# ETSI TR 103 351 V1.1.1 (2017-07)

TECHNICAL REPORT

**Satellite Earth Stations and Systems (SES);**
**Multi-link routing scheme in hybrid access network with heterogeneous links**

Reference
DTR/SES-00373

Keywords
broadband, satellite, terrestrial

***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

## Essential patents

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 (https://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.

## 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.

# Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Satellite Earth Stations and Systems (SES).

# 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.

# 1 Scope

The present document proposes and analyses a traffic distribution architecture for hybrid access networks combining one or several terrestrial access technologies (fixed or mobile service) together with a satellite broadband access network (Fixed Satellite Service).

The traffic distribution architecture will enhance the end users' Quality of Experience by efficiently utilizing all available connections simultaneously using the Multipath TCP protocol. It allows for splitting traffic flows into smaller chunks, so-called objects, for which the most appropriated link is selected. The architecture is complemented by a Capacity and Link Status Estimation process that estimates link characteristics by passively monitoring TCP traffic, so that the Link Selection can be performed on a more informed basis.

The present document aims at:

- Defining the usage of the Multipath TCP protocol in Hybrid FSS satellite/terrestrial architecture.
- Proposing a method to split TCP traffic into connected chunks of traffic to ease the multipath routing.
- Proposing a routing scheme that distributes traffic intelligently among the available connections.
- Proposing a TCP-based link estimation method to passively determine available bandwidth and latency of a path.

# 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 referenced 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 TR 103 272: "Satellite Earth Stations and Systems (SES); Hybrid FSS satellite/terrestrial network architecture for high speed broadband access".

[i.2] IETF RFC 6824 (2013) (Ford A., Raiciu C., Handley M. & Bonaventure O.): "TCP Extensions for Multipath Operation with Multiple Addresses", (Experimental).

[i.3] EC FP7 Project: "Broadband Access Via Integrated Terrestrial & Satellite Systems (BATS)", D7.1 "Trial Evaluation", 2016/01/14.

[i.4] IETF RFC 3697: "IPv6 Flow Label Specification".

[i.5] IETF RFC 3917: "Requirements for IP Flow Information Export (IPFIX)".

# 3 Definitions and abbreviations

## 3.1 Definitions

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

**access link:** link established between the IUG and the ING via a satellite or a terrestrial network

NOTE: One access link corresponds to one network interface.

**application:** program running on a device that requests or generates data that will form a Traffic Flow through a Network Interface

**broadband access:** access network where the downlink service rate is greater than or equal to 2 Mbps

**high speed broadband:** access network where the downlink service rate is greater or equal to 30 Mbps (Target set by the Digital Agenda for Europe)

**hybrid access network:** access networks combining a satellite component and a terrestrial component in parallel where the delivery of a service using both the satellite component and the terrestrial component intelligently to maximize the Quality of Experience for end users in under-served areas

**Intelligent Network Gateway (ING):** counterpart device of the IUG in an hybrid access network

**Intelligent User Gateway (IUG):** home device providing broadband access, security, cached storage capacity and QoE provisioning in a hybrid access network

**network interface:** interface that connects the IUG or ING to an access link

**object:** data unit created or requested by an application

**Quality of Experience (QoE):** subjective measure of the user's experiences with a service or an application (e.g. web browsing, phone call, TV, call to a Call Centre)

**Quality of Service (QoS):** objective measure of a service delivered by a network

**service component:** set of traffic flows resulting from an application including, where applicable, the various traffic flows requested by multiple functions within the application

**traffic flows:** sequence of packets sent from a particular source to a particular unicast, anycast, or multicast destination that the source desires to label as a flow

NOTE 1: More specifically it refers to a set of IP packets passing an observation point in the network during a certain time interval (see IETF RFC 3917 [i.5]).

NOTE 2: See IETF RFC 3697 [i.4].

## 3.2 Abbreviations

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

| | |
|---|---|
| ADSL | Asymmetric Digital Subscriber Line |
| CAPEX | CAPital Expenditures |
| FSS | Fixed Satellite Service |
| GEO | GEOstationary satellite |
| IAT | Inter-Arrival Time |
| IAT-ING-InterObj | IAT-ING-Inter object |
| IAT-ING-IntraObj | IAT-ING-Intra object |
| IAT-ING-Thresh | IAT-ING-Inter Threshold |
| IAT-IUG-InterObj | IAT-IUG-Inter object |
| IAT-IUG-IntraObj | IAT-IUG-Intra object |
| IAT-IUG-Thresh | IAT-IUG-Inter Threshold |

| | |
|---|---|
| IDU | InDoor unit (or modem) |
| IETF | Internet Engineering Task Force |
| ING | Intelligent Network Gateway |
| IP | Internet Protocol |
| IUG | Intelligent User Gateway |
| LAN | Local Area Network |
| LEO | Low Earth Orbit (satellite) |
| LTE | Long Term Evolution |
| MP | Management Point |
| MPTCP | MultiPath TCP |
| NAT | Network Address Translation |
| OPEX | OPerational EXpenditures |
| PSBOL | Path Selection Based on Object Length |
| QoE | Quality of Experience |
| QoS | Quality of Service |
| RFC | Request For Comment (IETF document) |
| RTT | Round Trip Time |
| SSH | Secure SHell protocol |
| TCP | Transmission Control Protocol |
| UDP | User Datagram Protocol |
| VPN | Virtual Private Network |
| WAN | Wide Area Network |
| WRR | Weighted Round Robin |

# 4 Hybrid access network with heterogeneous links

## 4.1 Architecture overview

The present document assumes a hybrid access network delivering High speed broadband service such as the one depicted in Figure 1. The concepts and rational for this as well as further details are defined in ETSI TR 103 272 [i.1].

