Quality of Service (QoS) measurements for IP-based web content delivery services with reliable transport where primarily a visual presentation of the content on screen is available to the user. The presented test methodology is not limited to be applied to web browsing, or more general, to smartphone applications presenting web content visually on the screen. It can also be applied to all sorts of applications based on retrieving web conent. The present document does not target the qualitative evaluation of received content as video or audio.

The underlying test procedure consists of two phases: first request to resolve the initial URL and to contact the landing page, and second - if implemented - connect to embedded URLs to retrieve more and supplementary content. In the present document, web browsing serves as the default example, but the described QoS parameters can easily be applied to other IP-based web content delivery services based on reliable transport.

Furthermore, the present document also offers practical guidance for measurement execution and evaluation of HTTP/HTTPS content delivery QoS measurement.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the reference document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI TS 102 250-2: "Speech and multimedia Transmission Quality (STQ); QoS aspects for popular services in mobile networks; Part 2: Definition of Quality of Service parameters and their computation".

3 Definition of terms, symbols and abbreviations

3.1 Terms

Void.

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

CDN Content Delivery Network

DNS	Domain Name System
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
HTTPS	Secure HTTP
IP	Internet Protocol
OS	Operating System
QoE	Quality of Experience
QoS	Quality of Service
QUIC	Quick UDP Internet Connection
TCP	Transmission Control Protocol
TLS	Transport Layer Security
UDP	User Datagram Protocol
URL	Uniform Resource Locator

4 Quality of Service measurements for web content delivery techniques like web browsing

4.1 General aspects of web content delivery

4.1.1 Static and dynamic web pages

Web pages and other web content are usually not just a stored HTML file on a server that can be retrieved at once.

A static web page provides the same content to all users, and often the content is constant over time. From the implementation point of view, all content can be part of the landing page, for example index.html, but also fix linked resources such as images or (formatted) text in other linked HTML files can be part of the content. These resources can be located on the same physical server or on different ones. However, these links and contents are permanent and not subject to perpetual change and - importantly - all users get the same stored content.

Dynamic pages are characterized by personalized, regional or operator depending information or advertisements. This means that two users may not get exactly the same content. In addition, dynamic web pages are changing content on a short-term scale (e.g. perpetually updated content). Typical dynamic web pages are newspaper or online shop pages, where perpetually new content is made available and personalized advertisements, regional news or similar tailored information is presented.

For benchmarking, a mixture of many different web pages is recommended, where some static ones can be included. However, benchmarking should mainly be based on the most popular pages and those are dominantly dynamic.

Hosting web content is usually outsourced to specialized companies which take care of the accessibility of the content and the dynamic distribution to mirror servers close to the addressed users of the web page. As a consequence, the access of the same URL through networks of different operators may lead to the delivery of individual instances of this page. They can be hosted at different locations and can provide different (supplementary) content, like specialized advertisements. When using popular, dynamic hosted websites, the browsing test methodology measures performance as perceived by a user of this operator is experiencing the access of this page. Furthermore, the (supplementary) content can change from access to access.

In best practice, selected dynamic web pages for testing should provide the same amount of data when called at the same time using different operators, even if the content is partially different.

4.1.2 Transport over HTTP

HTTP is specified for transport of HTML, and evolved over time. HTTP defines the commands or expressions for the browser's or an app's communication with the server. HTTP/1.0 was first released in mid of the 1990s and was already made obsolete by HTTP/1.1 released end of the 1990s which is still in use. Subsequently, HTTP/2 was introduced in 2015 and is supported by all of web browsers existing at time of publication of the present document. HTTP/2 defines more efficient semantics and uses techniques previously implemented by Google's™ proprietary TCP-based SPDY protocol extension.

The HTTP versions are backwards compatible, but newer releases enable an extended set of expressions. Even though HTTP focuses on the definition of expressions, there is also a link to the underlaying transport layers like TLS and lower layers. HTTP/1.x and HTTP/2 are based on TCP.

