MPTCP Working Group
Internet-Draft
Intended status: Informational
Expires: January 6, 2017

B. Peirens
Proximus
G. Detal
S. Barre
O. Bonaventure
Tessares
July 05, 2016

Link bonding with transparent Multipath TCP
draft-peirens-mptcp-transparent-00

Abstract

This document describes the utilisation of the transparent Multipath TCP mode to enable network operators to provide link bonding services in hybrid access networks.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 6, 2017.

Copyright Notice

Table of Contents

1. Introduction

Internet Service Provider networks are composed of different parts : access networks, metropolitan and wide area networks. Given the growing demand for bandwidth, these networks must evolve. In the metropolitan and wide area parts, bandwidth increases thanks to the utilisation of optical fiber or through link aggregation. Increasing bandwidth in the core is not sufficient to allow all users to benefit from faster services. For many operators, the last-mile of the access network remains a bottleneck that is difficult to upgrade.

Many service providers do not rely on a single access network technology. They often have deployed different access networks that were initially targeted at different types of users and customers. Such access networks include xDSL, DOCSIS, FTTx and various wireless technologies (3G, 4G, Wimax, satellite, 5G, ...). With these different access networks, service providers have different ways to reach their customers and combining two access networks would enable them to deliver higher bandwidth services to their customers [I-D.zhang-banana-problem-statement].

In this document, we first describe in section Section 2 the hybrid access networks that are being deployed by various network operators. We focus on the aggregation of a fixed network (e.g. xDSL) with a cellular network (e.g. LTE). Many operators wish to use the bandwidth that is not consumed by the mobile devices on their cellular network to deliver better services to their fixed line customers. Section Section 3 lists the main requirements expressed by these operators. Section Section 4 briefly evaluates whether the main proposed bonding techniques meet those requirements. We then describe in section Section 5 how a transparent mode of operation for Multipath TCP [RFC6824] can be used to meet those operator requirements.

2. Reference architecture

Our reference architecture is shown in figure Figure 1. We use a similar terminology as in [WT-348] and consider the following elements :

- a single homed end host H that is attached through one interface (e.g. WiFi) to a Hybrid Customer Premisses Equipment (HCPE)
- a Hybrid Customer Premises Equipment (HCPE) that is connected to two different access networks. The solution proposed in this document support any number of access networks, but we restrict our examples to two.
- A Hybrid Aggregation Gateway (HAG) that is reachable over both access networks
- a regular server, S

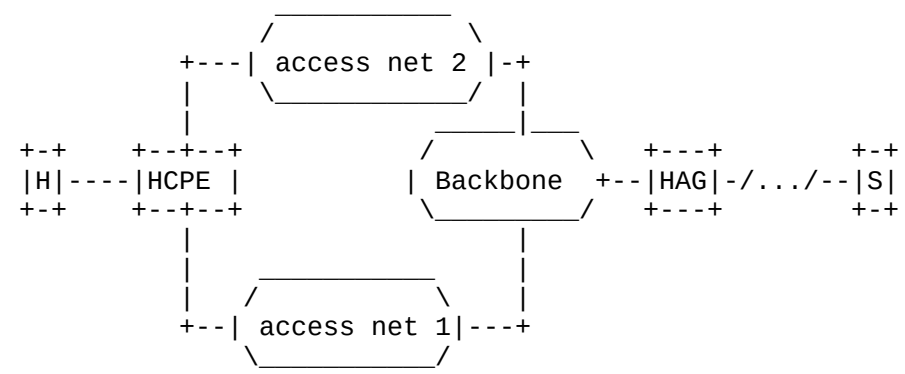

Figure 1: Hybrid access networks

We assume that IP addresses are assigned according to the best current practices, i.e. host H is allocated one IP address, and one IP address is assigned to each interface of the HCPE. Furthermore, BCP 38 [RFC2827] is used on the two access networks attached to the HCPE. The solution proposed in this document is agnostic of the IP version that is used. It operates equally well with both IPv4 and IPv6 and can use any mix of IPv4/IPv6. When writing IP addresses, we use the @ notation. For example, H@ is the IP address assigned to host H, HCPE@1 is the IP address assigned to the HCPE on access network 1,... For most network operators, the different access networks that need to be aggregated are not equivalent. One network, typically a fixed access network managed by the operator, is