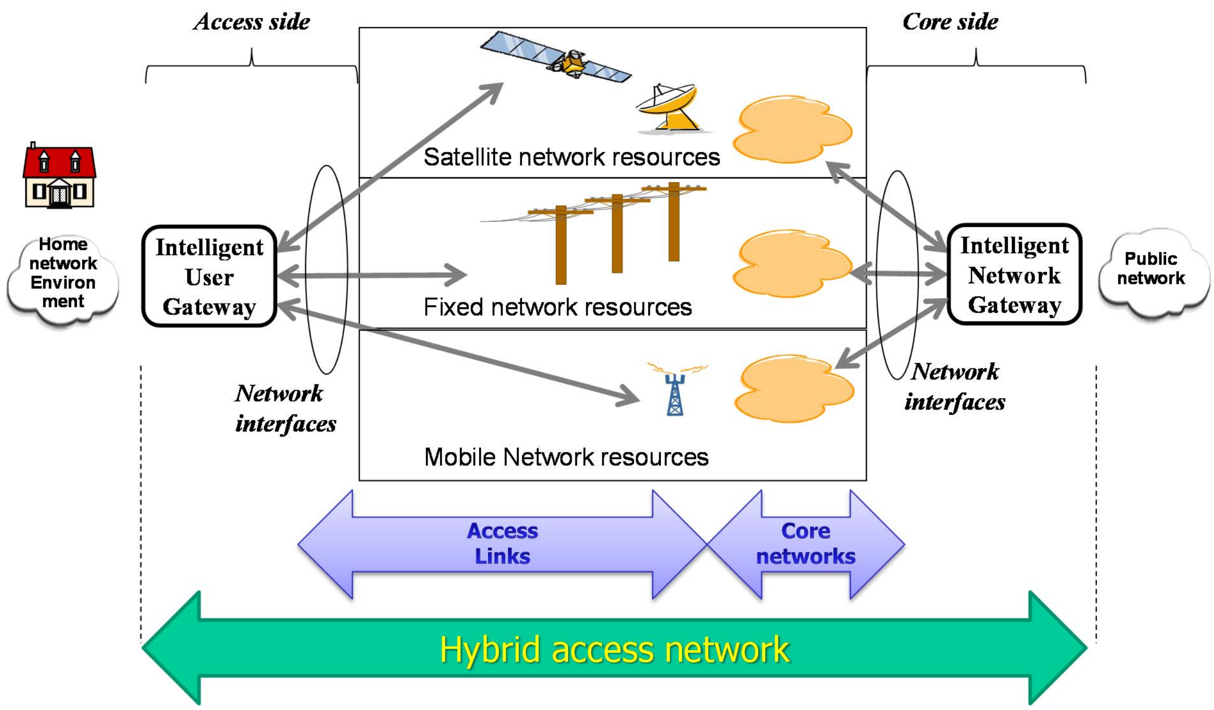


**Figure 1: Hybrid access network architecture**

## 4.2 IUG and ING functional architecture

The present document focuses on defining the key building blocks in both Intelligent User Gateway (IUG) and Intelligent Network Gateway (ING), which are the link estimation, the traffic splitting component and the link selection, as shown in Figure 2. The first is responsible for determining the characteristics of all available paths between the IUG and the ING, while the second splits the incoming traffic flows from the Home network environment or the public network, respectively, into smaller chunks of traffic, so-called objects. It is then the responsibility of the link selection component to distribute these chunks onto the available links based on their characteristics.

It should be noted the predominant traffic expected in this kind of hybrid access network is TCP/IP traffic. Hence, the present document focuses on optimizing TCP traffic handled in a hybrid FSS satellite terrestrial network. How other traffic is being handled is largely out of scope of the present document, although UDP handling is discussed in clause 5, as a proposal. Regular routing methods needs to be in place, which work independently of the mechanisms presented in the present document. The architecture in the present document neither harms nor optimizes the handling of non-TCP traffic.