Using TCP as transport protocol is the main difference to the latest proposal HTTP/3 which is supporting UDP/QUIC for an even more efficient transmission. First use of HTTP/3 in mobile browsing can be observed at time of publication of the present document but can be expected becoming more frequently used in following months and years.

4.1.3 Secured transmission

Plain, unsecured HTTP uses port 80 of web devices and was used for almost all exchange of web content in the past. At time of publication of the present document, unsecured transmission is hardly used anymore, as the vast majority of web pages uses securing mechanisms such as TLS in combination with HTTP/1.1 or onwards. This tendency to secure transport is a general trend, it also applies to video streaming, communication and every other sort of internet transport. Even when accessing URLs by unsecured HTTP, there usually is a redirection to the HTTPS instance of this page. Instead of port 80 as for HTTP, the secured HTTPS uses port 443.

Using HTTPS limits the possibilities for analysis on packet layer, as the content is encrypted by a previously exchanged symmetric key. Thus, many of the IP based and content-aware KPIs as in ETSI TS 102 250-2 [i.1] are not applicable under HTTPS. As the message flow between browser and server cannot be read in detail, the present document focuses on performance metrics obtained at upper layers and closer to the user's perception.

4.2 Phases and structure of web content retrieval

Common for all apps focusing on web content retrieval such as web browsing is the initial access to a URL leading to a landing page. After this initial IP access, the vast majority of IP-based web content retrievals is comprised of several stages where embedded links in HTML files are opened to retrieve the content in multiple and parallel stages. It can be imagined as sub-sequent downloads where in each stage multiple new links are opened and content is retrieved in parallel.

Figure 1 illustrates the sub-sequent and parallel delivery of web content to a requesting app or a browser.