considered to be the main access network. The other access network, possibly managed by another network operator, is used to provide additional capacity to cope with bandwidth limitations on the primary access network. We focus on this bandwidth aggregation scenario in this document. While the second access network can also be used in case of failure of the primary access network we currently leave it out of scope of the solution (existing solutions are already deployed by operators for this).

3. Operator requirements

Many operators have expressed their interest in efficiently supporting hybrid access networks. We list here some of the requirements that they have identified and have guided the design of the proposed solution.

- Req1: the bonding solution MUST support IPv6 and IPv4
- Req2: the bonding solution SHOULD minimize the additional delay that it introduces in the network
- Req3: the bonding solution MUST not expose multiple addresses for a given customer and the same address MUST be used for all transport protocols used by each customer
- Req4 : the bonding solution MUST not use more than one public IPv4 address per customer
- Req5 : the bonding solution SHOULD enable the operator to trace the connections created by a given customer
- Req6 : the bonding solution MUST monitor the quality of the different links and adapt the load distribution dynamically according to the load and the operator’s policies
- Req7 : the bonding solution MUST be decoupled from the underlying fixed/mobile access network
- Req8: the bonding solution MUST be able to efficiently load-balance the packets belonging to a single TCP connection over several access networks
- Req9: the bonding solution SHOULD not change the subscriber service attachment and authentication point in the network.

The second requirement reflects the importance of minimising the latency. Many applications, including HTTP, are affected by any increased latency. The third requirement reflects operational issues. Many applications expect that all the flows originated by a host will have the same source address, independently of the protocol used for each flow. A solution that would use different addresses for different transport protocols or for flows that do not benefit from hybrid access (e.g. by defined policy), would cause operational problems. The fifth requirement reflects the desire of the network operator to have some visibility of the flows that pass through its access network in order to apply filtering rules, log flows or provide QoS. The sixth requirement reflects the fact that the

bandwidth of the access networks that are aggregated can vary quickly. This is particularly the case for cellular networks where mobile devices could have priority over the bonding service. The last two requirements correspond to the utilisation of large TCP flows. Measurement studies in access networks show that TCP is the dominant protocol in these networks and that most of the data volume is carried by long TCP connections. Such connections must be load-balanced on a per packet basis to achieve a good aggregation.

4. Existing solutions

In this document, we focus on solutions that can combine very different access network technologies, typically a fixed line access network such as xDSL and a wireless access network such as LTE. We discuss only some of the proposed techniques. A complete overview of all the available solutions is outside the scope of this document.

4.1. Datalink solutions for hybrid access networks

Some datalink technologies, such as Multilink PPP [RFC1990], can load balance packets over different links. Unfortunately, they cannot easily be used in hybrid access networks that rely on different types of datalinks.

4.2. Network layer solutions for hybrid access networks

Various solutions exist in the network layer. A first possibility is to assign the same address to the HCPE (and thus the hosts behind it) over the different access networks. This requires a specific configuration of the routing in the access network and some network operators have deployed such solutions. Per-flow and per-packet load balancing are possible with this approach. Unfortunately, it has a number of important drawbacks :

- it is difficult for the HAG/HCPE to measure the performance of the different access networks in to adjust their utilisation in realtime (Req6)

- assigning the same address to the HCPE over different networks requires integration on the subscriber attachment points for both the fixed and mobile network (e.g. BNG & P-GW) for the bonding solution which might not be desirable (Req7)

- if packets from a transport connection are spread over different access networks, they experience different delays and different levels of congestion, but the transport protocol is unaware of those different networks. The net result is a lower throughput since the congestion control scheme adapts the throughput to the access network offering the lowest performance (Req8).