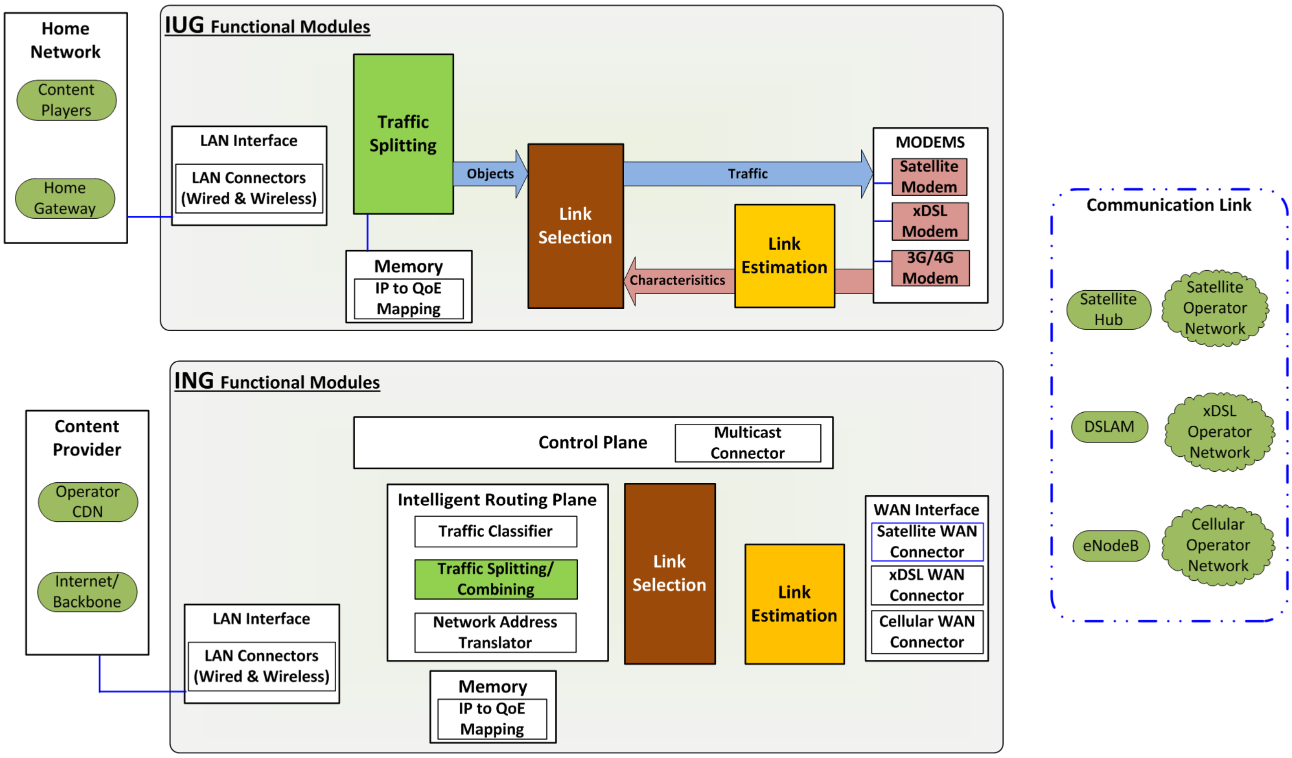


**Figure 2: Key building blocks in IUG and ING**

# 5 Multi link routing with traffic dichotomy

## 5.1 Introduction

In its current version, the Multi-Link Architecture relies on:

- A traffic Classification or "traffic dichotomy" between TCP traffic and non TCP traffic (such as UDP traffic).
- Multipath-TCP (MPTCP) [i.2] as its basic multipath technology between the IUG and the ING, to efficiently exploit the multiple paths between the IUG and ING. Hence, MPTCP is not used end-to-end. Instead, a MPTCP proxy is running on the IUG and ING, which breaks the TCP end-to-end paradigm. The connection from the client to the server is intercepted by the MPTCP proxies on the IUG and the ING, so that a single TCP connection can be split into three (MP)TCP connection, namely between end host and IUG, IUG and ING, where MPTCP is used, and ING and the other end host of the connection.

Between those two MPTCP proxies, tunnels are established in order to operate independently of the underlying networks. Between the end systems and the IUG or ING, respectively, regular TCP will be used.

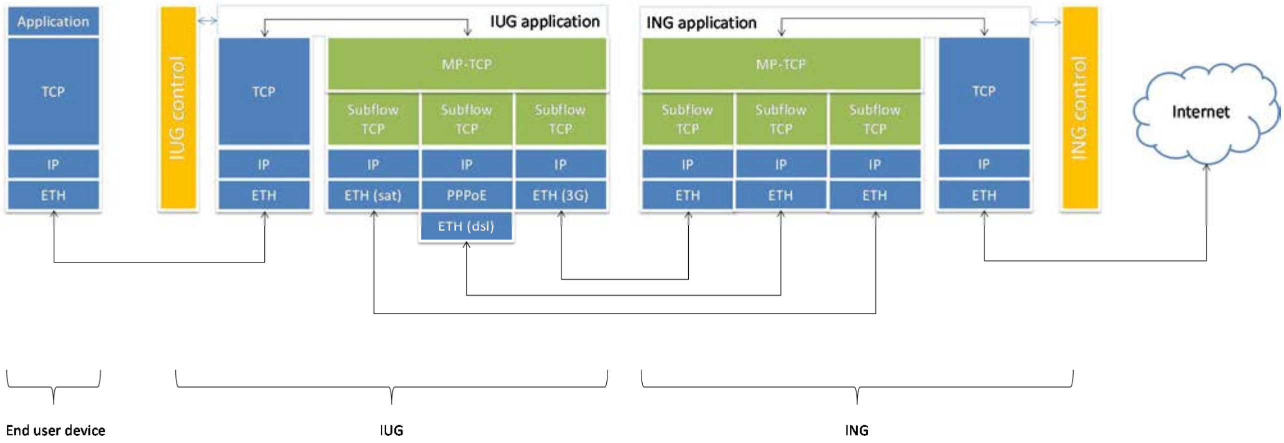


**Figure 3: MPTCP proxy architecture**

As depicted in Figure 3, between IUG and ING multiple MPTCP subflows are established. To be precise, the MPTCP proxy on the IUG creates for each TCP flow as many subflows as there are different Physical or Logical links available on the IUG side. In Figure 3 it can be seen that 3 logical links available on the IUG: One ADSL link thru a native ADSL physical interface, one Satellite link accessible thru an external IDU via a physical Ethernet interface and one 3G/4G Link accessible thru an external modem via the same or another physical Ethernet interface. The IUG and ING control are responsible for link estimation, link selection and traffic splitting to each TCP subflow. Hence, the link selection process running on IUG and ING needs to distribute the traffic on the available subflows.

It is important to note that the MPTCP standard [i.2] does not specify how the traffic is distributed among the available subflows.

## 5.2 TCP traffic splitting and recombining plus UDP traffic forwarding

### 5.2.1 Overview

For TCP traffic, the Link Selection operates on TCP objects. An object is a sequence of TCP segments belonging to the same flow, i.e. same source and destination IP and same source and destination port, which are sent within a given time frame. Figure 4 below gives an overview of the algorithm. As indicated long objects are routed over the highest bandwidth link available while short objects are routed over the link with the lowest RTT.

The Link for UDP Traffic may be selected as the one minimizing a link cost function, defined as a combination of weighted criteria such as Link Reliability, One Way Maximum Latency, Available Bandwidth, Bandwidth cost and OPEX/CAPEX considerations. The criteria and their weight should be configurable, in order to provide a flexible deployment for operator's needs. In a 1$^{st}$ approach, the One Way Maximum Latency may be measured or estimated per Access Network (ADSL, Cellular and Satellite), and this performance estimation could be provisioned as input of the link selection for UDP traffic.