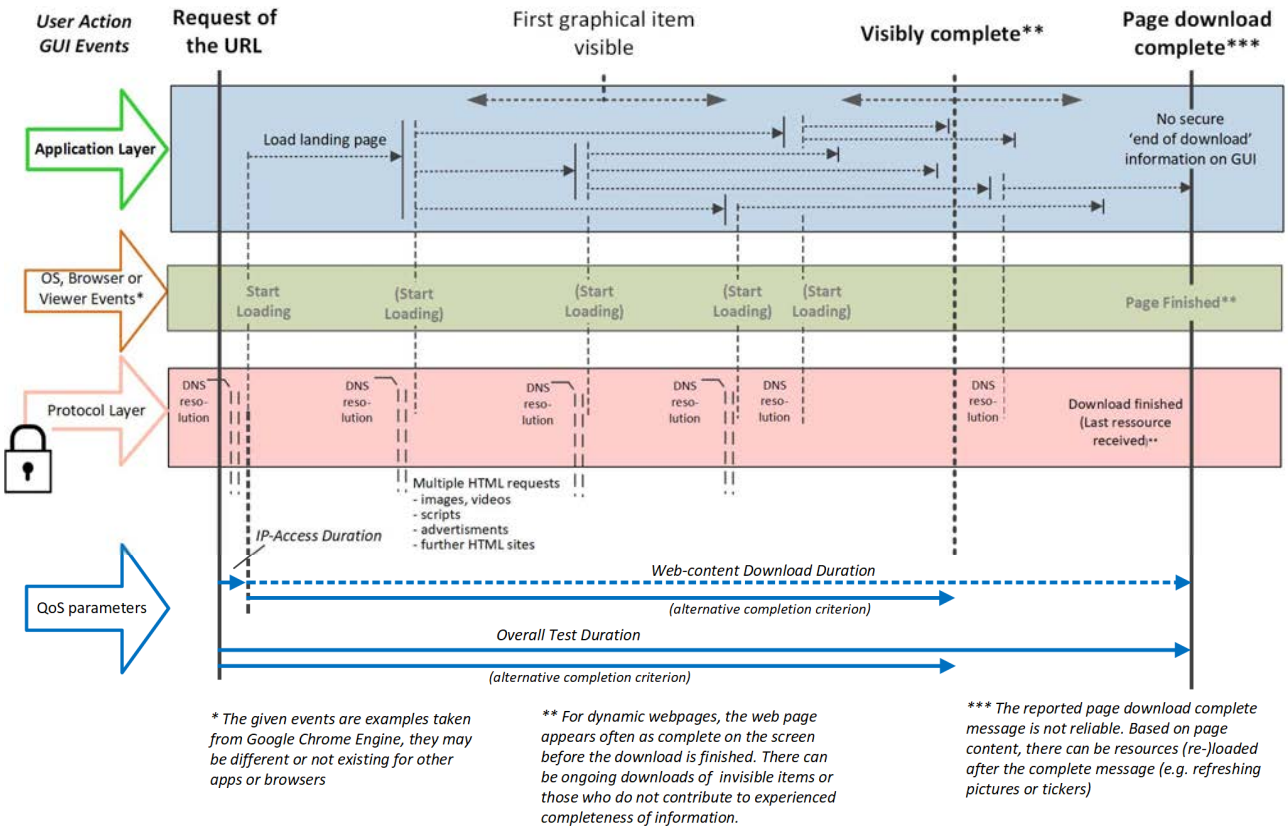

Figure 1: Typical phases of IP-based web content delivery

Even though HTTP and HTML are specified languages and protocols, the transported payload is usually secured and not accessible while browsing or downloading web content by an application and can barely be used for analysis or reliable definition of trigger points. Also, all web browsers and apps can interpret HTTP and HTML correctly, however there are many individual implementations regarding priority management when downloading and parallelizing tasks and rendering information. Neither 3rd party web browsers nor apps provide reliable and specified information about the download progress or potential errors. In addition, more detailed progress information would require a special, non-obligatory design of the websites to provide information about completeness and/or size information to the browser.

Measuring performance of web content delivery is therefore restricted to the few events visible, observation of transport layers to record the amount of transferred data, and even a mere assumption about completeness of transfer.

4.3 QoS aspects of web content delivery

4.3.1 Scope of aspects

Testing function and performance of web content delivery as in browsing requires the definition of objectively measurable impairments, such as:

- Failures to start.
- Failures to download requested content.
- Failures to download within a defined time-out period.

NOTE: Web pages and other web content usually consists of multiple individual files to be delivered from different data sources. The amount of data and individual components forming a website or a complete delivery is highly individual and has strong influence on the test results.

4.3.2 Initial IP Access

After a user requests a URL through the browser or an app, this initial URL is resolved by the DNS and the underlying IP address is successfully contacted.

4.3.3 Landing page access and linked resources

When contacting the hosting server by HTTP or HTTPS the incoming request is directed to the landing page, often named 'index.html'. Depending on implementation, the landing page is usually just used for initial information exchange and/or some basic content and mainly provides links to additional content. Specifically, dynamic commercial websites use HTML to request basic information from the calling browser on Operating System (OS), browser or app version information as well as installed capabilities such as fonts and screen size. This is to apply a representation of the web content matching with the capabilities of the receiving device.

There is neither reliable information provided by the browser nor a guarantee that first content is visible on the screen that can be used for detecting the complete download of the landing page. However, in today's realization of HTML, one or more DNS requests after the initial one can be used as an indication that the landing page is downloaded completely and embedded links to further resources have been requested.

4.3.4 Downloading web content

In principle, images and other objects can be directly embedded in HTML, even directly in the landing page, but almost all popular websites are far more complex.

Commonly, objects are stored somewhere as files and the HTML file contains only the link to these resources. The web content is - especially for dynamic, commercial web pages - distributed into many individual files. Those are not only images or graphical objects, but rather HTML files possibly containing further links. The content not only consists of objects becoming visible when rendered by the browser, but also of scripts and other invisible objects. Complex, dynamic websites of. When retrieving complex, dynamic websites, often > 100 individual resources are belonging to the site and are downloaded individually.

