ETSI TR 102 505 V1.3.1 (2012-11)

Technical Report

Speech and multimedia Transmission Quality (STQ); Development of a Reference Web page

Reference
RTR/STQ-00195m

Keywords
internet, quality

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 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is 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 (http://ipr.etsi.org).

Pursuant to the ETSI IPR Policy, no investigation, 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.

Foreword

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

Introduction

Web browsing is one of the most popular usages of the internet. Therefore, it is also an important target of QoS testing. While the basics of web browsing are simple (and assumed to be known for the purpose of the present document), proper performance testing of this service in mobile communications networks is not. There is a complex interplay between web content, web and fixed network infrastructure, air interface and end-user devices including actual web browser applications. If this interplay is not properly understood and taken care of, testing results may be meaningless at best and wrong at worst.

NOTE 1: In the following, the terms "http testing" and "web browsing" are used synonymously.

NOTE 2: For the purpose of the present document, we use the term "transport channel" to describe the entire path content has to pass from its origin (i.e. the http server) to its destination (the user's computer).

The goal of service testing is to get a quantitatively correct impression of the service usage experience from a typical user's point of view. For obvious reasons, neither third-party servers nor third-party web content should be used. Therefore, a specially assigned and well-controlled web server (reference server) will provide specially designed web content (reference web page) to facilitate the necessary control of environmental conditions and assurance of reproducibility.

It should be kept in mind that the scope of present work is usage of reference web pages in the context of QoS testing according to TS 102 250 [i.1]. Therefore, reference to QoS parameters implicitly refers to TS 102 250 [i.1].

The present document attempts to describe reference web page design in a generic way. In annexes A and B, examples are given. Annex A describes the workflow and parameters used to create the ETSI Copernicus. Annex B, likewise describes the background information used for creating Kepler.

1 Scope

The present document describes the way used by a task force within the STQ Mobile working group to create and validate the Copernicus and subsequently the Kepler and Kepler Smartphone reference web pages to be used for QoS testing of http services. This included acquisition of basic information on parameters of typical "real-world" web pages.

Whereas parts of the present document may appear to the reader as being a general guide, the present document should be understood primarily as a report on what was done.

It should be clearly mentioned that the issue at hand is not "exact science". We want to design a reference page having properties of a "typical" web page. However, such typical web pages follow trends, given by available design tools, design paradigms, and fashions. The limits in which web designers operate are also given by transport channel capabilities such as higher bandwidth or device capabilities to render multimedia content.

Given these facts, we should be aware that any concrete reference web page will need adjustment and reconsideration from time to time. QoS testing is about ranking. Given two systems under test having similar performance, choice of any particular reference web site may well determine ranking order, and open up room for discussions by people who do not like the results. Since this works in all directions, the conclusion suggested by the authors of the present document is to keep the qualitative nature of the subject well in mind and to refrain from taking the "winner/loser" aspect of benchmarking too serious.

2 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.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

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

2.1 Normative references

The following referenced documents are necessary for the application of the present document.

Not applicable.

2.2 Informative references

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 (all parts): "Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks".

NOTE: The Reference web sites worked out so far are to be found at http://docbox.etsi.org/STQ/Open/Kepler/.

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

object: single component of a web page, e.g. a jpg graphic

web page: overall subject of a single web site download, consisting of a main html document and a number of objects the page contains

3.2 Abbreviations

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

API	Application Programming Interface
CF	Compression Factor
CSS	Cascading Style Sheets
DSL	Digital Subscriber Line
DUN	Dial-Up Networking
IE	Internet Explorer
IP	Internet Protocol
LAN	Local Area Network
MSIE	Microsoft® Internet Explorer
OS	Operating System
QoS	Quality of Service
RAS	Remote Access Service
RSS	Really Simple Syndication
UA	User Agent

4 General aspects of reference web page design

4.1 Basic dimensions of typical web page content

4.1.1 General considerations

A reference web page is meant to represent "typical" content. The first and most important question is therefore: what does "typical" mean exactly? Technically, this question asks for the technical dimensions, or the parameter space, a web page has at all with respect to QoS testing.

According to the general framework and taxonomy defined in TS 102 250 [i.1], quality in data services usage is ultimately given - once the basic availability and sustainability has been secured - by the speed a user-initiated activity is performed, or, equivalently, by the time required to complete the activity, i.e. the download time, e.g. defined by SeT according to TS 102 250-2 [i.1] (we will however use the well-established general term "download time").

In packet data services, end-to-end speed is determined by two basic transport channel properties: throughput and packet delay (or roundtrip time). A transport channel able to deliver high throughput but having large delay may be inferior - from the customer's perspective - to another channel having lower throughput but smaller delay.

The degree of influence throughput and delay has done, however, strongly depend on content structure to be transferred. Two web pages having the same total size (sum of sizes for all objects that page is composed of) may have quite different loading speed, depending on the number of objects they contain. Also, since typical web browsers load content using multiple sockets in parallel, size distribution may have an impact on download time.

Therefore, the parameters or dimensions assumed to describe a web page are:

- Total size (sum of all objects the web page is composed of).
- Number of objects.
- Object size distribution.

Any web page can be described by a point in a space having these three dimensions. Graphically, a reference page would be represented as a point being in the - somehow defined - "centre" of the cloud of points describing a sufficiently large number of real-world web pages.

It should be noted here that many real-world web pages contain active content which is executed on the target computer. This may be simple visible (e.g. animated pictures or short movies) and audible effects (e.g. sound clips), but some pages also display dynamic content, i.e. news tickers or streaming video which constantly download new data from the web. Of course the concept of "download time" implies that there is a clear end to download. Therefore, it is assumed that any web page where "download time" is measured does either not perform content in aforementioned sense, or that a clear definition of "end of download" exists. By logic, treatment of dynamic content with a fixed end of download criterion does not add value to the task of reference web site, but may complicate things considerably and is therefore excluded from further considerations.

In the scope of the present document, multimedia content of static nature, such as sound clips or videos which are downloaded once, does not get special treatment: it is assumed that the related download dynamics are not significantly different to other content such as jpg pictures.

However, even if this approach proved useful and workable, it turned out that there are more aspects to be considered than just those described so far. The reason is that in the case of data transfer through mobile communication networks, the transport channel properties may become content-dependent. The impact of this fact is dealt with in the next clause.

4.1.2 Summary of typical parameters determined for Kepler

The following table shows the Copernicus parameters versus the proposed "Kepler" parameters determined from the survey described above and subsequent discussion and refinement steps in STQ Mobile.

Table 1

	Total size, bytes	% non-lossless compressible	Object count	Composition
Kepler	807 076 (Windows® size)	60	75	125 000 bytes in "flash video" surrogate, 3 jpg's of ~45 000 bytes each
Compare: Copernicus	200 000	57	38	Largest object: ~36 000 bytes, type jpg

4.1.3 Composition rule proposal applied for Kepler

The parameters total size, object count and percentage of "non-lossless compressible" allow a multitude of page designs. Apart from application of "common sense", there is no formal rule on how to distribute object sizes and in the author's opinion the effort to derive such rules may be unreasonably high, last but not least when considering the fact that web pages constantly evolve. However, some transparency on how the final page is composed appeared to be desirable. Therefore, the following rule set was proposed and positively discussed in the STQ MOBILE group:

- Apply the following rules for compressible and noncompressible objects separately.
- Define four size classes, each representing 25 % of the page's total size.
- Initially, use a "factor 3" object count scheme to populate each class with as many objects as to reach the total size in this class.
- To fine tune the desired object count, vary the number and individual size of objects in the "last" class.

- Example: Total size is 100 kBytes, target object count is 50:
  - 1 object of 25 kBytes.
  - 3 objects of ~8,3 kBytes.
  - 10 objects of 2,5 kBytes each.
  - The remaining objects (here: 36) share the "last" 25 kBytes which makes an average size per object of approximatively 690 bytes.

Remarks:

- Actual data shows that a typical site has quite a lot of very small objects (100 Bytes to 200 Bytes). Consider to apply above rule set to the last class again.
- Please note that the example if for 100 kBytes and actual sizes will have to be adjusted accordingly.

4.2 Accelerators: Basic considerations

In mobile communications, channel capacity is a precious resource. Therefore, the use of accelerators is widespread. While there are various types of accelerators, and their inner workings not publicly documented, there are some general principles that accelerators work on. Before going into details: Accelerators make reference web page design much more challenging and error-prone. A proper understanding of the differences between transport channels with and without acceleration is required. We will also show that with accelerators in the transport channel, the general concept of reference web page design has to be expanded; in particular, the server hosting the reference page needs to be taken into account too.

Given the fact that any acceleration can only be done by reducing the effective data volume to be transferred, acceleration can be basically achieved in two ways: either by lossless compression of data, or by reduction of data volume through removing information (lossy compression).

Basically there is not a real choice between these methods when it comes to efficiency of compression. Lossless compression works very well with data objects of "text" type (such as htm, js, css), but has a very small or even a negative effect on image-type objects such as gif or jpg which are already well compressed. For such objects, only lossy compression will give useful results.

NOTE: We assume that sound clips behave basically the same way as image-type objects. However no analysis has been made on sound clips so far.

Lossy compression may, however, affect quality. This will depend on the properties of the source object, and the way this object is presented on the target device. If a multi-mega pixel picture is displayed as a 10 cm x 10 cm image within a web site, a lot of compression can be made before there is a really visible effect on image quality. However, if the image is already optimized, any further reduction of object size will result in visible degradation of quality. According to our findings, typical web sites use quite well-optimized images, which is reasonable because they lead to less server traffic and provide a better download experience both of which are in the commercial interest of web site owners.

Actually, there are good reasons to discuss the question if, the concept of determining web site download "quality" just by download time, continues to make sense at all, even if this question will not be treated further in the present document.

Reference web site design has to take accelerator effects into account, otherwise comparison between networks using different types of accelerators (or no acceleration) would not be possible. If, for example, a reference web page has the right size, object count and object size distribution, but contains non-optimized images, measurements using this page would vastly exaggerate the effect of accelerators, i.e. give accelerated networks an unrealistic "measured" advantage.