An alternative to assigning the same IP addresses on the different access networks is to use tunnels between the HCPE and the HAG. Various types of IP tunnels are possible [RFC2784] [I-D.zhang-gre-tunnel-bonding]. With such tunnels, the problems mentioned above remain and the tunneling protocol adds a per-packet overhead which may be significant in some environments. Extensions have been recently proposed to include flow control mechanisms in

some of these tunneling techniques [I-D.zhang-banana-tcp-in-bonding-tunnels] but this increases the complexity of the solution. Tunnel based solutions assign the external exposed customer address within the tunnel and change the subscriber service attachment point (Req9) which forces operators to re-implement authentication, logging and service definitions at a different location than the non-hybrid access customers. See also concerns listed in the next section {#transport}.

4.3. Transport layer solutions

The Multipath TCP plain mode option [I-D.boucadair-mptcp-plain-mode] was recently proposed as a solution to cope with some of the above problems of the network layer solutions. This solution is an extension of the TCP option proposed in [HotMiddlebox13b]. With the plain mode option, the HAG maintains a pool of public addresses that are used to translate the client addresses. From an addressing viewpoint, this is equivalent to the deployment of a carrier-grade NAT which leads to operational problems for the management of access-lists that are used to provide QoS, firewalling, but also for the collection of meta data about customer traffic, logs, ... With [I-D.boucadair-mptcp-plain-mode], all the TCP traffic in the access networks appears to be destined to the HAG.

While the Multipath TCP plain mode optionally allows transparency of the source address by using the option a second time with D-bit set to zero, it would require subscriber session information from the network element that assigned the now embedded source address to correctly implement BCP-38 [RFC2827] validation when restoring this at the HAG.

4.4. Application layer solutions

The SOCKS protocol [RFC1928] was designed to enable clients behind a firewall to establish TCP connections through a TCP proxy running on the firewall. A possible deployment in hybrid access networks is to use the HAG as a SOCKS server over Multipath TCP to benefit from its aggregation capabilities. Since regular hosts usually do not use a SOCKS client and do not support Multipath TCP, the HCPE needs to act as a transparent TCP/Multipath-TCP+SOCK proxy.

Compared with the network layer solutions and [I-D.boucadair-mptcp-plain-mode], the SOCKS approach has several drawbacks from an operational viewpoint :

- the HAG must maintain a pool of public addresses

- to establish a TCP connection through a SOCKS server running on the HAG, the HCPE must first perform the three-way handshake and then exchange SOCKS messages to authenticate the client (the number of messages is function of the SOCKS authentication scheme that is used). This increases the establishment time for each TCP connection by one or more additional round-trip times (Req2).

5. The transparent MPTCP mode

The transparent MPTCP mode proposed in this documented was designed under the assumption that in many hybrid access networks, there is a primary access network and the other access network(s) that are combined are used to (virtually) increase the capacity of the primary access network. In such networks, operators usually expect that the secondary access networks will only be used if the primary access networks does not have sufficient capacity to handle the load.

The solution is targeted at TCP traffic only. Non TCP traffic is sent over the primary access network. Support for other transport protocols over the secondary access networks is outside the scope of this document.