This granular design makes it very easy for the owner of a website to exchange content and individual objects such as images or graphics. Furthermore, it enables the host of the website to distribute and store the individual resources on different locations optimized for regional access or similar.

Technically, the retrieval of such interlaced content leads to a multiplication of IP connections opened and sources of content. The content is downloaded by many individual threads from different locations and is composed to a graphical presentation after retrieval. However, there are no reliable rules regarding at which point of reception the graphical interface is populated with information. Except for very small web pages, it is almost never the case that the content appears at once on the screen when the download is complete.

It is common practice that at a certain stage the main information is made visible on the screen and step by step complementary information is added. Often, the page appears visibly complete, but there is still data downloaded in the background that belongs to the website.

To calculate QoS metrics, mainly the observation of the data transported in the individual threads can be used while delivering web content to the device.

4.3.5 Completion criterion based on complete download

- Page downloaded event by browser or viewer component.
- Tracing all connections to resources during the download period:
  - Separation from background traffic.
  - Perpetual reloads of content.

4.3.6 Completion criterion based on amount of downloaded data

- Assumption of 'visible complete' after a pre-defined of retrieved content (data):
  - Focus on most relevant phase of download (initial access, ramp-up, multiple parallel downloads).
  - No test result dependency on size of web page.
  - (Much) higher efficiency of test campaigns.

4.4 QoS parameters for web content delivery

4.4.1 Parameter and trigger points

In this clause, a set of QoS parameters based on the streaming QoS parameters as defined in ETSI TS 102 250-2 [i.1] is proposed for measuring HTTP browsing.

Table 1 provides an overview of the proposed QoS parameters and a mapping of these parameters to the phases introduced in clause 4.3. Furthermore, a parameter type is assigned for each QoS parameter in order to determine the calculation method to be used for the respective parameter.

Table 1: Overview of QoS parameters and mapping to typical phases of the web content delivery

Related Phase(s)	QoS parameter name	QoS parameter type
(initial) IP Access	IP Access Failure Ratio	Failure Ratio
(initial) IP Access	IP Access Duration	Duration
Content Delivery	Content Delivery Cut-off Ratio	Cut-off Ratio
Content Delivery	Content Delivery Duration	Duration
(initial) IP Access, Content Delivery	Overall Web Delivery Failure Ratio	Failure Ratio
(initial) IP Access, Content Delivery	Overall Web Delivery Duration	Duration

Within ETSI TS 102 250-2 [i.1] the following QoS parameter types are defined:

- Duration.
- Cut-off Ratio.
- Failure Ratio.

The type "Duration" is assigned to QoS parameters where the QoS parameter represents an expected or an actual time period between a start trigger point and a stop trigger point, both observed during a single measurement. The following equation defines the abstract equation to be used to calculate such a parameter:

$$\text{Duration}[s] = \left(t_{\text{stop trigger}} - t_{\text{start trigger}}\right)[s]$$

The type "{Failure | Cut-off} Ratio" is assigned to QoS parameters representing a failure or Cut-off ratio. The following equation defines the abstract equation to be used to calculate such a QoS parameter. Here, the term "unsuccessful attempt" should be understood in the way that, during a single measurement, the stop trigger point of the QoS parameter has not been observed within a given time after having observed the respective start trigger point:

$$\{\text{Failure}|\text{Cut-off}\}\text{Ratio}[\%] = \frac{\text{unsuccessful attempts}}{\text{all attempts}} \times 100$$

Table 2 gives an overview of the trigger points used for the QoS parameter definition. For each trigger point, an ID is introduced. This ID will later be used as a reference within the QoS parameter definitions.