Additionally, the traffic - TCP and UDP - may be split into critical and not critical traffic. The critical traffic may be routed towards the links that have the highest reliability, in terms of 'service continuity' or 'link availability along time'. In this case, the weights of the combined criteria have to be changed:

- Critical applications based on TCP would use reliable links as main criteria, if such links are available. These links are associated to MPTCP subflows.
- Critical application based on UDP would use reliable links (as main criteria), selected amongst ADSL, Cellular, Satellite reliable links. In this case, these links are not associated to MPTCP subflows. It does not prevent from using other secondary criteria, such as OneWay Maximum Latency and/or Available Bandwidth.

The algorithm described in Figure 4 handles Link selection for both TCP and UDP traffics and is given as an example:

- It only handles Bandwidth and RTT criteria for TCP traffic.
- It assumes that the ADSL Link has a lower cost than the Cellular Link, which is less expensive than the Satellite link. Moreover, the same route is selected for all the UDP traffic.

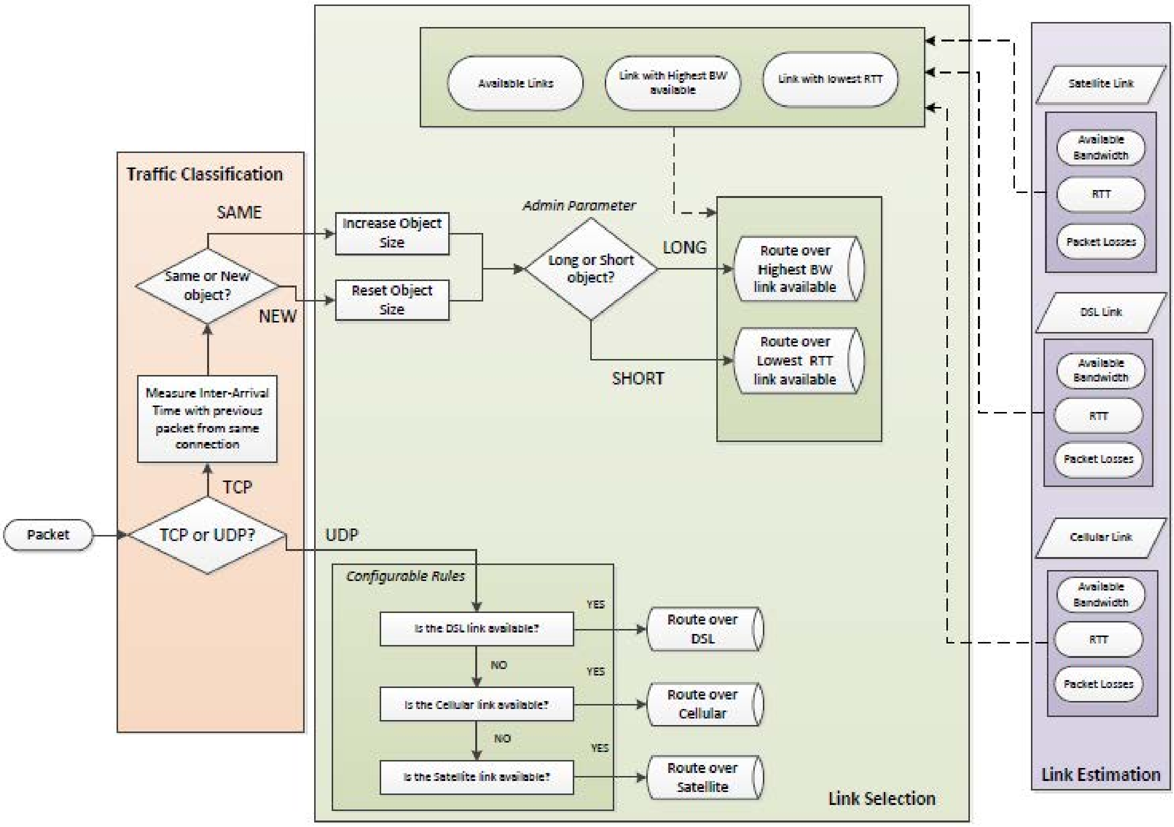


**Figure 4: Algorithm Flow Chart**

## 5.2.2 Concept of a TCP object

A typical client-server dialog is depicted in Figure 5. A client sends an object O1 to a server that replies with an object O2, then the client sends an object O3 to the server and receives an object O4.