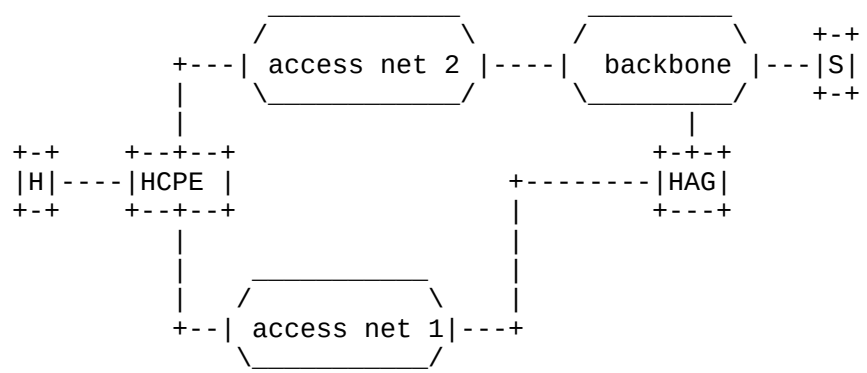

Figure 2: Reference architecture

We consider the network environment shown in figure Figure 2. Access net 1 is the primary network. This figure reflects the specific network configuration that is required for the transparent Multipath TCP mode. The HAG MUST reside on the path followed by the packets sent to/from the HCPE that it serves. This can be achieved, by e.g. using a specific mobile APN that has restricted routing, using service chaining at BNG/BRAS, using specific BNG/BRAS domains or AAA/ RADIUS triggered policy routing at BNG/BRAS. Several vendors have implemented such solutions and they are deployed in various networks.

A HAG typically serves a group of HCPEs and several HAGs can be deployed by an operator. Note that the requirement of placing the HAG on the path of the HCPE that it serves only applies to the primary access network. The other access networks only need to be able to reach the HAG. They do not need direct Internet access.

The HCPE has two IP addresses (or IP prefixes in the case of IPv6

prefix delegation) : HCPE@1 and HCPE@2. HCPE@1 is the primary address prefix assigned to the HCPE and host H uses one address from this prefix as its public address when contacting remote servers (we assume IPv6 in this description. With IPv4, the HCPE will assign a private IPv4 address to the hosts that it serves and will perform NAT). The HAG has one IP address that is reachable from the secondary network, identified as HAG@2. This is illustrated by the vertical link on the HAG in Figure 2.

Both the HCPE and the HAG include a transparent Multipath-TCP/TCP proxy. Various forms of TCP proxies have been defined and are deployed [RFC3135]. The HCPE uses its TCP/Multipath TCP proxy to convert the regular TCP connections initiated by the client host, H, into Multipath TCP connections towards the distant server. However, these Multipath TCP connections do not directly reach the distant server. They are converted into regular TCP connections by the Multipath-TCP/TCP proxy running on the HAG. This is illustrated in figure Figure 3.

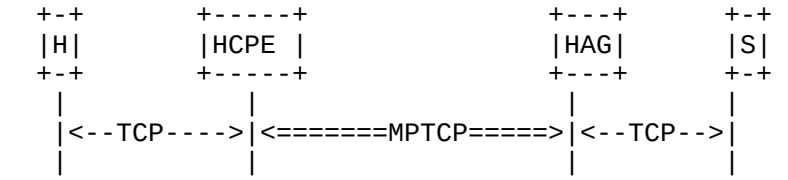

Figure 3: The TCP<->Multipath TCP proxies used on the HCPE and the HAG

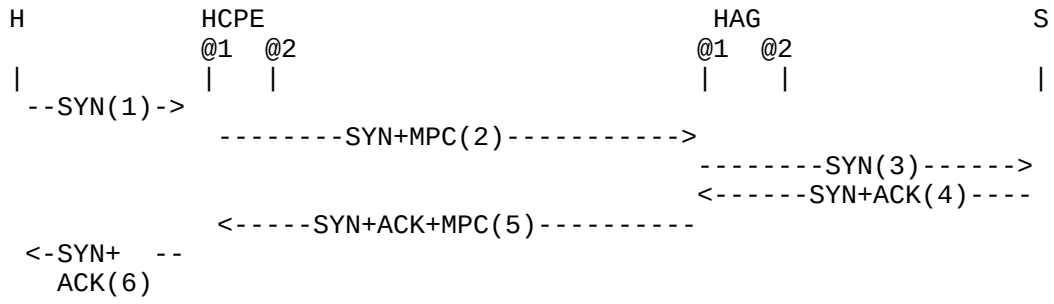

Figure 4: Creation of the initial subflow with the transparent mode

The operation of the transparent mode is illustrated in figure Figure 4. We consider the establishment of one TCP connection from host H (using address H@) to a distant server, S@. The following

packets are exchanged :

- (1) H sends a SYN towards S@.