Table 2: Overview of the trigger points used for the QoS parameter definition

Trigger ID	Abstract description	Technical description
tr-1	Request of the web page	The corresponding event in the app or browser
tr-2*	App or device starts to receive payload	First Byte of the landing page (in case of unencrypted transmission) or establishment of the TLS or QUIC connection to the delivering server
tr-3a**	Pre-defined amount of data retrieved	Passing a pre-defined threshold of downloaded data related to the approached web-site
tr-3	Web page download completed	Download complete reported by app or browser, all content assigned to the web page received
NOTE 1: (*) First received Byte of payload cannot be identified in case of secured transmission of content. Alternatively, the successful establishment of the TLS or QUIC (and first received packet with payload) connection can be used. NOTE 2: (**) This trigger point can be assigned as an alternative 'page download complete' criterion.		

The test is conducted from the viewpoint of a user who anticipates an existing internet connection, i.e. the web page request is made in any case. If no internet connection exists at this point in time, a valid tr1 is recorded and tr2 is not reached.

From these observable events in connection with the observable phases, the following QoS parameters can be identified.

4.4.2 IP Access Failure Ratio [%]

Table 3

QoS parameter description	Start trigger ID	Stop trigger ID
The overall failure ratio of the IP access to the hosting web server.	tr-1	tr-2

4.4.3 IP Access Duration [s]

Table 4

QoS parameter description	Start trigger ID	Stop trigger ID
The time it took to successfully connect to the web server.	tr-1	tr-2

4.4.4 Content Delivery Cut-Off Ratio [%]

Table 5

QoS parameter description	Start trigger ID	Stop trigger ID
The ratio of web content delivery that could not be completed.	tr-2	tr-3 alt. tr-3a

4.4.5 Content Delivery Duration [s]

Table 6

QoS parameter description	Start trigger ID	Stop trigger ID
The time it took to download the web content completely or to reach a pre-defined data threshold.	tr-2	tr-3 alt. tr-3a

4.4.6 Overall Web Content Delivery Test Failure Ratio [%]

Table 7

QoS parameter description	Start trigger ID	Stop trigger ID
The overall ratio of web content delivery attempts that failed to successfully reach the defined end trigger.	tr-1	tr-3 alt. tr-3a

4.4.7 Overall Web Content Delivery Test Duration [s]

Table 8

QoS parameter description	Start trigger ID	Stop trigger ID
The time it took from requesting the URL to reach the download completion criterion.	tr-1	tr-3 alt. tr-3a

4.5 Recommended supplementary information for web content delivery

4.5.1 Amount of data received

The amount of received data is a supplementary information in case the web content is downloaded completely. It specifies the total amount of data belonging to requested web source. It does not apply when using tr-3a as completion criterion.

Table 9

QoS parameter description	Start trigger ID	Stop trigger ID
The amount of data in Bytes downloaded until the web delivery is completed.	tr-2	tr-3

As additional information, the download process can be divided into individual IP-connections for further analysis. This would also give possibility to separate content belonging to the web content requested from the background traffic of the device. This specific analysis can require data analysis on IP-layer, as browsing or viewing components usually does not provide this information on application layer. Notice that there can be restrictions due to encryption at IP-layer.

4.6 Configuration aspects including timeout recommendations for web content delivery

4.6.1 Consideration of specific web browser and app implementations

Web-browsing by a human user on a smartphone is done by a mobile browser application. Those mobile browsers retrieve and render the content visually. However, individual browser implementations are processing the received data differently and with individual efficiency, they use or do not use natively supported viewer components and combine them with individually implemented functionality. Considering real mobile browsers includes the application performance itself into the test approach and the received results. It should be noted as well that mobile browser applications do not provide detailed information about the download progress. The visual feedback by e.g. progress bars is not seen as reliable in practice. However, the alternative 'page download complete' criterion by passing a pre-defined amount of downloaded data can be used for calculating the KPIs on IP-layer while running a real browser application.

In principle, this approach is also applicable for other applications than mobile web browsers. Instead of a web browser retrieving a web page, another mobile application retrieving web content (e.g. map services, newspaper apps and similar) can be started and a pre-defined amount of downloaded data used as 'task completion criterion'.

