

QoS Parameters and the Measurement Methodology:

1.1 Specifications:

The measurement setup to be used to conduct test calls for measuring the various Quality of Service (QoS) parameters such as successful data transmission download and upload ratio, download speed and latency etc. is given in Figure 2.1. The basic measurement set-up consists of a Test-Device and a Test-Server with specified software and hardware. Test calls have to be established between the Test-Device and Test-Server and measurements must be made for the respective QoS parameters. Also the necessary number of test calls and their distribution are defined for this purpose.

To assess the quality of the connection between an end user and an (Internet Access provider) IAP, ideally the Test-Server should be placed as near as possible to the gateway providing the interconnection between access network and IAP network. The location of the test-server as near as possible to the gateway providing the interconnection between access network and IAP network implies that the measurements will not reflect the influence in the QoS of the IAP network, between that gateway and the gateway interconnecting with the Internet.

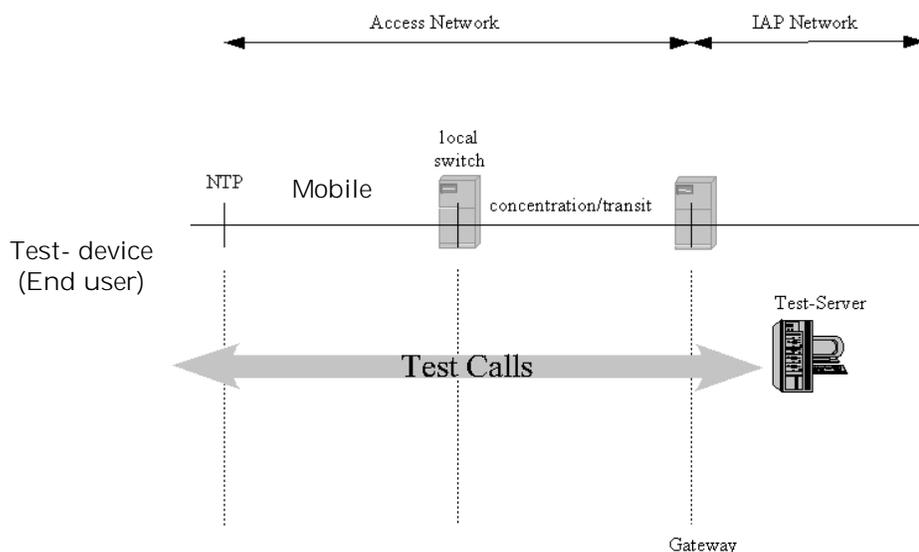


Figure 1.1: Measurement set-up

1.2 Requirements for the Test-Server:

A dedicated test server should be used as a well-defined reference for all the tests which are required to be done under test measurement set up. Under no circumstances should be a commercial server (e.g. www.yahoo.com) used, since the test conditions for such a server may change over time. This makes later reproduction of the results impossible.

The test server should be identified by an IP address and not by its FQDN (Fully Qualified Domain Name) in order to avoid issues with DNS (Domain Name Server) lookup and including the DNS caching strategies of the used operating system into the measurement.

The TCP (Transmission Control Protocol) settings of the server tested against should also be recorded. Since the number of host operating systems for internet servers is larger than on the client side, no detailed recommendation concerning the TCP settings of the server is given. However, the TCP stack of the reference server should at least be capable of the following:

- Maximum Segment Size between 1380 Bytes and 1460 Bytes.
- TCP RX Window Size > 4096 Bytes
- SACK (Selective Acknowledgement) enabled.
- TCP Fast Retransmit.
- TCP Fast Recovery enabled.
- Delayed ACK enabled (200 ms).

1.3 Test File:

The test file should consist of incompressible data i.e. a data file that is already compressed, e.g. like a zip or jpg file. The test file should have at least twice the size (in Kbit) of the theoretically maximum data transmission rate per second (in Kbit/s) of the Internet access under consideration.

- FTP (File Transfer Protocol): A large test file (5MB) may be used to upload from a workstation to a FTP server or download files from a FTP server to a workstation.
- HTTP (Hyper Text Transfer Protocol): A small test file (500 KB) may be used to transfer from a Web server onto a browser in order to view a Web page that is on the Internet.

1.4 Representativeness or Number of test calls: The choice of adequate test calls, i.e. geographical locations of origin and destination of calls as well as traffic variations, is a crucial point with respect to the comparability and validation of the statistics to be calculated for the measured parameters. For each parameter, it is to be ensured that the samples are aggregated over all classes of customers for fairness in reflecting the QoS actually perceived by the user and the statistics shall be preserved to substantiate the same.

The necessary number of samples (test calls) with a given precision may be calculated as described in the following text. It can be calculated by the formula:

$$n = \frac{z_{1-\alpha/2}^2}{a^2} \cdot \left(\frac{s}{\text{mean}(x)} \right)^2$$

Whereas:

n: is the number of samples.