Also, since in general the compression effect is different, for same-sized objects, in lossless and lossy compression, the fraction of text-type and image-type objects needs to be considered.

Therefore, the parameter space introduced in the preceding clause needs to be expanded by two additional dimensions:

- Fraction of text-type objects in relation to total size.
- Degree of optimization in source objects of image type.

While these considerations are necessary, they turned out not to be not sufficient in the presence of accelerators. There is one more aspect to take into consideration, which is the server hosting the reference content.

When a browser requests a web page, it may do so indicating its capability to support content compression (encoding). If the browser indicates that it supports encoding, and the server is also supporting this, certain object types such as htm (however probably not js or css) are transmitted compressed. If this is the case, any accelerator downstream in the transport chain will not be able to significantly reduce object size further. This means that if the server supports compression, the accelerator effect will be much smaller than otherwise.

In other words: it is not sufficient to focus on reference web pages for meaningful http tests; the server is part of the signal chain, and needs to be considered in test case design.

There is no such thing as an isolated "reference web page" which can just be placed on any server and then be expected to produce meaningful results.

Different methods of acceleration also are located in different places in the transport chain.

Lossy compression (quality reduction) of image-type takes place on the network infrastructure side and does not require any special software on the client/browser side.

Lossless compression may take place in three basically different ways:

- If the browser indicates support of compression and the server does not, the accelerator could compress objects (acting as a kind of proxy). This is, however, considered to be a rather weak form of acceleration since most servers already support compression.

- The accelerator can modify the original html page to download an applet which will then act as counterpart for accelerator-specific lossless compression. This raises the question of how the download of such an applet is to be treated, since its size may be non-negligible in relation to total data size per web page download. If every download is to be treated as an independent, complete entity, the applet would have to be downloaded every time. It is more realistic to assume that such an applet is only downloaded once and therefore would effectively disappear from the equation. This point requires additional discussion.

- In a typical real-world situation, a user would purchase and install a network-specific driver and front end package together with a wireless data card. This package may also contain a de-compression utility, acting basically the same way as an applet, downloaded through the accelerator, but does not create traffic.

In effect, the first approach is assumed to be unrealistic, while the second and third are considered to be equivalent, provided the applet download is taken out of the size consideration.

For completeness it should be mentioned that another acceleration method is optimization of htm text, e.g. be restructuring it in a more space-saving way or to remove non-functional content such as comments. It is assumed, however, that professional web pages are already optimized in that way and therefore this aspect can be neglected.

4.3 Selection of a web page sample set

The approach of how to select the sample set has been discussed in an ad-hoc working group within the ETSI STQ MOBILE group quite extensively, with the results put to discussion by the whole group. After some plausibility checks on available ranking lists created the impression that they are not trustworthy or reliable enough, the approach of manual selection was chosen. A web page appears to be candidate if it can reasonably be assumed that a large number of customers will know and use the respective site. Candidate branches included:

- Mobile network operators.

- Popular newspapers and journals.

- Government authorities and agencies (e.g. labour agency).

- Large holiday travel agencies, public transport, airlines, and airports.

- Car makers.

- Banks and insurance companies.

- Web sites of large companies with internet-related fields of work.
- Search engines and popular web sites such as Wikipedia®.
- Web shops and popular trade platforms (e.g. eBay®).

Result of above candidate identification leads to a candidate list, which then has to be screened for suitability. Pages containing dynamic content have to be removed from the list.

A concrete list of web pages will of course be different for each country. For the work being performed, German sites have been selected; it would be an interesting topic in its own right to compare, in the future, results for other countries. After having performed the first surveys, with a time of about 6 months between the first and the second one, it became, however, clear that the set of typical parameters underlies changes anyway which would make the search for precision below, say, 10 % an academic exercise anyway. We believe that there is a clear and, for the foreseeable time, unbroken trend towards larger total size and a reduction in object count. This trend is easily understandable and appears to stay unbroken for a longer period of time. Internet connections tend to become more broadband, so web designers create more sophisticated and feature-rich pages, which also have more active content.

4.4 Statistical and methodological considerations

If a reference web page is described by a number of parameters, each parameter value should represent a typical value for a sufficiently large set of real-world pages. There are a number of ways to determine the target values; we will discuss briefly the most common methods and explain the selection being made.

The criterion used to determine the appropriate method is assumed to be stability of result against changes in the set of values used to derive the typical value from. Addition or removal of a small fraction of values should not affect the typical value much. Most of the parameters describing a web page, as outline in the preceding text, can have extreme values in single web sites. Averaging of some type (linear average, root mean square, etc.) is quite sensitive to single extreme values and therefore appears to be not a good choice. In contrast, percentile-based methods are much more stable in the aforementioned sense. Therefore, we will select the typical values for object count and total size as the 50 % point of distribution of respective sample values, i.e. as the value where approximately half of the values taken from the basic set are below and above, respectively.

Web site content may change within relatively short periods of time, in particular for web sites of newspapers and similar sources. Therefore, downloads for comparison between accelerated and non-accelerated channels should be made as simultaneous as possible. It is assumed that it is sufficient to have a maximum time difference shift between download time windows of less than 10 minutes, which should be validated by analysis of respective timestamps. To increase security here, web site component object names can be compared for the downloads through different channels.

4.5 Validation of reference pages for acceleration effects

In order to validate that a reference page actually is representative, it needs to behave approximately the same way with accelerators as typical web sites. The full approach here would be to download each sample web page over the same transport channel with and without accelerator, and compare the download times to create an "acceleration factor". This approach is, however, de facto impossible or at least very difficult to achieve. It would require an "accelerator owner" providing a switch which enables or disables the accelerator for testing purposes, or to provide two otherwise transport channels, one of them with and the other one without acceleration.

Of course some tricks could be devised, such as using basically different transport channels (e.g. accelerated network vs. DSL) and performing some calibration-style numerical scaling. Assuming this works, there is still the issue that in packet-switched data transport, there is always the possibility that throughput is affected by some third party competing for the same resource. In summary, download time measurements are considered to be sensitive to overall transport channel characteristics to a degree which makes them unusable in practice.

The practical solution found here is rather simple. Instead of measuring download times, download sizes are measured. The temporal acceleration factor based on times:

$$A(t) = T1/T2$$

(where T1 is the download time in the non-accelerated transport channel and T2 is the download time in the accelerated transport channel) is therefore replaced by the size-related acceleration factor:

$$A(s) = S2/S1$$

It is assumed that these two values are sufficiently equivalent. This assumption is backed by comparing e.g. benchmarking results measured by P3 Solutions made in German 2G and 3G networks, to the values determined by equivalent measurements, as well as by various surveys on typical reference web sites performed by Focus Infocom™.

Therefore, the need to measure times is replaced by the need to measure data sizes, establishing practicability of measurements.

It should be mentioned that a size-related view, while useful for reference page design, may create expectations on the effect of accelerators which can not be observed in real life. The reason is that for web browsing, typical object sizes are small, which means that roundtrip time becomes the dominant factor in download time rather than throughput.

EXAMPLE: Figure 1 shows effective throughput for a RTT of 50 ms and two raw throughput values (TP1: 50 kbit/s, TP2: 200 kbit/s). Obviously for data object sizes below approx. 1 kByte, peak throughput does not really matter. This "critical size" where throughput becomes relevant scales linearly with RTT. If we compare these values to typical object sizes, we find that in today's web page designs approx. nearly 50 % of all objects have sizes of less than 1 kByte, and approx. 22 % of objects have sizes of less than 200 bytes.

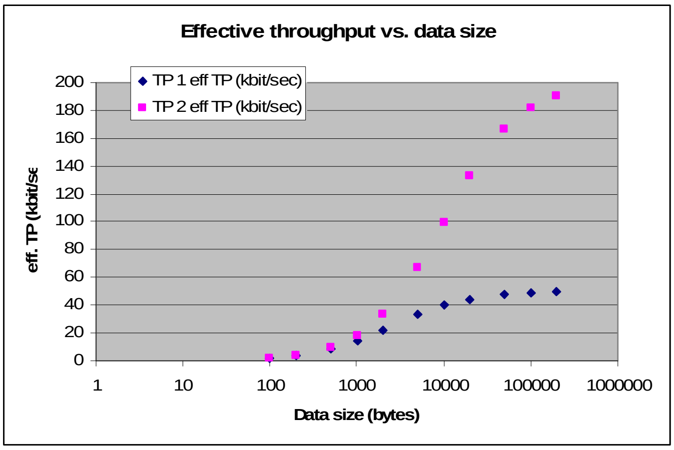

Figure 1

4.6 Impact of server-side compression on http test results

It has been mentioned before that the issue of reference web pages is not an "exact science" because the subject of our considerations constantly changes with web design "fashion" and the trend towards more multimedia in commercial web sites.

Actually server behaviour is the single most critical factor affecting test results in the presence of accelerators. For objects of type "image", further reduction of size can only be achieved by quality reduction; therefore the potential is limited. In practice, a reduction to about 70 % appears to be a realistic value. In contrast, lossless compression of text-type objects can reduce data volume by a factor of 5 to 10 for typical object sizes.

As we have seen in previous clauses, acceleration comes from two different sources: quality reduction and compression. The potential for quality reduction appears to be limited for obvious reasons. Therefore, the "substance" accelerators work on will foremost be lossless compression.

To highlight this issue let us make an estimation based on the following assumptions. A web page has 50 % text-type content which is compressible by a factor of 10; image-type content may be quality-reduced to 70 % of original size.

The overall acceleration factor is therefore 2,5. There is not much more to gain from more compression of text objects; even an infinite compression will only raise the acceleration factor to approximately 2,86.

If we now assume that the web server already uses lossless compression, the potential is drastically reduced. Given the compression factor above, text-type objects effectively disappear from the equation; overall acceleration practically depends only on the amount of quality reduction. Using the numbers from above, acceleration saturates at approximately 1,43.

