

Network Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: September 10, 2017

M. Boucadair, Ed.  
C. Jacquenet, Ed.  
Orange  
O. Bonaventure, Ed.  
Tessares  
D. Behaghel  
OneAccess  
S. Secci  
UPMC  
W. Henderickx, Ed.  
Nokia/Alcatel-Lucent  
R. Skog, Ed.  
Ericsson  
S. Vinapamula  
Juniper  
S. Seo  
Korea Telecom  
W. Cloetens  
SoftAtHome  
U. Meyer  
Vodafone  
LM. Contreras  
Telefonica  
B. Peirens  
Proximus  
March 9, 2017

Extensions for Network-Assisted MPTCP Deployment Models  
draft-boucadair-mptcp-plain-mode-10

Abstract

Because of the lack of Multipath TCP (MPTCP) support at the server side, some service providers now consider a network-assisted model that relies upon the activation of a dedicated function called MPTCP Conversion Point (MCP). Network-Assisted MPTCP deployment models are designed to facilitate the adoption of MPTCP for the establishment of multi-path communications without making any assumption about the support of MPTCP by the communicating peers. MCPs located in the network are responsible for establishing multi-path communications on behalf of endpoints, thereby taking advantage of MPTCP capabilities to achieve different goals that include (but are not limited to) optimization of resource usage (e.g., bandwidth aggregation), of resiliency (e.g., primary/backup communication paths), and traffic offload management.

This document specifies extensions for Network-Assisted MPTCP deployment models.

#### Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

#### 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 September 10, 2017.

#### Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
2. Terminology . . . . .	5
3. Target Use Cases . . . . .	6
3.1. Multipath Client . . . . .	6
3.2. Multipath CPE . . . . .	7
4. The MP_PREFER_PROXY MPTCP Option . . . . .	8
4.1. Option Format . . . . .	8
4.2. Option Processing . . . . .	8
5. Supplying Data to MCPs . . . . .	9
5.1. The MP_CONVERT Information Element . . . . .	9
5.2. Processing an MP_CONVERT Information Element . . . . .	11
6. MPTCP Connections from a Multipath TCP Client . . . . .	13
6.1. Description . . . . .	13
6.2. Theory of Operation . . . . .	14
7. MPTCP Connections Between Single Path Client and Server . . . . .	16
7.1. Description . . . . .	16
7.2. Theory of Operation . . . . .	17
7.2.1. Downstream MCP . . . . .	17
7.2.2. Upstream MCP . . . . .	17
8. Interaction with TFO . . . . .	19
9. IANA Considerations . . . . .	20
10. Security Considerations . . . . .	21
10.1. Privacy . . . . .	21
10.2. Denial-of-Service (DoS) . . . . .	21
10.3. Illegitimate MCP . . . . .	21
11. Acknowledgements . . . . .	21
12. References . . . . .	22
12.1. Normative References . . . . .	22
12.2. Informative References . . . . .	22
Authors' Addresses . . . . .	23

## 1. Introduction

The overall quality of connectivity services can be enhanced by combining several access network links for various purposes - resource optimization, better resiliency, etc. Some transport protocols, such as Multipath TCP [RFC6824], can help achieve such better quality, but failed to be massively deployed so far.

The support of multipath transport capabilities by communicating hosts remains a privileged target design so that such hosts can directly use the available resources provided by a variety of access networks they can connect to. Nevertheless, network operators do not control end hosts while the support of MPTCP by content servers remains close to zero.

Network-Assisted MPTCP deployment models are designed to facilitate the adoption of MPTCP for the establishment of multi-path communications without making any assumption about the support of MPTCP capabilities by communicating peers. Network-Assisted MPTCP deployment models rely upon MPTCP Conversion Points (MCPs) that act on behalf of hosts so that they can take advantage of establishing communications over multiple paths. MCPs can be deployed in CPEs (Customer Premises Equipment), as well as in the provider's network. MCPs are responsible for establishing multi-path communications on behalf of endpoints. Further details about the target use cases are provided in Section 3.

Most of the current operational deployments that take advantage of multi-interfaced devices rely upon the use of an encapsulation scheme (such as [I-D.zhang-gre-tunnel-bonding], [TR-348]). The use of encapsulation is motivated by the need to steer traffic towards the concentrator and also to allow the distribution of any kind of traffic besides TCP (e.g., UDP) among the available paths without requiring any advanced traffic engineering tweaking technique in the network to intercept traffic and redirect it towards the appropriate MCP.

Current operational MPTCP deployments by network operators are focused on the forwarding of TCP traffic. The design of such deployments sometimes assumes the use of extra signalling provided by SOCKS [RFC1928], at the cost of additional management complexity and possible service degradation (e.g., up to 6 SOCKS messages may have to be exchanged between two MCPs before actual payload data to be transferred, thereby yielding several tens of milliseconds of extra delay before the connection is established) .

To avoid the burden of encapsulation and additional signalling between MCPs, this document explains how a plain transport mode is enabled, so that packets are exchanged between a device and its upstream MCP without requiring the activation of any encapsulation scheme (e.g., IP-in-IP [RFC2473], GRE [RFC1701]). This plain transport mode also avoids the need for out-of-band signalling, unlike the aforementioned SOCKS context.

The solution described in this document also works properly when NATs are present in the communication path between a device and its upstream MCP. In particular, the solution in this document accommodates deployments that involve CGN (Carrier Grade NAT) upstream the MCP.

Network-Assisted MPTCP deployment and operational considerations are discussed in [I-D.nam-mptcp-deployment-considerations].

The plain transport mode is characterized as follows:

- o 0-RTT proxy.
- o No encapsulation required (no tunnels, whatsoever).
- o No out-of-band signaling for each MPTCP subflow is required.
- o Targets both on-path and off-path MCPs.
- o Avoids interference with native MPTCP connections.
- o Assists MPTCP connections even if endpoints are MPTCP-capable.
- o Accommodates various deployment contexts, such as those that require the preservation of the source IP address and others characterized by an address sharing design. In particular:
  - \* This solution is compatible with IPv4/IPv6.
  - \* This solution does not impose any constraint on the addressing scheme to be used by the client.
  - \* This solution does not require nor exclude the use of distinct IP prefix pools for network-assisted MPTCP deployments.
  - \* This solution supports both transparent and non-transparent operations.

## 2. Terminology

The reader should be familiar with the terminology defined in [RFC6824].

This document makes use of the following terms:

- o Client: an endhost that initiates transport flows forwarded along a single path. Such endhost is not assumed to support multipath transport capabilities.
- o Server: an endhost that communicates with a client. Such endhost is not assumed to support multipath transport capabilities.
- o Multipath Client: a Client that supports multipath transport capabilities.
- o Multipath Server: a Server that supports multipath transport capabilities. Both the client and the server can be single-homed or multi-homed. However, for the use cases discussed in this document, the number of interfaces available at the endhosts is not relevant.
- o Transport flow: a sequence of packets that belong to a unidirectional transport flow and which share at least one common characteristic (e.g., the same destination address). TCP and SCTP flows are composed of packets that have the same source and

destination addresses, the same protocol number and the same source and destination ports.

- o Multipath Conversion Point (MCP): a function that terminates a transport flow and relays all data carried in the flow into another transport flow.

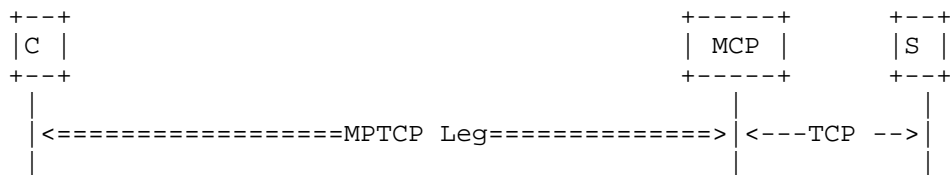
MCP is a function that converts a multipath transport flow and relays it over a single path transport flow and vice versa.

### 3. Target Use Cases

We consider two important use cases in this document. We briefly introduce them in this section and leave the details to Section 6 and Section 7. The first use case is a Multipath Client that interacts with a remote Server through a MCP (Section 3.1). The second use case is a multi-homed CPE that includes a MCP and interacts with a remote Server through a downstream MCP (Section 3.2).

#### 3.1. Multipath Client

In this use case, the Multipath Client would like to take advantage of MPTCP even if the Server does not support MPTCP. A typical example is a smartphone that could use both WLAN and LTE access networks to reach a server in order to achieve higher bandwidth or better resilience.



Legend:  
 C: Client  
 MCP: Multipath Conversion Point  
 S: Server

Figure 1: Network-assisted MPTCP (Host-based Model)

In reference to Figure 1, the MCP terminates the MPTCP connection established by the client and binds it to a TCP connection towards the remote server. Two deployments of this use case are possible.

A first deployment is when the MCP is on the path between the Multipath Client and the Server. In this case, the MCP can terminate the MPTCP connection initiated by the Client and binds it to a TCP

connection that the MCP establishes with the Server. When the MCP is not located on all default forwarding paths, the MPTCP connection must be initiated by using the path where the MCP is located.

A second deployment is when the MCP is not on the path between the Multipath Client and the Server. In this case, the Client must first initiate a connection towards the MCP and request it to initiate a TCP connection towards the Server. This is what the SOCKS protocol performs by exchanging control messages to create appropriate mappings to handle the connection. Unfortunately, this requires additional round-trip-time that affects the performance of the end-to-end data transfer, in particular for short-lived connections.

This document specifies the MP\_CONVERT Information Element that is carried in the SYN segment of the initial subflow. This SYN segment is sent towards the MCP. The MP\_CONVERT Information Element contains the destination address (and optionally a port number) of the Server. Thanks to this information, the MCP can immediately establish the TCP connection with the Server without any additional round-trip-time, unlike a SOCKS-based MPTCP design.

### 3.2. Multipath CPE

In this use case, neither the Client nor the Server support MPTCP. Two MCPs are used as illustrated in Figure 2. The upstream MCP is embedded in the CPE while the downstream MCP is located in the provider's network. The CPE is attached to multiple access networks (e.g., xDSL and LTE). The upstream MCP transparently terminates the TCP connections initiated by the Client and converts them into MPTCP connections.

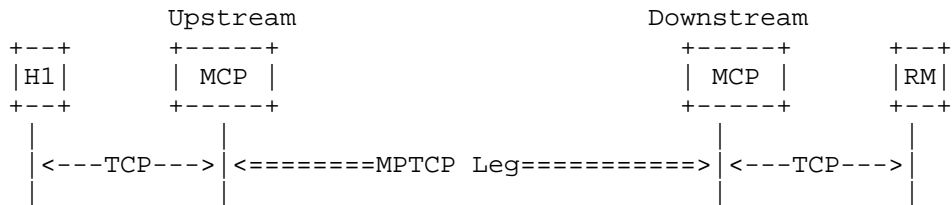


Figure 2: Network-assisted MPTCP (CPE-based Model)

The same considerations detailed in Section 3.1 apply for the insertion of the downstream MCP in an MPTCP connection.

4. The MP\_PREFER\_PROXY MPTCP Option

The implicit mode assumes that the MCP is located on a default forwarding path (Section 5.2.2 of [I-D.nam-mptcp-deployment-considerations]). In such mode, the first subflow must always be placed over that primary path so that the MCP can intercept MPTCP flows. Once intercepted, the MCP advertises its reachability information by means of MPTCP signals (MP\_JOIN or ADD\_ADDR).

In order to distinguish native MPTCP connections from proxied ones, a new MPTCP option, called MP\_PREFER\_PROXY, is defined. This option is meant to inform an on-path MCP that the connection should be proxied. The absence of the MP\_PREFER\_PROXY option is an indication that the corresponding MPTCP connection is native: an on-path MCP must not be involved in such connection. If no explicit signal is included in the initial SYN message, the MCP cannot distinguish "native" MPTCP connections from "proxied" ones.

4.1. Option Format

The format of the MP\_PREFER\_PROXY is shown in Figure 3.

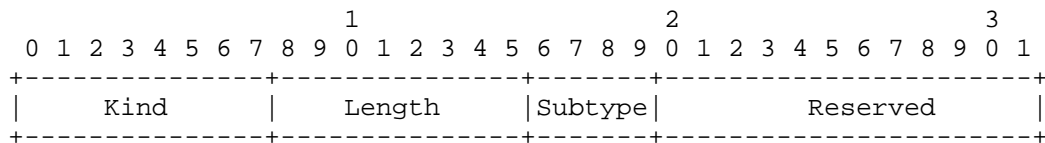


Figure 3: MP\_PREFER\_PROXY MPTCP Option

- o Kind and Length: are the same as those defined in Section 3 of [RFC6824]. The size of this option is 4 bytes.
- o Subtype: must be allocated by IANA (Section 9).
- o "Reserved" bits: are reserved bits for future assignment as additional flag bits. These additional flag bits MUST each be set to zero and MUST be ignored upon receipt.

4.2. Option Processing

The MP\_PREFER\_PROXY option MUST only appear in the SYN message used to create the initial subflow of a Multipath TCP connection.

If the MP\_PREFER\_PROXY appears in either a SYN segment that does not include the MP\_CAPABLE option or a segment whose SYN flag is unset,



it MUST be ignored. An implementation MAY log this event since it likely indicates an operational issue.

The sender inserts the MP\_PREFER\_PROXY option for MPTCP connections that it wants to be proxied by an on-path MCP. Such insertion is possible only when there is enough space left in the dedicated TCP option space.

Upon receipt of a SYN message with an MP\_CAPABLE, the MCP MUST check whether an MP\_PREFER\_PROXY option is present:

- o If no such option is included, the MCP MUST NOT interfere with that MPTCP connection (that is, it must not track this MPTCP connection). Processing subsequent subflows of this connection will be handled directly by the endpoints.
- o If the MP\_PREFER\_PROXY option is present, the MCP MUST track the establishment of the connection. That means that the MCP must be prepared to insert itself for the establishment of subsequent subflows, in particular.

Section 5.2.2.1 of [I-D.nam-mptcp-deployment-considerations] details the use of the MP\_PREFER\_PROXY option.

## 5. Supplying Data to MCPs

This section focuses mainly on the explicit mode (Section 5.2.1 of [I-D.nam-mptcp-deployment-considerations]) which assumes that the IP reachability information of an MCP is explicitly configured on a device, e.g., by means of a specific DHCP option [I-D.boucadair-mptcp-dhc].

### 5.1. The MP\_CONVERT Information Element

In order to avoid extra delays when establishing a proxied MPTCP connection, specific information are provided to an MCP during the 3WHS. Such information is meant to help the MCP instantiate the required states to process the connection upstream. The supply of such information is achieved by means of an object called the MP\_CONVERT (MC) Information Element (IE). This information element typically carries the source/destination IP addresses and/or port numbers of the used by the source and destination endpoints. Other information may also be supplied to an MCP; future extensions may be defined.

The format of the MP\_CONVERT Information Element is shown in Figure 4.

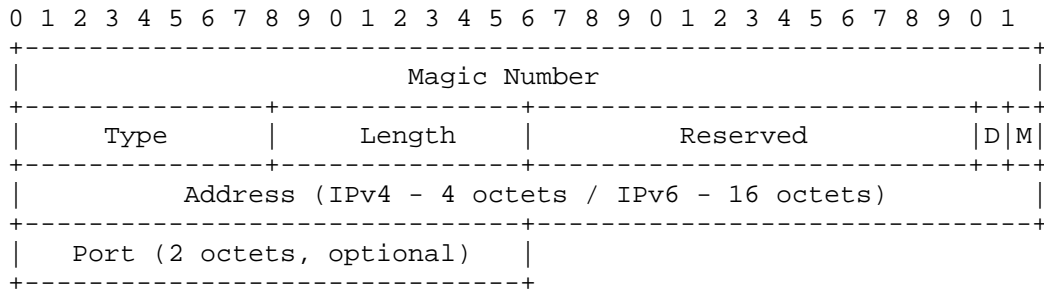


Figure 4: MP\_CONVERT Information Element

The description of the fields is as follows:

- o Magic Number: This field MUST be set to "0xFAA8 0xFAA8" to indicate this is an MP\_CONVERT Information Element. This field is meant to unambiguously distinguish any data supplied by an application from the one injected by an MCP. Other magic numbers are considered by the authors (e.g., 64 bits that include in addition to "0xFAA8 0xFAA8" 32 bits to enclose the RFC number).
- o Type: This field indicates the type of the MP\_CONVERT Information Element. It MUST be set to 0 to indicate this element includes an IP address and, eventually, a port number. Other type values MAY be defined in the future.
- o Length: Indicates, in bytes, the length of MP\_CONVERT Information Element. The minimum size of this option is 4 bytes.
- o "Reserved" bits: are reserved bits for future assignment as additional flag bits. These additional flag bits MUST each be set to zero and MUST be ignored upon receipt.
- o D-bit (Direction bit): this flag indicates whether the enclosed IP address (and port number) reflects the source or the destination IP address (and port number). When the D-bit is set, the enclosed IP address must be interpreted as the source IP address. When the D-bit is unset, the enclosed IP address must be interpreted as the destination IP address.
- o M-bit (More bit): When the M-bit is unset, it indicates that another MP\_CONVERT IE is included. When the M-bit is set, it indicates this is the last MP\_CONVERT IE included in the payload; if any data is placed right after this MP\_CONVERT IE, it is application data.

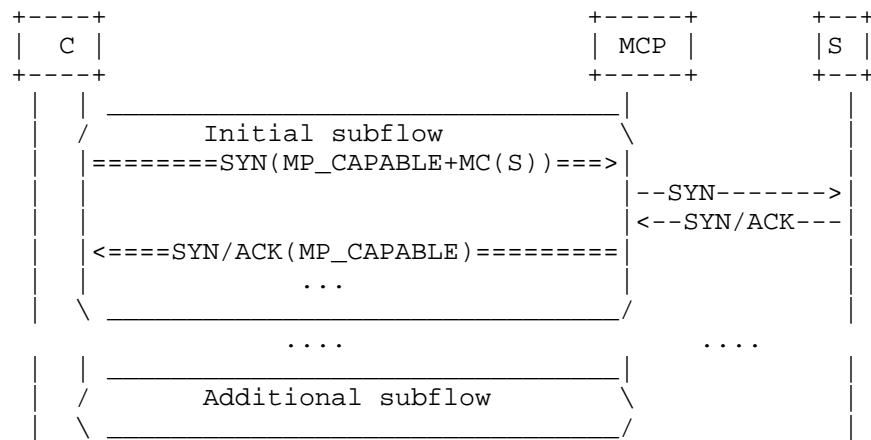
- o Address: includes a source or destination IP address. The address family is determined by the "Length" field. Concretely, a MP\_CONVERT Information Element that carries an IPv4 address has a Length field of 8 bytes (or 10, if a port number is included). A MP\_CONVERT Information Element that carries an IPv6 address has a Length of 20 bytes (or 22, if a port number is included).
- o Port: If the D-bit is set (resp. unset), a source (resp. destination) port number may be associated with the IP address. This field is valid for protocols that use a 16 bit port number (e.g., UDP, TCP, SCTP). This field is optional.

If the length of MP\_CONVERT Information Element is not a multiple of 4 bytes, padding MUST be added to preserve 32 bits boundaries.

## 5.2. Processing an MP\_CONVERT Information Element

The MP\_CONVERT Information Element is a variable length object that MUST NOT be used in TCP segments whose SYN flag is unset. This IE can only appear in the TCP control messages with SYN flag set. The information carried in the MP\_CONVERT IE is used by an MCP to create the initial subflow of a Multipath TCP connection (see the example in Figure 5).

Up to two MP\_CONVERT Information Elements with type set to zero can appear inside a SYN segment. If two MP\_CONVERT Information Elements with type zero are included, these options MUST NOT have the same D-bit value.



Legend:

<====>: MPTCP leg

<--->: TCP leg

MC(): MP\_CONVERT Information Element

Figure 5: Carrying the MP\_CONVERT Information Element

The MP\_CONVERT Information Element MUST be included in the payload of a TCP segment whose SYN flag is set.

If the MP\_CONVERT Information Element appears in either a SYN segment that does not include the MP\_CAPABLE option or a segment whose SYN flag is reset, it MUST be ignored. An implementation MAY log this event since it likely indicates an operational issue.

If the original SYN message contains data in its payload (e.g., [RFC7413]), that data MUST be placed right after the MP\_CONVERT IEs when generating the SYN in the MPTCP leg.

An implementation MUST ignore MP\_CONVERT Information Elements that include multicast, broadcast, and host loopback addresses [RFC6890]. Concretely, an implementation that receives an MP\_CONVERT Information Element with such addresses MUST silently tear down the MPTCP connection.

An implementation that supports the MP\_CONVERT Information Element with type zero MUST echo in the SYN/ACK the instances of the MP\_CONVERT Information Elements included in a SYN received from the sender. A sender that does not receive in a SYN/ACK a copy of the MP\_CONVERT Information Elements it included in a SYN message MUST terminate the MPTCP connection and falls back to TCP or native MPTCP connection. Furthermore, the sender MUST add an entry to its local

cache to record the MCPs that do not support the MP\_CONVERT Information Element. This cache MUST be flushed out under the following conditions: a new network attachment is detected by the host, a new MCP is configured, the host gets a new IP address/prefix, or a TTL has expired. Subsequent connections to an MCP in the cache MUST NOT be placed using the explicit proxy mode. This procedure is denoted as MCP capability discovery.

In the following sections, MP\_CONVERT Information Element is used to refer to the MP\_CONVERT Information Element with the type field set to zero. Future documents will specify the exact behavior of processing MP\_CONVERT Information Elements with a non zero type field.

## 6. MPTCP Connections from a Multipath TCP Client

### 6.1. Description

The simplest usage of the MP\_CONVERT Information Element is when a Multipath TCP Client wants to use MPTCP to efficiently utilise different network paths (e.g., WLAN and LTE from a smartphone) to reach a server that does not support Multipath TCP. The basic operation is illustrated in Figure 6.

To use its multipath capabilities to establish an MPTCP connection over the available networks, the Client splits its end-to-end connection towards the TCP Server into two:

- (1) An MPTCP connection, that typically relies upon the establishment of one subflow per network path, is established between the client and the MCP.
- (2) A TCP connection that is established by the MCP with the server.

Any data that is eligible to be transported over the MPTCP connection is sent by the Client towards the MCP over the MPTCP connection. The MCP then forwards these data over the regular TCP connection until they reach the server. The same forwarding principle applies for the data sent by the Server over the TCP connection with the MCP.

```

C <=====>MCP <-----> S
+<=====>+

```

Legend:

<====>: subflows of the upstream MPTCP connection  
 <---->: downstream TCP connection

Figure 6: A Multipath TCP Client interacts with a Server through a Multipath Conversion Point

## 6.2. Theory of Operation

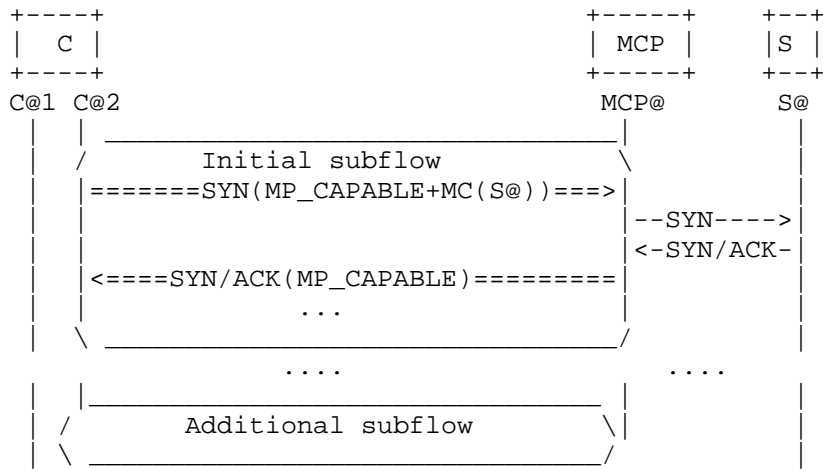
We assume in this section that the Multipath TCP Client has been configured with the IP address of one or more MCPs which convert the Multipath TCP connection into a regular TCP connection. The address of such MCPs can be statically configured on the Client, dynamically provisioned to the MPTCP Client by means of a DHCP option [I-D.boucadair-mptcp-dhc], or by any other means that are outside the scope of this document.

Conceptually, the MCP acts as a relay between an upstream MPTCP connection and a downstream TCP connection. The MCP has at least a single IP address that is reachable from the Multipath TCP Client. It may be assigned other IP addresses. For the sake of simplicity, we assume in this section that the MCP has a single IP address denoted MCP@. Similarly, we assume that the client has two addresses C@1 and C@2 while address S@ is assigned to the server.