- (2) The HCPE intercepts the SYN of the initial handshake. It creates some state for a regular TCP connection between H@ and itself acting as a transparent proxy for S@ and a Multipath TCP connection towards S@. These two connections are linked together and any data received over one connection is forwarded over the other. The HCPE then sends a SYN with the MP_CAPABLE option towards S@ over its primary access network to create a Multipath TCP connection to the HAG. Over the primary access network, this SYN appears as originating from H@ and being sent to S@.

- (3) The HAG acts as a transparent proxy for S@ and intercepts the SYN that contains the MP_CAPABLE option. It creates some state for this Multipath TCP connection and initiates a regular TCP connection towards S@. It should be noted that neither the HCPE nor the HAG perform address translation. The distant server receives the SYN from the client as originating from address H@.

- (4) The server replies with a SYN+ACK to confirm the establishment of the connection.

- (5) The HAG intercepts the returning SYN+ACK. The HAG then sends a SYN+ACK with the MP_CAPABLE option to confirm the establishment of the Multipath TCP connection that is proxied by the HCPE.

- (6) The HCPE sends a SYN+ACK to the client host to confirm the establishment of the regular TCP connection

At this point, the establishment of the three connections can be finalised by sending a third ACK. Data can be exchanged by the client and the server through the proxied connections.

The end-to-end connection between the client host (H) and the server (S) is composed of three TCP connections : a regular TCP connection between the host and the HCPE, a Multipath TCP connection between the HCPE and the HAG and a regular TCP connection between the HAG and the remote server. All the packets sent on these three connections contain the H@ and S@ addresses in their IP header.

To use another access network, the HAG simply advertises its address reachable through this access network (HAG@2) on the initial subflow by sending an ADD_ADDR option (1). This triggers the establishment of an additional subflow from the HCPE over the second access network (arrows (2), (3) and (4) in figure Figure 5). The endpoints of this subflow are the IP address of the HCPE on the second access network,

i.e. HCPE@2, and the IP address of the HAG, i.e. HAG@2. Note that the ADD_ADDR option shown in figure Figure 5 is optional. If the HCPE already knows, e.g. by configuration or through mechanisms such as [I-D.boucadair-mptcp-radius] or [I-D.boucadair-mptcp-dhc], the IP address of the HAG, it can crate the additional subflow without waiting for the ADD_ADDR option.

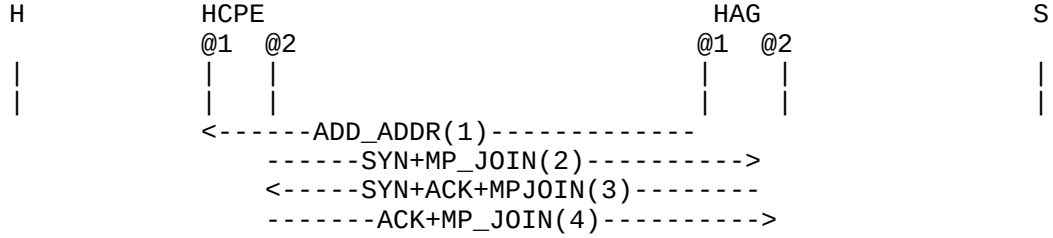

Figure 5: Creation of the second subflow by the HCPE with the transparent MPTCP mode

At this point, any data received from the host by the HCPE or from the server by the HAG can be transported over any of the established subflows. Both the HAG and the HCPE select the most appropriate subflow based on their policies and the current network conditions that are automatically measured by Multipath TCP.

This is not the only way to create additional subflows. The HAG may also create additional subflows. This is illustrated in figure Figure 6 where we assume that the HAG already knows the IP address of the HCPE and thus does not wait for the reception of an ADD_ADDR option from the HCPE to create the additional subflow.