The general insight can be summarized as followed: Assuming that typical servers support compression, effective sizes of text-type objects are negligible, any acceleration factor is strongly dominated by the degree of image quality reduction.

In this context also the question of different treatment of certain text-type objects needs to be mentioned. Tests have shown that some servers supporting lossless compression will only compress objects of type htm (i.e. the main page), but not other text-type objects, e.g. js or css; other servers also compress these objects. This issue needs further consideration. At present, it is assumed that all text-type objects are equally compressed. Consequences of changes to this assumptions would require re-design of the Copernicus web page since the current page appears to have an untypically small css. Also, more insight to the decision mechanisms in web servers with respect to selection of encoding would be required.

4.7 Handling precautions for reference web site file sets

To set up a reference web server typically involves handling the file set comprising the reference web site. It is critical to understand and observe the following precautions; otherwise, erroneous measurements may result.

In the following, it is assumed that the reference web site is represented by an archive containing all relevant files.

Please note that all the text files included in the archive are in a format conforming the Linux® operating system. Thus, it might occur that the size and the content of the related files will change when copying these files between operating systems using for example a FTP client application.

This is due to the fact that, depending on the application used, the sequence of control codes included in the text files will get adapted to the target operating system without further notification of the user.

Typically, the control codes indicating the end of a line will get adapted. For Microsoft® Windows operating systems, the related control code sequence consists of two control codes, namely the 'Carriage Return' character and the 'Line Feed' character. For Unix and its derivate operating systems as well as for Mac operating systems, the related control sequence consists only of a single character, namely the 'Line Feed' character (Unix) or the 'Carriage Return' character (Mac OS), respectively.

In order to avoid measuring against a modified Copernicus web page, great care should be taken to ensure that such replacement does not occur. For example, if files are to be transferred utilizing a FTP client, it is recommended to use the binary transfer mode.

5 Measurement of web page characteristics

After having defined which parameters are to be used, and after having determined a candidate set of sample web pages, the question arises how the desired parameters should be measured.

While all parameters can be determined using IP trace, this method involves either considerable programming effort, or a nontrivial amount of hand work. It showed that there is a workable alternative, namely integrating a standard web browser into a software application running tests automatically based on a simple list of web site URL. We selected the Microsoft® Internet Explorer (MSIE) which creates the information required on its API.

These values have been validated against IP traces taken with Ethereal®. It turned out that web page structure and size assessment need to be based on information output by MSIE during download. Approaches have been heard of to determine web site structure from a post-download cache analysis. It was found that MSIE alters the page structure when storing a downloaded page so this kind of analysis would produce incorrect results.

Some care has to be taken regarding MSIE settings. Cache usage needs to be disabled to make sure that each download is actually happening via the Internet. Settings affecting cookie, Pop-up window and Java Script treatment need to be controlled, i.e. they need to be recorded as part of the measurement documentation and it be made sure that they are not altered within a measurement run for the same sample web page set.

5.1 Measurement of object count

MSIE creates a message through its API for each object being downloaded, which can be logged. Total object count for each web page is therefore simply determined by log evaluation and counting these objects.

5.2 Measurement of total size

It turned out that with some network-specific accelerators, objects sizes of lossless-compressible objects are reported incorrectly by MSIE, the values neither represent the original nor the actually transferred size. Also, for gzip-compressed objects MSIE reports the uncompressed size rather than the actual (transferred) size.

Another method of determining size had to be found. As mentioned before IP trace evaluation should be avoided if possible because it is considered not sufficiently automatable. The solution was to use Windows® performance monitoring information which is available through a simple API and provides the necessary information both for RAS connections and for LAN/WLAN.

5.3 Measurement of object size distribution

This information can be created easily by evaluation of the web page object logs mentioned above.

5.4 Measurement of web page composition

Basically the evaluation can also be made from the web page objects log, by using the object's file extension to categorize into text or image type. With reference to clause 4.6, we decided to treat objects of type css and js as text-type objects having the same basic compressibility as htm objects.

5.5 Measurement of image optimization

In principle, the degree of size reduction for each image-type object can be determined by comparing the sizes of each image, as provided by the web element log, for transfer through the accelerated vs. non-accelerated transport channel. Tests have shown, however, that even within the same web page size reduction for image-type objects is not uniform and does not simply correlate with object size. The approach chosen is to assume an average size reduction and create the reference page's image objects by iteratively running some original pictures through web optimizers which are part of many web page authoring software packages.

6 Work flow description

In the previous clause, the necessary framework has been established, and the parameters to be taken into account have been identified. This clause will deal with the general work flow to create and validate a reference web site according to the following simple step by step schematic:

1) Select a set of sample web pages using the rules outlined in clause 4.3.

2) Run downloads through a non-accelerated transport channel, and through one or more representative accelerated transport channels. Eventually, validate the assumption that typical commercial web pages are hosted on servers supporting lossless compression, by running IP trace evaluation.

3) Evaluate the static web page parameters described in clause 4.1 and the fraction of text-type objects described in clause 4.2.

4) Create a "step zero" reference site according to the values determined. For images, use original images pre-optimized by a ratio which should come from experience; a value of 15 % pre-compression (relative to original size for a typical Megapixel camera) has shown to be a good starting point.

5) Determine acceleration-factor computations on the measurement data achieved in step 2, and for the draft reference web site. Select the reference server's support of lossless compression according to the result from step 2: if typical servers support lossless compression, the reference server needs to support this too and vice versa.

6) Adjust the reference web site's design by modifying the image-type objects optimization 8 and eventually re-adjusting the other parameters to keep them in the typical range until the compression factor is in the band determined to be typical.

Annex A: Example workflows used in creation of the Copernicus reference web page

Starting point for the Copernicus reference web page was the former ETSI Galileo web page.

A.1 Determination of parameters for the Copernicus web page

A.1.1 Selection of samples

Web sites chosen are from the following market segments:

- Banks.
- Insurance companies.
- Car makers.
- Travel agencies, airlines, airports, railway companies, car rental companies.
- Newspapers and journals.
- Online shops.
- TV stations.
- Mobile network operators and service providers.

From a total of approximately 190 sites, those having dynamic content were removed, resulting in a set of 167 web sites having the necessary basic properties.

A.1.2 Determining typical parameters

Sample pages were downloaded using DSL as a non-accelerated transport channel and using Vodafone® D2 network as a known network using a rather aggressive acceleration with respect to image size and quality reduction.

From the measurement data set, further pages were removed due to findings that these sites are highly dynamic, leaving a set of 155 pages.

A.1.2.1 Compression Factors

As reported, it has been found that object sizes reported by MSIE are not reliable. This fact had not been known at the time of previous surveys. In order to establish numerical relationship to these previous surveys, figure A.1 show results for both methods.

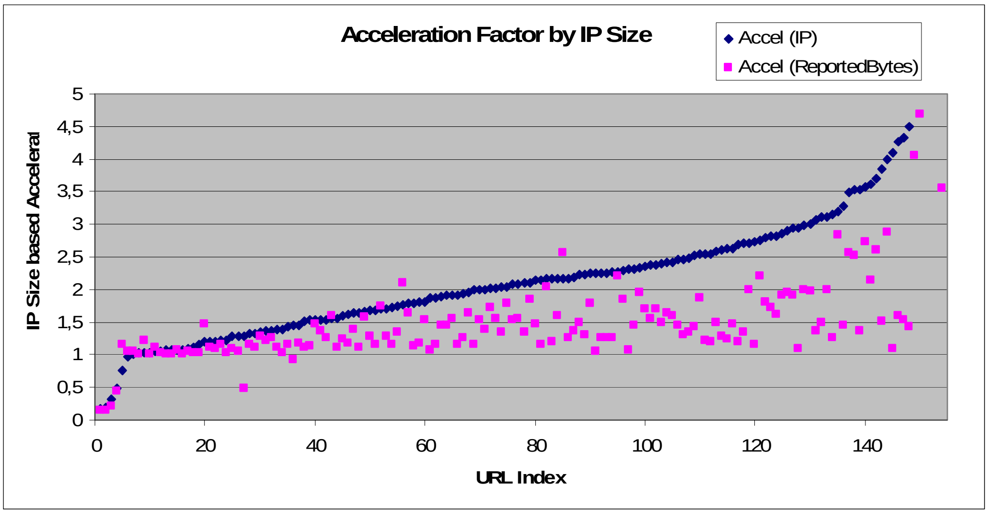

Figure A.1

Figure A.1 shows, in curve "Accel (IP)", the sorted result for acceleration factor determined from IP level (RAS or LAN adapter, respectively) with respect to above (RAS vs DUN). The 50 % point is approximately at a value of 2. The second curve, "Accel (Reported Bytes)" shows that the previously used acceleration factor is regularly lower, which is in line with expectations. Also interesting is the fact that reported size and IP size are not fully coupled. It is interesting to note that acceleration factors for the two methods are not fully synchronous. While there is a rather clear relationship between both value spaces, values for one and the same site may be quite different.

Figure A.2 shows, for clarity, the line-up for acceleration factor based on MSIE-reported sizes. The 50 % point is at 1,34 here.