The MCP maps an upstream MPTCP connection (and its associated subflows) onto a downstream TCP connection. On the MCP, an established Multipath TCP connection can be identified by the local Token that was assigned upon reception of the SYN segment.

This Token is guaranteed to be unique on the MCP (provided that it has a single IP address) during the entire lifetime of the MPTCP connection. The 4-tuple (IP src, IP dst, Port src, Port dst) is used to identify the downstream TCP connection.

To initiate a connection to a remote server S, the Multipath TCP Client sends a SYN segment towards the MCP that includes the MP\_CONVERT Information Element described in Figure 4. The destination address of the SYN segment is the IP address of the MCP. The MP\_CONVERT Information Element included in the SYN contains the IP address and optionally the destination port of the Server (see Figure 7).



Legend:  
 <====>: MPTCP leg  
 <---->: TCP leg

Figure 7: Single-ended MCP Flow Example

The MCP processes this SYN segment as follows. First, it generates the local key and a unique Token for the Multipath TCP connection. This Token identifies the MPTCP connection. It is passed to the MCP together with the contents of the MP\_CONVERT Information Element (i.e., the address of the destination server) and the destination port.

The MCP then establishes a TCP connection with the destination server. If the received MP\_CONVERT Information Element contains a port number, it is used as the destination port of the outgoing TCP connection that is being established by the MCP. Otherwise, the destination port of the upstream MPTCP connection is used as the destination port of the downstream TCP connection. The MCP creates a flow entry for the downstream TCP connection and maps the upstream MPTCP connection onto the downstream TCP connection.

The downstream TCP connection is considered to be active upon reception of the SYN/ACK segment sent by the destination server. The reception of this segment triggers the MCP that confirms the establishment of the upstream MPTCP connection by sending a SYN/ACK segment towards the Multipath TCP Client (including MP\_Convert).

At this point, there are two established connections. The endpoints of the upstream Multipath TCP connection are the Multipath TCP Client

and the MCP. The endpoints of the downstream TCP connection are the MCP and the Server. These two connections are bound by the MCP.

All the techniques defined in [RFC6824] can be used by the upstream Multipath TCP connection. In particular, the subflows established over the different network paths can be controlled by either the Multipath TCP Client or the MCP. It is likely that the network operators that deploy MCPs will define policies for the utilisation of the MCP. These policies are discussed in Section 5.6 of [I-D.nam-mptcp-deployment-considerations].

Any data received by the MCP on the upstream Multipath TCP connection will be forwarded by the MCP over the bound downstream TCP connection. The same applies for data received over the downstream TCP connection which will be forwarded by the MCP over the upstream Multipath TCP connection.

One of the functions of the MCP is to maintain the binding between the upstream Multipath TCP connection and the downstream TCP connection. If the downstream TCP connection fails for some reason (excessive retransmissions, reception of a RST segment, etc.), then the MCP SHOULD force the teardown of the upstream Multipath TCP connection by transmitting a FASTCLOSE. Similarly, if the upstream Multipath TCP connection fails for some reason (e.g., reception of a FASTCLOSE), the MCP SHOULD tear the downstream TCP connection down and remove the flow entries.

The same reasoning applies when the upstream Multipath TCP connection ends with the transmission of DATA\_FINs. In this case, the MCP SHOULD also terminate the bound downstream TCP connection by using FIN segments. If the downstream TCP connection terminates with the exchange of FIN segments, the MCP SHOULD initiate a graceful termination of the bound upstream Multipath TCP connection.

An MCP SHOULD associate a lifetime with the Multipath TCP and TCP flow entries. In this case, it SHOULD use the same lifetime for each pair of bounded connections.

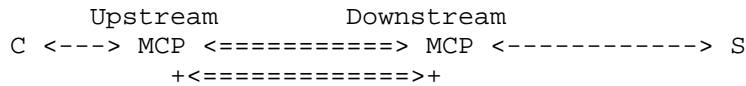
## 7. MPTCP Connections Between Single Path Client and Server

### 7.1. Description

There are situations where neither the client nor the server can use multipath transport protocols albeit network providers would want to optimize network resource usage by means of multi-path communication techniques. Hybrid access service offerings are typical business incentives for such situations, where network operators combine a fixed network (e.g., xDSL) with a wireless network (e.g., LTE). In



this case, as illustrated in Figure 8, two MCPs are used for each flow. The first MCP, located downstream of the client, converts the single path TCP connection originated from the client into a Multipath TCP connection established with a second MCP. The latter will then establish a TCP connection with the destination server.



Legend:

```

<====>: MPTCP leg
<---->: TCP leg

```

Figure 8: A Client interacts with a Server through an upstream and a downstream Multipath Conversion Points

## 7.2. Theory of Operation

### 7.2.1. Downstream MCP

The downstream MCP can be deployed on-path or off-path. If the downstream MCP is deployed off-path, its behavior is described in Section 6.2.

If the downstream MCP is deployed on-path, it only terminates MPTCP connections that carry an empty `MP_PREFER_PROXY` option inside their SYN (i.e., no address is conveyed). If the MCP receives a SYN segment that contains the `MP_CAPABLE` option but no `MP_PREFER_PROXY`, it MUST forward the SYN to its final destination without any modification.

### 7.2.2. Upstream MCP

The upstream and downstream MCPs cooperate. The upstream MCP may be configured with the addresses of downstream MCPs. If the downstream MCP is deployed on-path, the upstream MCP inserts an `MP_PREFER_PROXY` option.

In this section, we assume that the upstream MCP has been configured with one address of the downstream MCP. This address can be configured statically, dynamically distributed by means of a DHCP option [I-D.boucadair-mptcp-dhc], or by any other means that are outside the scope of this document.

We assume that the upstream MCP has two addresses `uMCP@1` and `uMCP@2` while the downstream MCP is assigned a single IP address `dMCP@`.

The upstream MCP maps an upstream TCP connection onto a downstream MPTCP connection (and its associated subflows) . On the upstream MCP, an established MPTCP connection can be identified by the local Token that was assigned upon reception of the SYN segment from the Client.

The Client sends a SYN segment addressed to the Server and it is intercepted by the upstream MCP which in turns initiates an MPTCP connection towards its downstream MCP that includes the MP\_CONVERT Information Element described in Figure 4. The destination address of the SYN segment is the IP address of the downstream MCP. The MP\_CONVERT Information Element included in the SYN contains the IP address and optionally the destination port of the Server; this information is extracted from the SYN message received over the upstream TCP connection.

Concretely, the upstream MCP processes the SYN segment received from the Client as follows.

First, it generates the local key and a unique Token for the Multipath TCP connection to identify the MPTCP connection. It extracts the destination IP address and, optionally, the destination port that will then be carried in a MP\_CONVERT Information Element. The upstream MCP establishes an MPTCP connection with the downstream MCP. The upstream MCP creates a flow entry for the downstream MPTCP connection and maps the upstream TCP connection onto the downstream MPTCP connection.

The downstream MPTCP connection is considered to be active upon reception of the SYN+ACK segment from the downstream MCP. The reception of this segment triggers the upstream MCP that confirms the establishment of the upstream TCP connection by sending a SYN+ACK segment towards the TCP Client.

At this point, there are two established connections maintained by the upstream MCP:

- (1) The endpoints of the upstream TCP connection are the Client and the upstream MCP.
- (2) The endpoints of the downstream MPTCP connection are the upstream MCP and the downstream MCP.

These two connections are bound by the upstream MCP. An example is shown in Figure 9.

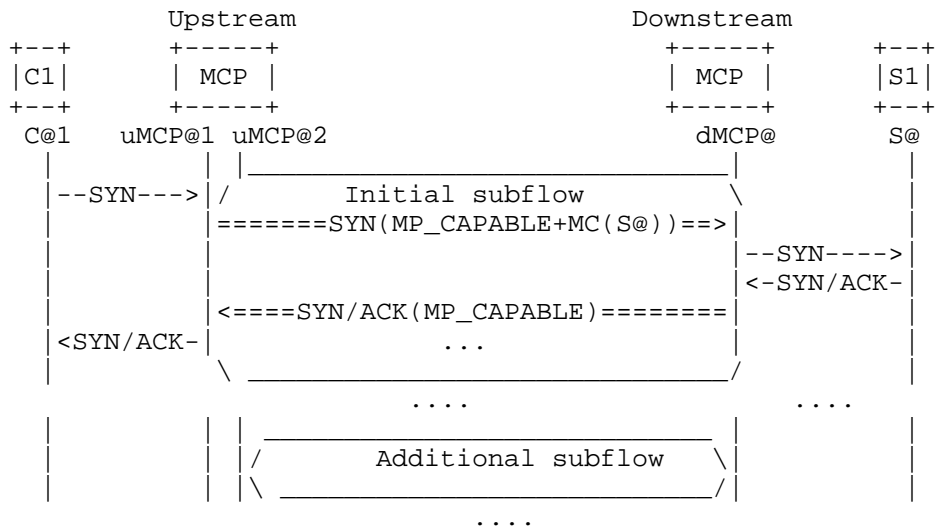


Figure 9: Dual-Ended MCP Flow Example

All the techniques defined in [RFC6824] can be used by the MPTCP connection. In particular, the utilisation of the different network paths can be controlled by one MCP or the other.

Any data received by the upstream MCP over the upstream TCP connection will be forwarded by the MCP over the bound downstream MPTCP connection, assuming such data are eligible to MPTCP transport. The same applies for data received over the downstream MPTCP connection which will be forwarded by the upstream MCP over the upstream TCP connection.

The same considerations as in Section 6.2 apply for the maintenance of the connections by the upstream MCP.

### 8. Interaction with TFO

This section discusses the implications of using MP\_CONVERT Information Elements with TCP Fast Open (TFO). We distinguish between TFO negotiation (i.e., a Fast Open option with an empty cookie field to request a cookie) and TFO data (i.e., SYN with data and the cookie in the Fast Open option).

This section focuses on the implications of using MP\_CONVERT Information Element on TFO efficiency. Implications related to MPTCP options and TFO negotiation are not specific to this document; the reader may refer to [I-D.barre-mptcp-tfo].

Distinct implications are assessed depending whether TFO negotiation and usage occurs before MCP capability discovery phase is completed or not (Section 5.2). Concretely, the following cases are discussed:

1. MCP capability discovery was already completed prior to receiving a message with TFO negotiation or TFO data: For this case, the host has already contacted its MCP in the context of a prior connection. The outcome of such connections is used to determine the capabilities of its MCP (Section 5.2).
  - A. The MCP supports MP\_CONVERT Information Element: Any information provided to an MCP to facilitate MPTCP operation is unambiguously distinguished from TFO data that are also included in the SYN payload. An upstream MCP will remove the MP\_CONVERT Information Elements before relaying the SYN message (with TFO data) to the next hop.
  - B. The MCP does not support MP\_CONVERT Information Element: No additional issue is raised for obvious reasons.
2. MCP capability discovery is not completed prior to receiving a message with TFO negotiation or TFO data.
  - A. If the same message is used to negotiate TFO and to retrieve the capabilities of the MCP, extra delay may be observed before negotiating TFO if the MCP does not support the MP\_CONVERT Information Element. Obviously, no concern is raised when the MCP supports the MP\_CONVERT Information Element.
  - B. If the same message includes TFO data and is used to retrieve the capabilities of the MCP, extra delay may be observed before negotiating TFO if the MCP does not support the MP\_CONVERT Information Element. Obviously, no concern is raised when the MCP supports the MP\_CONVERT Information Element.

To mitigate cases where extra delays are experienced when TFO is present, it is RECOMMENDED to not proxy connections with TFO before the MCP capability discovery procedure is completed.

## 9. IANA Considerations

This document requests an MPTCP subtype code for this option:

- o MP\_PREFER\_PROXY

## 10. Security Considerations

MPTCP-related security threats are discussed in [RFC6181] and [RFC6824]. Additional considerations are discussed in the following sub-sections.

### 10.1. Privacy

The MCP may have access to privacy-related information (e.g., IMSI, link identifier, subscriber credentials, etc.). The MCP **MUST NOT** leak such sensitive information outside a local domain.

### 10.2. Denial-of-Service (DoS)

Means to protect the MCP against Denial-of-Service (DoS) attacks **MUST** be enabled. Such means include the enforcement of ingress filtering policies at the network boundaries [RFC2827].

In order to prevent the exhaustion of MCP resources by establishing a great number of simultaneous subflows for each MPTCP connection, the MCP administrator **SHOULD** limit the number of allowed subflows per CPE for a given connection. Means to protect against SYN flooding attacks **MUST** also be enabled ([RFC4987]).

Attacks that originate outside of the domain can be prevented if ingress filtering policies are enforced. Nevertheless, attacks from within the network between a host and an MCP instance are yet another actual threat. Means to ensure that illegitimate nodes cannot connect to a network should be implemented.

### 10.3. Illegitimate MCP

Traffic theft is a risk if an illegitimate MCP is inserted in the path. Indeed, inserting an illegitimate MCP in the forwarding path allows traffic intercept and can therefore provide access to sensitive data issued by or destined to a host. To mitigate this threat, secure means to discover an MCP should be enabled.

## 11. Acknowledgements

Many thanks to Chi Dung Phung, Mingui Zhang, Rao Shoaib, Yoshifumi Nishida, and Christoph Paasch for their valuable comments.

Thanks to Ian Farrer, Mikael Abrahamsson, Alan Ford, Dan Wing, and Sri Gundavelli for the fruitful discussions in IETF#95 (Buenos Aires).

Special thanks to Pierrick Seite, Yannick Le Goff, Fred Klamm, and Xavier Grall for their inputs.

Thanks also to Olaf Schleusing, Martin Gysi, Thomas Zasowski, Andreas Burkhard, Silka Simmen, Sandro Berger, Michael Melloul, Jean-Yves Flahaut, Adrien Desportes, Gregory Detal, Benjamin David, Arun Srinivasan, and Raghavendra Mallya for the discussion.

The design approach adopted in -10 is the outcome of fruitful discussions with Alan Ford. Many thanks Alan.

## 12. References

### 12.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [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>>.
- [RFC6890] Cotton, M., Vegoda, L., Bonica, R., Ed., and B. Haberman, "Special-Purpose IP Address Registries", BCP 153, RFC 6890, DOI 10.17487/RFC6890, April 2013, <<http://www.rfc-editor.org/info/rfc6890>>.

### 12.2. Informative References

- [I-D.barre-mptcp-tfo] Barre, S., Detal, G., and O. Bonaventure, "TFO support for Multipath TCP", draft-barre-mptcp-tfo-01 (work in progress), January 2015.
- [I-D.boucadair-mptcp-dhc] Boucadair, M., Jacquenet, C., and T. Reddy, "DHCP Options for Network-Assisted Multipath TCP (MPTCP)", draft-boucadair-mptcp-dhc-06 (work in progress), October 2016.
- [I-D.nam-mptcp-deployment-considerations] Boucadair, M., Jacquenet, C., Bonaventure, O., Henderickx, W., and R. Skog, "Network-Assisted MPTCP: Use Cases, Deployment Scenarios and Operational Considerations", draft-nam-mptcp-deployment-considerations-01 (work in progress), December 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-05 (work in progress), December 2016.
- [RFC1701] Hanks, S., Li, T., Farinacci, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 1701, DOI 10.17487/RFC1701, October 1994, <<http://www.rfc-editor.org/info/rfc1701>>.
- [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>>.
- [RFC2473] Conta, A. and S. Deering, "Generic Packet Tunneling in IPv6 Specification", RFC 2473, DOI 10.17487/RFC2473, December 1998, <<http://www.rfc-editor.org/info/rfc2473>>.
- [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>>.
- [RFC4987] Eddy, W., "TCP SYN Flooding Attacks and Common Mitigations", RFC 4987, DOI 10.17487/RFC4987, August 2007, <<http://www.rfc-editor.org/info/rfc4987>>.
- [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>>.
- [RFC7413] Cheng, Y., Chu, J., Radhakrishnan, S., and A. Jain, "TCP Fast Open", RFC 7413, DOI 10.17487/RFC7413, December 2014, <<http://www.rfc-editor.org/info/rfc7413>>.
- [TR-348] BBF, "Hybrid Access Broadband Network Architecture", July 2016.

## Authors' Addresses

Mohamed Boucadair (editor)  
Orange  
Rennes 35000  
France

Email: mohamed.boucadair@orange.com

Christian Jacquenet (editor)  
Orange  
Rennes  
France

Email: christian.jacquenet@orange.com

Olivier Bonaventure (editor)  
Tessares  
Belgium

Email: olivier.bonaventure@tessares.net

Denis Behaghel  
OneAccess

Email: Denis.Behaghel@oneaccess-net.com

Stefano Secci  
UPMC

Email: stefano.secci@lip6.fr

Wim Henderickx (editor)  
Nokia/Alcatel-Lucent  
Belgium

Email: wim.henderickx@alcatel-lucent.com

Robert Skog (editor)  
Ericsson

Email: robert.skog@ericsson.com



Suresh Vinapamula  
Juniper  
1137 Innovation Way  
Sunnyvale, CA 94089  
USA

Email: Sureshk@juniper.net

SungHoon Seo  
Korea Telecom  
Seoul  
Korea

Email: sh.seo@kt.com

Wouter Cloetens  
SoftAtHome  
Vaartdijk 3 701  
3018 Wijgmaal  
Belgium

Email: wouter.cloetens@softathome.com

Ullrich Meyer  
Vodafone  
Germany

Email: ullrich.meyer@vodafone.com

Luis M. Contreras  
Telefonica  
Spain

Email: luismiguel.contrerasmurillo@telefonica.com

Bart Peirens  
Proximus

Email: bart.peirens@proximus.com

DMM WG  
Internet-Draft  
Intended status: Standards Track  
Expires: January 26, 2017

P. Seite  
Orange  
A. Yegin  
Samsung  
S. Gundavelli  
Cisco  
July 25, 2016

MAG Multipath Binding Option  
draft-ietf-dmm-mag-multihoming-02.txt

Abstract

The document [RFC4908] proposes to rely on multiple Care-of Addresses (CoAs) capabilities of Mobile IP [RFC6275] and Network Mobility (NEMO; [RFC3963]) to enable Multihoming technology for Small-Scale Fixed Networks. In the continuation of [RFC4908], this document specifies a multiple proxy Care-of Addresses (pCoAs) extension for Proxy Mobile IPv6 [RFC5213]. This extension allows a multihomed Mobile Access Gateway (MAG) to register more than one proxy care-of-address to the Local Mobility Anchor (LMA).

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 26, 2017.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	2
2. Conventions and Terminology . . . . .	4
2.1. Conventions . . . . .	4
2.2. Terminology . . . . .	4
3. Overview . . . . .	5
3.1. Example Call Flow . . . . .	5
3.2. Traffic distribution schemes . . . . .	6
4. Protocol Extensions . . . . .	7
4.1. MAG Multipath-Binding Option . . . . .	7
4.2. MAG Identifier Option . . . . .	9
4.3. New Status Code for Proxy Binding Acknowledgement . . . . .	10
5. IANA Considerations . . . . .	10
6. Security Considerations . . . . .	11
7. Acknowledgements . . . . .	11
8. References . . . . .	11
8.1. Normative References . . . . .	11
8.2. Informative References . . . . .	13
Authors' Addresses . . . . .	13

## 1. Introduction

Using several links, the multihoming technology can improve connectivity availability and quality of communications; the goals and benefits of multihoming are as follows:

- o Redundancy/Fault-Recovery
- o Load balancing
- o Load sharing
- o Preferences settings

According to [RFC4908], users of Small-Scale Networks can take benefit of multihoming using mobile IP [RFC6275] and Network Mobility (NEMO) [RFC3963] architecture in a mobile and fixed networking environment. This document is introducing the concept of multiple Care-of Addresses (CoAs) [RFC5648] that have been specified since then.

In the continuation of [RFC4908], a Proxy Mobile IPv6 [RFC5213] based multihomed achitecture could be defined. The motivation to update [RFC4908] with proxy Mobile IPv6 is to leverage on latest mobility working group achievements, namely:

- o using GRE as mobile tuneling, possibly with its key extension [RFC5845] (a possible reason to use GRE is given on Section 3.2).
- o using UDP encapsulation [RFC5844] in order to support NAT traversal in IPv4 networking environment.
- o Prefix Delegation mechanism [RFC7148].
- o Using the vendor specific mobility option [RFC5094], for example to allow the MAG and LMA to exchange information (e.g. WAN interface QoS metrics) allowing to make appropriate traffic steering decision.

Proxy Mobile IPv6 (PMIPv6) relies on two mobility entities: the mobile access gateway (MAG), which acts as the default gateway for the end-node and the local mobility anchor (LMA), which acts as the topological anchor point. Point-to-point links are established, using IP-in-IP tunnels, between MAG and LMA. Then, the MAG and LMA are distributing traffic over these tunnels. All PMIPv6 operations are performed on behalf of the end-node and its corespondent node, it thus makes PMIPv6 well adapted to multihomed architecture as considered in [RFC4908]. Taking the LTE and WLAN networking environments as an example, the PMIPv6 based multihomed architecture is depicted on Figure 1. Flow-1,2 and 3 are distributed either on Tunnel-1 (over LTE) or Tunnel-2 (over WLAN), while Flow-4 is spread on both Tunnel-1 and 2.

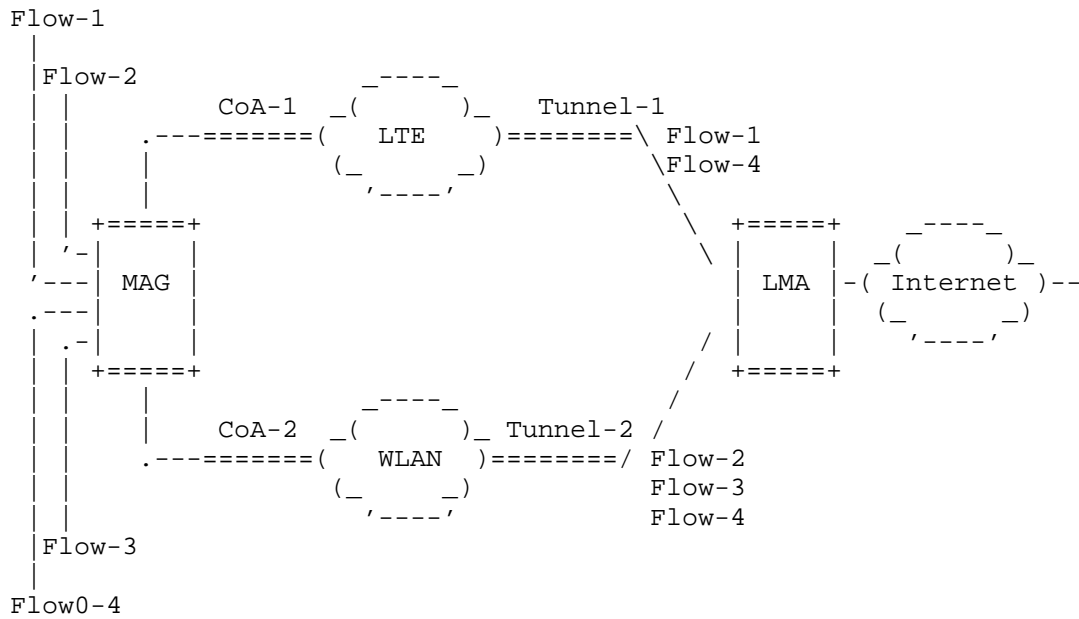


Figure 1: Multihomed MAG using Proxy Mobile IPv6

The current version of Proxy Mobile IPv6 does not allow a MAG to register more than one proxy Care-of-Adresse to the LMA. In other words, only one MAG/LMA link, i.e. IP-in-IP tunnel, can be used at the same time. This document overcomes this limitation by defining the multiple proxy Care-of Addresses (pCoAs) extension for Proxy Mobile IPv6.

## 2. Conventions and Terminology

### 2.1. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

### 2.2. Terminology

All mobility related terms used in this document are to be interpreted as defined in [RFC5213], [RFC5844] and [RFC7148]. Additionally, this document uses the following terms:

IP-in-IP

IP-within-IP encapsulation [RFC2473], [RFC4213]

### 3. Overview

#### 3.1. Example Call Flow

Figure 2 is the callflow detailing multi-access support with PMIPv6. The MAG in this example scenario is equipped with both WLAN and LTE interfaces and is also configured with the multihoming functionality. The steps of the callflow are as follows:

Steps (1) and (2): the MAG attaches to both WLAN and LTE networks; the MAG obtains respectively two different proxy care-of-addresses (pCoA).