Focusing on testing performance of web-content delivery does not require dedicated mobile browser applications. In such scenario, generic viewing components like HTTP receiver and HTML interpreter, or viewer components provided by the OS of the device, are an appropriate alternative. These generic viewing components are often the underlying components for mobile browser applications but can provide intermediate information. Generic receiver components are also not bound on content for visualization, rather can retrieve other sorts of web content too.

Both approaches are applicable, the choice depends on the aim of the test. Using mobile browser applications restricts the analysis widely to the IP-layer but includes the individual performance of the selected browser. The use of generic viewing components can provide more intermediate insights and focuses more on network delivery performance, since the individual optimization of a browser does not influence the results.

4.6.2 Handling of privacy agreements, cookies, advertisements and cache

The cache of the browser application or the generic viewing components should be cleared before each measurement. This will delete the previously downloaded web content and prevent inadequate transfer times for sub-sequent measurements due to the use of previously downloaded and cached content. However, this cleaning step will also delete previously stored cookies.

Due to legal privacy constraints most websites or other retrieved content initially present overlay dialogues to agree to given privacy settings and acceptance of cookies. Those overlays can become active prior to download, which means that retrieving of the content might not start before there is agreement by the user (or a key-press emulation). More common for websites is to start the download in the background while the overlay dialogue stays in the foreground.

Besides other guidance to selecting websites, those sites should be chosen where the content retrieval is not paused by the mentioned overlay but continues in the background.

In case mobile browser applications are used to retrieve web content, there should not be an active ad-blocker. Active ad-blockers are recognized by many websites and there is no delivery of content in these cases.

4.6.3 Handling of non-accessible resources

A website is typically comprised of content stored on multiple locations. Such individual contents can be text and images but especially advertisements or other components like activity trackers and similar which are often located on remote servers.

In the best case, all these resources are available and reachable by the retrieving device. Potentially, redirections are taking place after resource requests. Once the HTTP frame containing the resource is downloaded completely, it is acknowledged with HTTP 200 OK. However, it can happen that especially remote resources like advertisements, not belonging to the content as such, are not accessible. This leads to the typical HTTP 4xx, like e.g. HTTP 404 (Not Found) and 5xx response messages for content delivery failures.

The 4xx and 5xx response messages are valid responses and should not cause a test failure criterion, since the focus is set on the performance of web content delivery, meaning the transport of available resources. However, reporting unavailable resources or unresponsive servers can provide deeper insights into potential problems of internet connectivity.

More problematic are unavailable resources without a proper return message, as in such case browser applications do internally apply predefined time-outs. However, these time-outs and long waiting times for individual non-available resources influence measures based on tr-3 according to clause 4.4. However, the alternative tr-3a as 'page complete' criterion avoids dependencies on unavailable resources.

4.6.4 Best practice website selection criteria

Selection of websites or general web content delivery sites in the CDN can impact on measured performance due to complexity and connectivity of the servers and content. To ensure a representative performance comparison the following information and advice should be considered:

- For sufficient diversity and robustness of results describing the network's performance, consideration of more than a single website is recommended to retain diversity. Exception is the dedicated end-to-end performance measurement of a single website.

- It is recommended to select popular sites according to their relevance to end customers. As an example for popular ranking of websites per country Alexa™ Internet, Inc (www.alexa.com) and similar services can be considered. If possible, pages should be selected from Top 50 list, where an extension of that range is justifiable if not enough suitable pages exist within the Top 50.

- Internationally popular pages and country dependent pages may be used in reasonable proportion (e.g. 10 pages - 4 international, 6 local).

- Pages should usually be dynamic and not static as is the state-of-the-art technology.

- In case the complete download of the web page(s) is used as end criterion (tr-3), the size of the web pages should be the same or at least comparable for all operators in a measurement campaign. It should be avoided that the size of the web page is different due to different advertisement content per operator.