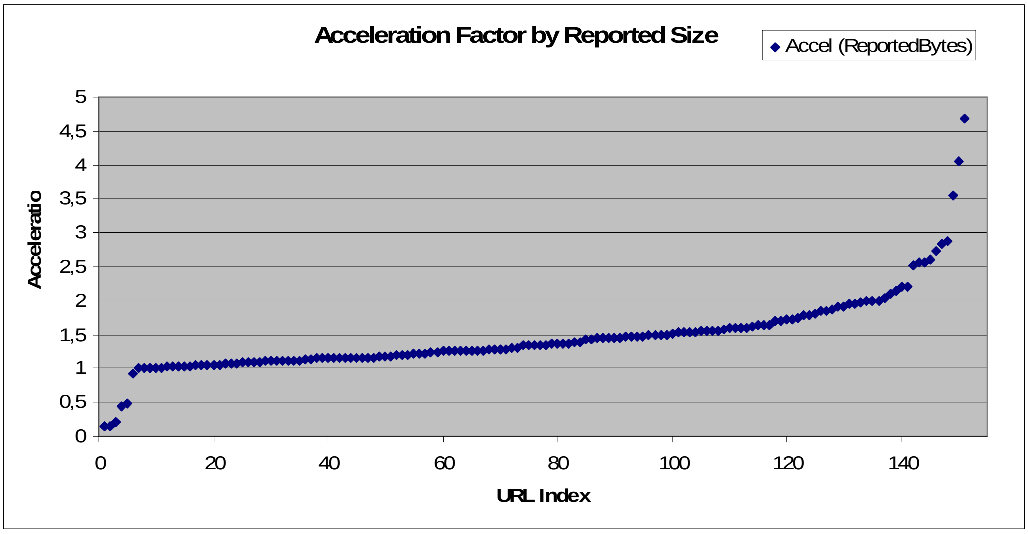

Figure A.2

As mentioned, acceleration factor values for the same web site are not fully correlated. Figure A.3 highlights these findings further by showing correlation between MSIE-reported sizes and IP sizes.

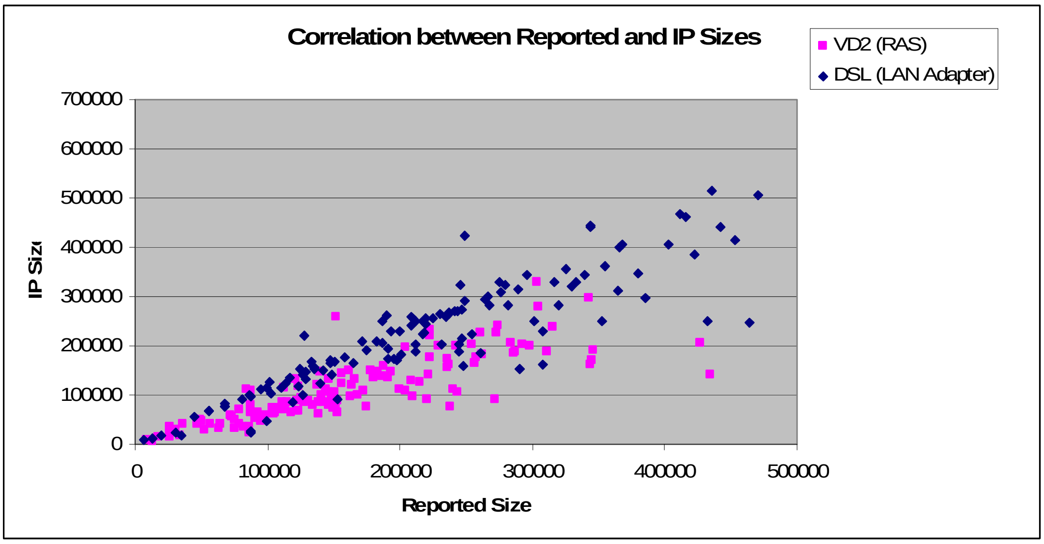

Figure A.3

It appears that the correlation, while generally quite "noisy", is weaker for accelerated downloads than for non-accelerated ones.

For comparison with previous measurements. Figure A.4 show results from a survey which was conducted using MSIE-reported sizes. Also, these older data used the "Compression Factor" (CF) which is simply the inverse of the acceleration factor. The value of 1,34 mentioned above computes to a CF of 74 %, slightly higher than for previous surveys which is, however, assumed to be within the expected error margin.

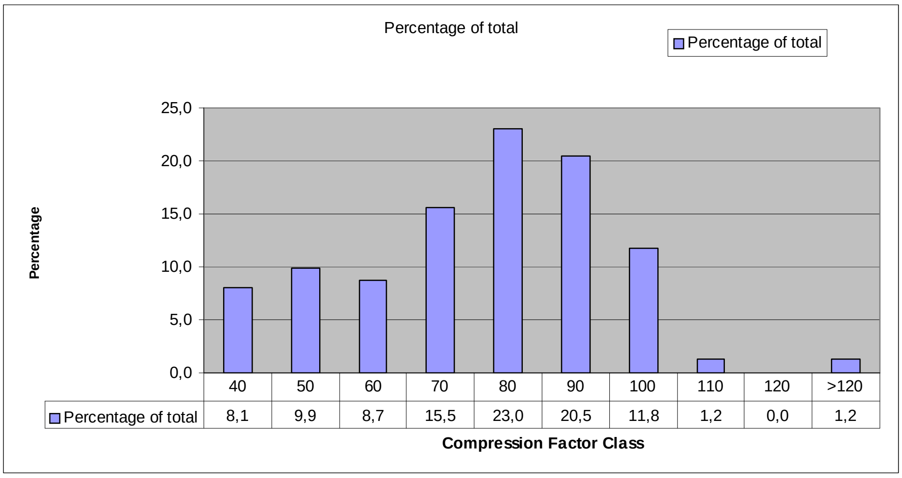

Figure A.4

Compression factor is shown in % here. In the example shown here, the accelerator-related applet was taken into account which may lead to compression-factor values exceeding 100 %.

Figures A.5 and A.6 show compression factors for the whole set of samples, and the reduced set with five extremes removed at either side. It can clearly be seen that the 50 % point does not change significantly in the latter case. Therefore it can be concluded that the set is representative.

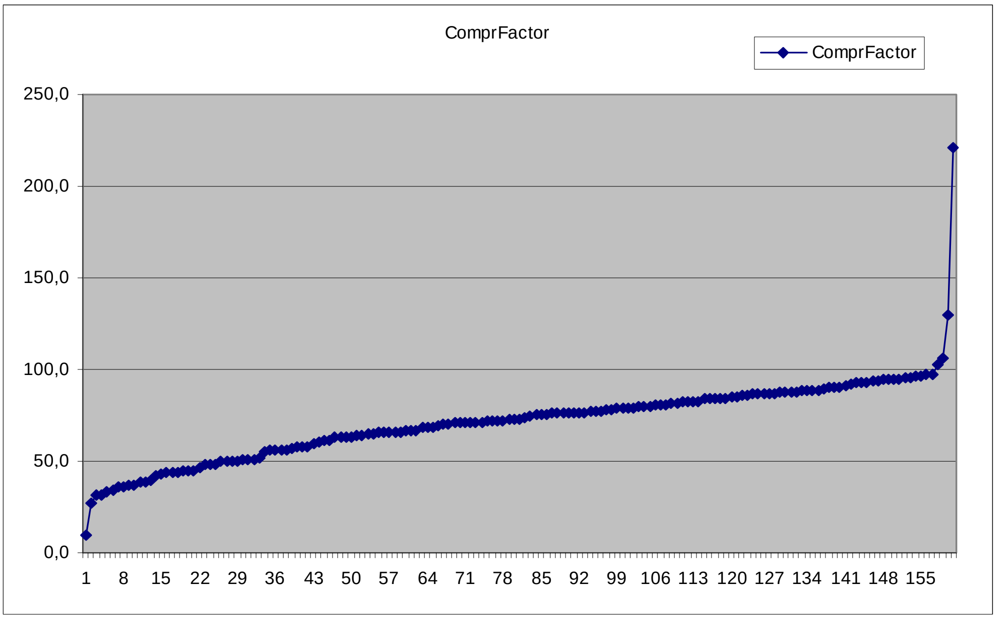

Figure A.5

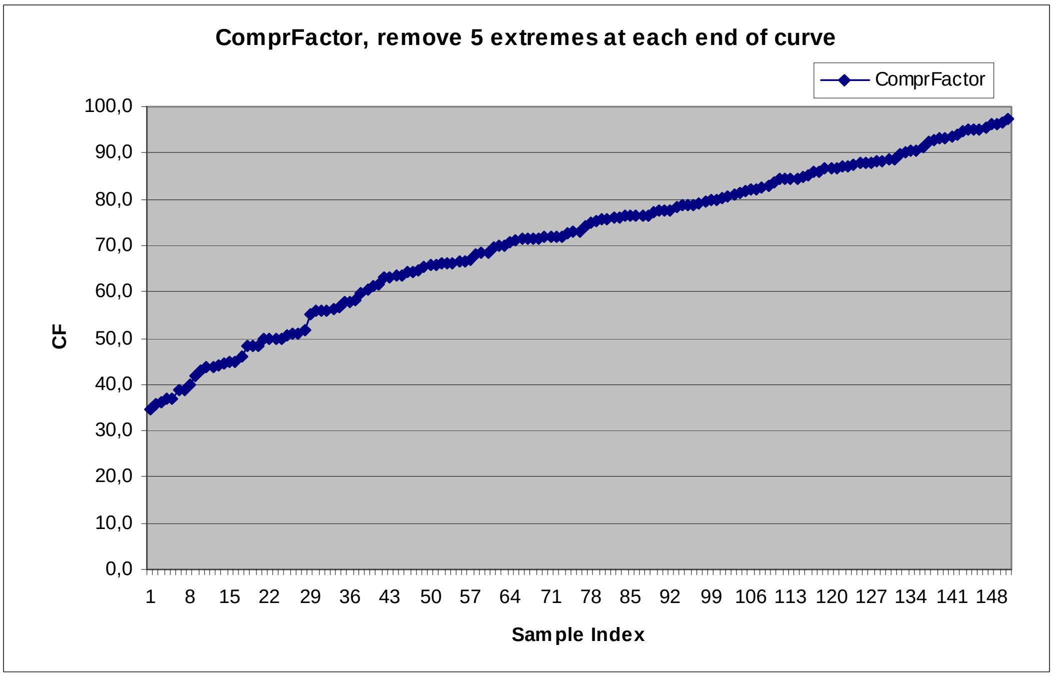

Figure A.6

A.1.2.2 Object count and Size

Figures A.7 and A.8 are done in order to check for point density distribution. It shows the expected even distribution with a "spread" towards higher sizes where naturally more room for variation of site structure can be expected.