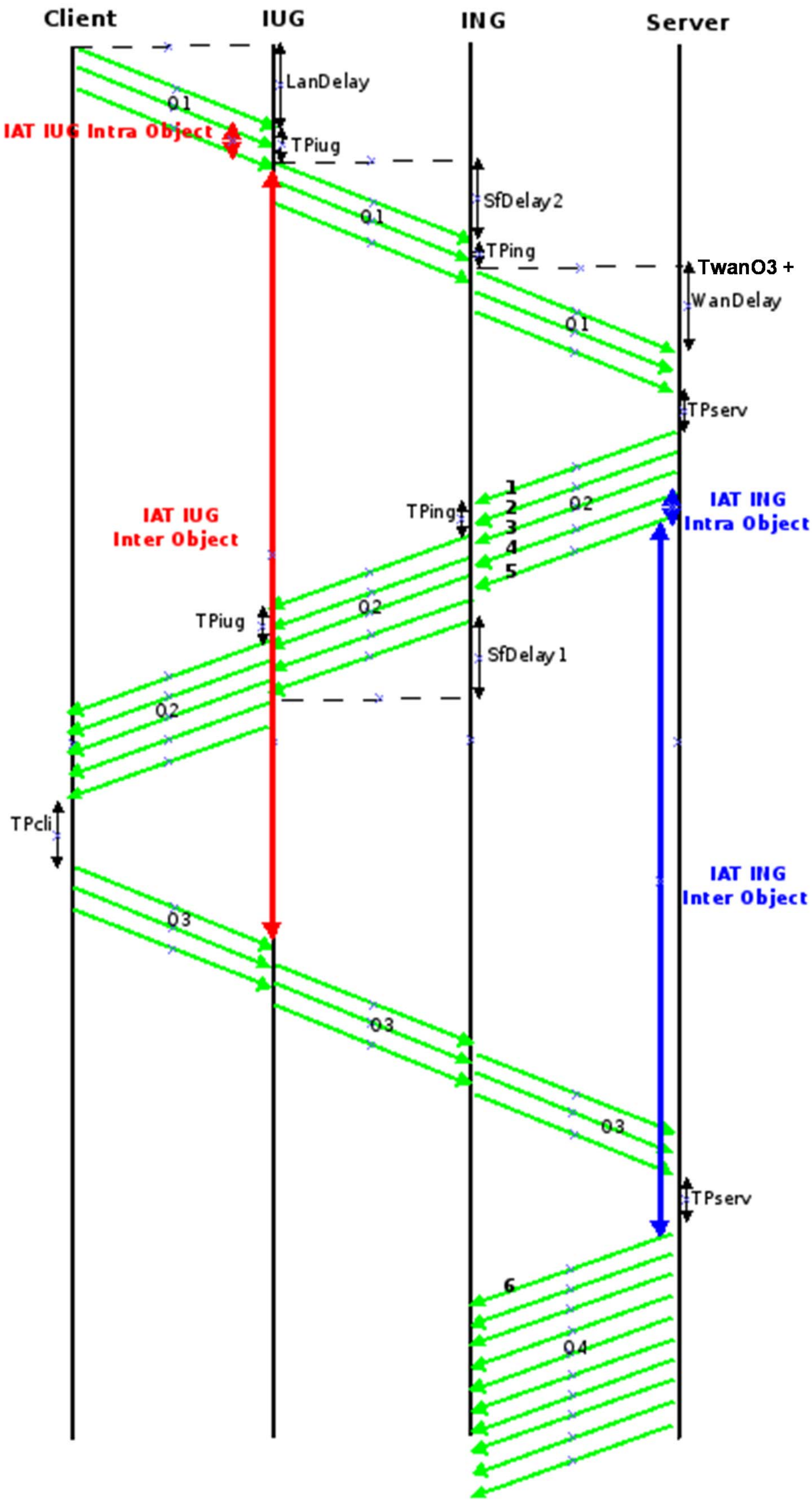


**Figure 5: TCP traffic splitting into objects**

The respective objects sizes for O1, O2, O3 are: 3, 5 and 3 TCP non-null segments.

The Inter arrival time for objects at the Server side is computed as following:

IatINGInterObj = TPing + SfDelay1 + TPiug + LanDelay + TPclient + LanDelay + TPiug + SfDelay2 + TPing + WanDelay + TwanO3 + TPserv + WanDelay

Where:

- TPing = Latency introduced by the ING proxy.
- SfDelay1 = ING to IUG delay. This delay is, depending on the current object size, either the delay of the highest bandwidth link or the delay of the lowest delay link.
- TPiug = Latency introduced by the IUG proxy.
- LanDelay = IUG to Client Lan delay.
- TPclient = Processing time by the client, may include user waiting time.
- SfDelay2 = IUG to ING delay. This delay is, depending on the size of the last object sent by the IUG, either the delay of the highest bandwidth link, or the delay of the lowest delay link.
- WanDelay = Propagation time from the ING to the Server.
- TwanO3 = Time to transmit O3 from the ING to the Server, over symbols, according to the symbol rate.
- TPserv = Server processing time.

On Figure 5, it can be seen that:

- The client sends the segments of objects O1 and O3 over the LAN.
- The IUG forwards the corresponding data to the ING and introduce latency: TPiug.
- The ING forwards these data to the server between the ING and the server. The ING adds latency: TPing.
- The server process the data of object O1 and after a processing time TPserv, it sends the data of object O2 to the ING, which forwards it to the IUG.
- The IUG forwards O2 to the client that in turn, after processing time TPclient, issues the object O3. O3 is sent in the same way as the object O1 to the server.
- The server issues the response O4.

Moreover, the following time periods are defined:

- The time period between two consecutive TCP segments, which belong to the same flow, is referred to as Inter-Arrival Times (IAT).
- The time period between two consecutive TCP segments belonging to the same object and measured at the ING is referred to as IAT Intra object (IAT-ING-IntraObj), e.g. TCP segments 1 to 5 of object O2.
- The time period between two consecutive TCP segments not belonging to the same object and measured at the ING is referred to as IAT Inter object (IAT-ING-InterObj), e.g. last segment of O1 and first segment of O3.

## 5.2.3 Detecting an object

### 5.2.3.1 Overview

Since the Link Selection is being performed on IUG and ING the detection of an object is done on the IUG for the traffic from the end user devices in the home network environment towards the public network and on the ING for the traffic in the opposite direction.

To allow for a simple and scalable implementation, only locally available parameters should be used, available on IUG and ING, respectively.

The proposed detection process takes place on both ING and IUG and analyses IATs of consecutive TCP segments belonging to the same TCP connection. Thus, pure TCP messages like SYNs, ACKs, and FINs are excluded from the decision process.

On the ING, the process analyse IATs for TCP segments received from an arbitrary server in the public network and on the IUG the process analyse IATs for TCP segments received from devices in the home network environment.