$z_{1-\alpha/2}$: is the $1-\alpha/2$ -percentile of the standard normal distribution.

s: is the expected standard deviation

mean(x): is the expected mean value .

a: is the relative accuracy.

The number of observations must be chosen such that an absolute accuracy with a confidence level of 95% is achieved.

The following table gives the resulting values where:

- $z_{1-\alpha/2} = 1,96$ for a confidence level of 95 %; and
- $a = 2\%$.

s/mean(x)	observations
<0,1	100
0,1 to 0,3	1000
>0,3 to 0,5	2500
>0,5 to 0,7	5000
>0,7 to 0,9	7500
>0,9	10 000

2. Frequency of measurements:

For measurement of QoS parameters, it will be suitable to take measurements on a monthly basis during Time Consistent Busy Hour (TCBH). The number of test probes (Test PCs) used in the area of coverage of the system of measurements should be enough to guarantee that the measurements are representative of the coverage area and sufficient from a statistical perspective. In order to guarantee the spatial representation of the measurements, the operators should establish and deploy test probes depending on the number of active clients in each region according to rules, which should take into account the market size and demographics in the license service area concerned.

In some cases disasters, freak weather, etc. may distort measured QoS figures. Such occurrences may not necessarily damage a network, but could degrade QoS by inducing exceptional traffic levels etc. In these cases, service providers should provide the measured QoS and may additionally provide a second figure which excludes the effects of the exceptional circumstances. A note clearly explaining the difference should also be provided.

3. The QoS parameters:

3.1 Service activation/ provisioning:

This refers to service activation and involves programming various network databases with the customer's information, for various

technologies permitted under the license networks of an existing customer, and any gateways to standard Internet chat or mail services. The service provider typically sends these settings to the subscriber's handset using SMS or WAP.

3.2 Successful data transmission download attempts:

The successful data download attempts is defined as the ratio of successful data downloads to the total number of data download attempts in a specified time period. A data transmission is successful if a test file is downloaded completely and with no errors.

Measurement:

The percentage that is the sum total of successful data downloads, divided by the sum total of all attempts to download a test file should be provided. The statistics should be calculated from test calls made according to the measurement set-up and taking into account the representativeness requirements. The successful data download is measured by downloading a test file. An attempt to transmit the test file should be considered unsuccessful if it takes longer than 60 seconds.

NOTE: The threshold of 60 seconds refers to the limit for acceptable performance for bulk data transmission/retrieval of ITU-T Recommendation G.1010.

The transmission time is the time period starting when the access network has received the necessary information to start the transmission and ending when the last bit of the test file has been received.

3.3 Successful data transmission upload attempts:

The successful data upload attempts is defined as the ratio of successful data uploads to the total number of data upload attempts in a specified time period. A data upload is successful if a test file is uploaded completely and with no errors.

Measurement:

The percentage that is the sum total of successful data uploads, divided by the sum total of all attempts to upload a test file should be provided. The statistics should be calculated from test calls made according to the measurement set-up and taking into account the representativeness requirements. The successful data upload is measured by uploading a test file.

3.4 Minimum download speed:

The download speed is defined as the data transmission rate that is achieved separately for downloading a specified test file from a test server to a user's device.

Measurement:

The statistics should be calculated from test calls made according to the measurement set-up. Make test calls at regular intervals and to weigh the results according to the patterns of real traffic. The necessary number of samples (test calls) with a given precision may be calculated as described in representativeness.

This parameter is to be measured and reported to TRAI for all the mobile data services or technologies.

3.5 Average Throughput for Packet data:

This is one of the most important parameter from customer perspective and it is defined as the rate at which packets are transmitted in a network. It can be expressed as an average rate. In a mobile network the download speed varies depending on the number of users in a particular location. Even though a service provider may be advertising certain speed, the actual speed may vary as per the number of users in the network and there could be customer dissatisfaction on account of the slow speed. Hence, there is a need to prescribe an average throughput to protect the interest of consumers. The service providers need to

constantly upgrade their network to meet average throughput benchmark.

Equation: The equation for calculating the Average Throughput is as given below:

$$\text{Average Throughput(kbps)} = \frac{\sum_{\text{No of calls}} \text{Volume of data transferred (kbits)}}{\sum_{\text{No of calls}} \text{Session duration (seconds)}}$$

This equation is valid for uplink and downlink.

Definition: The Key Performance Indicator Terms for Average Throughput are defined in the table given below:

Key Performance Indicator Term	Definition
Volume of Data Transferred (kbits)	FTP download/ upload in kilo-bit during one session.
Session Duration (sec)	Time duration (seconds) to download/upload a file.
Average Throughput (kbps)	Packet data throughput in kbps

Measurement:

The statistics should be calculated from test calls made according to the measurement set-up. Make test calls at regular intervals and to weigh the results according to the patterns of real traffic. The necessary number of samples (test calls) with a given precision may be calculated as described in representativeness. Throughput shall be calculated using data from all such calls.