Step (3): The MAG sends, over the WLAN access, a Proxy Binding Update (PBU) message, with the new MAG Multipath Binding (MMB) and MAG Identifier (MAG-NAI) options to the LMA. A logical-NAI (MAG-NAI) with ALWAYS-ON configuration is enabled on the MAG. The mobility session that is created (i.e. create a Binding Cache Entry) on the LMA is for the logical-NAI. The LMA and allocates a Home Network Prefix (HNP), that shall be delegated to mobile nodes, to the MAG.

Step (4): the LMA sends back a Proxy Binding Acknowledgement (PBA) including the HNP allocated to the MAG.

Step (5): IP tunnel (IP-in-IP, GRE ...) is created over the WLAN access.

Steps (6) to (8): The MAG repeats steps (3) to (5) on the LTE access. The MAG includes the HNP, received on step (4) in the PBU. The LMA update its binding cache by creating a new mobility session for this MAG.

Steps (9) and (10): The IP hosts MN\_1 and MN\_2 are assigned IP addresses from the mobile network prefix delegated by the MAG.

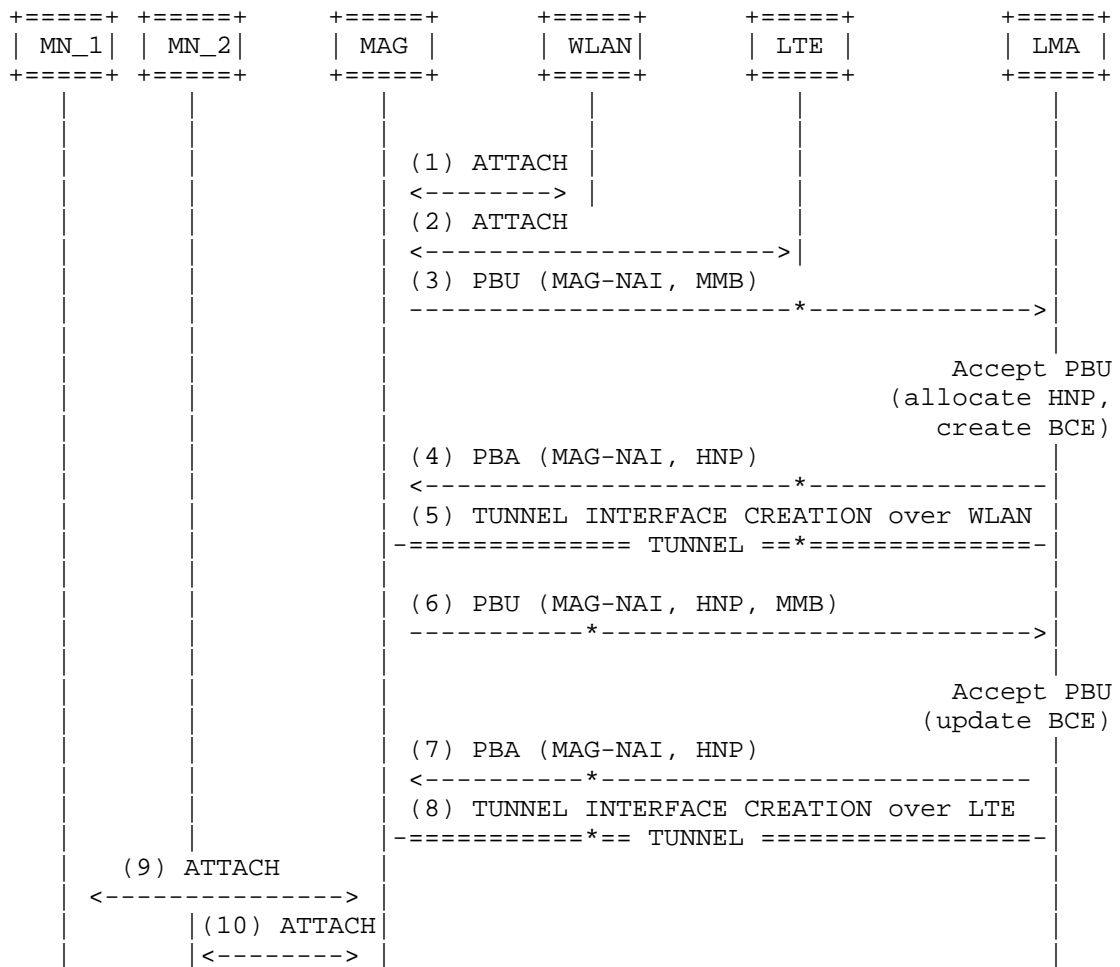


Figure 2: Functional Separation of the Control and User Plane

### 3.2. Traffic distribution schemes

When receiving packets from the MN, the MAG distributes packets over tunnels that have been established. Traffic distribution can be managed either on a per-flow or on a per-packet basis:

- o Per-flow traffic management: each IP flow (both upstream and downstream) is mapped to a given tunnel, corresponding to a given WAN interface. Flow binding extension [RFC6089] is used to exchange, and synchronize, IP flow management policies (i.e. rules associating traffic selectors [RFC6088] to a tunnel).

- o Per-packet management: the LMA and the MAG distribute packets, belonging to a same IP flow, over more than one bindings (i.e. more than one WAN interface). When operating at the IP packet level, different packets distribution algorithms are possible. For example, the algorithm may give precedence to one given access: the MAG overflows traffic from the primary access, e.g. WLAN, to the second one, only when load on primary access reaches a given threshold. The distribution algorithm is left to implementer but whatever the algorithm is, packets distribution likely introduces packet latency and out-of-order delivery. LMA and MAG shall thus be able to make reordering before packets delivery. Sequence number can be used for that purpose, for example using GRE with sequence number option [RFC5845]. However, more detailed considerations on reordering and IP packet distribution scheme (e.g. definition of packets distribution algorithm) are out the scope of this document.

Because latency introduced by per-packet can cause injury to some application, per-flow and per-packet distribution schemes could be used in conjunction. For example, high throughput services (e.g. video streaming) may benefit from per-packet distribution scheme, while latency sensitive applications (e.g. VoIP) are not spread over different WAN paths. IP flow mobility extensions, [RFC6089] and [RFC6088], can be used to provision the MAG with such flow policies.

#### 4. Protocol Extensions

##### 4.1. MAG Multipath-Binding Option

The MAG Multipath-Binding option is a new mobility header option defined for use with Proxy Binding Update and Proxy Binding Acknowledgement messages exchanged between the local mobility anchor and the mobile access gateway.

This mobility header option is used for requesting multipath support. It indicates that the mobile access gateway is requesting the local mobility anchor to register the current care-of address associated with the request as one of the many care-addresses through which the mobile access gateway can be reached. It is also for carrying the information related to the access network associated with the care-of address.

The MAG Multipath-Binding option has an alignment requirement of  $8n+2$ . Its format is as shown in Figure 3:



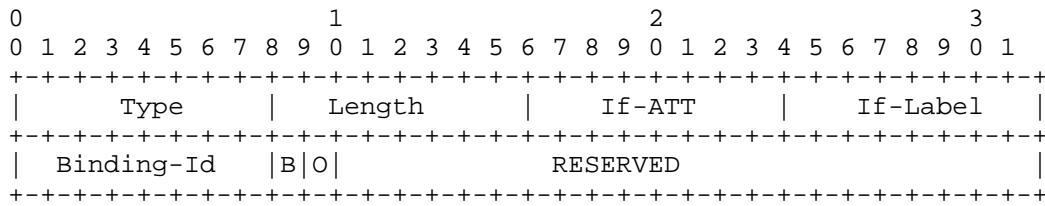


Figure 3: MAG Multipath Binding Option

Type

<IANA-1> To be assigned by IANA.

Length

8-bit unsigned integer indicating the length of the option in octets, excluding the type and length fields.

Interface Access-Technology Type (If-ATT)

This 8-bit field identifies the Access-Technology type of the interface through which the mobile node is connected. The permitted values for this are from the Access Technology Type registry defined in [RFC5213].

Interface Label (If-Label)

This 8-bit field represents the interface label represented as an unsigned integer. The MAG identifies the label for each of the interfaces through which it registers a pCoA with the LMA. When using static traffic flow policies on the mobile node and the home agent, the label can be used for generating forwarding policies. For example, the operator may have policy which binds traffic for Application "X" needs to interface with Label "Y". When a registration through an interface matching Label "Y" gets activated, the home agent and the mobile node can dynamically generate a forwarding policy for forwarding traffic for Application "X" through mobile IP tunnel matching Label "Y". Both the home agent and the mobile node can route the Application-X traffic through that interface. The permitted values for If-Label are 1 through 255.

Binding-Identifier (BID)

This 8-bit field is used for carrying the binding identifier. It uniquely identifies a specific binding of the mobile node, to which this request can be associated. Each binding identifier is

represented as an unsigned integer. The permitted values are 1 through 254. The BID value of 0 and 255 are reserved. The mobile access gateway assigns a unique value for each of its interfaces and includes them in the message.

Bulk Re-registration Flag (B)

This flag, if set to a value of (1), is to notify the local mobility anchor to consider this request as a request to update the binding lifetime of all the mobile node's bindings, upon accepting this specific request. This flag MUST NOT be set to a value of (1), if the value of the Registration Overwrite Flag (O) is set to a value of (1).

Binding Overwrite (O)

This flag, if set to a value of (1), notifies the local mobility anchor that upon accepting this request, it should replace all of the mobile node's existing bindings with this binding. This flag MUST NOT be set to a value of (1), if the value of the Bulk Re-registration Flag (B) is set to a value of (1). This flag MUST be set to a value of (0), in de-registration requests.

Reserved

This field is unused in this specification. The value MUST be set to zero (0) by the sender and MUST be ignored by the receiver.

4.2. MAG Identifier Option

The MAG Identifier option is a new mobility header option defined for use with Proxy Binding Update and Proxy Binding Acknowledgement messages exchanged between the local mobility anchor and the mobile access gateway. This mobility header option is used for conveying the MAG's identity.

This option does not have any alignment requirements.

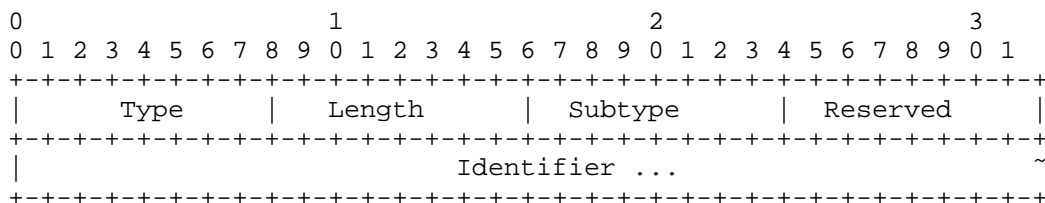


Figure 4: MAG Identifier Option

#### Type

<IANA-2> To be assigned by IANA.

#### Length

8-bit unsigned integer indicating the length of the option in octets, excluding the type and length fields.

#### Subtype

One byte unsigned integer used for identifying the type of the Identifier field. Accepted values for this field are the registered type values from the Mobile Node Identifier Option Subtypes registry.

#### Reserved

This field is unused in this specification. The value MUST be set to zero (0) by the sender and MUST be ignored by the receiver.

#### Identifier

A variable length identifier of type indicated in the Subtype field.

### 4.3. New Status Code for Proxy Binding Acknowledgement

This document defines the following new Status Code value for use in Proxy Binding Acknowledgement message.

CANNOT\_SUPPORT\_MULTIPATH\_BINDING (Cannot Support Multipath Binding):  
<IANA-4>

## 5. IANA Considerations

This document requires the following IANA actions.

- o Action-1: This specification defines a new mobility option, the MAG Multipath-Binding option. The format of this option is described in Section 4.1. The type value <IANA-1> for this mobility option needs to be allocated from the Mobility Options registry at <<http://www.iana.org/assignments/mobility-parameters>>. RFC Editor: Please replace <IANA-1> in Section 4.1 with the assigned value and update this section accordingly.
- o Action-2: This specification defines a new mobility option, the MAG Identifier option. The format of this option is described in

Section 4.2. The type value <IANA-2> for this mobility option needs to be allocated from the Mobility Options registry at <<http://www.iana.org/assignments/mobility-parameters>>. RFC Editor: Please replace <IANA-2> in Section 4.2 with the assigned value and update this section accordingly.

- o Action-3: This document defines a new status value, CANNOT\_SUPPORT\_MULTIPATH\_BINDING (<IANA-3>) for use in Proxy Binding Acknowledgement message, as described in Section 4.3. This value is to be assigned from the "Status Codes" registry at <<http://www.iana.org/assignments/mobility-parameters>>. The allocated value has to be greater than 127. RFC Editor: Please replace <IANA-4> in Section 4.3 with the assigned value and update this section accordingly.

## 6. Security Considerations

This specification allows a mobile access gateway to establish multiple Proxy Mobile IPv6 tunnels with a local mobility anchor, by registering a care-of address for each of its connected access networks. This essentially allows the mobile node's IP traffic to be routed through any of the tunnel paths and either based on a static or a dynamically negotiated flow policy. This new capability has no impact on the protocol security. Furthermore, this specification defines two new mobility header options, MAG Multipath-Binding option and the MAG Identifier option. These options are carried like any other mobility header option as specified in [RFC5213]. Therefore, it inherits security guidelines from [RFC5213]. Thus, this specification does not weaken the security of Proxy Mobile IPv6 Protocol, and does not introduce any new security vulnerabilities.

## 7. Acknowledgements

The authors of this draft would like to acknowledge the discussions and feedback on this topic from the members of the DMM working group.

## 8. References

### 8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.

- [RFC3963] Devarapalli, V., Wakikawa, R., Petrescu, A., and P. Thubert, "Network Mobility (NEMO) Basic Support Protocol", RFC 3963, DOI 10.17487/RFC3963, January 2005, <<http://www.rfc-editor.org/info/rfc3963>>.
- [RFC5094] Devarapalli, V., Patel, A., and K. Leung, "Mobile IPv6 Vendor Specific Option", RFC 5094, DOI 10.17487/RFC5094, December 2007, <<http://www.rfc-editor.org/info/rfc5094>>.
- [RFC5213] Gundavelli, S., Ed., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, DOI 10.17487/RFC5213, August 2008, <<http://www.rfc-editor.org/info/rfc5213>>.
- [RFC5648] Wakikawa, R., Ed., Devarapalli, V., Tsirtsis, G., Ernst, T., and K. Nagami, "Multiple Care-of Addresses Registration", RFC 5648, DOI 10.17487/RFC5648, October 2009, <<http://www.rfc-editor.org/info/rfc5648>>.
- [RFC5844] Wakikawa, R. and S. Gundavelli, "IPv4 Support for Proxy Mobile IPv6", RFC 5844, DOI 10.17487/RFC5844, May 2010, <<http://www.rfc-editor.org/info/rfc5844>>.
- [RFC5845] Muhanna, A., Khalil, M., Gundavelli, S., and K. Leung, "Generic Routing Encapsulation (GRE) Key Option for Proxy Mobile IPv6", RFC 5845, DOI 10.17487/RFC5845, June 2010, <<http://www.rfc-editor.org/info/rfc5845>>.
- [RFC6088] Tsirtsis, G., Giarreta, G., Soliman, H., and N. Montavont, "Traffic Selectors for Flow Bindings", RFC 6088, DOI 10.17487/RFC6088, January 2011, <<http://www.rfc-editor.org/info/rfc6088>>.
- [RFC6089] Tsirtsis, G., Soliman, H., Montavont, N., Giaretta, G., and K. Kuladinithi, "Flow Bindings in Mobile IPv6 and Network Mobility (NEMO) Basic Support", RFC 6089, DOI 10.17487/RFC6089, January 2011, <<http://www.rfc-editor.org/info/rfc6089>>.
- [RFC6275] Perkins, C., Ed., Johnson, D., and J. Arkko, "Mobility Support in IPv6", RFC 6275, DOI 10.17487/RFC6275, July 2011, <<http://www.rfc-editor.org/info/rfc6275>>.
- [RFC7148] Zhou, X., Korhonen, J., Williams, C., Gundavelli, S., and CJ. Bernardos, "Prefix Delegation Support for Proxy Mobile IPv6", RFC 7148, DOI 10.17487/RFC7148, March 2014, <<http://www.rfc-editor.org/info/rfc7148>>.

## 8.2. Informative References

- [RFC2473] Conta, A. and S. Deering, "Generic Packet Tunneling in IPv6 Specification", RFC 2473, DOI 10.17487/RFC2473, December 1998, <<http://www.rfc-editor.org/info/rfc2473>>.
- [RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, DOI 10.17487/RFC4213, October 2005, <<http://www.rfc-editor.org/info/rfc4213>>.
- [RFC4908] Nagami, K., Uda, S., Ogashiwa, N., Esaki, H., Wakikawa, R., and H. Ohnishi, "Multi-homing for small scale fixed network Using Mobile IP and NEMO", RFC 4908, DOI 10.17487/RFC4908, June 2007, <<http://www.rfc-editor.org/info/rfc4908>>.

## Authors' Addresses

Pierrick Seite  
Orange  
4, rue du Clos Courtel, BP 91226  
Cesson-Sevigne 35512  
France

Email: [pierrick.seite@orange.com](mailto:pierrick.seite@orange.com)

Alper Yegin  
Samsung  
Istanbul  
Turkey

Email: [alper.yegin@partner.samsung.com](mailto:alper.yegin@partner.samsung.com)

Sri Gundavelli  
Cisco  
170 West Tasman Drive  
San Jose, CA 95134  
USA

Email: [sgundave@cisco.com](mailto:sgundave@cisco.com)

DMM WG  
Internet-Draft  
Intended status: Standards Track  
Expires: March 29, 2018

P. Seite  
Orange  
A. Yegin  
Actility  
S. Gundavelli  
Cisco  
September 25, 2017

MAG Multipath Binding Option  
draft-ietf-dmm-mag-multihoming-07.txt

Abstract

This specification defines extensions to the Proxy Mobile IPv6 protocol for allowing a mobile access gateway to register more than one proxy care-of-address with the local mobility anchor and to simultaneously establish multiple IP tunnels with the local mobility anchor. This capability allows the mobile access gateway to utilize all the available access networks for routing mobile node's IP traffic.

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 March 29, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
2. Conventions and Terminology . . . . .	4
2.1. Conventions . . . . .	5
2.2. Terminology . . . . .	5
3. Overview . . . . .	5
3.1. Example Call Flow . . . . .	5
3.2. Traffic distribution schemes . . . . .	7
4. Protocol Extensions . . . . .	8
4.1. MAG Multipath-Binding Option . . . . .	8
4.2. MAG Identifier Option . . . . .	10
4.3. New Status Code for Proxy Binding Acknowledgement . . . . .	11
4.4. Signaling Considerations . . . . .	11
5. IANA Considerations . . . . .	12
6. Security Considerations . . . . .	13
7. Acknowledgements . . . . .	13
8. References . . . . .	13
8.1. Normative References . . . . .	13
8.2. Informative References . . . . .	14
Authors' Addresses . . . . .	15



## 1. Introduction

Multihoming support on IP hosts can greatly improve the user experience. With the simultaneous use of multiple access networks, multihoming brings better network connectivity, reliability and improved quality of communication. Following are some of the goals and benefits of multihoming support:

- o Redundancy/Fault-Recovery
- o Load balancing
- o Load sharing
- o Preferences settings

According to [RFC4908], users of Small-Scale Networks can take benefit of multihoming using mobile IP [RFC6275] and Network Mobility (NEMO) [RFC3963] architecture in a mobile and fixed networking environment. This document is introducing the concept of multiple Care-of Addresses (CoAs) [RFC5648] that have been specified since then.

The motivation for this work is to extend Proxy Mobile IPv6 protocol with multihoming extensions [RFC4908] for realizing the following capabilities:

- o using GRE as mobile tunneling, possibly with its key extension [RFC5845] (a possible reason to use GRE is given on Section 3.2).
- o using UDP encapsulation [RFC5844] in order to support NAT traversal in IPv4 networking environment.
- o Prefix Delegation mechanism [RFC7148].
- o Using the vendor specific mobility option [RFC5094], for example to allow the MAG and LMA to exchange information (e.g. WAN interface QoS metrics) allowing to make appropriate traffic steering decision.

Proxy Mobile IPv6 (PMIPv6) relies on two mobility entities: the mobile access gateway (MAG), which acts as the default gateway for the end-node and the local mobility anchor (LMA), which acts as the topological anchor point. Point-to-point links are established, using IP-in-IP tunnels, between MAG and LMA. Then, the MAG and LMA are distributing traffic over these tunnels. All PMIPv6 operations are performed on behalf of the end-node and its correspondent node, it thus makes PMIPv6 well adapted to multihomed architecture as

considered in [RFC4908]. Taking the LTE and WLAN networking environments as an example, the PMIPv6 based multihomed architecture is depicted on Figure 1. Flow-1,2 and 3 are distributed either on Tunnel-1 (over LTE) or Tunnel-2 (over WLAN), while Flow-4 is spread on both Tunnel-1 and 2.

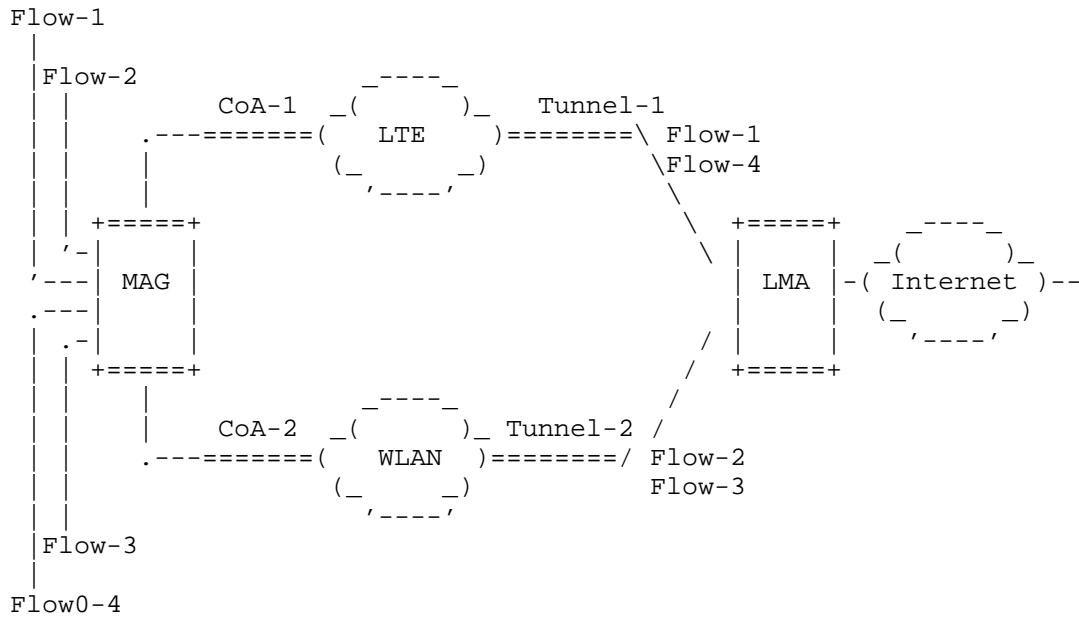


Figure 1: Multihomed MAG using Proxy Mobile IPv6

The current version of Proxy Mobile IPv6 does not allow a MAG to register more than one proxy Care-of-Adresse to the LMA. In other words, only one MAG/LMA link, i.e. IP-in-IP tunnel, can be used at the same time. This document overcomes this limitation by defining the multiple proxy Care-of Addresses (pCoAs) extension for Proxy Mobile IPv6.

## 2. Conventions and Terminology

## 2.1. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

## 2.2. Terminology

All mobility related terms used in this document are to be interpreted as defined in [RFC5213], [RFC5844] and [RFC7148]. Additionally, this document uses the following terms:

IP-in-IP

IP-within-IP encapsulation [RFC2473], [RFC4213]

## 3. Overview

### 3.1. Example Call Flow

Figure 2 is the callflow detailing multi-access support with PMIPv6. The MAG in this example scenario is equipped with both WLAN and LTE interfaces and is also configured with the multihoming functionality. The steps of the callflow are as follows:

Steps (1) and (2): the MAG attaches to both WLAN and LTE networks; the MAG obtains respectively two different proxy care-of-addresses (pCoA).

Step (3): The MAG sends, over the WLAN access, a Proxy Binding Update (PBU) message, with the new MAG Multipath Binding (MMB) and MAG Identifier (MAG-NAI) options to the LMA. The request can be for a physical mobile node attached to the MAG, or for a logical mobile node configured on the mobile node. A logical mobile node is ALWAYS-ATTACHED mobile node configuration enabled on the MAG. The mobility session that is created (i.e. create a Binding Cache Entry) on the LMA will be marked with multipath support.