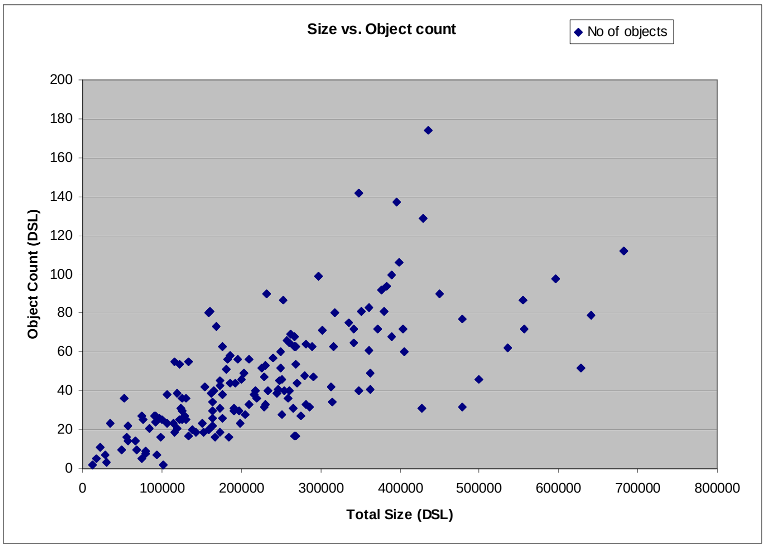

Figure A.7

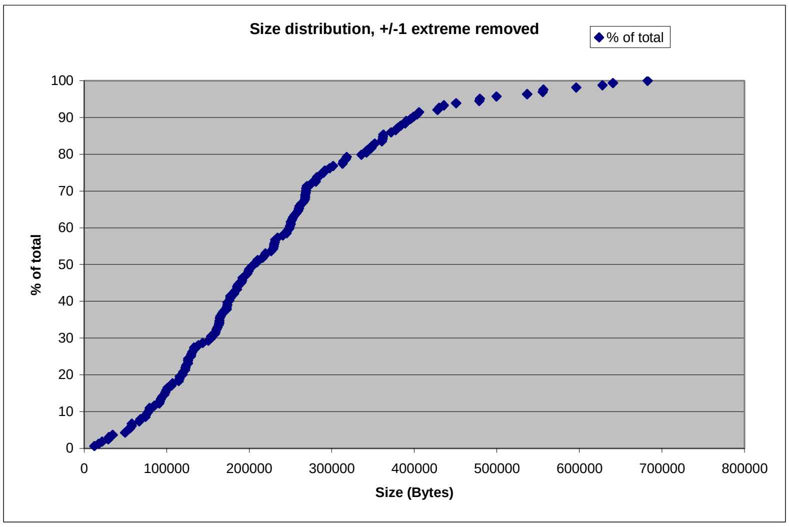

Figure A.8

The "50 %" point for total size is at ~205 000 bytes. For figure A.9, the lower and upper extreme value (at approximately 7 Kbytes and approximately 1,1 Mbytes) have already been removed.

Comparison to the "typical size" of approximately 160 Kbytes (from 50 % distribution) from the February/March 2005 survey using nearly the same URL list clearly shows a trend towards "heavier" sites. Assuming that commercial web site operators have an interest in reasonable download times, this points to reliance on higher available throughputs; it is likely that this trend will continue.

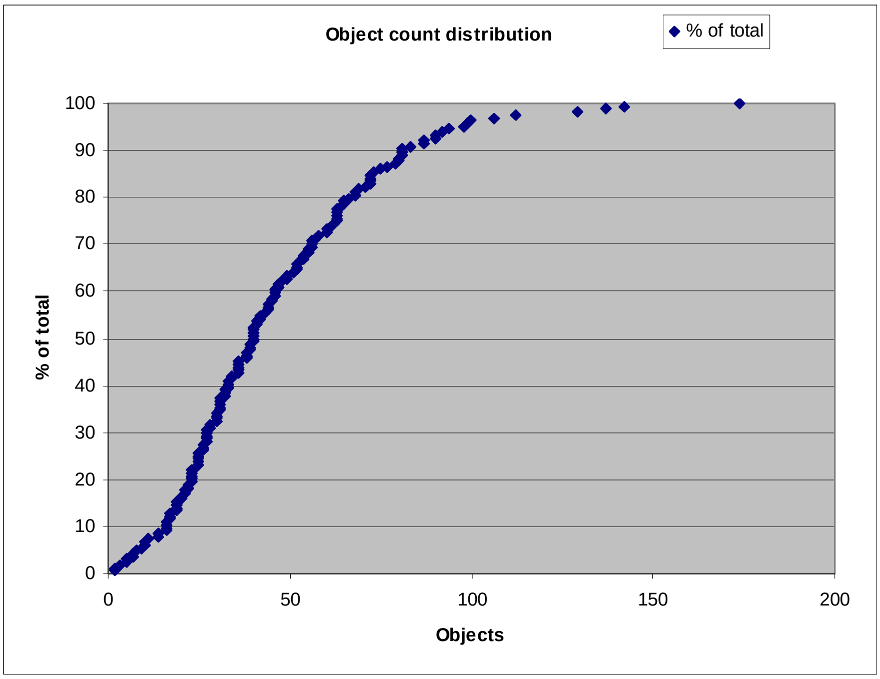

Figure A.9

Comparing to the February/March 2005 results, it appears that object count has slightly decreased (from 45 objects/50 % point to 40 objects). However, distribution is so broad that it is difficult to mark a clear trend here.

A.1.2.3 Object types

Figure A.10 shows distribution of object types for the sample set. Figure A.10 shows raw types from primary data object evaluation.

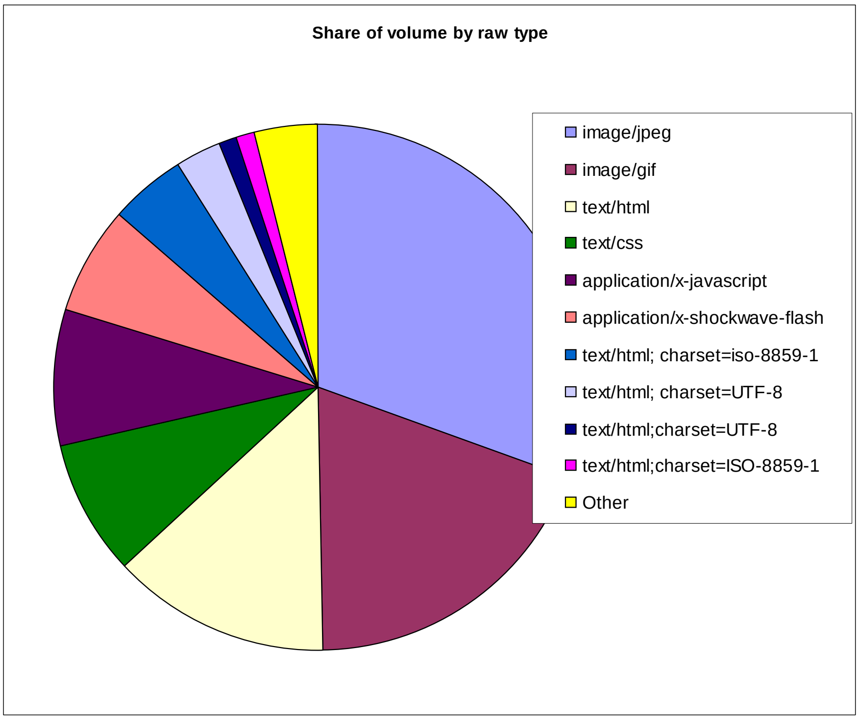

Figure A.10

Only a handful of raw-category types dominate (5 types make up 80 %, 10 types make up 96 % of all volume).

Aggregation of similar types yield the refined result shown in figure A.11.

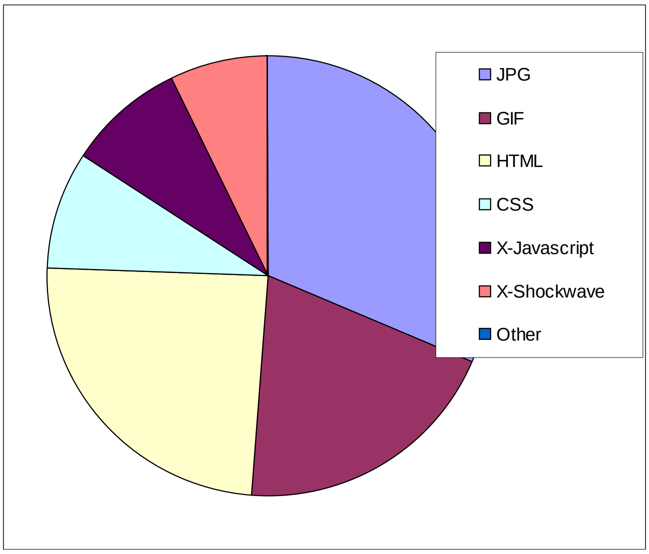

Figure A.11

Type Category	Total Counts	Total Size	Percent of total	Size share of major type
Total	7 854	35 454 800	100,0	
JPG	1 554	10 932 512	30,8	98,8
GIF	4 600	6 825 366	19,3	99,7
HTML	423	8 460 233	23,9	56,5
CSS	354	3 026 321	8,5	97,1
X-JavaScript	598	2 944 423	8,3	99,7
X-Shockwave	112	2 535 038	7,2	92,7
Other			2,1	

Objects of type HTML, CSS and JavaScript are considered to be of text type. They make up.