The service provider will advertise the throughput being offered to its customers as per their category or plan.

3.6 Percentage of Node B/ BTS carrying less than 80% of the average throughput in a license service area:

In mobile data network, the most important parameter for the user is the throughput available to him for data usage. The throughput available to a customer will depend on the number of users served by a BTS/Node B. In case the number of users is more the throughput will be less, leading to customer dissatisfaction. In many cases the throughput is affected due to limitations in the backhaul like less number of E1s connecting Node B to RNCs/ BTS to BSCs etc. The operators may have to augment this to maintain certain level of throughput. As such, it is necessary to measure the average throughput during TCBH for Node B/ BTS in the licensed service area. The average throughput for Node B/ BTS can be averaged over a period of one month and subsequently average throughput per Node B for the licensed service area for a month period may be calculated.

The numbers of Node B/BTS below say 80% of such calculated average throughput per Node B for the licensed service area (LSA) shall give an indication of the performance of the network with regard to throughput available to customers.

3.7 Latency:

Latency is the amount of time taken by a packet to reach the receiving endpoint after being transmitted from the sending endpoint. This time period is termed the "end-to-end delay" occurring along the transmission path. Latency generally refers to network conditions, such as congestion, that may affect the overall time required for transit.

Measurement:

Latency shall be measured with the destination server for ping connected directly to the server on the same Intranet domain. It shall be measured in a stationary mode. Latency shall be measured for conversational/real time services using different mediums like audio, video, data and data (interactive).

3.8 PDP Context Activation Success Rate:

A Packet Data Protocol (PDP) context specifies access to an external packet-switching network. The data associated with the PDP context contains information such as the type of packet-switching network, the MS PDP (Mobile Station PDP) address that is the IP address, the reference of GGSN (Gateway GPRS Support Node), and the requested QoS. A PDP context is handled by the MS, SGSN (Serving GPRS Support Node), and GGSN and is identified by a mobile's PDP address within these entities. Several PDP contexts can be activated at the same time within a given MS. A PDP context activation procedure is used to create a PDP context. This procedure may be initiated either by the MS or by the network. The MS is always GPRS attached before PDP context negotiation. The PDP context activation may be performed:

- Automatically, if it is generated during a given procedure to perform a GPRS data transfer;
- Manually, if it is generated by user intervention.

A PDP context deactivation procedure is used to remove a PDP context. This procedure may be initiated either by the MS or by the network (SGSN or GGSN). The PDP context deactivation may be generated either during an application deactivation or during the GPRS detach or delete subscriber data procedures.

Measurement:

This measurement provides the number of successfully completed PDP context activations. For these context activations, the GGSN is updated successfully. A report of PDP context activation success is generated at GGSN.

PDP Context Activation Success Rate(%)

$$= \frac{\text{number of successfully completed PDP context activations}}{\text{Total attempts of context activation}} \times 100$$

3.9 Drop Rate:

It measures the inability of Network to maintain a connection and is defined as the ratio of abnormal disconnects w.r.t. all disconnects (both normal and abnormal). An abnormal disconnect may happen because of Radio Link Failures, Uplink (UL) or Downlink (DL) interference, bad coverage, unsuccessful handovers or any other reason. For example, when associated HS-DSCH is released abnormally by either UTRAN or CN, HSDPA session is taken as dropped.

Equation: The equation for HSDPA Drop Rate is given by the following formula:

$$\text{HSDPA Drop Rate (Through Counters)} = \frac{\# \text{HS - DSCH Release due to Failure}}{\# \text{HS - DSCH Allocation Success}}$$

HSDPA Drop Rate (Through Drive Test) =

$$\text{HSDPA_DropRate} = 100 * \frac{\sum_{\text{HSDPA_Call}} \text{HS - DSCH_Drop}}{\sum_{\text{HSDPA_Call}} \text{HS - DSCH_AllocationSuccess}}$$

Definition: Key Performance Indicator Terms for HSDPA Drop are defined in the following table.

Key Performance Indicator Term	Definition
HS-DSCH Release Due To Failure	Number of HS-DSCH allocation releases due to radio link and other failures
HS-DSCH Allocation Success	Number of allocations when the RNC has received RRC: Radio Bearer Re-Configuration Complete message from the UE after successful HS-DSCH MAC-d flow setup

Measurement:

HSDPA Drop Rate shall be measured by data session activation (HSDPA) followed by ftp download. This test is done with a single UE performing HSDPA call in the Cell under test with 384/64 Kbps associated DCH on the uplink and HS-DSCH in the DL with downgrade/upgrade to/from R99 RAB. This parameter will be measured through drive test and counter values.