Step (4): the LMA sends back a Proxy Binding Acknowledgement (PBA) including the HNP and other session parameters allocated for that mobility session.

Step (5): IP tunnel (IP-in-IP, GRE ...) is created over the WLAN access.

Steps (6) to (8): The MAG repeats steps (3) to (5) on the LTE access. The MAG includes the HNP, received on step (4) in the PBU. The LMA

update its binding cache by creating a new mobility session for this MAG.

Steps (9) and (10): The IP hosts MN\_1 and MN\_2 are assigned IP addresses from the mobile network prefix delegated by the MAG.

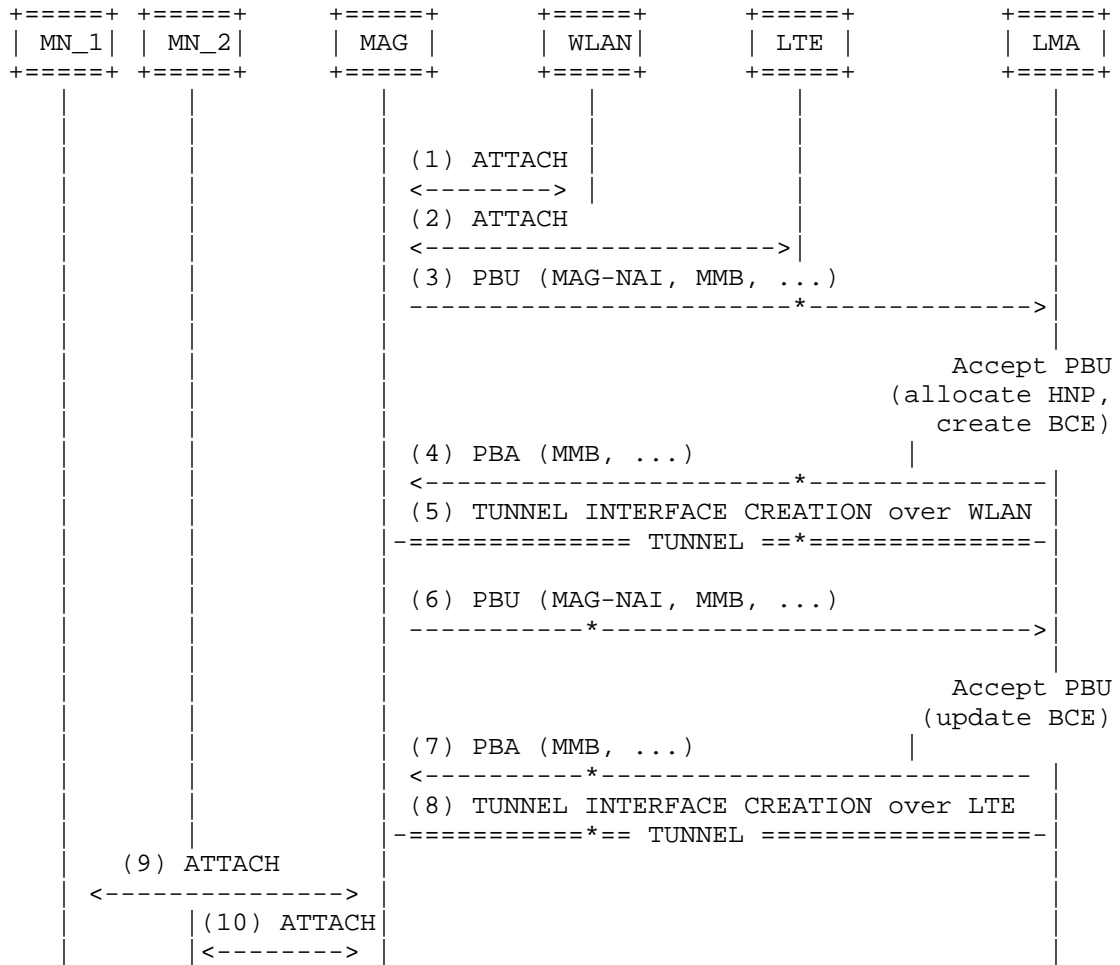


Figure 2: Functional Separation of the Control and User Plane

### 3.2. Traffic distribution schemes

When the MAG has registered multipath binding with the LMA, there will be multiple established overlay tunnels between them. The MAG and the LMA can use any one, or more of the available tunnels paths for routing the mobile node's IP traffic. This specification does not recommend, or define any specific traffic distribution scheme, however it identifies two well-known approaches that implementations can potentially use. These approaches are, Per-flow and Per-packet Traffic distribution schemes.

#### Per-Flow Traffic Distribution:

- o In this approach the MAG and the LMA associate each of the IP flows (upstream and downstream) to a specific tunnel path. The packets in a given IP flow are always routed on the same overlay tunnel path; they are never split and routed concurrently on more than one tunnel path. It is possible a given flow may be moved from one tunnel path to another, but the flow is never split. The decision to bind a given IP flow to a specific tunnel path is based on traffic distribution policy. This traffic distribution policy is either statically configured on both the MAG and the LMA, or dynamically negotiated over Proxy Mobile IPv6 signaling. The Flow Binding extension [RFC6089] and Traffic Selectors for Flow Bindings [RFC6088] defines the mechanism and the semantics for exchanging the traffic policy between two tunnel peers and the same mechanism and the mobility options are used here.

#### Per-Packet Traffic Distribution:

- o In this approach, packets belonging a given IP flow will be split and routed across more than one tunnel paths. The exact approach for traffic distribution, or the distribution weights is outside the scope of this specification. In a very simplistic approach, assuming the established tunnel paths have symmetric characteristics, the packets can be equally distributed on all the available tunnel paths. In a different scenario when the links have different speeds, the chosen approach can be based on weighted distribution (Ex: n:m ratio). However, in any of these chosen approaches, implementations have to be sensitive to issues related to asymmetric link characteristics and the resulting issues such as re-ordering, buffering and the impact to the application performance. Care must be taken to ensure there is no negative impact to the application performance due to the use of this approach.

4. Protocol Extensions

4.1. MAG Multipath-Binding Option

The MAG Multipath-Binding option is a new mobility header option defined for use with Proxy Binding Update and Proxy Binding Acknowledgement messages exchanged between the local mobility anchor and the mobile access gateway.

This mobility header option is used for requesting multipath support. It indicates that the mobile access gateway is requesting the local mobility anchor to register the current care-of address associated with the request as one of the many care-addresses through which the mobile access gateway can be reached. It is also for carrying the information related to the access network associated with the care-of address.

The MAG Multipath-Binding option has an alignment requirement of 8n+2. Its format is as shown in Figure 3:

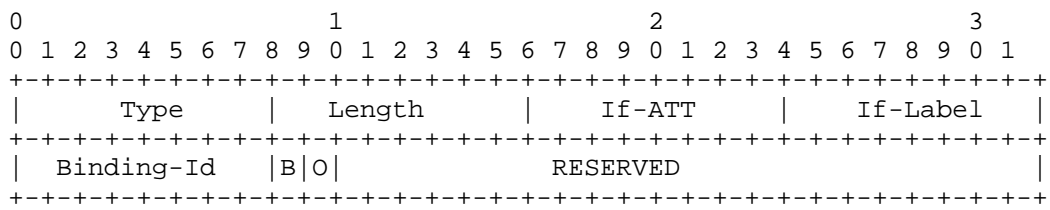


Figure 3: MAG Multipath Binding Option

Type

<IANA-1> To be assigned by IANA.

Length

8-bit unsigned integer indicating the length of the option in octets, excluding the type and length fields.

Interface Access-Technology Type (If-ATT)

This 8-bit field identifies the Access-Technology type of the interface through which the mobile node is connected. The permitted values for this are from the Access Technology Type registry defined in [RFC5213].

Interface Label (If-Label)

This 8-bit unsigned integer represents the interface label.

The interface label is an identifier configured on the WAN interface of the MAG. All the WAN interfaces of the MAG that are used for sending PBU messages are configured with a label. The labels merely identify the type of WAN interface and are primarily used in Application routing policies. For example, a Wi-Fi interfaces can be configured with a label RED and a LTE interface with a label BLUE. Furthermore, the same label may be configured on two WAN interfaces of similar characteristics (Ex: Two Ethernet interfaces with the same label).

Interfaces labels are signaled from the MAG to LMA in the PBU messages and both the LMA and MAG will be able to mark each of the dynamically created Binding/Tunnel with the associated label. These labels are used in generating consistent application routing rules on the both the LMA and the MAG. For example, there can be a policy requiring HTTP packets to be routed over interface that has Label RED, and if any of the RED interfaces are not available, the traffic needs to be routed over the BLUE interface. The MAG and the LMA will be able to apply this Routing Rule with the exchange of Labels in PBU messages and by associating the application flows to tunnels with the matching labels.

#### Binding-Identifier (BID)

This 8-bit unsigned integer is used for identifying the binding. The permitted values are 1 through 254. The values, 0 and 255 are reserved.

The MAG identifies each of the mobile node's binding with a unique identifier. The MAG includes the identifier in the PBU message and when the PBU request is accepted by the LMA, the resulting Binding is associated with this binding identifier.

#### Bulk Re-registration Flag (B)

This flag, if set to a value of (1), is to notify the local mobility anchor to consider this request as a request to update the binding lifetime of all the mobile node's bindings, upon accepting this specific request. This flag MUST NOT be set to a value of (1), if the value of the Registration Overwrite Flag (O) is set to a value of (1).

#### Binding Overwrite (O)

This flag, if set to a value of (1), notifies the local mobility anchor that upon accepting this request, it should replace all of

the mobile node's existing bindings with this binding. This flag MUST NOT be set to a value of (1), if the value of the Bulk Re-registration Flag (B) is set to a value of (1). This flag MUST be set to a value of (0), in de-registration requests.

Reserved

This field is unused in this specification. The value MUST be set to zero (0) by the sender and MUST be ignored by the receiver.

4.2. MAG Identifier Option

The MAG Identifier option is a new mobility header option defined for use with Proxy Binding Update and Proxy Binding Acknowledgement messages exchanged between the local mobility anchor and the mobile access gateway. This mobility header option is used for conveying the MAG's identity.

This option does not have any alignment requirements.

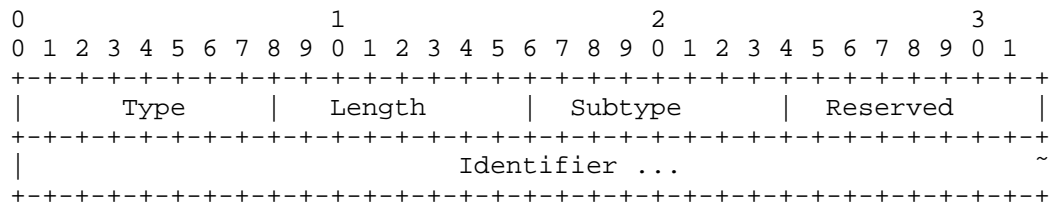


Figure 4: MAG Identifier Option

Type

<IANA-2> To be assigned by IANA.

Length

8-bit unsigned integer indicating the length of the option in octets, excluding the type and length fields.

Subtype

One byte unsigned integer used for identifying the type of the Identifier field. Accepted values for this field are the registered type values from the Mobile Node Identifier Option Subtypes registry.

Reserved



This field is unused in this specification. The value MUST be set to zero (0) by the sender and MUST be ignored by the receiver.

#### Identifier

A variable length identifier of type indicated in the Subtype field.

### 4.3. New Status Code for Proxy Binding Acknowledgement

This document defines the following new Status Code value for use in Proxy Binding Acknowledgement message.

The LMA SHOULD use this error code when rejecting a Proxy Binding Update message from a MAG requesting a multipath binding. Following is the potential reason for rejecting the request:

- o The LMA does not support multipath binding.

CANNOT\_SUPPORT\_MULTIPATH\_BINDING (Cannot Support Multipath Binding):  
<IANA-4>

### 4.4. Signaling Considerations

- o The MAG when requesting multipath support MUST include the MAG Multipath Binding Option (Section 4.1) in each of the PBU messages that it sends through the different WAN interfaces. The inclusion of this option serves as a hint that the MAG is requesting Multipath support. Furthermore, the MAG Identifier option MUST also be present in the PBU message.
- o If the MAG is aware that the LMA supports the multipath feature defined in this specification and if it chooses to enable multiple path feature, then it can send the PBU packets for each of the paths, either sequentially, or concurrently. However, if the MAG is not aware of the LMA capability, then it should first discover the LMA capability by sending PBU packets with multipath on only one path first. This will ensure the LMA will not be over-writing the binding of one path with the other path.
- o If the LMA supports multipath capability as defined in this specification and if it enables the same for a mobile node's session per the MAG's request, then the LMA MUST include the Multipath Binding Option (Section 4.1), without the MAG NAI Option Section 4.2 in the corresponding PBA reply.
- o If the LMA is a legacy LMA that does not support this specification, the LMA will skip the MAG Multipath Binding option

(and MAG NAI option) and process the rest of the message as specified in the base Proxy Mobile IPv6 specification ([RFC5213]). Furthermore, the LMA will not include the MAG Multipath Binding option (or the MAG NAI Option) in the PBA message. The MAG on receiving the PBA message without the MAG Multipath Binding option SHOULD disable Multipath support for the mobile node.

- o If the mobile node is not authorized for Multipath support, then the LMA will reject the request by sending a PBA message with the Status field value set to CANNOT\_SUPPORT\_MULTIPATH\_BINDING (Section 4.3). The LMA will echo the MAG Multipath Binding option and the MAG NAI option in the PBA message. The MAG on receiving this message SHOULD disable Multipath support for the mobile node.
- o If the request for multipath support is accepted, then the LMA SHOULD enable multipath support for the mobile node and SHOULD also echo the MAG Multipath Binding option and the MAG NAI option in the corresponding PBA message.

## 5. IANA Considerations

This document requires the following IANA actions.

- o Action-1: This specification defines a new mobility option, the MAG Multipath-Binding option. The format of this option is described in Section 4.1. The type value <IANA-1> for this mobility option needs to be allocated from the Mobility Options registry at <<http://www.iana.org/assignments/mobility-parameters>>. RFC Editor: Please replace <IANA-1> in Section 4.1 with the assigned value and update this section accordingly.
- o Action-2: This specification defines a new mobility option, the MAG Identifier option. The format of this option is described in Section 4.2. The type value <IANA-2> for this mobility option needs to be allocated from the Mobility Options registry at <<http://www.iana.org/assignments/mobility-parameters>>. RFC Editor: Please replace <IANA-2> in Section 4.2 with the assigned value and update this section accordingly.
- o Action-3: This document defines a new status value, CANNOT\_SUPPORT\_MULTIPATH\_BINDING (<IANA-3>) for use in Proxy Binding Acknowledgement message, as described in Section 4.3. This value is to be assigned from the "Status Codes" registry at <<http://www.iana.org/assignments/mobility-parameters>>. The allocated value has to be greater than 127. RFC Editor: Please replace <IANA-3> in Section 4.3 with the assigned value and update this section accordingly.

## 6. Security Considerations

This specification allows a mobile access gateway to establish multiple Proxy Mobile IPv6 tunnels with a local mobility anchor, by registering a care-of address for each of its connected access networks. This essentially allows the mobile node's IP traffic to be routed through any of the tunnel paths based on the negotiated flow policy. This new capability has no impact on the protocol security. Furthermore, this specification defines two new mobility header options, MAG Multipath-Binding option and the MAG Identifier option. These options are carried like any other mobility header option as specified in [RFC5213]. Therefore, it inherits security guidelines from [RFC5213]. Thus, this specification does not weaken the security of Proxy Mobile IPv6 Protocol, and does not introduce any new security vulnerabilities.

## 7. Acknowledgements

The authors of this draft would like to acknowledge the discussions and feedback on this topic from the members of the DMM working group. The authors would also like to thank Jouni Korhonen, Jong Hyouk Lee, Dirk Von-Hugo, Seil Jeon, Carlos Bernardos, Robert Sparks, Adam Roach, Kathleen Moriarty, Hilarie Orman, Ben Campbell, Warren Kumari, for their review feedback. Special thanks to Mirja Kuehlewind for a very thorough review and suggesting many text improvements.

## 8. References

### 8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC3963] Devarapalli, V., Wakikawa, R., Petrescu, A., and P. Thubert, "Network Mobility (NEMO) Basic Support Protocol", RFC 3963, DOI 10.17487/RFC3963, January 2005, <<https://www.rfc-editor.org/info/rfc3963>>.
- [RFC5094] Devarapalli, V., Patel, A., and K. Leung, "Mobile IPv6 Vendor Specific Option", RFC 5094, DOI 10.17487/RFC5094, December 2007, <<https://www.rfc-editor.org/info/rfc5094>>.
- [RFC5213] Gundavelli, S., Ed., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6",

- RFC 5213, DOI 10.17487/RFC5213, August 2008, <<https://www.rfc-editor.org/info/rfc5213>>.
- [RFC5648] Wakikawa, R., Ed., Devarapalli, V., Tsirtsis, G., Ernst, T., and K. Nagami, "Multiple Care-of Addresses Registration", RFC 5648, DOI 10.17487/RFC5648, October 2009, <<https://www.rfc-editor.org/info/rfc5648>>.
- [RFC5844] Wakikawa, R. and S. Gundavelli, "IPv4 Support for Proxy Mobile IPv6", RFC 5844, DOI 10.17487/RFC5844, May 2010, <<https://www.rfc-editor.org/info/rfc5844>>.
- [RFC5845] Muhanna, A., Khalil, M., Gundavelli, S., and K. Leung, "Generic Routing Encapsulation (GRE) Key Option for Proxy Mobile IPv6", RFC 5845, DOI 10.17487/RFC5845, June 2010, <<https://www.rfc-editor.org/info/rfc5845>>.
- [RFC6088] Tsirtsis, G., Giarreta, G., Soliman, H., and N. Montavont, "Traffic Selectors for Flow Bindings", RFC 6088, DOI 10.17487/RFC6088, January 2011, <<https://www.rfc-editor.org/info/rfc6088>>.
- [RFC6089] Tsirtsis, G., Soliman, H., Montavont, N., Giaretta, G., and K. Kuladinithi, "Flow Bindings in Mobile IPv6 and Network Mobility (NEMO) Basic Support", RFC 6089, DOI 10.17487/RFC6089, January 2011, <<https://www.rfc-editor.org/info/rfc6089>>.
- [RFC6275] Perkins, C., Ed., Johnson, D., and J. Arkko, "Mobility Support in IPv6", RFC 6275, DOI 10.17487/RFC6275, July 2011, <<https://www.rfc-editor.org/info/rfc6275>>.
- [RFC7148] Zhou, X., Korhonen, J., Williams, C., Gundavelli, S., and CJ. Bernardos, "Prefix Delegation Support for Proxy Mobile IPv6", RFC 7148, DOI 10.17487/RFC7148, March 2014, <<https://www.rfc-editor.org/info/rfc7148>>.

## 8.2. Informative References

- [RFC2473] Conta, A. and S. Deering, "Generic Packet Tunneling in IPv6 Specification", RFC 2473, DOI 10.17487/RFC2473, December 1998, <<https://www.rfc-editor.org/info/rfc2473>>.
- [RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, DOI 10.17487/RFC4213, October 2005, <<https://www.rfc-editor.org/info/rfc4213>>.

[RFC4908] Nagami, K., Uda, S., Ogashiwa, N., Esaki, H., Wakikawa, R., and H. Ohnishi, "Multi-homing for small scale fixed network Using Mobile IP and NEMO", RFC 4908, DOI 10.17487/RFC4908, June 2007, <<https://www.rfc-editor.org/info/rfc4908>>.

#### Authors' Addresses

Pierrick Seite  
Orange  
4, rue du Clos Courtel, BP 91226  
Cesson-Sevigne 35512  
France

Email: [pierrick.seite@orange.com](mailto:pierrick.seite@orange.com)

Alper Yegin  
Actility  
Turkey

Email: [alper.yegin@actility.com](mailto:alper.yegin@actility.com)

Sri Gundavelli  
Cisco  
170 West Tasman Drive  
San Jose, CA 95134  
USA

Email: [sgundave@cisco.com](mailto:sgundave@cisco.com)



Homenet Working Group  
Internet-Draft  
Intended status: Informational  
Expires: January 9, 2017

M. Cullen  
Painless Security  
M. Zhang  
Huawei Technologies  
July 8, 2016

Considerations for Bandwidth Aggregation  
draft-mrc-banana-considerations-01

Abstract

This document lists a number of architectural and technical topics that should be considered in the design and implementation of Bandwidth Aggregation mechanisms.

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 9, 2017.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	2
2. What is Bandwidth Aggregation? . . . . .	3
3. Taxonomy of Solutions . . . . .	3
3.1. Tunnel-Based Solutions . . . . .	3
3.2. Per-Packet vs. Per-Flow Multiplexing . . . . .	3
4. Considerations for All Solutions . . . . .	4
4.1. Link Characteristics and Performance . . . . .	4
4.2. Bypass Traffic . . . . .	4
4.3. Capped or Tariffed Interfaces . . . . .	4
4.4. Learning from History (Multilink PPP) . . . . .	4
5. Considerations for Tunnel-Based Solutions . . . . .	5
5.1. Tunnel Overhead . . . . .	5
5.2. MTU Issues . . . . .	5
5.2.1. Fragmentation Issues . . . . .	5
5.2.2. Issues with MTU Changes . . . . .	5
6. Considerations for Per-Packet Solutions . . . . .	5
6.1. Packet Ordering . . . . .	5
6.2. Transport Layer Algorithms . . . . .	5
7. Considerations for Per-Flow Solutions . . . . .	6
7.1. Granularity Issues . . . . .	6
7.2. Aggregated Flows . . . . .	6
7.3. Encrypted Traffic . . . . .	7
8. Practical Considerations . . . . .	7
8.1. Use Available Information . . . . .	7
8.2. Theory is No Substitute for Experience . . . . .	7
9. Security Considerations . . . . .	7
9.1. Binding Tunnel Endpoints . . . . .	7
10. Appendix A: List of Solutions . . . . .	7
10.1. Multilink PPP . . . . .	8
10.2. GRE Tunnel Binding . . . . .	8
10.3. LISP-Based Solution . . . . .	8
10.4. MIP-Based Solution . . . . .	8
10.5. MP-TCP-Based Solution . . . . .	8
11. Informative References . . . . .	8
Authors' Addresses . . . . .	8

## 1. Introduction

There are currently several bandwidth aggregation solutions being discussed within the IETF or other parts of the Internet industry. This document discusses a number of technical and architectural facts that should be considered in the design and implementation of those solutions. This document is intended to provide useful information to the community, not to state requirements or advocate for a particular solution.



There is one simple thought underlying many of the considerations in this document: the goals of bandwidth aggregation are to increase the effective bandwidth available to customers and improve the reliability of customers' Internet access by using all of the available links, not just one of them. Intuitively, two links should have more bandwidth and reliability than one link, but experience shows that it is actually quite hard to design a bandwidth aggregation solution that will achieve the desired goals in all cases, and quite easy to design a solution that will reduce the effective bandwidth or decrease the reliability of Internet access in an unacceptably high number of cases. Many of the considerations in this document are intended to point out why that happens, so that solutions and implementations can avoid known pitfalls in this area.

[Note: This document is a work in progress. Feedback on the existing content is welcome, as well as feedback on other considerations that should be included. Please send any feedback to the Bandwidth Aggregation mailing list: [banana@ietf.org](mailto:banana@ietf.org)]

## 2. What is Bandwidth Aggregation?

[TBD]

## 3. Taxonomy of Solutions

This section attempts to categorize bandwidth aggregation solutions along several axes, providing a taxonomy that we can use to describe and reason about individual solutions. [Note: This section is largely TBD.]

### 3.1. Tunnel-Based Solutions

Many of the Bandwidth Aggregations currently under discussion are tunnel-based solutions. They tunnel traffic over the links that are being aggregated, and recombine the traffic on the remote end.

[Insert ASCII image of tunnel-based approach.]

There is at least one proposal for Bandwidth Aggregation (the MP-TCP-based approach) that does not use tunnels. The considerations for tunnel-based solutions listed below may not apply to non-tunnel-based solutions.

### 3.2. Per-Packet vs. Per-Flow Multiplexing

The solutions currently under discussion use several different methods to determine which traffic will be sent over which interface.

These methods can be grouped into two categories: per-packet multiplexing and per-flow multiplexing.