A.1.2.4 Object size distribution by type

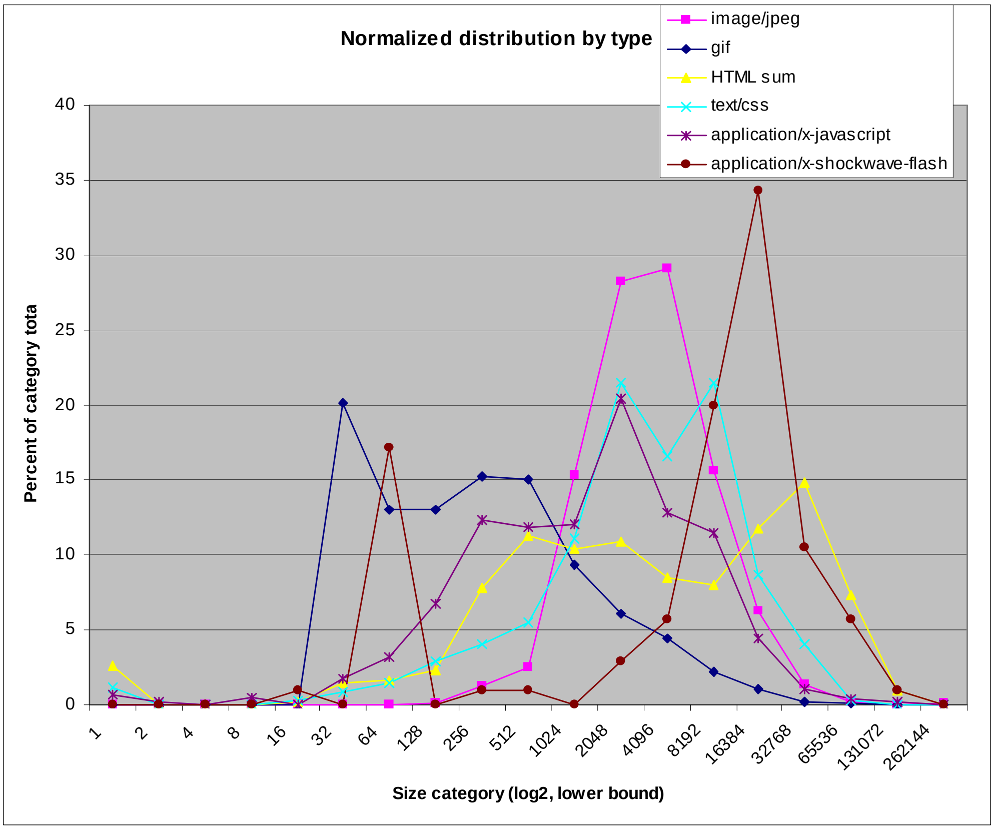

Figure A.12

A.1.2.5 Compression factor for text-type objects

These tests were made using the gzip application available in the web.

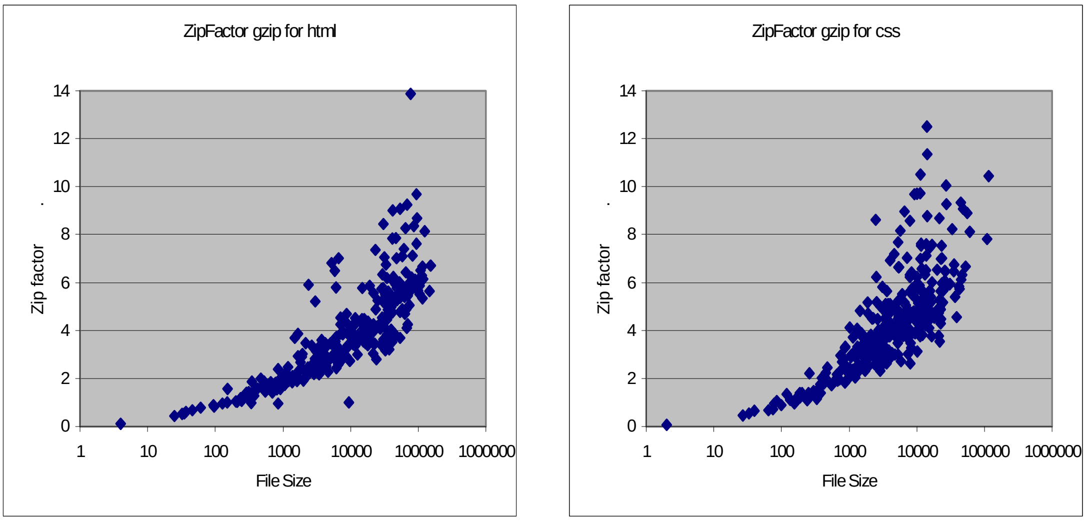

Figure A.13

It appears that css files are even a bit better compressible than htm objects. In any case overall compressibility for typical sizes (see previous clause) is in a range where small variations do not matter much anyway. With servers supporting compression, effective sizes are - before or, at latest - after compression, small against those of image-type objects.

A.1.3 Summary of typical parameters determined

For reference, values of the previous Galileo web page are also shown. Please note that several Galileo subvariants are in circulation. Versions previous to 4.2 had some image objects with very low optimization, so image size reduction yielded unrealistically high gains. The results shown for Galileo 4.2 also demonstrate which huge error can occur if the wrong type of server is used.

Table A.1

Property	Method	Typical (measurement)	Galileo 4.2	Copernicus target corridor
Total size, bytes (from reference download)	50 % point	205 000	180 280	200 k to 210 k
Number of objects	50 % point	40	51	38 to 42
Acceleration factor based on reported sizes (from comparison reference vs VD2)	50 % point	1,3	2,3 on server not supporting content encoding gzip	1,25 to 1,43 on server supporting content encoding gzip
Acceleration factor from IP sizes	50 % point	2,0	3,23	1,9 to 2,1 on server supporting "content encoding gzip"
Image object size, % of total size	By-type Evaluation	57	51	55 to 59

A.2 Parameter assessment of the current Copernicus page

NOTE 1: For the acceleration factor, sizes were determined from Windows® Performance Monitor measurements. Due to the change in methodology results are hence not completely back-trackable.

Download transport channels were DSL and Vodafone® D2.

Table A.2

Property	Method	Copernicus target corridor	Copernicus result
Total size, bytes (from reference download)	50 % point	200 k to 210 k	208 600
Number of objects	50 % point	38 to 42	38
Acceleration factor from reported sizes (from comparison reference vs VD2)	50 % point	1,25 to 1,43 on server supporting "content encoding gzip"	1,53
Acceleration factor from IP sizes	50 % point	1,9 to 2,1 on server supporting "content encoding gzip"	2,09
Image object size, % of total size	By-type Evaluation	55 to 59	57

NOTE 2: The acceleration factor given in this table was, for downward-compatibility reasons, computed using the MSIE reported values which were in the meantime determined to be incorrect (see previous clauses). With the method based on sizes reported by Performance Monitor (WLAN resp. RAS), the factor is 2,09. Measurements on the reference set using the new method are under way.

Conclusion:

Except for the "old" acceleration factor the Copernicus parameters are well within the target corridor. Since the old Acceleration Factor method had methodological errors anyway due to MSIE limitations, this fact is not considered critical. Conclusion is, therefore, that the Copernicus web page is sufficient for usage.

Annex B: Example workflows used in creation of the Kepler reference web page

As an ongoing process web pages are growing both in size and complexity. So it is clear that a reference page has a limited timespan of usefulness. Consequently, work on a follow up reference page started which led to the release of the Kepler web page in 2011. The following describes how this was done.

While the basic workflows are the same as those leading to the Copernicus page (see annex A), analysis of popular web pages showed that the parameters describing a web page needed to be extended.

It is important to keep in mind that the term "typical web site" needs to be used with some care. Real web pages can be very different from each other. One could argue that there may be no real web page having the same structure as a reference page, therefore putting the term "typical" in question. Of course this issue is somewhat philosophic; however, some people may feel more comfortable to see a reference page - in terms of parameters describing its structure - as an "average" web page.

Also, many web pages are hosted on multiple servers or may contain dynamic elements such as videos or RSS feeds. For practical reasons, a reference page will also be limited in some respect. Therefore, one should remember that a reference page has the purpose of providing a relatively simple way to capture the essence of web browsing in QoS measurements and should not be expected to deliver answers to all questions which could be asked about special cases.

The basic process was the same as for the Copernicus page, so the chapter structure of annex B is the same as of annex A and references for corresponding subsections were omitted for the sake of readability.

B.1 Determination of parameters for the Kepler web page

The same basic steps were done as those for Copernicus described in annex A. Building on the knowledge already gained, some steps were not done in a less extensive way by just making sure that the basics are still valid. Instead, new aspects reflecting increased relevance of video-type content were analyzed. For details see the rest of this clause.

B.1.1 Selection of samples

Basically the same considerations as for Copernicus (see clause A.1.1) were made and the same list of web pages used; some sites had to be removed since they went out of operation since then. Totally 165 pages were analyzed.

Most of the survey was made in November 2009, with some after-analysis extending into January-February 2010.

B.1.2 Determining typical parameters

Sample pages were downloaded using DSL as a non-accelerated transport channel.

B.1.2.1 Compression Factors

No special analysis on compression was made, assuming that the general findings from Copernicus are still valid.

B.1.2.2 Object count and Size

It was found that since the last survey, both total size and object count have considerably increased, see figure B.1.