- In case a fix amount of retrieved data is used as end criterion (tr-3a), the selected websites should exceed a minimum size larger than the defined data threshold (e.g. 1 Mbyte) for all operators in a measurement campaign. This is to cover the minimum amount of data in case the download of a predefined data amount is used as success criterion. The page size needs to be observed on a daily basis throughout the measurements. In case of severe size changes, a reaction may be needed (see clause 4.6.5).

- Ad blockers should not be used in case of real web browser apps on Smartphones, tablets or PCs.

- A web-page selection that is hosted pre-dominantly by one CDN should be avoided.

- Websites of services that are predominantly accessed via a dedicated app on a smartphone should not be selected. For example, Facebook™, YouTube™ and similar websites/services are typically not accessed via a mobile browser and should therefore not be used as websites for HTTP browsing tests in mobile benchmarking campaigns. However, those sites can be tested by using the dedicated app (e.g. Facebook™) by applying the same test procedures as described in the present document. However, the analysis might be limited to a completion criterion as defined by tr-3a. where the app has to retrieve a minimum of data to exceed the defined data threshold.

- No additional website should be selected that is a sub-page/site of another already selected website.

- No website should be selected where the content is legally suspicious or contains harming, racist or sexist content.

4.6.5 Monitoring websites during measurement campaigns

Live web pages may change content regularly. It applies not only to main content such e.g. articles in a newspaper, but also complementary and advertisement information. Depending on content (e.g. pictures and their size, animations or videos), the amount of data on a website can change significantly.

These changes in size have a dramatic influence on complete download of the website (tr-3). In such case, the page size has a direct influence on download duration and download success ratio. In case a fix amount of downloaded data is used as completion criterion, the influence of changing content is lesser, as long as the minimum amount of data to pass the defined data threshold is ensured.

To avoid issues such as inconsistent or even invalid results during a measurement campaign, it is recommended to monitor the used websites on a daily basis:

- End criterion 'Web page download completed' (tr-3): Check for size changes of the selected web-pages, i.e. compare page size (downloaded data) to page size in the history (e.g. the day before). In case the page sizes changes, it is recommended to check the following:

  - Is the new size still in the acceptable range and does not change the desired size distribution of the selected web-pages or the targeted size? If not, the web page should be changed to one that matches the history in its size according to the rules in clause 4.6.4.

  - Is the change of size comparable for all operators in a benchmarking campaign? It is necessary to avoid that by a change, the size of the web page becomes different for the individual operators (e.g. different advertisements). In case there are considerable differences in web page size among operators, even on the same day, the web page should be changed to one that matches the history in its size according to the rules in clause 4.6.3.

  - Is the change of size in an acceptable narrow range during a benchmarking campaign? In general, dynamic websites changing their size widely (e.g. one day 3 Mbyte, next day 7 Mbyte) should be changed, even if the change is the same for all operators. This is to avoid the risk that temporarily very large (or small) sizes of a single page have a considerable impact on the statistics of the campaign.

- End criterion 'Pre-defined amount of data retrieved' (tr-3a): Check for size changes of the selected web pages, to verify that the amount of downloaded data is still reaching the pre-defined amount of data for success. In practice, there no complete download should be signalized before the data threshold is passed. This proof has to be done for each operator in the measurement campaign, because size changes can be different for each individual operators (e.g. in case of different advertisements). In case a web page consists of less data than required for passing the pre-defined data threshold for one of the operators, a new web page has to be selected according to clause 4.6.3.

4.6.6 Best-practice timeouts for web content delivery tests

The following are best practice examples for the configuration of testing web content delivery based on trigger points and measures as introduced in clause 4.4:

- Completion criteria:
  - tr-3: Page complete.
  - tr-3a: 1 Mbyte downloaded.
  - There can be individually defined success criteria under consideration of additional minimum requirements. An example is a composite success criterion, where in addition to the download of a pre-defined amount of Bytes of data within a given period of time, a minimum bitrate during active transmission has to be reached.
- Best practice Overall Web Content Delivery Test Timeout:
  - tr-3: 15s.
  - tr-3a: 8s.

History

Document history		
V1.1.1	March 2022	Publication