The IATs of each segment are compared to an IAT Threshold, which are referred to as IAT-ING-Threshold (IAT-ING-Thresh) and IAT-IUG-Threshold (IAT-IUG-Thresh). On the ING IAT-ING-Thresh should be greater than IAT-ING-IntraObj and smaller than IAT-ING-InterObj:

$$\text{IAT-ING-IntraObj} < \text{IAT-ING-Thresh} < \text{IAT-ING-InterObj} \quad \text{(seconds)}$$

Similarly, on the IUG, it should be greater than IAT-IUG-IntraObj and smaller than IAT-IUG-InterObj:

$$\text{IAT-IUG-IntraObj} < \text{IAT-IUG-Thresh} < \text{IAT-IUG-InterObj} \quad \text{(seconds)}$$

Consequently, each processed segment is considered as a new object if the IAT of this segment is greater than IAT-IUG-Thresh or IAT-ING-Thresh, respectively, and is considered as the same object is the IAT is lower than IAT-IUG-Thresh or IAT-ING-Thresh, respectively.

False detections may occur in the cases where the IAT of a segment logically belonging to the same object is greater than IAT-ING-Thresh or IAT-IUG-Thresh, respectively. In this case a new object is wrongly detected. Similarly, the first segment of a logically new object might be detected falsely if the IAT of this segment is smaller than IAT-ING-Thresh or IAT-IUG-Thresh, respectively.

#### 5.2.3.2 Calculation of IAT-ING-Thresh

The IAT-ING-Thresh should be computed as follows:

$$\text{IAT-ING-Thresh} = (\text{SfD1} + \text{SfD2} + 2 \times \text{WanD}) \times \text{alpha}$$

With:

- Alpha represents a configuration parameter for future use cases. Currently should be set to $\geq 1$.
- SfD1 represents the latency on the connection between the IUG and ING. Since multiple connections exists it is either equal half of the roundtrip time (RTT) of the link with the lowest RTT, if the current object size is lower than the threshold differentiating long objects and short objects (see clause 5.4) , or half of the RTT of the link with the highest available bandwidth.
- SfD2 represents half of the RTT of the link with the lowest RTT.
- WanD represents the RTT between the ING and the server in the public network

#### 5.2.3.3 Calculation of IAT-IUG-Thresh

The IAT-IUG-Thresh should be computed as follows:

$$\text{IAT-IUG-Thresh} = (\text{SfD1} + \text{SfD2} + 2 \times \text{LanD}) \times \text{alpha}$$

With:

- Alpha represents a configuration parameter for future use cases. Currently should be set to $\geq 1$.
- SfD1 represents half of the RTT of the link with the lowest RTT.
- SfD2 represents the latency on the connection between the IUG and ING. Since multiple connections exists it is either equal half of the roundtrip time (RTT) of the link with the lowest RTT, if the current object size is lower than the threshold differentiating long objects and short objects (see clause 5.4) , or half of the RTT of the link with the highest available bandwidth.
- LanD represents the RTT between the end host in the home network environment and the IUG.

### 5.2.4 Classification of objects

Objects can be classified into long objects and short objects. An object is considered as long if it consists of more bytes than the long object threshold $\theta$, otherwise it is considered as short.

The value for $\theta$ needs to be adjusted based on the actual link configuration. An exact algorithm focused on how $\theta$ is determined is outside of the scope of the present document.

## 5.3 Link selection for TCP traffic

The link selection algorithm determining which subflow and, hence, which link is used, is operating on TCP objects. Long-objects benefit most from high-bandwidth links, which reduce the total time to transmit the object, while short-objects are most optimally sent via the link with the lowest latency. Thus, for long-objects the link with the highest available bandwidth is selected, whereas for short-objects the link with the lowest latency is selected.

Under normal conditions this implies that long-objects are sent via the satellite link while short-objects are sent terrestrially.

The identification of the link with the highest available bandwidth and the lowest latency is described in clause 5.4.

It should be noted that this is a process, which is started upon a packet arrival. That is, an object consisting of several segments might first be classified as short, when the first segment arrives, but once more segments are received on IUG or ING and, thus, the object size increases, the classification might change. Hence, segments belonging to the same object might be routed differently.

It should be noted also that this method based on packet arrival times apply to all application protocols , the data exchanged being encrypted or not at transport layer.

It is possible to increase more the user experience in using the aggregated bandwidth of some, or all, links.

Usually, to take advantage of the cumulated bandwidth of several links, algorithms such as weighted round robin or offload mode are used.

In weighted round robin each link has a weight proportional to its bandwidth and consecutive segments are sent on each link according to its weight

In offload mode consecutive segment of a flow (or an object) are sent on the first link if until its bandwidth is reached and then the following are sent on the second link and so on for the third link if any.

In these modes the bandwidths and weights are given by the dynamic link estimation described here after.

The user experience of latency T can be represented as:

$$T = Ts + Tl$$

where Ts and Tl are respectively the sum of the times spent for transmitting short and long objects. Ts is driven by the delays, and Tl by the bandwidths over which the data is sent.

The user experience will be maximized (i.e. T minimized) by sending short objects on the link having the lowest RTT and long objects on all the links in weighted round robin offload mode to provide the maximum bandwidth.

## 5.4 Link estimation for TCP traffic

A passive link status estimation process should be used. This mechanism estimates the path characteristics between the IUG and the ING. It estimates the path capacity, the packet loss and the round trip time without querying any device on the path, so that the Link selection can operate accordingly. The estimation is based on traffic analyses.

The link estimation module provides information on the path characteristics and should be implemented on both IUG and ING. Figure 6 depicts the chain of the processing modules.

In this figure relative to an IUG having a LAN and 3 link Ethernet interfaces, the link interfaces are connected respectively to a 3G/4G modem (cellular modem), an ADSL modem and a satellite modem. The 3G/4G, ADSL and Satellite modems hardware could be integrated to the IUG appliance.

A generic transparent proxy intercepts the TCP connections from the LAN (in TCP0) and split them on a 1st dedicated link interface (TCP1 handles a connection on the3G/4G interface). It is assumed that TCP1 has the MPTCP (MP-capable) option. This connection is intercepted by the remote ING.