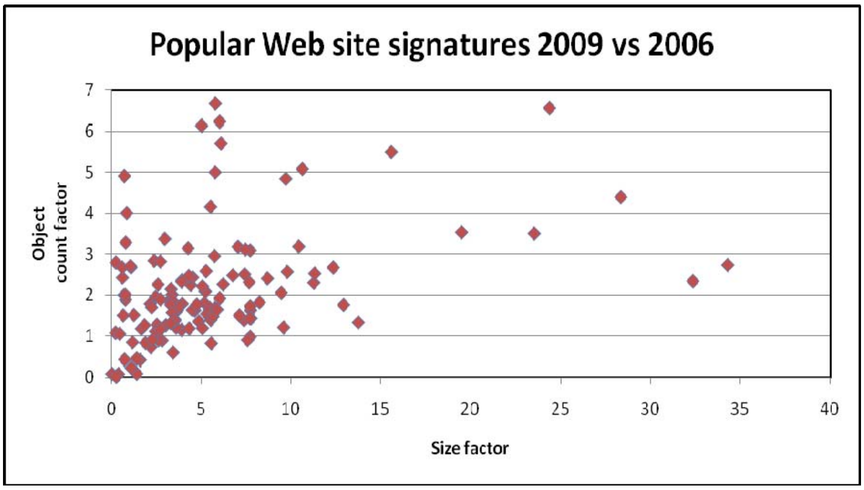

Figure B.1

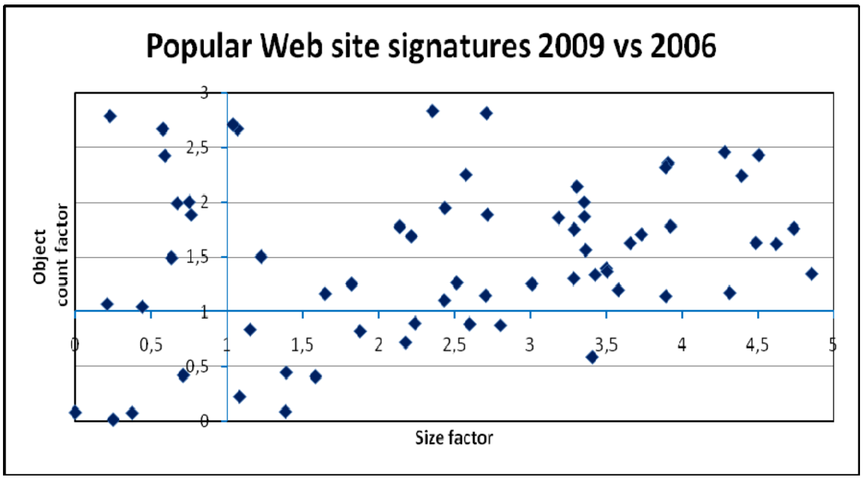

Figure B.2

Figure B.1 shows the increase by the "2009 vs 2006 factor". Size is determined by the total number of bytes downloaded. In many cases, pages contained "streaming" content where content was downloaded all the time and the technical concept of "end of download" does not hold. Figure B.2 therefore shows the area assumed to be typical for "static" pages, with object count factors up to 3 and size factors up to 5.

We found that news-related web pages typically have an object count higher than the average of all sites, while sites with a high degree of multimedia content typically have an object count which is lower than this average.

Figure B.3 shows absolute values of total sizes by their distribution. The new median size was determined to be 820 kBytes.

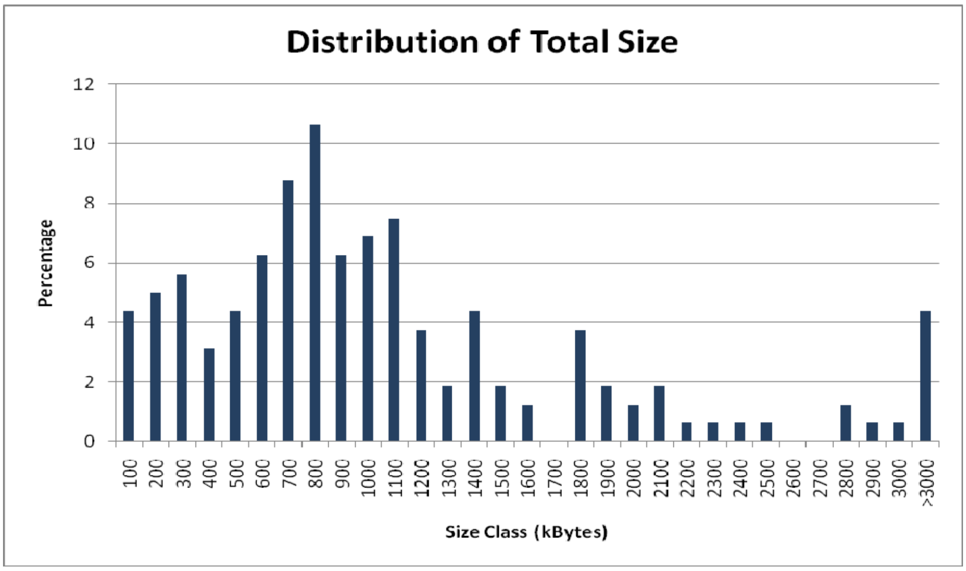

Figure B.3

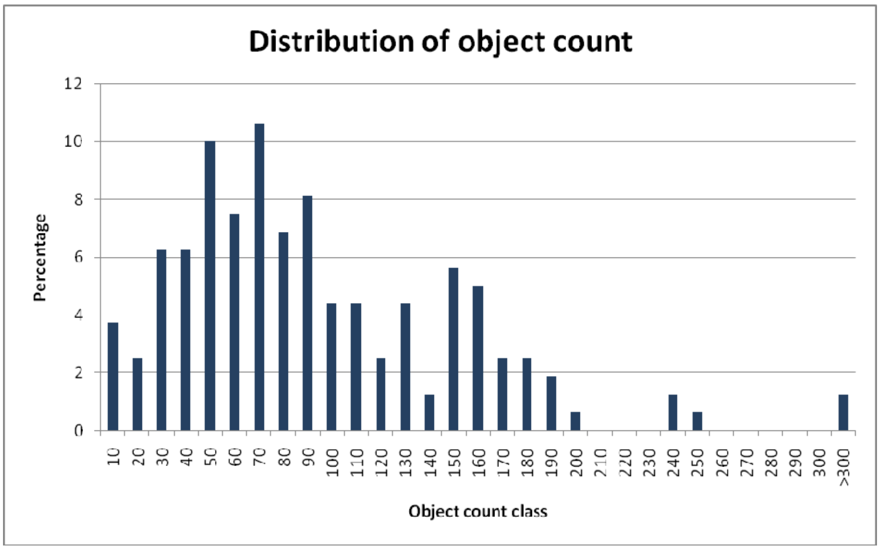

Figure B.4

Figure B.4 shows the new object count distribution, giving a median value of 74.

Figure B.5 shows the site in a co-ordinate system of total size and object count with the "median point" for both values shown in red.

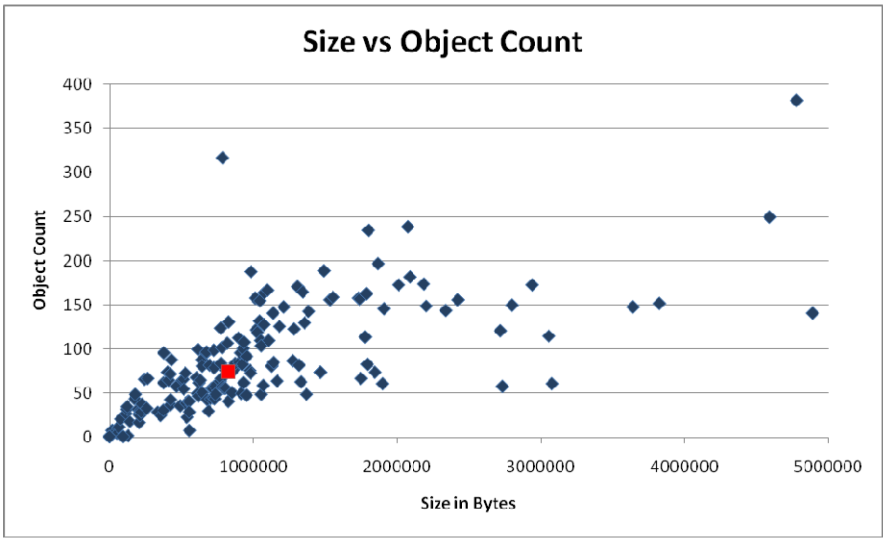

Figure B.5

B.1.2.3 Object types

Figure B.6 shows the distribution of (losslessly) "compressible objects" along the definitions made in previous clauses.

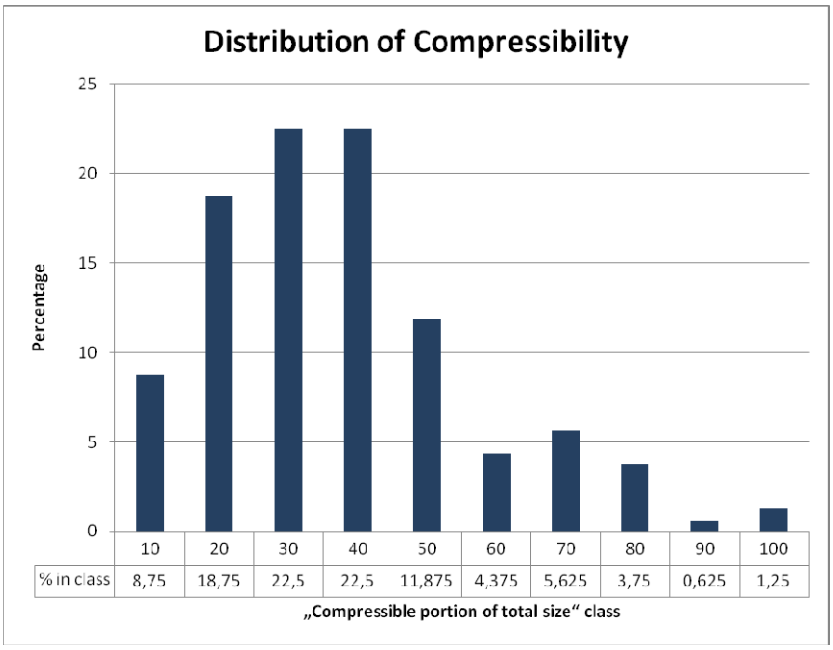

Figure B.6

Nowadays, many sites contain objects which are some kind of movie clip, e.g. embedded YouTube® videos. This led to an important new sub-categorization of "uncompressible" objects. The line of thought is this: for objects representing images (e.g. jpg) acceleration can be done by lossy compression, i.e. by reducing image quality to reduce object size.