Per-packet multiplexing aggregates the bandwidth by sending the desired proportion of packets over each interface. In these solutions, packets from single flow (such as a TCP connection) may be split across multiple interfaces and will need to be recombined at the remote end. However, the ability to multiplex on a per-packet basis makes it possible to most precisely apportion traffic across the available bandwidth.

Per-flow multiplexing involves choosing a single interface for each flow (i.e. TCP connection or application session) and sending all of the packets for a single flow across that interface. In these solutions, the flow do not need to be combined on the remote end. However, the ability to balance traffic between multiple links may be limited if there are only a small number of traffic flows active.

#### 4. Considerations for All Solutions

This section describes potential issues that should be considered in the design and implementation of all bandwidth aggregation solutions.

##### 4.1. Link Characteristics and Performance

##### 4.2. Bypass Traffic

##### 4.3. Capped or Tariffed Interfaces

In some cases, bandwidth aggregation may be performed between dedicated links and links that have traffic caps or tariffs associated with additional use. In these cases, customer may want to use bandwidth aggregation to increase the performance of some applications, while other applications (e.g. firmware upgrades or content downloads) may be limited to using the dedicated link. Solutions that wish to support this capability will need to support having a set of traffic that will be distributed using the bandwidth aggregation algorithms, and a set of traffic that will not.

##### 4.4. Learning from History (Multilink PPP)

The IETF has a venerable, standard, implemented solution to this sort of problem: Multilink PPP. Unfortunately, it is commonly said that experience with Multilink PPP did not find that it increased the effective bandwidth when it was used to share two identical ISDN lines, compared to the bandwidth that was achieved from using only one line...

[Note: We should attempt to determine if this is true and, if so, find any research papers or other documentation that might help us understand why this was true, so that we might learn from history.]

## 5. Considerations for Tunnel-Based Solutions

### 5.1. Tunnel Overhead

Tunneling involves more overhead than sending non-tunnelled traffic for two reasons: the extra IP and tunnel headers that must be included in each packet, and any tunnel management traffic that must be exchanged. This means that, in order to achieve increased effective bandwidth by aggregating traffic across more than one link, the raw bandwidth across multiple links must be higher than the bandwidth on a single link by a large enough margin to compensate for the tunnel overhead, so that increased effective bandwidth will result.

### 5.2. MTU Issues

There are a number of MTU Issues associated with all tunneling mechanisms, and there is a different set of MTU issues associated with any mechanism that changes the MTU of packets within a given flow.

[Note: This section is TBD.]

#### 5.2.1. Fragmentation Issues

#### 5.2.2. Issues with MTU Changes

## 6. Considerations for Per-Packet Solutions

### 6.1. Packet Ordering

### 6.2. Transport Layer Algorithms

There are transport layer congestion control algorithms implemented in every TCP/IP stack. It is the purpose of these algorithms to ramp up the speed of a TCP connection slowly, and to back off at the first sign of congestion (i.e. packet loss). There are also algorithms which are designed to detect packet loss as quickly as possible by analyzing the protocol round-trip times, and deciding that a packet has been lost whenever there is a longer delay than expected before an acknowledgement is received. Per-packet solutions run the risk of interacting pathologically with these algorithms.

For example, if traffic from a single flow is being demultiplexed across two links with significantly different round-trip times (i.e. different latencies), the TCP retransmission algorithms may be triggered for packets that traverse the higher latency link. This may cause the TCP congestion control algorithms to inaccurately detect congestion (even when neither link is congested) and slow down the speed of the TCP connection. In these cases, the throughput of each TCP connection may be reduced, thus reducing the performance of a customer's applications to the point where their applications would have run faster over a single link.

This problem can potentially be avoided by avoiding aggregation of links with significantly different latencies. However, it may be desirable to perform bandwidth aggregation across those links in some cases.

## 7. Considerations for Per-Flow Solutions

This section describes some potential issues that should be considered in the design of per-flow bandwidth aggregation solutions.

### 7.1. Granularity Issues

Per-Flow demultiplexing is in widespread use for traffic engineering and load balancing in carrier and corporate networks. Within those networks, there are a very large number of flows, so being able to direct traffic on a per-flow basis will generally be sufficient to achieve acceptable load-balancing or link aggregation.

However, the number of flows generated by a single home or small office might not provide sufficient granularity to achieve the desired level of bandwidth aggregation. Also some flows, such as streaming video flows, might use far more bandwidth than other, such as downloading a single image on a web page. It is not always possible to predict which flows will be high-bandwidth flows, and which will require less bandwidth.

### 7.2. Aggregated Flows

Some Internet flows are aggregated into single, larger flows at the end-nodes. This would include VPN traffic that is tunneled to a corporate intranet, or other tunneled traffic such as Teredo traffic for IPv6. Use of these mechanisms can prevent proper classification of traffic into separate flows, thus exacerbating the granularity issues described above.

### 7.3. Encrypted Traffic

In some cases such as secure VPN traffic, the contents of packets may be encrypted in a way that does not allow a middlebox to see the transport-layer flow information (such as TCP or UDP ports). In these cases, it might not be possible to properly separate multiple flows between a single set of endpoints. This can exacerbate the granularity issues described above.

## 8. Practical Considerations

### 8.1. Use Available Information

In many of the environments in which these mechanisms will be deployed, there is already considerable information available about link quality, lost packets, traffic loads and effective bandwidth. It is possible to use that information to actively tune a bandwidth aggregation solution to achieve optimal effective bandwidth. This information can also be used to detect situations in which the link quality of a secondary link is not sufficient to provide enough additional bandwidth to compensate for the bandwidth aggregation overhead.

### 8.2. Theory is No Substitute for Experience

Because of the complexity of the algorithms implemented at multiple layers of the TCP/IP Stack, many things that would appear to work in theory or in limited simulation do not have the expected results when deployed in a real-world environment. Therefore, it would be highly desirable to have real-world experience running a bandwidth aggregation mechanism in an operational network before standardizing it within the IETF.

## 9. Security Considerations

### 9.1. Binding Tunnel Endpoints

## 10. Appendix A: List of Solutions

This is a (possibly incomplete) list of current or proposed solutions for Bandwidth Aggregation. The descriptions in this section (when present) were provided by the proponents of each solution. This list is provided only as a source of information about possible solutions, not as a recommendation for or against any of these solutions.

[Note: Insert information from Google Doc in this section.]

- 10.1. Multilink PPP
- 10.2. GRE Tunnel Binding
- 10.3. LISP-Based Solution
- 10.4. MIP-Based Solution
- 10.5. MP-TCP-Based Solution
- 11. Informative References

[RFC6126] Chroboczek, J., "The Babel Routing Protocol", RFC 6126,  
DOI 10.17487/RFC6126, April 2011,  
<<http://www.rfc-editor.org/info/rfc6126>>.

Authors' Addresses

Margaret Cullen  
Painless Security  
14 Summer Street, Suite 202  
Malden, MA 02148  
USA

Phone: +1 781 405-7464  
Email: [margaret@painless-security.com](mailto:margaret@painless-security.com)  
URI: <http://www.painless-security.com>

Mingui Zhang  
Huawei Technologies

Email: [zhangmingui@huawei.com](mailto:zhangmingui@huawei.com)

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

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as

described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
2. Reference architecture . . . . .	4
3. Operator requirements . . . . .	6
4. Existing solutions . . . . .	8
4.1. Datalink solutions for hybrid access networks . . . . .	8
4.2. Network layer solutions for hybrid access networks . . . . .	8
4.3. Transport layer solutions . . . . .	9
4.4. Application layer solutions . . . . .	9
5. The transparent MPTCP mode . . . . .	11
6. Security considerations . . . . .	15
7. IANA Considerations . . . . .	16
8. Conclusion . . . . .	17
9. References . . . . .	18
9.1. Normative References . . . . .	18
9.2. Informative References . . . . .	18
Authors' Addresses . . . . .	21



## 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 :

- o a single homed end host H that is attached through one interface (e.g. WiFi) to a Hybrid Customer Premises Equipment (HCPE)
- o 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.
- o A Hybrid Aggregation Gateway (HAG) that is reachable over both access networks
- o a regular server, S

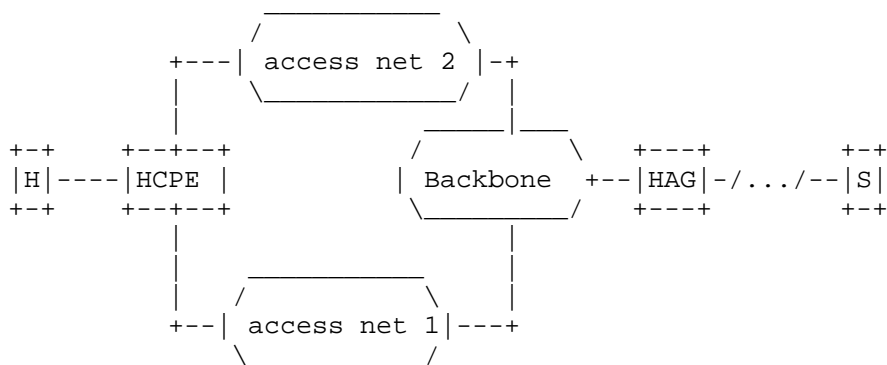


Figure 1: Hybrid access networks

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.

- o Req1: the bonding solution MUST support IPv6 and IPv4
- o Req2: the bonding solution SHOULD minimize the additional delay that it introduces in the network
- o 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
- o Req4 : the bonding solution MUST not use more than one public IPv4 address per customer
- o Req5 : the bonding solution SHOULD enable the operator to trace the connections created by a given customer
- o 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
- o Req7 : the bonding solution MUST be decoupled from the underlying fixed/mobile access network
- o Req8: the bonding solution MUST be able to efficiently load-balance the packets belonging to a single TCP connection over several access networks
- o 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 :

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

- o the HAG must maintain a pool of public addresses

- o 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.

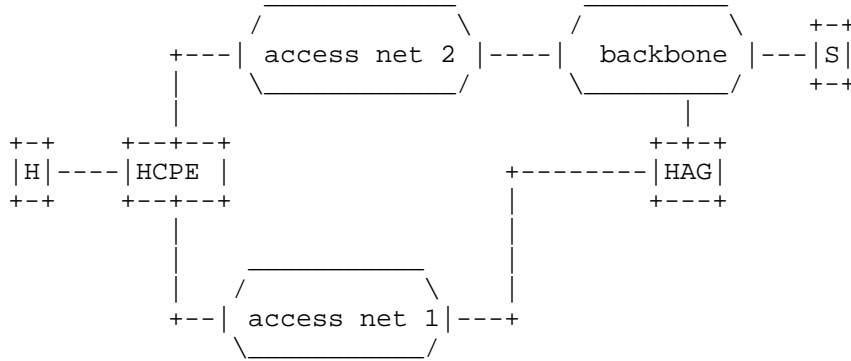


Figure 2: Reference architecture

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.

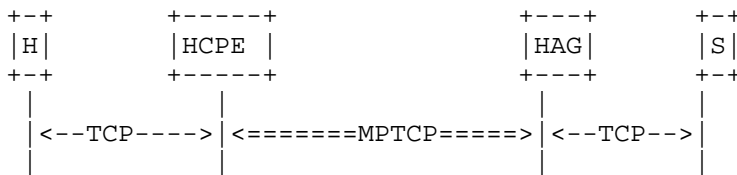


Figure 3: The TCP<->Multipath TCP proxies used on the HCPE and the HAG

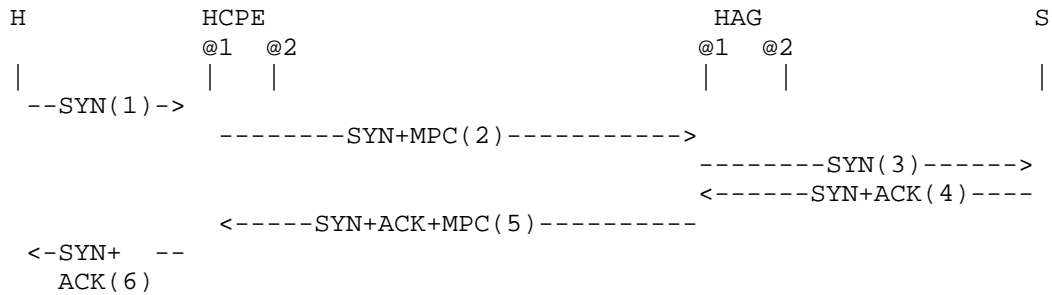


Figure 4: Creation of the initial subflow with the transparent mode

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 :

- o (1) H sends a SYN towards S@.
- o (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@.
- o (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@.
- o (4) The server replies with a SYN+ACK to confirm the establishment of the connection.
- o (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.
- o (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 create the additional subflow without waiting for the ADD\_ADDR option.

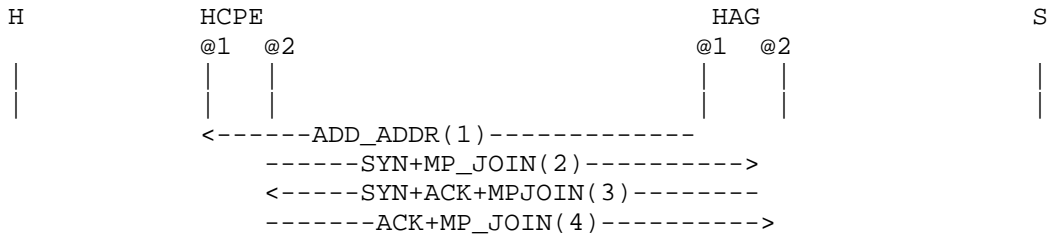


Figure 5: Creation of the second subflow by the HCPE with the transparent MPTCP mode

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.

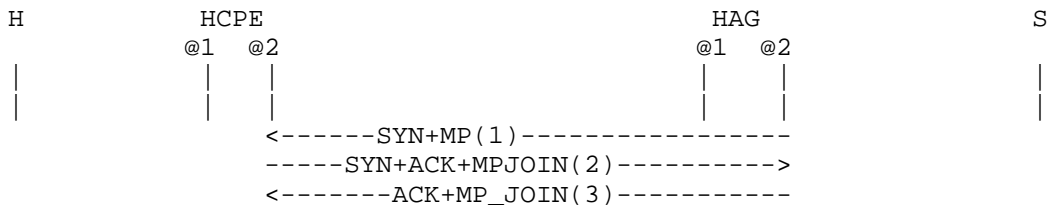


Figure 6: Creation of the second subflow by the HAG with the transparent MPTCP mode

## 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-model], 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



INTERNET-DRAFT  
Intended Status: Informational

M. Cullen  
Painless Security  
N. Leymann  
C. Heidemann  
Deutsche Telekom AG  
M. Boucadair  
France Telecom  
H. Deng  
China Mobile  
M. Zhang  
B. Sarikaya  
Huawei  
October 31, 2016

Expires: May 4, 2017

Problem Statement: Bandwidth Aggregation for Internet Access  
draft-zhang-banana-problem-statement-03.txt

#### Abstract

Bandwidth aggregation capabilities for Internet access services can significantly improve end user's Quality of Experience. Such capabilities are especially attractive in environments where multi-interfaced boxes become commonplace and can technically connect to various access networks, both wired and wireless.

This document describes the problems with bandwidth aggregation for Internet access. A set of requirements are derived from the said problems.

#### Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at  
<http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at  
<http://www.ietf.org/shadow.html>

## Copyright and License Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
2. Acronyms and Terminology . . . . .	3
3. Generic Reference Model . . . . .	4
4. Problem Areas . . . . .	4
4.1. Addressing . . . . .	4
4.2. Traffic Classification . . . . .	5
4.3. Traffic Distribution . . . . .	5
4.4. Traffic Recombination . . . . .	6
4.4.1. Reordering Buffer . . . . .	6
4.5. Bypass . . . . .	7
4.6. Measurement . . . . .	8
4.7. Policy Control . . . . .	8
5. Requirements . . . . .	9
6. Related IETF Work . . . . .	10
6.1. GRE Tunnel Bonding . . . . .	10
6.2. LISP . . . . .	11
6.3. Mobile IP . . . . .	11
6.4. Multipath TCP Proxy . . . . .	11
7. Security Considerations . . . . .	11
8. IANA Considerations . . . . .	12
9. Acknowledgements . . . . .	12
10. References . . . . .	12
10.1. Normative References . . . . .	12
10.2. Informative References . . . . .	13
Appendix A. Additional Requirements . . . . .	14
Author's Addresses . . . . .	16

## 1. Introduction

Use cases of BANDwidth Aggregation for interNet Access (BANANA, a.k.a., Hybrid Access) are described in the Technical Report [TR-348] published by Broadband Forum: by providing Hybrid Access, Service Providers can provide customers with increased access bandwidth and higher access reliability; Service delivery may also be fostered to access the Internet by means of providing a LTE (Long Term Evolution) connection while the wired line is being constructed.

Although host-based Hybrid Access is possible, the scope of this document is restricted to be network-based only. Host-based might be standardized in other places, such as the MIF Working Group.

Design techniques that are being investigated, developed and sometimes deployed to facilitate bandwidth aggregation and improve the resiliency of access conditions raise several problems from various standpoints: traffic routing and forwarding, congestion control, security, etc. This document aims at presenting these problems regardless of the nature of the design technique. It also intends to derive requirements accordingly, and which should be addressed by any bandwidth aggregation technique. Typically, this is one of the inputs that are expected from the IETF community.

A common framework will be sketched, including required functional modules and interactions. The various solution proposals (e.g., GRE, LISP, MIP, MPTCP) can be viewed as applicability assessments of the framework. To support BANANA, the problems to be addressed includes: addressing, traffic classification, distribution and recombination, bypassing, measurement and policy control. To address these problems, we may work as a group to

- specify the encapsulation format;
- develop a common control plane;
- and define or suggest approaches to address BANANA problems developed in this document.

## 2. Acronyms and Terminology

**Hybrid Access:** The coordinated and simultaneous use of two heterogeneous access paths (e.g., DSL and LTE) [TR-348].

**CPE:** Customer Premises Equipment. An equipment which is the property of the network operator and located on the customer premises.

**HG:** Home Gateway. A CPE device that is enhanced to support the simultaneous use of both fixed broadband and 3GPP access connections.

HAAP: Hybrid Access Aggregation Point. A logical function in Operator's network, terminating bonded connections while offering high speed Internet.

PDP: Packet Data Protocol. A packet transfer protocol used in wireless GPRS (General Packet Radio Service)/HSDPA (High Speed Downlink Packet Access) networks.

PPPoE: Point-to-Point Protocol over Ethernet is a network protocol for encapsulating PPP frames inside Ethernet frames.

DHCP: Dynamic Host Configuration Protocol [RFC2131].

DNS: Domain Name System [RFC1035].

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

### 3. Generic Reference Model

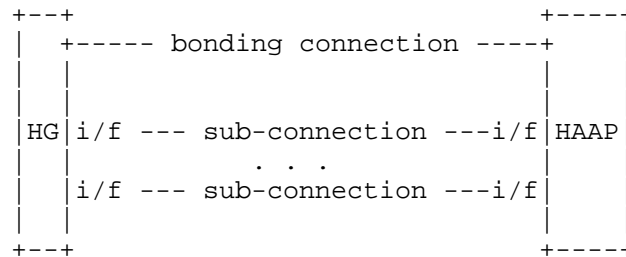


Figure 3.1: Reference model of the Hybrid Access

Customers access the Internet using the Hybrid Access which comprises of several key component functions as shown in Figure 3.1: the Home Gateway (HG) as one peer, the Hybrid Access Aggregation Point (HAAP) as the other peer, the bonding connection between the two peers and the sub-connections that logically make up the bonding connection.

### 4. Problem Areas

#### 4.1. Addressing

At the HG side, interface addresses of sub-connections are locally acquired upon the bootstrap of the system by means of certain existing protocols such as Point-to-Point Protocol over Ethernet (PPPoE) [RFC2516] and Packet Data Protocol (PDP). At the HAAP side,



interface addresses are usually pre-configured by operators. HG and HAAP will rely on the control protocol that is to be developed to exchange these addresses. Afterwards, sub-connections are de-multiplexed by their interface addresses. Both IPv4 and IPv6 should be supported.

End users behind the HG box will regard the bonding connection as a traditional connection to the Internet. With the established sub-connections, connectivity between the HG and HAAP has been built up, therefore endpoint addresses for this bonding connection can be obtained from existing protocols, e.g., DHCP and DNS.

#### 4.2. Traffic Classification

Traffic classification occurs before the flows or packets are distributed among the connections. HG and HAAP should support the classification function that marks a flow or packets which are to be further processed by the traffic distribution function or bypass the Hybrid Access (See Section 4.5). Classification criteria include IP addresses, port numbers, etc. Traffic classification policies can be defined by end users and service providers and must be enforced by the HG and HAAP equipments.

#### 4.3. Traffic Distribution

For traffic that is to be distributed across multiple sub-connections, equal load balancing generally applies, possibly inferred by the bandwidth that is available in each link that supports sub-connection. Unequal load balancing should be supported as well. Traffic may be distributed across sub-connections as a function of their available bandwidth. Traffic may also be split in such a way that whenever one sub-connection is saturated, then traffic is forwarded over a secondary sub-connection.

There are two kinds of traffic distribution methods for the Hybrid Access: per-flow load balancing and per-packet load sharing. The per-flow load balancing method is used to be widely adopted in core IP networks. It is suitable for the scenario where there are a large number of flows to be distributed. For end users, usually there are few number of applications to be transmitted over the bonded sub-connections. Per-flow load balance techniques might not be able to achieve a fine grained load distribution, so that the per-packet load sharing is necessary.

For the per-flow load balancing, the load can be distributed using hashing methods. For the per-packet load splitting, the coloring mechanism specified in [RFC2698] can be used to classify customer's IP packets, both upstream and downstream, and packets will then be

forwarded over the appropriate sub-connections. For example, packets colored as green are forwarded over one sub-connection and packets colored as yellow are forwarded over another sub-connection. For scenarios that rely upon more than two sub-connections, multiple levels of coloring mechanism could be implemented.

#### 4.4. Traffic Recombination

For the packet-based traffic distribution, the recombination function at the receiver sides must reorder packets to preserve the integrity of the communication. The sender needs to mark each packet with a sequence number. The sequence number are set per the whole bonding connection rather than per sub-connection so that all packets forwarded over several sub-connections actually share the same reordering buffer.

##### 4.4.1. Reordering Buffer

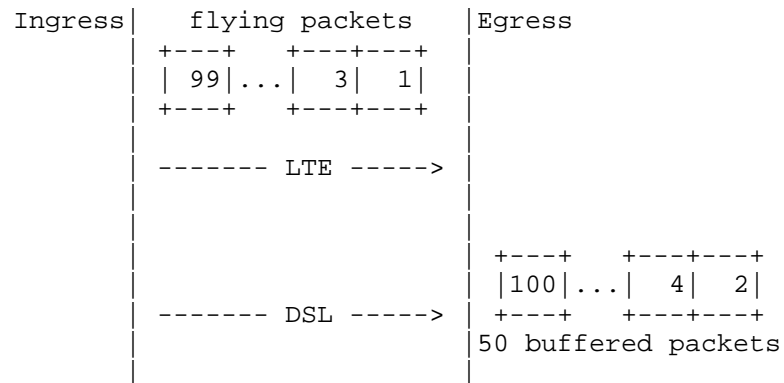


Figure 4.1: Minimizing the reordering buffer

Deployment experiences show that a secondary sub-connection might suffer from large latency, jitter and high packet loss rate. For packet-based traffic distribution, packets are distributed onto those sub-connections at the ingress and then recombined again in a buffer at the egress. If the secondary sub-connection suffers, the entire bonding connection will suffer as well due to the recombination function. For example, assume packet 1,3,...,99 are distributed onto the secondary sub-connection while packet 2,4,...,100 are distributed onto the primary sub-connection. If packet 1 is delayed by 100 ms, even packet 2 arrives at the recombination buffer at the first millisecond, it has to remain in the buffer awaiting for packet 1 as long as 99 ms. Packets distributed onto the primary sub-connection,

which arrive after packet 2, have to be buffered. This can easily lead to the overflow of the reordering buffer and the user's TCP throughput of the bonding connection might be greatly reduced.

Latency of each sub-connection will be monitored. For example, HG and HAAP may calculate the Adaptive Acknowledgment Time-Out of each sub-connection as specified in [RFC2637]; HG and HAAP may periodically exchange control messages to detect the RTT of each sub-connection [FLARE]; Packet loss rate of each sub-connection may be monitored as well [BondL3]. If the difference of the monitored latency exceeds a predefined threshold or the secondary sub-connection exhibits a too high packet loss rate, attached HG and HAAP will stop distributing traffic onto this sub-connection.