Then, when receiving the SynAck with the MP Capable option from the ING, the IUG establishes MPTCP sub flows (handled by TCP2, TCP3) on the other dedicated link interfaces.

The link interfaces are configured with one VPN per link interface. These VPNs, that can be seen as basic IP tunnels, are established by the ING and are used to route the traffic through the ING. This is useful in multi operator situations where the ING cannot be placed on a shared Path from the clients to all the Internet servers. In this case, the use of a NAT service in the ING as for result having the responses from the servers coming to the ING.

This tunnel architecture is very useful in situations where the operator proposing the multilink service does not own all the links. It also handles cases where the multilink service is proposed by a virtual operator.

Once all the MPTCP flows have been established, the link selection algorithm can operate and route the data on the best suited subflow, thanks to the indications provided by the link estimation module.

```
+-------------------------------------------------------+
| IUG appliance      +-----------------+                |
|                    |  Generic Proxy  |                |
|                    |+--------------+ |                |
|                    ||Link Selection+-----+            |
|                    |+--------------+ |   |            |
|                    +-------+---------+   |            |
|                            |             |            |
|    +-----------------------+-------------|-------+    |
|    |                              +------+-----+ |    |
|    |                              |   Link     | |    |
|    |                              | Estimation | |    |
|    |                              +------------+ |    |
|    |       +------+------+------+--------+       |    |
|    |       | TCP1 | TCP2 | TCP3 |  TCP0  |       |    |
|    |       |      |      |      +--------+       |    |
|    |   TCP |      |      |      |QoS Tag |       |    |
|    |       |      |      |      +--------+       |    |
|    |       |      |      |      |  DPI   |       |    |
|    |       +------+------+------+--------+       |    |
|    |    IP | VPN1 | VPN2 | VPN3 | IP LAN |       |    |
|    |       +--------------------+--------+       |    |
|    +-------+------+------+------+--------+-------+    |
|      Phys. | eth1 | eth2 | eth3 | eth0   |            |
|            +------+------+------+--------+            |
+---------------+------+------+-------+-----------------+
                |      |      |       |
               WAN1   WAN2   WAN3    LAN
                |      |      |       |
                |      |      |       |
               3/4G   DSL    Sat.
```

**Figure 6: Link Status Estimation process**

The Link Estimation process operates on the transport layer and delivers real time estimates of the path characteristics.

The link estimation process analyses TCP and non-TCP traffic. The link estimation should estimate the bandwidth of TCP and non-TCP traffic. The sum of both estimations being the estimation of the overall link capacity in use.

The link estimation measures the path characteristics in real time between each IUG and ING, based on information gathered from TCP connections. The link estimation module updates the statistics each time a TCP packet is sent or received.

For each TCP connection, the following information should be available:

- Peer ID
- state: ESTABLISHED, FIN_WAIT, CLOSE_WAIT, etc.
- snd_wnd: system maximum send window
- snd_cwnd: current congestion window
- srtt: smoothed round trip time
- snd_ssthresh: slow start threshold. If (snd_cwnd < snd_ssthresh) then {slow start phase} else {congestion avoidance phase}
- sb_space: amount of available buffer in send buffer. If full, the connection is considered as active
- sb_cc: character count in send buffer. Data waiting to be sent out or unacknowledged data
- tcps_sack_rexmits: number of retransmitted packets from start of connection
- tcps_sack_rexmit_bytes: number of retransmitted bytes from start of connection

The link estimation is based on analysing real time traffic. An algorithm is defined to evaluate the bandwidth from the internal TCP connection states and variables. An initial version of the bandwidth algorithm is specified in Figure 7.

```
For each outgoing TCP segment on WAN {

   if (sb_cc < snd_cwnd)
      /* Handle low bw connections that do not fill cwmd */
      buf_sz = so->so_snd.sb_cc;
   else
      buf_sz = tp->snd_cwnd;

      avg_srtt = MOBILE_AVG_100( srtt );
      avg_cwnd = MOBILE_AVG_100( buf_sz );

   throughput = (avg_cwnd * 8 * nb_conn_estab * 1000) / avg_srtt
}
```

**Figure 7: Algorithm for bandwidth estimation based on TCP**

The bandwidth estimation is the product of the average of the congestion windows times the number of concurrent connections divided by the smoothed round trip time average of the path. If the amount of in-flight data is less than the congestion window, the server delivers data at lower bandwidth than the link capacity and, then, the amount data in the buffer instead of the congestion window size is considered.

The parameters avg_srtt and avg_cwnd should be mobile averages of the srtt and congestion windows based on the last 100 samples. This calculation should be performed for each outgoing packet on the all interfaces connecting IUG and ING for any TCP connections, the mobile average provides smooth changes in the srtt and cwnd estimations.

The use of the smoothed round trip time in the equation allows computing a stable throughput during congestion phases. Indeed, during congestion and prior a packet loss, the srtt increases as well as the congestion windows but the throughput on the network remains constant. The congestion phase corresponds to the phase where queues in the network are being filled until a packet is dropped in those queues. The algorithm provides a constant bandwidth, which is in-line with the throughput observed on the network.

This algorithm is expected to work better when the link capacity is fully loaded with TCP traffic. Each TCP connection reaches its maximum throughput until a loss is detected. The congestion window is reduced and does not increase continuously. The algorithm takes advantage of the evolution of the congestion windows that increases and decreases to adjust the throughput. It can then compute the bandwidth estimation close to the real network throughput.

The round trip time (RTT) and packet loss estimations can be computed precisely whatever the traffic pattern, even during low TCP transfer rate. The RTT and the packet loss rate are updated per TCP segment that are acknowledged:

- RTT is computed from smooth RTT (srtt) in TCP kernel control block.
- Loss rate between IUG and ING (or vice versa) is computed from segment retransmission counters in TCP kernel control block.