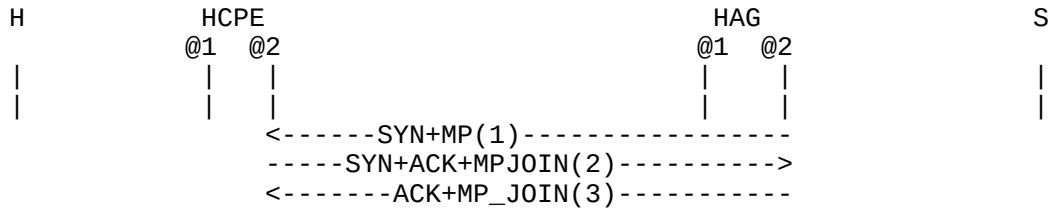

Figure 6: Creation of the second subflow by the HAG with the transparent MPTCP mode

6. Security considerations

Providing a bonding service through different access networks introduces new capabilities, but also new threats to the network. We focus in this section on the threats that are specific to the bonding service and assume that the CPE devices implement the recommendations listed in [RFC6092]. For the HAG, since it operates on the path like a router, many of the the security considerations for routers apply.

When Multipath TCP is used over different paths, the security threats listed in [RFC6181] and [RFC7430] need to be considered. Some of these can be mitigated through proper configuration of the HCPEs, HAGs and access networks.

An important security threat with Multipath TCP is if an attacker were able to inject data on an existing Multipath TCP by associating an additional subflow. Such an attack is already covered by the utilisation of keys in the Multipath TCP handshake. Thanks to the utilisation of the tokens and the HMAC in the MP_JOIN option, the HAG and the HCPE will refuse additional subflows created by an attacker who did not observe the initial SYN packets. Note that since the keys are only exchanged on the first access network, this attacker would have to reside on this access network.

Since the HAG and the HCPE only create subflows among themselves, it is possible for an operator to configure those devices to only accept SYN packets with the MP_CAPABLE or MP_JOIN option to those prefixes. Furthermore, the second access network does not need to be connected to the Internet. This implies that an attacker would need to reside on this network to send packets towards the visible address of the HAG. Ingress filtering and uRPF should be deployed on the access networks to prevent spoofing attacks.

If TCP connections originating from the Internet are accepted on the HCPEs, then the HAG must be secured against denial of service attacks since it will be involved in the processing of all incoming SYN packets.

7. IANA Considerations

There are no IANA considerations in this document.

8. Conclusion

In this document, we have proposed the transparent mode for Multipath TCP and described its utilisation in hybrid access networks where a secondary access network is used to complement a primary access network. Our solution leverages the flow and congestion control capabilities of Multipath TCP to allow an efficient utilisation of the different access networks, even if their capacity fluctuates.

Compared with network layer solutions such as [I-D.zhang-gre-tunnel-bonding], the transparent mode does not introduce any per-packet overhead and does not require any form of network address translation. Compared with the plain mode Multipath TCP proposed in [I-D.boucadair-mptcp-plain-mode], our solution does not require any form of network address translation which has clear operational benefits.

9. References

9.1. Normative References

[RFC6824] Ford, A., Raiciu, C., Handley, M., and O. Bonaventure, "TCP Extensions for Multipath Operation with Multiple Addresses", RFC 6824, DOI 10.17487/RFC6824, January 2013, <http://www.rfc-editor.org/info/rfc6824>.

9.2. Informative References

[HotMiddlebox13b] Detal, G., Paasch, C., and O. Bonaventure, "Multipath in the Middle(Box)", HotMiddlebox'13 , December 2013, <http://inl.info.ucl.ac.be/publications/multipath-middlebox>.

[I-D.boucadair-mptcp-dhc] Boucadair, M., Jacquenet, C., and T. Reddy, "DHCP Options for Network-Assisted Multipath TCP (MPTCP)", draft-boucadair-mptcp-dhc-05 (work in progress), May 2016.