At the time the first results of the Kepler survey were presented in STQ Mobile, assumption was that for such movie clips (in the following simply called "videos", in graphics also subsumed under "Flash" video type), no lossy compression was available so this kind of content actually represented a third type of handling content (lossless compression, lossy compression, and "no touching" at all. In the ongoing discussion expert opinions evolved by having to concede that even if with today's state of technology compression of videos may not be possible or may not be done due for other reasons, this is not something which one should rely on. With solid commercial motives on one hand (both reduction of network load with smaller data volumes, and of possible QoE increase as long as the reduction is done with some care), it was determined that it may just be a question of time until such content is also "accelerated".

Figure B.7 shows the distribution of video-type content in the set of web pages surveyed. While about 50 % of the pages did not contain a significant amount of video content (less than 5 % of total size), video content made up more than 25 % of total size for about 14 % of all pages. It can be safely assumed that at the time this text is written (August 2011), importance of video-type content has further increased.

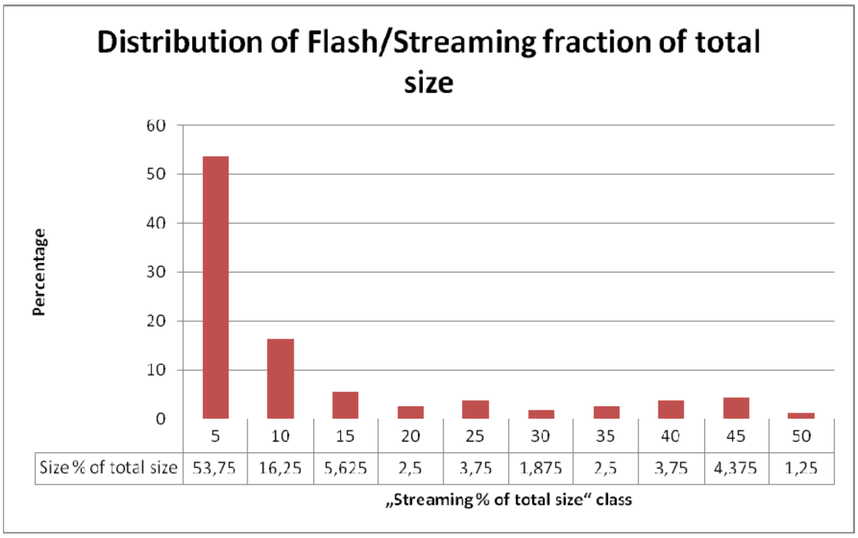

Figure B.7

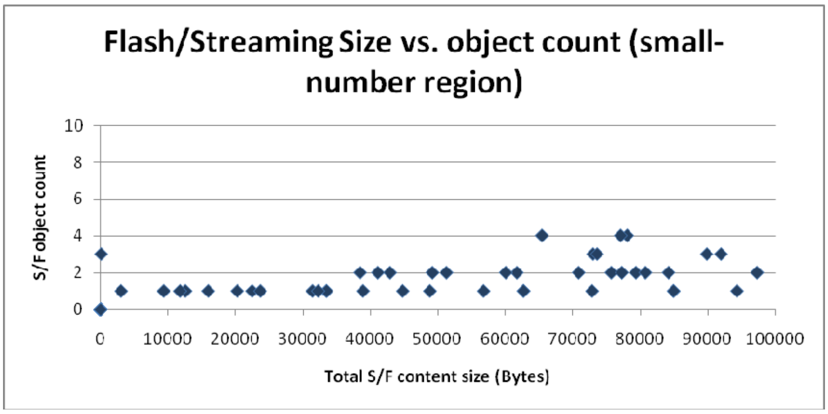

Figure B.8

Figure B.8 shows a distribution of "video" type sizes versus their count in a page. It can be seen that most pages have 1 or 2 videos.

B.1.2.4 Special considerations on "Flash video" objects

Following a thorough analysis where also a manufacturer of performance enhancement proxies was involved, it was determined that the type "non-compressible content for simulating flash" is best represented by non-compressible text, i.e. by textual representation (hexy dump) of random numbers.

Annex C: Example workflows used in creation of the Kepler smartphone reference web page

With the growing popularity of smartphones the nature of web usage changed considerably. As a first step, many web sites have "mobile" versions available which are mostly adaptations of the basic pages optimized for typical smartphone screen sizes and resolutions.

Consequently, another evaluation was done focussing on mobile pages. Initially, this was kept generic (using the term "Kepler Small Screen"). This evaluation started with the same list of approximately 150 web sites which was used to determine the structural parameters of "typical" web sites. The basic steps were very similar to those done for the standard Kepler page and therefore should not be treated in the same detail.

However, it became clear quite soon that some methods which worked very well for standard web sites - mainly, using the Internet Explorer (IE), in this case with a modified User Agent (UA), was not sufficient. The simple reason was that many web sites are using advanced detection methods to determine the type of device. Without referring to details here - new techniques had to be developed which mainly involved analysis of IP trace data to determine web site structures.

As a first step the IE was used with replaced user agent information in order to emulate a Samsung Galaxy S1. To obtain a usable list of "mobile" pages, the following methods were used in combination:

- Use the original URL, select those URL which deliver a page different from the one with a PC browser UA, assuming that there automatic transfer to a mobile version based on the UA was done.
- View the page and use a URL mentioned as the "mobile" version of that page.
- Modify the URL by inserting "m." to the site's name (this approach worked in about 30 % of all cases).

The actual outcome was 37 web pages. Their structural parameters are shown in figures C.1 and C.2.

Figure C.1

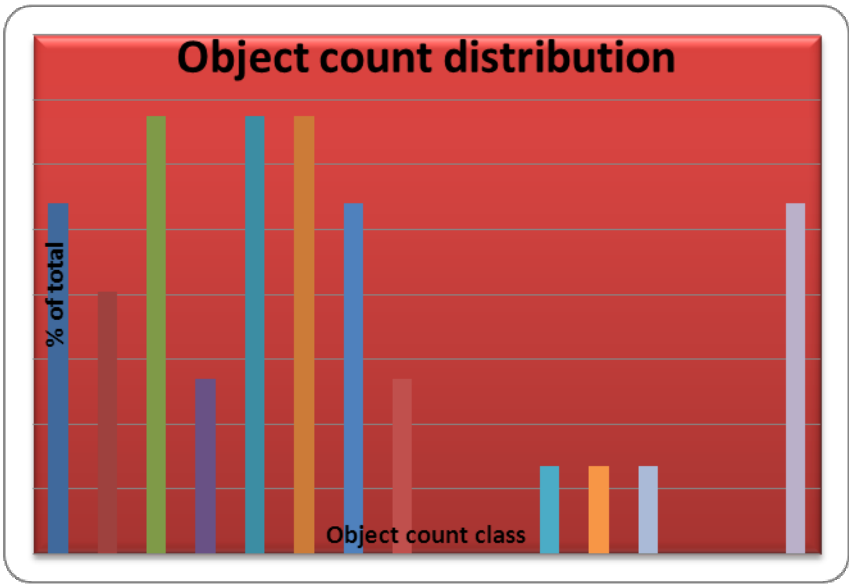

Figure C.2

By closer examination, mobile pages roughly fell into three categories:

- **"Icon collection":** Screen is just starting point for function-oriented further branching.
- **"Log-in/entry screen":** The screen displayed initially is a landing page to access further content.
- **"Functional web site":** This type of page actually serves a functional purpose.

As a reference web page should cover the use case "web browsing", it was decided that a reference page should relate to the "functional web site" category. With "icon collection" or "Log-in/entry screen" sites it was assumed that any meaningful usage of such pages involves another step done by the user to access actual functionality.

In a next step, functional pages were analyzed for their structural parameters. It turned out that these pages were actually quite large (typically around 600 kByte to 800 kByte, but closer inspection showed that a great portion of this size was due to a high portion of text in HTML, CSS and Javascript elements. The conclusion was drawn that this content was more the result of automatic creation or measures to make the pages work on a large number of devices.

The actual sizes relevant for data traffic were of course smaller due to applicable lossless compression but still typical traffic volume was in the order of 350 kByte to 450 kByte for a mobile page.

At this point of evaluation, there was another trend becoming more and more obvious. More and more service providers started to offer apps, motivating users to move from mobile pages towards these apps. This makes sense: Even with a mobile page, a web browser offers less functionality and usability than an app does.

So the trend towards apps was obvious, enhanced by the rise of app-generating software. So two strong forces towards replacement of browser based towards app based content presentation exist which essentially mean that for the use case "user views web site content". In consequence, analyzing the content typical apps transfer was added to the survey.

For a couple of arbitrarily selected apps, traffic was recorded on IP Trace level (smartphone based tcpdump/pcap files) and subsequently analyzed by a proprietary tool developed for this purpose.

By comparison of some sites where both a mobile page and an app exist, it turned out that a) apps actually tended to create very "browser like" traffic and that typical app-related content size was in the 120 kByte to 150 kByte category, i.e. a functional site in the above sense was about ¼ of the size of a mobile page. This was also consistent with expectations based on screen size, resolution, image count and image size on such pages. In short, it can be stated that a "web page replacement app" delivers approximately the same content with about 25 % of content size (all these figures relate to equivalent uncompressed sizes).

It has to be stated here that the process of obtaining data was moving towards a smaller number of sample sites simply because the amount of manual work increased in this process - the solution initially employed where a URL sequencer was used to scan a large number of sites could not be used anymore.

Finally, the following structural parameters were chosen - along design principles outlined in other parts of the present document - for a Kepler SP page proposal:

- Total size approx. 145 000 bytes.
- 22 objects.
- 12 non-compressible objects having a total size of approximately 88 000 Byte (60,5 % of total size).

The Kepler SP page can be downloaded from the ETSI Portal.

C.1 Multiple reference pages for different usage profiles

During the discussions leading to the design of the Kepler reference web page, it became clear that in the future a single reference page may not be enough. To the perception of STQ Mobile, mobile internet usage is developing into different directions represented by different classes of mobile devices (smartphones, tablet PC devices, etc.) having also different usage profiles resulting into different ways to perceive quality. This is currently subject to further study.

History

Document history		
V1.1.1	December 2006	Publication
V1.2.1	December 2011	Publication
V1.3.1	November 2012	Publication