Even if the latency of the two sub-connections are comparable and the packet loss rate of the secondary sub-connection is fine so that the reordering buffer does not overflow, it's still worthy to design solutions to minimize the usage of the reordering buffer. In order to realize this goal, the traffic distribution at the ingress should be manipulated. For example, the idea of [FLARE] might be borrowed: basically, a traffic flow would be split into "flowlets" by the gaps between the arriving packets. Packets of a specific flowlet is solely distributed onto one sub-connection. In this way, reordering is avoided or minimized. The load-balancing method of MPTCP [RFC6824] could be used as well: packets are always distributed to the sub-connection with the least congestion level and/or latency [MPscheduler].

#### 4.5. Bypass

Service Providers may require some traffic to bypass the Hybrid Access. For example, some delay sensitive applications such as live TV broadcasting carried over a lossy sub-connection would impair customers' Quality of Experience. Service providers could then make sure that such traffic is forwarded over a set of wired sub-connections only, thereby disregarding low-rate mobile connections, for example.

There are two types of bypass: the bypassing traffic are transmitted on a sub-connection out of all the sub-connections between HG and HAAP; the bypassing traffic is still transmitted on a sub-connection between HG and HAAP, just that the occupied bandwidth of the bypassing traffic on this sub-connection can not be used for bandwidth aggregation. In either case, the bypassing traffic would not be under control of the Hybrid Access scheme.

HG and HAAP needs to exchange information about bypassing through the control protocol, such as the application types that need to bypass

the Hybrid Access and the bandwidth occupied by the bypassing traffic (See also Section 4.6).

#### 4.6. Measurement

HG and HAAP need to measure and exchange performance parameters of the Hybrid Access, including performance parameters that pertain to each sub-connection that belongs to the same connection. Such parameters include (but are not necessarily limited to):

- Operating state: The operating state has to be measured by control messages. When a sub-connection fails, this sub-connection has to be removed from the bonding connection.
- Round Trip Time (RTT): The measurement of this parameter is used by flow and congestion control algorithms for per-flow and per-packet distribution purposes. The probing packet could be piggy-backed by data packets or could be carried by control messages.
- Maximum sub-connection bandwidth: This parameter may be used to determine the amount of traffic that can be distributed across all or a subset of sub-connections.
- Bypassing bandwidth: This parameter has to be periodically monitored. Usually, this is measured for the opposite end to figure out the available sending bandwidth. For example, the HG reports the downloading bypassing bandwidth used in a sub-connection so that the HAAP can calculate the remaining downloading bandwidth of this sub-connection.

#### 4.7. Policy Control

Operators and customers may control the Hybrid Access with policies. These policies will be instantiated into parameters or actions that impact traffic classification, distribution, combination, measurement and bypassing. Such policies may consist in:

- Defining traffic filter lists used by the traffic classification function.
- Per-flow or per-packet forwarding policies.
- Operators may specify maximum value of the difference between two measured one-way transit delays. Operators may also specify the maximum allowed packet loss rate of a sub-connection.
- Operators may determine the maximum allowed size (See MAX\_PERFLOW\_BUFFER in [RFC2890]) of the buffer for reordering.

Operators may also define the maximum time (See OUTFORDER\_TIMER in [RFC2890]) that a packet can stay in the buffer for reordering. These parameters may pact traffic recombination.

- Operators and customers may specify the service types to bypass the Hybrid Access.
- Operators may specify the frequency for detecting a sub-connection and the detection retry times before a sub-connection can be declared as "failed".

## 5. Requirements

Requirements for the Hybrid Access are described in this section. Also, some additional requirements are listed for discussion in the Appendix.

The solution MUST apply for both IPv4 and IPv6 traffic.

The solution MUST NOT require any new capability to support Hybrid Access from the host.

In the Hybrid Access, forwarding traffic flows over various interfaces may have a negative impact on customers' experience (e.g., hazardous log outs, broken HTTPS associations, etc.). The solution should be carefully designed, so that

- activating the solution MUST NOT impact the stability, availability of the services delivered to the customer compared to customers who access the same service whose traffic is forwarded along a single path.

"Roles" (primary or backup) should be assigned to each supported network interface type (e.g., fixed or mobile access). This role is assigned by the network operator (e.g., fixed access can be set as the "primary"). Note that there may be more than two sub-connections to support primary and backup roles. A default setting can be considered.

- Setting of the role of the sub-connections SHOULD NOT be changed by the user.

The solution should provide Service Providers with means to enforce policy control of the Hybrid Access. For example,

- the solution MUST allow to rate limit the traffic on a per-interface basis.

- the solution MUST support means to enable/disable the activation of multiple interfaces at the HG.
- the solution MUST support means to instruct the HG about the logic for mounting interfaces.
- the solution MUST support means to bind a given traffic to a subset of interfaces.

For the sake of policy enforcement or analytics at the network side,

- the solution MAY ease correlating flows, conveyed over multiple access networks, and which belong to the same customer.

Some services such as VoIP may be available over all active interfaces, but distinct logins and credentials may be used.

- The HG SHOULD be provided with clear instructions about which account to use to place outgoing sessions. For the sake of simplicity, it is RECOMMENDED to use the login/credentials that are independent of the underlying access network used to access the service.

## 6. Related IETF Work

Hybrid Access designs can rely upon several solutions. The following subsections recap the work that has been or is being conducted by the IETF in this area. Note that solutions are listed in an alphabetic order. No preference order should be assumed by the reader.

### 6.1. GRE Tunnel Bonding

GRE Tunnel Bonding [GRE-HA] uses per-packet traffic distribution to achieve a fine-grained load sharing among the sub-connections. Out-of-sequence packets may be received at the egress so that reordering function is provided. IP packets are encapsulated in the GRE header which is in turn encapsulated in an outer IP header and forwarded over the sub-connections. The Sequence Number field of the GRE header is used to number the packets at the sender side, while the receiver uses of this sequence number to reorder the packets.

A new control plane is defined. Control messages are transported in the same GRE tunnels that are used to transport data packets. The control messages and data packets are distinguished by the GRE Protocol Type 0xB7EA.

GRE tunnel bonding has been deployed by Deutsche Telekom AG and Austria Telekom.

## 6.2. LISP

LISP has the basic capability to support the Hybrid Access [LISP-HA] [ILNP]. LISP can be used to enforce traffic engineering based upon static load balancing that is not agnostic to link characteristics.

Packet-based traffic distribution has been considered in [LISP-HA] as well. The detail specification of the reordering mechanism based on a "Reorder" flag is left for future work.

## 6.3. Mobile IP

Mobile IP [RFC3775] and Network Mobility (NEMO; [RFC3963]) used to handle multiple L3 connectivity to the Internet via multiple ISPs for a multi-homed end user [RFC4908]. By treating Hybrid Access as a special scenario, some existing capabilities of Mobile IP and NEMO could be reused to realize Hybrid Access. Take [MIP-HA] as an example, rely on the multiple Care-of Addresses (CoAs) capability [RFC5648] [RFC6275], the "addressing" problem of BANANA could be settled. Currently, per-flow traffic distribution has already been supported by Mobile IP and NEMO ([RFC6088], [RFC6089]) while packet-based traffic distribution is left for future work [MIP-HA].

## 6.4. Multipath TCP Proxy

MPTCP provides the ability to establish a communication over multiple paths, by means of sub-flow establishment and operation [RFC6824]. However, the traditional MPTCP is a host-based technology therefore it's out the scope of this document. What is considered as a candidate technology to support the Hybrid Access is the MPTCP proxy mechanism. There are some implementations and deployments.

The MPTCP proxy operates at the transport layer and locates in the operator's network. A transparent MPTCP mode is proposed in [MPTCP-trans]: a MPTCP proxy terminates a user's TCP flow and reinitiates MPTCP sub-flows towards the other MPTCP proxy; The other MPTCP proxy will terminate the MPTCP sub-flows and restore the user's TCP flow; The MPTCP protocol suite provides features to manage the state of sub-flows between the two proxies. [MPTCP-plain] discusses MPTCP proxy (i.e., transparent MPTCP mode) deployment concerns and also specifies an extension to transport UDP datagrams in MPTCP packets. UDP traffic can therefore be forwarded over a MPTCP connection.

## 7. Security Considerations

Hybrid Access might introduce new threats to the network. For example, traffic sent on unsecured sub-connections would be easily intercepted by an attacker who performs man-in-the-middle attack. The

multi-path communication may be leveraged to perform Denial of Service (DoS) attack or Distributed Denial of Service (DDoS) attack (e.g., based upon flooding traffic) that may jeopardize the aggregation gateway as well as the access equipment and end station operation.

These kind of new security issues should be carefully considered in designing solutions that aim to address the problems of Bandwidth Aggregation for Internet Access.

## 8. IANA Considerations

No IANA action is required in this document.

## 9. Acknowledgements

Authors would like to thank the comments and suggestions from Christian Jacquenet and Li Xue.

## 10. References

### 10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [TR-348] Broadband Forum, "Technical Report on Hybrid Access Broadband Network Architecture", July, 2016, <<https://www.broadband-forum.org/technical/download/TR-348.pdf>>.
- [RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, DOI 10.17487/RFC2131, March 1997, <<http://www.rfc-editor.org/info/rfc2131>>.
- [RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, DOI 10.17487/RFC1035, November 1987, <<http://www.rfc-editor.org/info/rfc1035>>.
- [RFC2516] Mamakos, L., Lidl, K., Evarts, J., Carrel, D., Simone, D., and R. Wheeler, "A Method for Transmitting PPP Over Ethernet (PPPoE)", RFC 2516, DOI 10.17487/RFC2516, February 1999, <<http://www.rfc-editor.org/info/rfc2516>>.
- [RFC2689] Bormann, C., "Providing Integrated Services over Low-bitrate Links", RFC 2689, DOI 10.17487/RFC2689, September



1999, <<http://www.rfc-editor.org/info/rfc2689>>.

[RFC2890] Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, DOI 10.17487/RFC2890, September 2000, <<http://www.rfc-editor.org/info/rfc2890>>.

## 10.2. Informative References

[RFC2637] Hamzeh, K., Pall, G., Verthein, W., Taarud, J., Little, W., and G. Zorn, "Point-to-Point Tunneling Protocol (PPTP)", RFC 2637, DOI 10.17487/RFC2637, July 1999, <<http://www.rfc-editor.org/info/rfc2637>>.

[BondL3] Maciej Bednarek, Guillermo Barrenetxea, Mirja Mirja Kuehlewind and Brian Trammell, "Multipath bonding at Layer 3", Applied Networking Research Workshop, July 16, 2016, Berlin, Germany

[FLARE] Srikanth Kandula, Dina Katabi, Shantanu Sinha, and Arthur Berger, "Dynamic Load Balancing Without Packet Reordering", ACM SIGCOMM Computer Communication Review, April 2007.

[MPScheduler] Hyunwoo Nam, Doru Calin and Henning Schulzrinne, "Towards Dynamic MPTCP Path Control Using SDN", IEEE NetSoft Conference and Workshops (NetSoft), June 2016.

[GRE-HA] N. Leymann, C. Heidemann, M. Zhang, et al, "GRE Tunnel Bonding", draft-zhang-gre-tunnel-bonding, work in progress.

[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>>.

[MPTCP-tans] B. Peirens, G. Detal, S. Barre and O. Bonaventure, "Link bonding with transparent Multipath TCP", draft-peirens-mptcp-transparent, work in progress.

[MPTCP-plain] M. Boucadair and C. Jacquenet, "An MPTCP Option for Network-Assisted MPTCP Deployments: Plain Transport Mode", draft-boucadair-mptcp-plain-mode, work in progress.

[MIP-HA] P. Seite, A. Yegin and S. Gundavelli, "Multihoming support for Residential Gateways", draft-seite-dmm-rg-multihoming, work in progress.

[RFC5213] Gundavelli, S., Ed., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, DOI

10.17487/RFC5213, August 2008, <<http://www.rfc-editor.org/info/rfc5213>>.

[RFC6088] Tsirtsis, G., Giarreta, G., Soliman, H., and N. Montavont, "Traffic Selectors for Flow Bindings", RFC 6088, DOI 10.17487/RFC6088, January 2011, <<http://www.rfc-editor.org/info/rfc6088>>.

[RFC6089] Tsirtsis, G., Soliman, H., Montavont, N., Giaretta, G., and K. Kuladinithi, "Flow Bindings in Mobile IPv6 and Network Mobility (NEMO) Basic Support", RFC 6089, DOI 10.17487/RFC6089, January 2011, <<http://www.rfc-editor.org/info/rfc6089>>.

[802.1AX] IEEE, "IEEE Standard for Local and Metropolitan Area Networks - Link Aggregation", 802.1AX-2014, 24 December 2014.

[LISP-HA] M. Menth, A. Stockmayer and M. Schmidt, "LISP Hybrid Access", draft-menth-lisp-ha, work in progress.

[ILNP] "ILNP - Identifier-Locator Network Protocol", online available: <http://ilnp.cs.st-andrews.ac.uk/>

#### Appendix A. Additional Requirements

The following requirements are listed as record and may subject to change.

- The solution MUST be valid for any kinds of interfaces that need to be aggregated. No dependency to the underlying media should be assumed.
- The solution MUST comply with servers policy regarding IP addresses that are bound to (HTTP session) cookies.
- The solution MUST NOT break TLS associations.
- Activating the solution MUST NOT have negative impacts on the service usability (including the HG management).
- Service degradation MUST NOT be observed when enabling the solution.
- Enabling the solution MUST increase the serviceability of the HG. In particular, the solution MUST allow for the HG to always establish a network attachment when the primary connectivity is out of service.

- The solution SHOULD NOT alter any mechanism, to aggregate available resources or to ensure a service continuity among multiple access points, that is supported by end-devices connected to the HG.
- The HG MUST bind the DNS server(s) discovered during the network attachment phase to the interface from which the information was received.
- The HG MUST bind the service information (e.g., SIP Proxy Server) discovered during the network attachment phase to the interface from which the information was received.
- When sending the traffic via a given interface, the HG MUST use as source address an address (or an address from a prefix) that was assigned for that interface.
- For protocols such as RTP/RTCP, the same IP address MUST be used for both RTP and RTCP sessions.
- Dedicated tools SHOULD be provided to the customer to assess the aggregated capacity (instead of link-specific). This can be included as part of the HG UI, a dedicated portal, etc.

## Author's Addresses

Margaret Cullen  
Painless Security  
14 Summer St. Suite 202  
Malden, MA 02148 USA

EMail: [margaret@painless-security.com](mailto:margaret@painless-security.com)

Nicolai Leymann  
Deutsche Telekom AG  
Winterfeldtstrasse 21-27  
Berlin 10781  
Germany

Phone: +49-170-2275345  
EMail: [n.leymann@telekom.de](mailto:n.leymann@telekom.de)

Cornelius Heidemann  
Deutsche Telekom AG  
Heinrich-Hertz-Strasse 3-7  
Darmstadt 64295  
Germany

Phone: +4961515812721  
EMail: [heidemannc@telekom.de](mailto:heidemannc@telekom.de)

Mohamed Boucadair  
France Telecom  
Rennes 35000  
France

EMail: [mohamed.boucadair@orange.com](mailto:mohamed.boucadair@orange.com)

Hui Deng  
China Mobile  
53A,Xibianmennei Ave.,  
Xuanwu District,  
Beijing 100053  
China

EMail: [denghui@chinamobile.com](mailto:denghui@chinamobile.com)

Mingui Zhang  
Huawei Technologies  
No.156 Beiqing Rd. Haidian District,  
Beijing 100095 P.R. China

EMail: zhangmingui@huawei.com

Behcet Sarikaya  
Huawei USA  
5340 Legacy Dr. Building 3  
Plano, TX 75024

EMail: sarikaya@ieee.org

Independent Submission  
Internet Draft  
Intended Category: Informational

N. Leymann  
C. Heidemann  
Deutsche Telekom AG  
M. Zhang  
B. Sarikaya  
Huawei  
M. Cullen  
Painless Security  
December 21, 2016

Expires: June 24, 2017

Huawei's GRE Tunnel Bonding Protocol  
draft-zhang-gre-tunnel-bonding-05.txt

#### Abstract

There is an emerging demand for solutions that provide redundancy and load-sharing across wired and cellular links from a single service provider, so that a single subscriber is provided with bonded access to heterogeneous connections at the same time.

In this document, GRE (Generic Routing Encapsulation) Tunnel Bonding is specified as an enabling approach for bonded access to a wired and a wireless network in customer premises, e.g. homes. In GRE Tunnel Bonding, two GRE tunnels, one per network connection, are set up and bonded together to form a single GRE tunnel for a subscriber. Compared with each composing connection, the bonded connections promise increased access capacity and improved reliability. The solution described in this document is currently implemented by Huawei and deployed by Deutsche Telekom AG. Publication of this document is to enable other developers to build interoperable implementations.

#### Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at

<http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at  
<http://www.ietf.org/shadow.html>

#### Copyright and License Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

#### Table of Contents

1. Introduction . . . . .	3
2. Acronyms and Terminology . . . . .	4
3. Use Case . . . . .	6
4. Overview . . . . .	6
4.1. Control Plane . . . . .	7
4.2. Data Plane . . . . .	7
4.3. Traffic Classification and Distribution . . . . .	7
4.4. Traffic Recombination . . . . .	8
4.5. Bypass . . . . .	8
4.6. Measurement . . . . .	9
4.7. Policy Control Considerations . . . . .	9
5. Control Protocol Specification (Control Plane) . . . . .	9
5.1. GRE Tunnel Setup Request . . . . .	11
5.1.1. Client Identification Name . . . . .	12
5.1.2. Session ID . . . . .	12
5.1.3. DSL Synchronization Rate . . . . .	13
5.2. GRE Tunnel Setup Accept . . . . .	13
5.2.1. H IPv4 Address . . . . .	14
5.2.2. H IPv6 Address . . . . .	14
5.2.3. Session ID . . . . .	15
5.2.4. RTT Difference Threshold . . . . .	15
5.2.5. Bypass Bandwidth Check Interval . . . . .	16
5.2.6. Active Hello Interval . . . . .	16
5.2.7. Hello Retry Times . . . . .	17

5.2.8. Idle Timeout	17
5.2.9. Bonding Key Value	18
5.2.10. Configured DSL Upstream Bandwidth	19
5.2.11. Configured DSL Downstream Bandwidth	19
5.2.12. RTT Difference Threshold Violation	20
5.2.13. RTT Difference Threshold Compliance	20
5.2.14. Idle Hello Interval	21
5.2.15. No Traffic Monitored Interval	22
5.3. GRE Tunnel Setup Deny	22
5.3.1. Error Code	22
5.4. GRE Tunnel Hello	23
5.4.1. Timestamp	23
5.4.2. IPv6 Prefix Assigned by HAAP	24
5.5. GRE Tunnel Tear Down	25
5.6. GRE Tunnel Notify	25
5.6.1. Bypass Traffic Rate	25
5.6.2. Filter List Package	26
5.6.3. Switching to DSL Tunnel	29
5.6.4. Overflowing to LTE Tunnel	29
5.6.5. DSL Link Failure	30
5.6.6. LTE Link Failure	30
5.6.7. IPv6 Prefix Assigned to Host	30
5.6.8. Diagnostic Start: Bonding Tunnel	31
5.6.9. Diagnostic Start: DSL Tunnel	31
5.6.10. Diagnostic Start: LTE Tunnel	32
5.6.11. Diagnostic End	32
5.6.12. Filter List Package ACK	33
5.6.13. Switching to Active Hello State	33
5.6.14. Switching to Idle Hello State	34
5.6.15. Tunnel Verification	34
6. Tunnel Protocol Operation (Data Plane)	35
6.1. The GRE Header	36
6.2. Automatic Setup of GRE Tunnels	36
7. Security Considerations	38
8. IANA Considerations	38
9. Contributors	38
10. References	39
10.1. Normative References	39
10.2. Informative References	40
Author's Addresses	41

## 1. Introduction

Service providers used to provide subscribers with separate access to their fixed networks and mobile networks. It has become desirable to bond these heterogeneous networks together to offer access service to subscribers that offer increased access capacity and improved reliability.



This document focuses on the use case that DSL (Digital Subscriber Line) connection and LTE (Long Term Evolution) connection are bonded together. When the traffic volume exceeds the bandwidth of the DSL connection, the excess amount can be offloaded to the LTE connection. Home Gateway (HG) is a Customer Premises Equipment (CPE) initiating the DSL and LTE connections. Hybrid Access Aggregation Point (HAAP) is the network function that resides in the provider's networks to terminate these bonded connections. Note that if there were more than two connections that need to be bonded, the GRE Tunnel Bonding mechanism could support that scenario as well. However, support for more than two connections is out the scope of this document. Also, the protocol specified in this document is limited to the single-operator scenario only, i.e., the two peering boxes, HG and HAAP, are operated by a single provider. The adaptation of the GRE Tunnel Bonding protocol to the multi-provider scenario is left as future work.

This document bases the solution on GRE (Generic Routing Encapsulation [RFC2874] [RFC2890]) since GRE is widely supported in both fixed and mobile networks. Approaches specified in this document might as well be used by other tunneling technologies to achieve tunnel bonding. However, such kind of variants are out the scope of this document.

For each heterogeneous connection (DSL and LTE) between the HG and HAAP, one GRE tunnel is set up. HG and HAAP respectively serve as the common termination point of the two tunnels at both end. Those GRE tunnels are further bonded together to form a logical GRE tunnel for the subscriber. HG conceals the composing GRE tunnels from the end nodes, and end nodes simply treat the logical GRE tunnel as a single IP link. This provides an overlay: the users' IP packets (inner IP) are encapsulated in GRE which is in turn carried over IP (outer IP).

The GRE Tunnel Bonding Protocol is developed by Huawei and has been deployed in networks operated by Deutsche Telekom AG. Publication of this document is to make it open to the public and enable other developers to build interoperable implementations.

## 2. Acronyms and Terminology

GRE: Generic Routing Encapsulation [RFC2874] [RFC2890]

DSL: Digital Subscriber Line is a family of technologies that are used to transmit digital data over telephone lines

LTE: Long Term Evolution. A standard for wireless communication of high-speed data for mobile phones and data terminals. Commonly marketed as 4G LTE.

HG: Home Gateway. A CPE device that is enhanced to support the simultaneous use of both fixed broadband and 3GPP access connections.

HAAP: Hybrid Access Aggregation Point. A logical function in Operator's network, terminating bonded connections while offering high speed Internet.

CIR: Committed Information Rate [RFC2698]

RTT: Round Trip Time

AAA: Authentication, Authorization and Accounting [RFC6733]

SOAP: Simple Object Access Protocol. It is a protocol specification for exchanging structured information in the implementation of web services in computer networks.

FQDN: A Fully Qualified Domain Name (FQDN) is a domain name that includes all higher level domains relevant to the entity named. [RFC1594]

DSCP: The six-bit codepoint (DSCP) of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers [RFC2724].

BRAS: Broadband Remote Access Server. It routes traffic to and from broadband remote access devices such as Digital Subscriber Line Access Multiplexers (DSLAM) on an Internet service provider's (ISP) network.

PGW: Packet Data Network Gateway. In the Long Term Evolution (LTE) architecture for the Evolved Packet Core (EPC), the PGW acts as an anchor for user plane mobility.

PDP: Packet Data Protocol. A packet transfer protocol used in wireless GPRS (General Packet Radio Service)/HSDPA (High Speed Downlink Packet Access) networks.

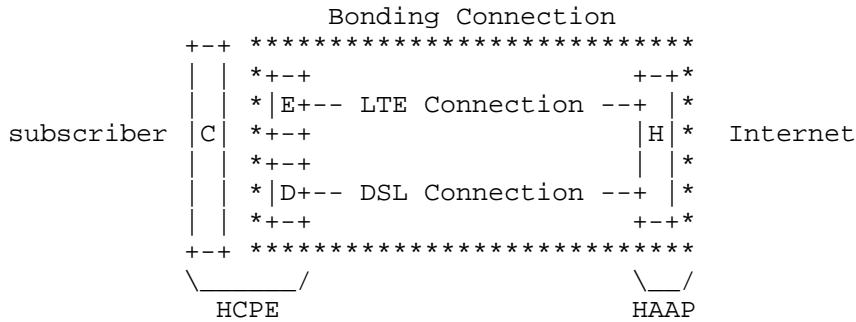
PPPoE: Point-to-Point Protocol over Ethernet is a network protocol for encapsulating PPP frames inside Ethernet frames.

DNS: Domain Name System is a hierarchical distributed naming system for computers, services, or any resource connected to the Internet or a private network.