[I-D.boucadair-mptcp-plain-mode] Boucadair, M., Jacquenet, C., Behaghel, D., stefano.secci@lip6.fr, s., Henderickx, W., Skog, R., Bonaventure, O., Vinapamula, S., Seo, S., Cloetens, W., Meyer, U., and L. Contreras, "An MPTCP Option for Network-Assisted MPTCP Deployments: Plain Transport Mode", draft-boucadair-mptcp-plain-mode-08 (work in progress), July 2016.

[I-D.boucadair-mptcp-radius] Boucadair, M. and C. Jacquenet, "RADIUS Extensions for Network-Assisted Multipath TCP (MPTCP)", draft-boucadair-mptcp-radius-01 (work in progress), January 2016.

[I-D.zhang-banana-problem-statement] Cullen, M., Leymann, N., Heidemann, C., Boucadair, M., Hui, D., Zhang, M., and B. Sarikaya, "Problem Statement: Bandwidth Aggregation for Internet Access", draft-zhang-banana-problem-statement-02 (work in progress), July 2016.

[I-D.zhang-banana-tcp-in-bonding-tunnels] Zhang, M., Leymann, N., Heidemann, C., and M. Cullen, "Flow Control for Bonding Tunnels", draft-zhang-banana-tcp-in-bonding-tunnels-00 (work in progress), March 2016.

[I-D.zhang-gre-tunnel-bonding]
Leymann, N., Heidemann, C., Zhang, M., Sarikaya, B., and M. Cullen, "Huawei's GRE Tunnel Bonding Protocol", draft-zhang-gre-tunnel-bonding-03 (work in progress), May 2016.

[RFC1928] Leech, M., Ganis, M., Lee, Y., Kuris, R., Koblas, D., and L. Jones, "SOCKS Protocol Version 5", RFC 1928, DOI 10.17487/RFC1928, March 1996, <http://www.rfc-editor.org/info/rfc1928>.

[RFC1990] Sklower, K., Lloyd, B., McGregor, G., Carr, D., and T. Coradetti, "The PPP Multilink Protocol (MP)", RFC 1990, DOI 10.17487/RFC1990, August 1996, <http://www.rfc-editor.org/info/rfc1990>.

[RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784, DOI 10.17487/RFC2784, March 2000, <http://www.rfc-editor.org/info/rfc2784>.

[RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, DOI 10.17487/RFC2827, May 2000, <http://www.rfc-editor.org/info/rfc2827>.

[RFC3135] Border, J., Kojo, M., Griner, J., Montenegro, G., and Z. Shelby, "Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations", RFC 3135, DOI 10.17487/RFC3135, June 2001, <http://www.rfc-editor.org/info/rfc3135>.

[RFC6092] Woodyatt, J., Ed., "Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service", RFC 6092, DOI 10.17487/RFC6092, January 2011, <http://www.rfc-editor.org/info/rfc6092>.

[RFC6181] Bagnulo, M., "Threat Analysis for TCP Extensions for Multipath Operation with Multiple Addresses", RFC 6181, DOI 10.17487/RFC6181, March 2011, <http://www.rfc-editor.org/info/rfc6181>.

[RFC7430] Bagnulo, M., Paasch, C., Gont, F., Bonaventure, O., and C. Raiciu, "Analysis of Residual Threats and Possible Fixes for Multipath TCP (MPTCP)", RFC 7430, DOI 10.17487/RFC7430, July 2015, <http://www.rfc-editor.org/info/rfc7430>.

[WT-348] Broadband Forum, ., "Hybrid Access for Broadband Network", 2014, <http://datatracker.ietf.org/liaison/1355/>.

Authors' Addresses

Bart Peirens
Proximus

Email: bart.peirens@proximus.com

Gregory Detal
Tessares

Email: Gregory.Detal@tessares.net

Sebastien Barre
Tessares

Email: Sebastien.Barre@tessares.net

Olivier Bonaventure
Tessares

Email: Olivier.Bonaventure@tessares.net