# 6 Conclusion

The present document provides a definition of functional modules for utilizing multipath TCP protocol in hybrid FSS satellite/terrestrial or terrestrial/terrestrial architectures. The method to split TCP traffic into objects and intelligently routing traffic chunks over multiple links has been described and apply to all the application protocols encrypted or not at transport layer A TCP based link estimation method to passively determine the available bandwidth and latency of each path has also been described.

According to the tests done, this method provides user experience gains that are very significant in a satellite/terrestrial architecture and significant in a terrestrial/terrestrial context. Some results for tests in a satellite/terrestrial architecture are presented in [i.3].

Current implementation of MPTCP uses all available subflows. However, the proposed Multilink TCP routing makes it possible to add and remove subflows based on the measured performance of each link due to its direct interaction with the network interfaces. A typical use case for this scenario is variable bitrate video streaming. Possible standardization actions include amendment to the IETF RFC 6824 [i.2] to consider traffic splitting based on Object length and consideration of high bandwidth, high RTT satellite links.

The Link for given UDP Traffic may be selected as the one minimizing a link cost function, defined as a combination of weighted criteria such as Link Reliability, One Way Maximum Latency, Available Bandwidth, Bandwidth cost and OPEX/CAPEX considerations.

# Annex A: Experimental Results

# A.1 Test description

The efficiency of the proposed method can be demonstrated by the following test using ssh.

This test consists, on a test set up like the one described in Figure 2, to launch from a client connected to the IUG Lan a script that performs a ssh connection to a remote Unix server on the Internet and execute the commands in a scenario file.

Three scenario files are tested: Short.scen, Long.scen, Mixed.scen:

- Short.scen: 70 short commands like ls, pwd, cd.; this scenario is expected to be well suited on a low delay link.
- Long.scen: cat 5 Mbytes remote file; this scenario is expected to be well suited on a high bandwidth link.
- Mixed.scen: all the commands in 2 above files; this scenario is expected to be well suited on a link having both an high bandwidth and a low RTT.

It is important to note that SSH communications are encrypted.

In this test a real prototype was used, connected to real networks and implementing the above described algorithms and methods.

The test was run in an ADSL+Satellite and an ADSL+4G architectures.

# A.2 GEO/LEO Satellite link combined with ADSL link in an hybrid architecture

This option clause encompasses either a GEO or a LEO satellite access.

In this test, the hybrid network is made of a 1 Mbps/70 ms RTT ADSL Link and a 8 Mbps/700 ms RTT Satellite link. The test is run on each link alone and on the two links using a path selection based on objects length + Weighted round robin algorithm.

That means that flows are split in objects: Short objects are sent on the lowest RTT link (hopefully ADSL) and long objects are sent in the weighted round robin mode on the ADSL (weight = 1) and Satellite (weight = 8) link.

The results are shown in Table A.1 and demonstrate as expected that this hybrid satellite + ADSL architecture performs as an equivalent 9 Mbps link with 70 ms RTT. This looks like the satellite being placed not on the GEO but near the ground.

The results for the long scenario in hybrid mode are better than those that would be seen with a simple bandwidth aggregation

This is due to the fact that the TCP establishment time + Ssh handshake take about 9 s on a satellite link and only about 2 s on a ADSL link. In the hybrid mode implementation the TCP establishment of the satellite subflow is done via the ADSL link and the ssh initialization messages are also sent as short objects via the ADSL link.

**Table A.1: Execution time of the three SSH based scenarios (Short, long and mixed) over different combinations of access links (ADSL, satellite or parallel combining of both)**

| Scenario execution time (seconds) | | | | | |
|---|---|---|---|---|---|
| | ADSL only | Satellite only | ADSL+Satellite (with PSBOL+WRR scheme) | Gain Vs ADSL | Gain vs Sat |
| **Short** | 10 | 79 | **10** | 0 % | 87 % |
| **Long** | 47 | 19 | **10** | 79 % | 47 % |
| **Mixed** | 56 | 70 | **17** | 70 % | 76 % |

# A.3 ADSL link combined with a 4G/LTE access in an hybrid architecture

In this test, the hybrid network is made of a 2,5 Mbs/70 ms RTT ADSL Link and a 15 Mbs/75 ms RTT LTE link. The test is run on each link alone and on the two links using a path selection based on objects length + Offload. That means that flows are split in objects: Short objects are sent on the lowest RTT link (hopefully ADSL) long objects are sent in offload mode on the ADSL until the 2,5 Mbs limit is reached and on the 4G/LTE link after.

The test is done on a loaded network .the load being obtained by download of a 20 MB file running in loop.

Only the results for the mixed scenario are shown due to a limited access time to the foreign test platform.

The results are shown in Table A.2 and demonstrate that in this pure terrestrial hybrid architecture the multilink network performs as an equivalent 17,5 Mbs/70 ms RTT and delivers substantial user experience gain compared to the non-hybrid situation with each link used alone.

Looking a bit more to the logs during the test it appears that the RTT of the 4G/LTE link increases from 75 ms up to 200 ms.

This is due to the offload algorithm that does not overload the ADSL link and keeps its RTT stable at the same time the data in excess are sent over LTE that is overloaded and the QoS queues are filed introducing some latency.

**Table A.2: Execution time of the mixed SSH based scenarios over different combinations of access links (ADSL, 4G/LTE or parallel combining of both) under typical network load conditions**

| Scenario execution time (seconds) | | | | | |
|---|---|---|---|---|---|
| | ADSL | LTE | ADSL+LTE PSBOL +Offload | Gain Vs ADSL | Gain Vs LTE |
| **Mixed** | 49 | 22 | **13,5** | 73 % | 39 % |

# History

| Document history | | |
|---|---|---|
| V1.1.1 | July 2017 | Publication |
| | | |
| | | |
| | | |
| | | |