DHCP: Dynamic Host Configuration Protocol. A standardized network protocol used on Internet Protocol (IP) networks for dynamically distributing network configuration parameters, such as IP addresses for interfaces and services.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Use Case



- C: The service endpoint of the bonding service at the HG.
- E: The endpoint of the LTE connection resides in HG.
- D: The endpoint of the DSL connection resides in HG
- H: The endpoint for each heterogeneous connection at HAAP.

Figure 3.1: Offloading from DSL to LTE, increased access capacity

If a Service Provider runs heterogeneous networks, such as fixed and mobile, subscribers eager to use those networks simultaneously for increased access capacity rather than just using a single network. As shown by the reference model in Figure 3.1, the subscriber expects a significantly higher access bandwidth from the bonding connection than from the DSL connection. In other words, when the traffic volume exceeds the bandwidth of the DSL connection, the excess amount may be offloaded to the LTE connection.

Compared to per-flow load balancing mechanisms which are widely used nowadays, the use case described in this document requires a per-packet offloading approach. For per-flow load-balancing, the maximum bandwidth that may be used by a traffic flow is the bandwidth of an individual connection. While for per-packet offloading, a single flow may use the added-up bandwidth of the two connections.

4. Overview

In this document, the widely supported GRE is chosen as the tunneling technique. With the newly defined control protocol, GRE tunnels are setup on top of the DSL and LTE connections which are ended at D and H or E and H, as shown in Figure 3.1. These tunnels are bonded

together to form a single logical bonding connection between HG and HAAP. Subscribers use this logical connection without knowing the composing GRE tunnels.

#### 4.1. Control Plane

A clean-slate control protocol is designed to manage the GRE tunnels that are setup per heterogeneous connection between HG and HAAP. The goal is to design a compact control plane for bonding access instead of reusing existing control planes.

In order to measure the performance of connections, control packets need to co-route the same path with data packets. Therefore, a GRE Channel is opened for the purpose of data plane forwarding of control plane packets. As shown in Figure 5.1, the GRE header ([RFC2784]) with the Key extension specified by [RFC2890] is being used. The GRE Protocol Type (0xB7EA) is used to identify this GRE Channel. A family of control messages are encapsulated with GRE header and carried over this channel. Attributes, formatted in Type-Length-Value style, are further defined and included in each control message.

With the newly defined control plane, the GRE tunnels between HG and HAAP can be established, managed and released without the involvement of operators.

#### 4.2. Data Plane

Using the control plane defined in Section 4.1, GRE tunnels can be automatically setup per heterogeneous connection between the HG and the HAAP. For the use case described in Section 3, one GRE tunnel is ended at the DSL WAN interfaces, e.g., DSL GRE tunnel, and another GRE tunnel is ended at the LTE WAN interfaces, e.g., LTE GRE tunnel. Each tunnel may carry user's IP packets as payload, which forms a typical IP-over-IP overlay. These tunnels are bonded together to offer a single access point to subscribers.

As shown in Figure 6.1, the GRE header ([RFC2784]) with the Key and Sequence Number extensions specified by [RFC2890] is used to encapsulate data packets. The Protocol Type is either 0x0800 [RFC2784] or 0x86DD [RFC7676], which indicates the inner packet is either an IPv4 packet or an IPv6 packet. The GRE Key field is set to a unique value for the entire bonding connection. The GRE Sequence Number field is used to maintain the sequence of packets transported in all GRE tunnels as a single flow between the HG and the HAAP.

#### 4.3. Traffic Classification and Distribution

For the offloading use case, the coloring mechanism specified in

[RFC2697] is being used to classify subscriber's IP packets, both upstream and downstream, into the DSL GRE tunnel or the LTE GRE tunnel. Packets colored as green or yellow will be distributed into the DSL GRE tunnel and packets colored as red will be distributed into the LTE GRE tunnel. For the scenario that requires more than two GRE tunnels, multiple levels of token buckets might be realized. However, that is out of the scope for this document.

The Committed Information Rate (CIR) of the coloring mechanism is set to the total DSL WAN bandwidth minus the bypass DSL bandwidth (See Section 4.4.). The total DSL WAN bandwidth MAY be configured, MAY be obtained from the management system (AAA server, SOAP server, etc.), or MAY be detected in real-time using ANCP [RFC6320].

#### 4.4. Traffic Recombination

For the offloading use case, the recombination function at the receiver provides in-order delivery of subscribers' traffic. The receiver maintains a small reordering buffer and orders the data packets in this buffer by the Sequence Number field [RFC2890] of the GRE header. All packets carried on GRE tunnels which belong to the same bonding connection go into a single reordering buffer.

Operators may configure the maximum allowed size (See `MAX_PERFLOW_BUFFER` in [RFC2890]) of the buffer for reordering. They may also configure the maximum time (See `OUTOFORDER_TIMER` in [RFC2890]) that a packet can stay in the buffer for reordering. The `OUTOFORDER_TIMER` must be configured carefully. Values larger than the difference of the normal Round-Trip Time (e.g., 100 ms) of the two connections are not recommended. Implementation and deployment experiences exhibits there is usually a large margin for the value of `MAX_PERFLOW_BUFFER`. Values larger than the multiply of the sum of the line rate of the two connections and the value of `OUTOFORDER_TIMER` should be used.

#### 4.5. Bypass

Service Providers provide some services that should not be delivered over the bonding connection. For example, Service Providers may not expect real-time IPTV to be carried by the LTE GRE tunnel. It is required that IPTV traffic bypasses the GRE Tunnel Bonding and uses the raw DSL bandwidth. Bypass traffic is not subject to the traffic classification and distribution specified above. The raw connection used for bypass traffic is not controlled by the HAAP. It may or may not go through device that HAAP resides in.

The HAAP may announce the service types that need to bypass the bonded GRE tunnels using the Filter List Package attribute as

specified in Section 5.6.2. The HG and the HAAP need to set aside the DSL bandwidth for bypassing. The available DSL bandwidth for GRE Tunnel Bonding is equal to the total DSL bandwidth minus the bypass bandwidth.

#### 4.6. Measurement

Since control packets are routed using the same paths as the data packets, the real performance of the data paths (e.g., the GRE tunnels) can be measured. The GRE Tunnel Hello messages specified in Section 5.3 are used to carry the timestamp information and the Round Trip Time (RTT) value can therefore be calculated based on the timestamp.

Besides the end-to-end delay of the GRE tunnels, the HG and the HAAP need to measure the capacity of the tunnels as well. For example, the HG is REQUIRED to measure the downstream bypassing bandwidth and report it to the HAAP in real time (See Section 5.6.1.).

#### 4.7. Policy Control Considerations

Operators and users may input policies into the GRE Tunnel Bonding. These policies will be interpreted into parameters or actions that impact the traffic classification, distribution, combination, measurement and bypass.

Operators and users may offer the service types that need to bypass the bonded GRE tunnels. Service types defined by operators will be delivered from the HAAP to the HG through the control plane (See Section 5.6.12.), and the HG will use the raw connection to transmit traffic for these service types. Users may as well define bypass service types on the HG. Bypass service types defined by users need not to be delivered to the HAAP.

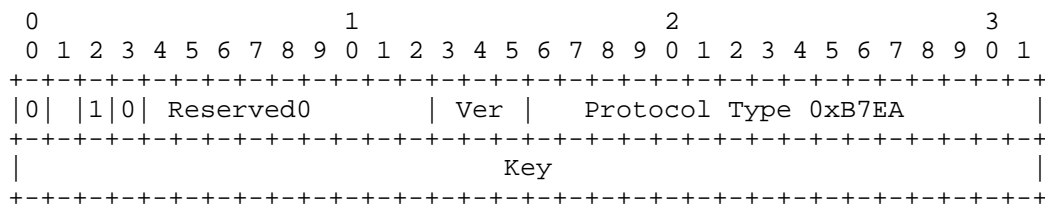
Operators may specify the interval for sending Hello messages and the retry times for the HG or the HAAP to send out Hello messages before the failure of a connection.

Since the GRE tunnels are setup on top of heterogeneous DSL and LTE connections, if the difference of the transmission delays of these connections exceeds a given threshold for a certain period, the HG and the HAAP should be able to stop the offloading behavior and fallback to a traditional transmission mode, where the LTE GRE tunnel is disused while all traffic is transmitted over the DSL GRE tunnel. Operators are allowed to define this threshold and period.

### 5. Control Protocol Specification (Control Plane)

Control messages are used to establish, maintain, measure and tear down GRE tunnels between the HG and the HAAP. Also, the control plane undertakes the responsibility convey traffic policies over the GRE tunnels.

For the purpose of measurement, control messages need to be delivered as GRE encapsulated packets and co-routed with data plane packets. The new GRE Protocol Type (0xB7EA) is allocated for this purpose and the standard GRE header as per [RFC2874] with the Key extension specified by [RFC2890] is used. The Checksum Present bit is set to zero. Key Present bit is set to 1. The Sequence Number Present bit is set to 0. So the format of the GRE header for control messages of the GRE Tunnel Bonding protocol is as follows:



#### Key

The Key field is used to carry a random number for the purpose of security. The random number is generated by the HAAP and informed to the HG. (See Section 5.2.9.)

The general format of the entire control message is as follows:

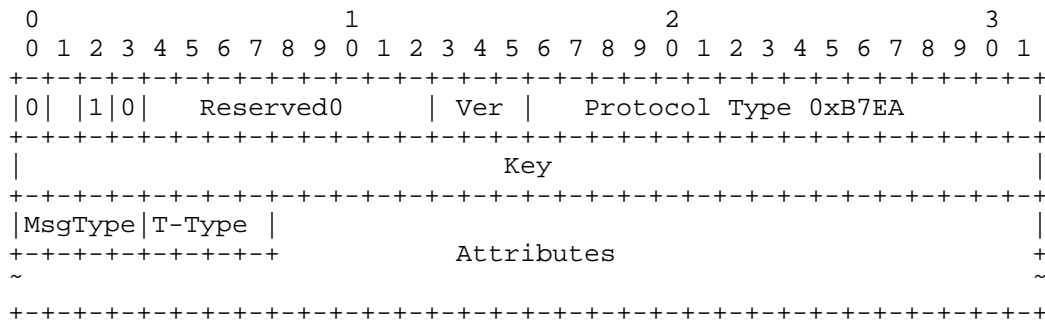


Figure 5.1: The format of control messages of GRE Tunnel Bonding

#### MsgType (4 bits)

Message Type. The control message family contains the following 6 types of control messages:

Control Message Family =====	Type =====
GRE Tunnel Setup Request	1
GRE Tunnel Setup Accept	2
GRE Tunnel Setup Deny	3
GRE Tunnel Hello	4
GRE Tunnel Tear Down	5
GRE Tunnel Notify	6
Reserved	0,7-15

#### T-Type (4 bits)

Tunnel Type. Set to 0001 if the control message is sent via the primary GRE tunnel (normally the DSL GRE tunnel). Set to 0010 if the control message is sent via the secondary GRE tunnel (normally the LTE GRE tunnel). Values 0000 and values from 0011 through 1111 are reserved for future use and MUST be ignored on receipt.

#### Attributes

The Attributes field includes the attributes that need to be carried in the control message. Each Attribute has the following format.

```

+-----+
|Attribute Type |                (1 byte)
+-----+-----+
| Attribute Length |          (2 bytes)
+-----+-----+
| Attribute Value   ~          (variable)
+-----+-----+

```

#### Attribute Type (1 octet)

The Attribute Type specifies the type of the attribute.

#### Attribute Length (2 octets)

Attribute Length indicates the length of the Attribute Value in octets.

#### Attribute Value (variable)

The Attribute Value includes the value of the attribute.

All control messages are sent in network byte order (high order octets first). Protocol Type carried in the GRE header for the control message is 0xB7EA. Based on this number, the receiver will determine to consume the GRE packet locally rather than further forwarding.

### 5.1. GRE Tunnel Setup Request



HG uses the GRE Tunnel Setup Request message to request that the HAAP establish the GRE tunnels. It is sent out from HG's LTE and DSL WAN interfaces separately. Attributes that need to be included in this message are defined in the following subsections.

#### 5.1.1.1. Client Identification Name

Operator uses the Client Identification Name (CIN) to identify the HG. The HG sends the CIN to the HAAP for authentication and authorization as specified in [TS23.401]. It is REQUIRED that the GRE Tunnel Setup Request message sent out from the LTE WAN interface contains the CIN attribute while the GRE Tunnel Setup Request message sent out from the DSL WAN interface does not contain this attribute.

The CIN attribute has the following format:

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+
| Client Identification Name (40 bytes) |
+-----+

```

Attribute Type  
CIN, set to 3.

Attribute Length  
Set to 40.

Client Identification Name  
This is a 40-byte string value encoded in UTF-8 and set by the operator. It is used as the identification of the HG in the operator's network.

#### 5.1.1.2. Session ID

This Session ID is generated by the HAAP when the LTE GRE Tunnel Setup Request message is received. The HAAP announces the Session ID to the HG in the LTE GRE Tunnel Setup Accept message. For those WAN interfaces that need to be bonded together, the HG MUST use the same Session ID. The HG MUST carry the Session ID attribute in each DSL GRE Tunnel Setup Request message. For the first time that the LTE GRE Tunnel Setup Request message is sent to the HAAP, the Session ID attribute need not be included. However, if the LTE GRE Tunnel fails and HG tries to revive it, the LTE GRE Tunnel Setup Request message MUST include the Session ID attribute.

The Session ID attribute has the following format:

```

+-----+
|Attribute Type |                (1 byte)
+-----+-----+
| Attribute Length |            (2 bytes)
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Session ID      |            (4 bytes) |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Attribute Type  
Session ID, set to 4.

Attribute Length  
Set to 4.

Session ID  
An unsigned integer generated by the HAAP. It is used as the identification of bonded GRE Tunnels.

### 5.1.3. DSL Synchronization Rate

The HG uses the DSL Synchronization Rate to notify the HAAP about the downstream bandwidth of the DSL link. The DSL GRE Tunnel Setup Request message MUST include the DSL Synchronization Rate attribute. The LTE GRE Tunnel Setup Request message SHOULD NOT include this attribute.

```

+-----+
|Attribute Type |                (1 byte)
+-----+-----+
| Attribute Length |            (2 bytes)
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| DSL Synchronization Rate |    (4 bytes) |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Attribute Type  
DSL Synchronization Rate, set to 7.

Attribute Length  
Set to 4.

DSL Synchronization Rate  
An unsigned integer measured in kbps.

### 5.2. GRE Tunnel Setup Accept

The HAAP uses the GRE Tunnel Setup Accept message as the response to

the GRE Tunnel Setup Request message. This message indicates acceptance of the tunnel establishment and carries parameters of the GRE tunnels. Attributes that need be to included in this message are defined below.

#### 5.2.1. H IPv4 Address

The HAAP uses the H IPv4 Address attribute to inform the HG of the H IPv4 address. The HG uses the H IPv4 address as the destination endpoint IPv4 address of the GRE tunnels (the source endpoint IPv4 addresses of the GRE tunnels are respectively DSL/LTE WAN interface IP address (D/E)). The LTE GRE Tunnel Setup Accept message MUST include the H IPv4 Address attribute.

```
+-----+
|Attribute Type |                (1 byte)
+-----+-----+
| Attribute Length |            (2 bytes)
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| H IPv4 Address |                (4 bytes) |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

Attribute Type  
H IPv4 Address, set to 1.

Attribute Length  
Set to 4.

H IPv4 Address  
Set to the pre-configured IPv4 address (e.g. an IP address of a Line Card in the HAAP) which is used as the endpoint IP address of GRE tunnels by the HAAP.

#### 5.2.2. H IPv6 Address

HAAP uses the H IPv6 Address attribute to inform the HG of the H IPv6 address. The HG uses the H IPv6 address as the destination endpoint IPv6 address of the GRE tunnels (the source endpoint IPv4 addresses of the GRE tunnels are respectively DSL/LTE WAN interface IP address (D/E)). The LTE GRE Tunnel Setup Accept message MUST include the H IPv6 Address attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+...+
| H IPv6 Address (16 bytes) |
+-----+...+

```

Attribute Type  
H IPv6 Address, set to 1.

Attribute Length  
Set to 16.

H IPv6 Address  
Set to the pre-configured IPv6 address (e.g. an IP address of a Line Card in the HAAP) which is used as the endpoint IP address of GRE tunnels by HAAP.

### 5.2.3. Session ID

The LTE GRE Tunnel Setup Accept message MUST include Session ID attribute as defined in Section 5.1.2.

### 5.2.4. RTT Difference Threshold

The HAAP uses the RTT Difference Threshold attribute to inform the HG of the acceptable threshold of RTT difference between the DSL link and the LTE link. If the measured RTT difference exceeds this threshold, the HG SHOULD stop offloading traffic to the LTE GRE tunnel. The LTE GRE Tunnel Setup Accept message MUST include the RTT Difference Threshold attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+...+
| RTT Difference Threshold (4 bytes) |
+-----+...+

```

Attribute Type  
RTT Difference Threshold, set to 9.

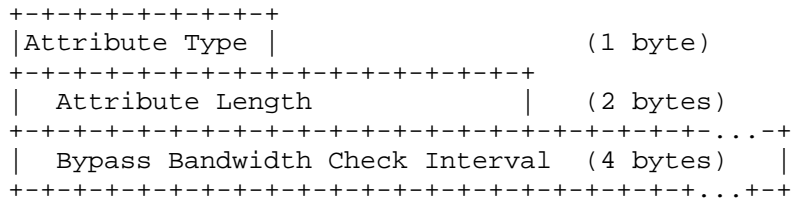
Attribute Length  
Set to 4.

RTT Difference Threshold

An unsigned integer measured in milliseconds. This value can be chosen in the range 0 through 1000.

5.2.5. Bypass Bandwidth Check Interval

The HAAP uses the Bypass Bandwidth Check Interval attribute to inform the HG of how frequently the bypass bandwidth should be checked. The HG should check the bypass bandwidth of the DSL WAN interface in each time period indicated by this interval. The LTE GRE Tunnel Setup Accept message MUST include the Bypass Bandwidth Check Interval attribute.



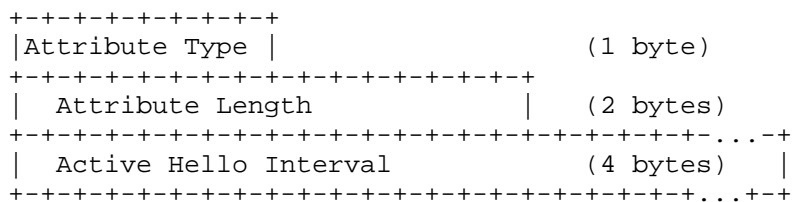
Attribute Type  
Bypass Bandwidth Check Interval, set to 10.

Attribute Length  
Set to 4.

Bypass Bandwidth Check Interval  
An unsigned integer measured in seconds. This value can be chosen in the range 10 through 300.

5.2.6. Active Hello Interval

The HAAP uses the Active Hello Interval attribute to inform the HG of the pre-configured interval for sending out GRE Tunnel Hellos. The HG should send out GRE Tunnel Hellos via both the DSL and LTE WAN interfaces in each time period as indicated by this interval. The LTE GRE Tunnel Setup Accept message MUST include the Active Hello Interval attribute.



Attribute Type

Active Hello Interval, set to 14.

Attribute Length

Set to 4.

Active Hello Interval

An unsigned integer measured in seconds. This value can be chosen in the range 1 through 100.

#### 5.2.7. Hello Retry Times

The HAAP uses the Hello Retry Times attribute to inform the HG of the retry times for sending GRE Tunnel Hellos. If the HG does not receive any acknowledgement from the HAAP for the number of GRE Tunnel Hello attempts specified in this attribute, the HG will declare a failure of the GRE Tunnel. The LTE GRE Tunnel Setup Accept message MUST include the Hello Retry Times attribute.

```

+++++
|Attribute Type | (1 byte)
+++++
| Attribute Length | (2 bytes)
+++++
| Hello Retry Times (4 bytes) |
+++++

```

Attribute Type

Hello Retry Times, set to 15.

Attribute Length

Set to 4.

Hello Retry Times

An unsigned integer, which takes values in the range 3 through 10.

#### 5.2.8. Idle Timeout

The HAAP uses the Idle Timeout attribute to inform the HG of the pre-configured timeout value to terminate the DSL GRE tunnel. When an LTE GRE Tunnel failure is detected, all traffic will be sent over the DSL GRE tunnel. If the failure of the LTE GRE tunnel lasts longer than the Idle Timeout, subsequent traffic will be sent over raw DSL rather than over a tunnel, and the DSL GRE tunnel SHOULD be terminated. The LTE Tunnel Setup Accept message MUST include the Idle Timeout attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+...+
| Idle Timeout (4 bytes) |
+-----+...+

```

Attribute Type  
Idle Timeout, set to 16.

Attribute Length  
Set to 4.

Idle Timeout  
An unsigned integer measured in seconds. It takes values in the range 0 through 172,800 with the granularity of 60. The default value is 1,440 (24 hours). The value 0 indicates the idle timer never expires.

#### 5.2.9. Bonding Key Value

The HAAP uses the Bonding Key Value attribute to inform the HG of the number which is to be carried as the Key of the GRE header for subsequent control messages. The Bonding Key Value is generated by the HAAP and used for the purpose of security.

The method used to generate this number is up to implementations. The Pseudo Random Number Generator defined in ANSI X9.31 Appendix A.2.4 is RECOMMENDED. Note that Random Number Generation collision is allowed in the GRE Tunnel Bonding protocol.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+...+
| Bonding Key Value (4 bytes) |
+-----+...+

```

Attribute Type  
Bonding Key Value, set to 20.

Attribute Length  
Set to 4.

Bonding Key Value  
A 32-bit random number generated by the HAAP.

5.2.10. Configured DSL Upstream Bandwidth

The HAAP obtains the upstream bandwidth of the DSL link from the management system and uses the Configured DSL Upstream Bandwidth attribute to inform the HG. The HG uses the received upstream bandwidth as the Committed Information Rate for the DSL link [RFC2697]. The DSL GRE Tunnel Setup Accept message MUST include the Configured DSL Upstream Bandwidth attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+
| Configured DSL Upstream Bandwidth (4 bytes) |
+-----+

```

Attribute Type  
Configured DSL Upstream Bandwidth, set to 22.

Attribute Length  
Set to 4.

Configured DSL Upstream Bandwidth  
An unsigned integer measured in kbps.

5.2.11. Configured DSL Downstream Bandwidth

The HAAP obtains the downstream bandwidth of the DSL link from the management system and uses the Configured DSL Downstream Bandwidth attribute to inform the HG. The HG uses the received downstream bandwidth as the base in calculating the bypassing bandwidth. The DSL GRE Tunnel Setup Accept message MUST include the Configured DSL Downstream Bandwidth attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+
|Configured DSL Downstream Bandwidth(4 bytes) |
+-----+

```

Attribute Type  
Configured DSL Downstream Bandwidth, set to 23.

Attribute Length  
Set to 4.



Configured DSL Downstream Bandwidth  
An unsigned integer measured in kbps.

#### 5.2.12. RTT Difference Threshold Violation

The HAAP uses the RTT Difference Threshold Violation attribute to inform the HG of the number of times in a row that the RTT Difference Threshold (See Section 5.2.4.) may be violated before the HG MUST stop using the LTE GRE Tunnel. If the RTT Difference Threshold is continuously violated for more than the indicated number of measurements, the HG MUST stop using the LTE GRE tunnel. The LTE GRE Tunnel Setup Accept message MUST include the RTT Difference Threshold Violation attribute.

```
+-----+
|Attribute Type |                (1 byte)
+-----+-----+
| Attribute Length |          (2 bytes)
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| RTT Diff Threshold Violation (4 bytes) |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

Attribute Type

RTT Difference Threshold Violation, set to 24.

Attribute Length

Set to 4.

RTT Difference Threshold Violation

An unsigned integer which takes values in the range 1 through 25.  
A typical value is 3.

#### 5.2.13. RTT Difference Threshold Compliance

The HAAP uses the RTT Difference Threshold Compliance attribute to inform the HG of the number of times in a row the RTT Difference Threshold (See Section 5.2.4.) must be compliant before use of the LTE GRE tunnel can be resumed. If the RTT Difference Threshold is continuously detected to be compliant across more than this number of measurements, the HG MAY resume using the LTE GRE tunnel. The LTE GRE Tunnel Setup Accept message MUST include the RTT Difference Threshold Compliance attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+
| RTT Diff Threshold Compliance (4 bytes) |
+-----+

```

## Attribute Type

RTT Diff Threshold Compliance, set to 25.

## Attribute Length

Set to 4.

## RTT Diff Threshold Compliance

An unsigned integer which takes values in the range 1 through 25.

A typical value is 3.

## 5.2.14. Idle Hello Interval

The HAAP uses the Idle Hello Interval attribute to inform the HG of the pre-configured interval for sending out GRE Tunnel Hellos when the subscriber is detected to be idle. The HG SHOULD begin to send out GRE Tunnel Hellos via both the DSL and LTE WAN interfaces in each time period as indicated by this interval, if the bonded tunnels have seen no traffic longer than the "No Traffic Monitored Interval" (See Section 5.2.15.). The LTE GRE Tunnel Setup Accept message MUST include the Idle Hello Interval attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+
| Idle Hello Interval (4 bytes) |
+-----+

```

## Attribute Type

Idle Hello Interval, set to 31.

## Attribute Length

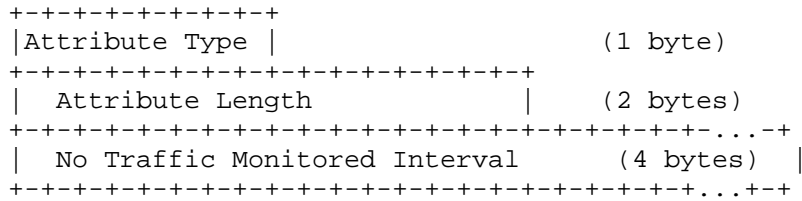
Set to 4.

## Idle Hello Interval

An unsigned integer measured in seconds. This value can be chosen from the range 100 through 86,400 (24 hours) with the granularity of 100. The default value is 1800 (30 minutes).

5.2.15. No Traffic Monitored Interval

The HAAP uses the No Traffic Monitored Interval attribute to inform the HG of the pre-configured interval for switching the GRE Tunnel Hello mode. If traffic is detected on the bonded GRE tunnels before this informed interval expires, the HG SHOULD switch to the Active Hello Interval. The LTE GRE Tunnel Setup Accept message MUST include the No Traffic Monitored Interval attribute.



Attribute Type  
No Traffic Monitored Interval, set to 32.

Attribute Length  
Set to 4.

No Traffic Monitored Interval  
An unsigned integer measured in seconds. This value is in the range 30 through 86,400 (24 hours). The default value is 60.

5.3. GRE Tunnel Setup Deny

HAAP MUST send the GRE Tunnel Setup Deny message to HG if the GRE tunnel setup request from this HG is denied. The HG MUST terminate the GRE tunnel setup process as soon as it receives the GRE Tunnel Setup Deny message.

5.3.1. Error Code

The HAAP uses the Error Code attribute to inform the HG of the error code. The error code depicts the reason why the GRE tunnel setup request is denied. Both the LTE GRE Tunnel Setup Deny message and the DSL GRE Tunnel Setup Deny message MUST include the Error Code attribute.

```

+-----+
|Attribute Type |                               (1 byte)
+-----+
| Attribute Length |                         (2 bytes)
+-----+-----+
| Error Code |                               (4 bytes) |
+-----+-----+

```

**Attribute Type**

Error Code, set to 17.

**Attribute Length**

Set to 4.

**Error Code**

An unsigned integer. The list of the codes are listed as follows.

- 1: The HAAP was not reachable over LTE during the GRE tunnel setup request.
- 2: The HAAP was not reachable via DSL during the GRE tunnel setup request.
- 3: The LTE GRE tunnel to the HAAP failed.
- 4: The DSL GRE tunnel to the HAAP failed.
- 5: The given DSL User ID is not allowed to use the GRE Tunnel Bonding service.
- 6: The given User Alias (TOID)/User ID (GUID) is not allowed to use the GRE Tunnel Bonding service.
- 7: The LTE and DSL User IDs do not match.
- 8: The HAAP denied the GRE tunnel setup request because a bonding session with the same User ID already exists.
- 9: The HAAP denied the GRE tunnel setup request because the user's CIN is not permitted.
- 10: The HAAP terminated a GRE Tunnel Bonding session for maintenance reasons.
- 11: There was a communication error between the HAAP and the management system during the LTE tunnel setup request.
- 12: There was a communication error between the HAAP and management system during the DSL tunnel setup request.

**5.4. GRE Tunnel Hello**

After the DSL/LTE GRE tunnel is established, the HG begins to periodically send out GRE Tunnel Hello messages via the tunnel, which the HAAP acknowledges by returning GRE Tunnel Hello messages back to the HG. This continues until the tunnel is terminated.

**5.4.1. Timestamp**

The HAAP uses the Timestamp attribute to inform the HG of the timestamp value that is used for RTT calculation. Both the LTE GRE Tunnel Hello message and DSL GRE Tunnel Hello message MUST include the Timestamp attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+...+
| Timestamp (8 bytes) |
+-----+...+

```

Attribute Type  
Timestamp, set to 5.

Attribute Length  
Set to 8.

Timestamp  
The time since the system restarts. The high-order 4 octets indicate an unsigned integer in units of one second; the low-order 4 octets indicate an unsigned integer in unit of one millisecond.

#### 5.4.2. IPv6 Prefix Assigned by HAAP

The HAAP uses the IPv6 Prefix Assigned by the HAAP attribute to inform the HG of the assigned IPv6 prefix. This IPv6 prefix is to be captured by the Lawful Interception. Both the LTE GRE Tunnel Hello message and the DSL GRE Tunnel Hello message MUST include the IPv6 Prefix Assigned by HAAP attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+...+
| IPv6 Prefix Assigned by HAAP (16 bytes) |
+-----+...+

```

Attribute Type  
IPv6 Prefix Assigned by HAAP, set to 13.

Attribute Length  
Set to 17.

IPv6 Prefix Assigned by HAAP  
The highest-order 16 octets encode an IPv6 address. The lowest-

order one octet encodes the prefix length. These two values are put together to represent an IPv6 prefix.

#### 5.5. GRE Tunnel Tear Down

The HAAP can terminate a DSL/LTE GRE tunnel by sending the GRE Tunnel Tear Down message to the HG via the tunnel. The Error Code attribute as defined in Section 5.3.1 MUST be included in this message. After receiving the GRE Tunnel Tear Down message, the HG removes the IP address of H which is the destination IP addresses of the DSL and LTE GRE tunnels.

#### 5.6. GRE Tunnel Notify

The HG and the HAAP use the GRE Tunnel Notify message which is transmitted either through the DSL GRE tunnel or LTE GRE tunnel to notify each other about their status regarding the DSL/LTE GRE tunnels, the information for the bonded tunnels, the actions that need to be taken, etc.

Usually, the receiver just sends the received attributes back as the acknowledgement for each GRE Tunnel Notify message. There is an exception for the Filter List Package. Since the size of the Filter List Package attribute can be very large, a special attribute is specified in Section 5.6.12 as the acknowledgement.

Attributes that need be to included in the GRE Tunnel Notify message are defined below.

##### 5.6.1. Bypass Traffic Rate

There are a few types of traffic that need to transmitted over the raw DSL WAN interface rather than the bonded GRE tunnels. The HG has to set aside bypass bandwidth on the DSL WAN interface for these traffic types. Therefore, the available bandwidth of the DSL GRE tunnel is the entire DSL WAN interface bandwidth minus the occupied bypass bandwidth.

The HG uses the Bypass Traffic Rate attribute to inform the HAAP of the downstream bypass bandwidth for the DSL WAN interface. The Bypass Traffic Rate attribute will be included in the DSL GRE Tunnel Notify message. The HAAP calculates the available downstream bandwidth for the DSL GRE tunnel as the Configured DSL Downstream Bandwidth minus this informed bypass bandwidth. The available DSL bandwidth will be used as the Committed Information Rate (CIR) of the coloring system [RFC2697].

```

+-----+
|Attribute Type |                (1 byte)
+-----+
| Attribute Length |            (2 bytes)
+-----+-----+
| Bypass Traffic Rate |        (4 bytes) |
+-----+-----+

```

## Attribute Type

Bypass Traffic Rate, set to 6.

## Attribute Length

Set to 4.

## Bypass Traffic Rate

An unsigned integer measured in kbps.

## 5.6.2. Filter List Package

The HAAP uses the Filter List Package attribute to inform the HG of the service types that need to bypass the bonded GRE tunnels. The full list of all filter items may be given by a series of Filter List Package attributes with each specifying a partial list. At the HG, a full list of filter items is maintained. Also, the HG needs to maintain an exception list of filter items. For example, the packets carrying the control messages defined in this document should be excluded from the filter list.

Incoming packets that match a filter item in the filter list while not matching any item in the exception list MUST be transmitted over the raw DSL rather than the bonded GRE tunnels. Both the LTE GRE Tunnel Notify message and GRE Tunnel Notify message MAY include the Filter List Package attribute. The DSL GRE Tunnel Notify message is preferred.

```

+-----+
|Attribute Type |                (1 byte)
+-----+
| Attribute Length |            (2 bytes)
+-----+-----+
| Filter List TLV |            (variable) ~
+-----+-----+

```

## Attribute Type

Filter List Package, set to 8.

## Attribute Length

The total length of the Filter List TLV. The maximum allowed





**Type**

The Type of the Filter Item. Currently, the following types are supported.

Filter Items	Type
=====	=====
FQDN [RFC1594]	1
DSCP [RFC2724]	2
Destination Port	3
Destination IP	4
Destination IP&Port	5
Source Port	6
Source IP	7
Source IP&Port	8
Source Mac	9
Protocol	10
Source IP Range	11
Destination IP Range	12
Source IP Range&Port	13
Destination IP Range&Port	14

Other values are reserved for future use and MUST be ignored on receipt.

**Length**

The length of the Filter Item in octets. Type and Length are excluded.

**Enable**

Whether the filter item is enabled. One means enabled and zero means disabled. Other possible values are reserved and MUST be ignored on receipt.

**Description Length**

The length of the Description Value in octets.

**Description Value**

A variable string value encoded in UTF-8 that describes the Filter List TLV (e.g., "FQDN").

**Value**

A variable string encoded in UTF-8 that specifies the value of the Filter Item (e.g. "www.yahoo.com"). As an example, Type = 1 and Value = "www.yahoo.com" means that packets whose FQDN field equals "www.yahoo.com" match the filter item. Values for "Mac" are specified using hexadecimal numbers. Port number are decimals as assigned by IANA in [Port-NO]. For the "Protocol" type, the value could either be a decimal or a keyword

specified by IANA in [Pro-NO]. The formats for "IP" and "IP Range" are defined in [RFC4632] and [RFC4291] for IPv4 and IPv6 respectively. When the filter item is a combination of two parameters (Type 5, 8 and 13), values for the two parameters are separated by a colon (":").

### 5.6.3. Switching to DSL Tunnel

If the RTT difference is continuously detected to violate the RTT Difference Threshold (See Section 5.2.4.) more than the times specified in the RTT Difference Threshold Violation (See Section 5.2.12.), the HG uses the Switching to DSL Tunnel attribute to inform the HAAP to use the DSL GRE tunnel only. When the HAAP receives this attribute, it MUST begin to transmit downstream traffic to this HG solely over the DSL GRE tunnel. The DSL GRE Tunnel Notify message MAY include the Switching to DSL Tunnel attribute.

```

+-----+
|Attribute Type |                               (1 byte)
+-----+-----+-----+-----+-----+
| Attribute Length           |               (2 bytes)
+-----+-----+-----+-----+-----+

```

Attribute Type  
Switching to DSL Tunnel, set to 11.

Attribute Length  
Set to 0.

### 5.6.4. Overflowing to LTE Tunnel

If the RTT difference is continuously detected to not violated the RTT Difference Threshold attribute (See Section 5.2.4.) more than the number of times specified in the RTT Difference Compliance attribute (See Section 5.2.13), the HG uses the Overflowing to LTE Tunnel attribute to inform HAAP that LTE GRE tunnel can be used again. The DSL GRE Tunnel Notify message MAY include the Overflowing to LTE Tunnel attribute.

```

+-----+
|Attribute Type |                               (1 byte)
+-----+-----+-----+-----+-----+
| Attribute Length           |               (2 bytes)
+-----+-----+-----+-----+-----+

```

Attribute Type  
Overflowing to LTE Tunnel, set to 12.

Attribute Length  
Set to 0.

5.6.5. DSL Link Failure

When the HG detects the DSL WAN interface status is down, it MUST tear down the DSL GRE tunnel. It informs HAAP about the failure using the DSL Link Failure attribute. The HAAP MUST tear down the DSL GRE tunnel upon the DSL Link Failure attribute is received. The DSL Link Failure attribute SHOULD be carried in the LTE GRE Tunnel Notify message.

```

+++++
|Attribute Type | (1 byte)
+++++
| Attribute Length | (2 bytes)
+++++

```

Attribute Type  
DSL Link Failure, set to 18.

Attribute Length  
Set to 0.

5.6.6. LTE Link Failure

When the HG detects the LTE WAN interface status is down, it MUST tear down the LTE GRE tunnel. It informs the HAAP about the failure using the LTE Link Failure attribute. HAAP MUST tear down the LTE GRE tunnel upon the LTE Link Failure attribute is received. The LTE Link Failure attribute SHOULD be carried in the DSL GRE Tunnel Notify message.

```

+++++
|Attribute Type | (1 byte)
+++++
| Attribute Length | (2 bytes)
+++++

```

Attribute Type  
LTE Link Failure, set to 19.

Attribute Length  
Set to 0.

5.6.7. IPv6 Prefix Assigned to Host

If the HG changes the IPv6 prefix assigned to the host, it uses the

IPv6 Prefix Assigned to Host attribute to inform the HAAP. Both the LTE GRE Tunnel Notify message and the DSL GRE Tunnel Notify message MAY include the IPv6 Prefix Assigned to Host attribute.

```

+-----+
|Attribute Type |                               (1 byte)
+-----+
| Attribute Length |                         (2 bytes)
+-----+
| IPv6 Prefix Assigned to Host (16 bytes) |
+-----+

```

Attribute Type

IPv6 Prefix Assigned to Host, set to 21.

Attribute Length

Set to 17.

IPv6 Prefix Assigned to Host

The highest-order 16 octets encode an IPv6 address. The lowest-order one octet encodes the prefix length. These two values are put together to represent an IPv6 prefix.

#### 5.6.8. Diagnostic Start: Bonding Tunnel

The HG uses the Diagnostic Start: Bonding Tunnel attribute to inform the HAAP to switch to diagnostic mode to test the performance of the entire bonding tunnel. The Diagnostic Start: Bonding Tunnel attribute SHOULD be carried in the DSL GRE Tunnel Notify message.

```

+-----+
|Attribute Type |                               (1 byte)
+-----+
| Attribute Length |                         (2 bytes)
+-----+

```

Attribute Type

Diagnostic Start: Bonding Tunnel, set to 26.

Attribute Length

Set to 0.

#### 5.6.9. Diagnostic Start: DSL Tunnel

The HG uses the Diagnostic Start: DSL Tunnel attribute to inform the HAAP to switch to diagnostic mode to test the performance of the DSL GRE tunnel. The Diagnostic Start: DSL Tunnel attribute SHOULD be carried in the DSL GRE Tunnel Notify message.

```

+-----+
|Attribute Type |                (1 byte)
+-----+
| Attribute Length          |    (2 bytes)
+-----+

```

## Attribute Type

Diagnostic Start: DSL Tunnel, set to 27.

## Attribute Length

Set to 0.

## 5.6.10. Diagnostic Start: LTE Tunnel

The HG uses the Diagnostic Start: LTE Tunnel attribute to inform the HAAP to switch to diagnostic mode to test the performance of the entire bonding tunnel. The Diagnostic Start: LTE Tunnel attribute SHOULD be carried in the DSL GRE Tunnel Notify message.

```

+-----+
|Attribute Type |                (1 byte)
+-----+
| Attribute Length          |    (2 bytes)
+-----+

```

## Attribute Type

Diagnostic Start: LTE Tunnel, set to 18.

## Attribute Length

Set to 0.

## 5.6.11. Diagnostic End

The HG uses the Diagnostic End attribute to inform th HAAP to stop operating in diagnostic mode. The Diagnostic End attribute SHOULD be carried in the DSL GRE Tunnel Notify message.

```

+-----+
|Attribute Type |                (1 byte)
+-----+
| Attribute Length          |    (2 bytes)
+-----+

```

## Attribute Type

Diagnostic End, set to 29.

## Attribute Length

Set to 0.

## 5.6.12. Filter List Package ACK

The HG uses the Filter List Package ACK attribute to acknowledge the Filter List Package sent by the HAAP. Both the LTE GRE Tunnel Notify message and the DSL GRE Tunnel Notify message MAY include the Filter List Package ACK attribute.

```

+-----+
|Attribute Type | (1 byte)
+-----+
| Attribute Length | (2 bytes)
+-----+
| Filter List Package ACK (5 bytes) |
+-----+

```

Attribute Type  
Filter List Package ACK, set to 30.

Attribute Length  
Set to 5.

Filter List Package ACK  
The highest-order 4 octets are the Commit\_Count as defined in Section 5.6.2. The lowest-order 1 octet encodes the following error codes:

- 0: The Filter List Package is acknowledged.
- 1: The Filter List Package is not acknowledged. The HG is a new subscriber and has not ever received a Filter List Package. In this case, the HAAP SHOULD tear down the bonding tunnels and force the HG to re-establish the GRE Tunnels.
- 2: The Filter List Package is not acknowledged. The HG has already got a valid Filter List Package. The filter list on the HG will continue to be used while HAAP need to do nothing.

## 5.6.13. Switching to Active Hello State

If traffic is being sent/received over the bonding GRE tunnels before the "No Traffic Monitored Interval" expires (See Section 5.2.15.), the HG sends to the HAAP a GRE Tunnel Notify message containing the Switching to Active Hello State attribute.

The HAAP will switch to active hello state and send the HG a GRE Tunnel Notify message carrying the Switching to Active Hello State attribute as the ACK.

When the HG receives the ACK, it will switch to active hello state, start RTT detection and start sending GRE Tunnel Hello messages with

the Active Hello Interval (See Section 5.2.6.).

```

+-----+
|Attribute Type |                (1 byte)
+-----+-----+
| Attribute Length          |    (2 bytes)
+-----+-----+

```

Attribute Type

Switching to Active Hello State, set to 33.

Attribute Length

Set to 0.

#### 5.6.14. Switching to Idle Hello State

The HG initiates switching to idle hello state when the bonding of GRE Tunnels is successfully established and the LTE GRE Tunnel Setup Accept message carrying the Idle Hello Interval attribute (See Section 5.2.14.) is received. The HG sends the HAAP a GRE Tunnel Notify message containing the Switching to Idle Hello State attribute.

The HAAP will switch to idle hello state, clear RTT state and send the HG a GRE Tunnel Notify message carrying the Switching to Idle Hello State attribute as the ACK.

When the HG receives the ACK, it will switch to idle hello state, stop RTT detection, clear RTT state as well and start sending GRE Tunnel Hello messages with the Idle Hello Interval (See Section 5.2.14).

```

+-----+
|Attribute Type |                (1 byte)
+-----+-----+
| Attribute Length          |    (2 bytes)
+-----+-----+

```

Attribute Type

Switching to Idle Hello State, set to 34.

Attribute Length

Set to 0.

#### 5.6.15. Tunnel Verification

The HAAP uses the Tunnel Verification attribute to inform the HG to verify whether an existing LTE GRE tunnel is still functioning. The

Tunnel Verification attribute SHOULD be carried in the LTE GRE Tunnel Notify message. It provides a means to detect the tunnel faster than the GRE Tunnel Hello, especially when the LTE GRE tunnel is in the Idle Hello state and it takes much longer time to detect this tunnel.

When the HAAP receives an LTE GRE Tunnel Setup Request and finds the requested tunnel is conflicting with an existing tunnel, the HAAP initiates the Tunnel Verification. The HAAP drops all conflicting LTE GRE Tunnel Setup Request messages and sends GRE Tunnel Notify messages carrying the Tunnel Verification attribute until the verification ends. The HG MUST respond to the HAAP with the same Tunnel Verification attribute as the ACK if the tunnel is still functioning.

If the ACK of the Tunnel Verification attribute is received from the HG, the HAAP judges that the existing tunnel is still functioning. An LTE GRE Tunnel Deny message (with Error Code = 8) will be sent to the HG. The HG SHOULD terminate the GRE tunnel setup request process immediately.

If the HAAP does not receive a Tunnel Verification ACK message after 3 attempts (1 initial attempt and 2 retries), it will regard the existing tunnel as failed and the LTE GRE Tunnel Setup Request will be accepted.

```

+-----+
|Attribute Type |                (1 byte)
+-----+-----+
| Attribute Length |            (2 bytes)
+-----+-----+

```

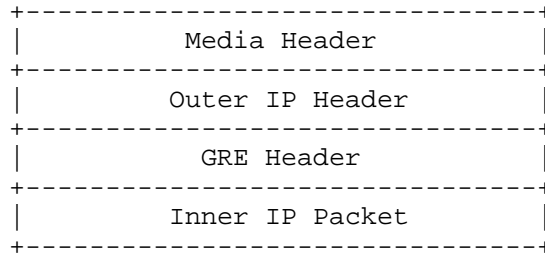
Attribute Type  
Tunnel Verification, set to 35.

Attribute Length  
Set to 0.

## 6. Tunnel Protocol Operation (Data Plane)

GRE tunnels are set up over heterogeneous connections, such as LTE and DSL, between the HG and the HAAP. Users' IP (inner) packets are encapsulated in GRE packets which in turn are carried in IP (outer) packets. The general structure of data packets of the GRE Tunnel Bonding protocol is shown as below.





### 6.1. The GRE Header

The GRE header is first standardized in [RFC2784]. [RFC2890] adds the optional Key and Sequence Number fields.

The Checksum and the Reserved1 fields are not used in the GRE Tunnel Bonding, therefore the C bit is set to zero.

The Key bit is set to one so that the Key field is present. The Key field is used as a 32-bit random number. It is generated by the HAAP per bonding connection and notified to HG (See Section 5.2.9).

The S bit is set to one, and the Sequence Number field is present and used for in-order delivery as per [RFC2890].

The Protocol Type field in the GRE header MUST be set to 0x0800 for IPv4 or 0x86DD for IPv6. So the GRE header used by data packets of the GRE Tunnel Bonding protocol have the following format.

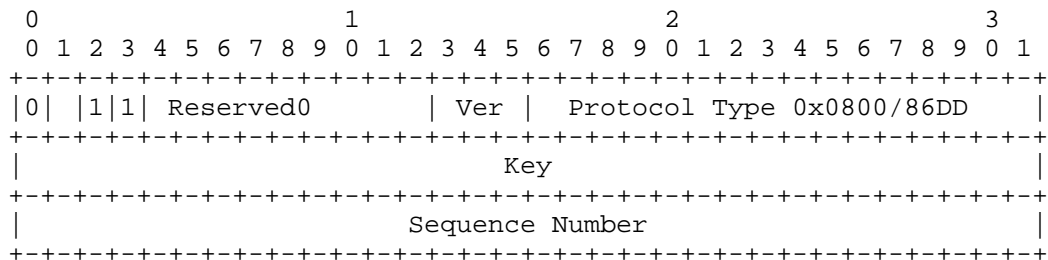


Figure 6.1 The GRE header for data packets of GRE Tunnel Bonding

### 6.2. Automatic Setup of GRE Tunnels

The HG gets the DSL WAN interface IP address (D) from the Broadband Remote Access Server (BRAS) via Point-to-Point Protocol over Ethernet (PPPoE), and gets the LTE WAN interface IP address (E) through Packet Data Protocol (PDP) from the Packet Data Network Gateway (PGW). The domain name of a HAAP group may be configured or obtained via the

DSL/LTE WAN interface based on gateway configuration protocols such as [TR-069], where the HAAP group comprises of one or multiple HAAPs. The Domain Name System (DNS) resolution of the HAAP group's domain name is requested via the DSL/LTE WAN interface. The DNS server will reply with an anycast HAAP IP address (G) which MAY be pre-configured by the operator.

After the interface IP addresses have been acquired, the HG starts the following GRE Tunnel Bonding procedure. It is REQUIRED that the HG first set up the LTE GRE tunnel and then set up the DSL GRE tunnel.

The HG sends the GRE Tunnel Setup Request message to the HAAP via the LTE WAN interface. The outer source IP address for this message is the LTE WAN interface IP address (E) while the outer destination IP address is the anycast HAAP IP address (G). The HAAP with the highest priority (e.g., the one that the HG has the least cost path to reach) in the HAAP group, which receives the GRE Tunnel Setup Request message, will initiate the Authentication and Authorization procedure, as specified in [TS23.401], to check whether the HG is trusted by the PGW.

If the Authentication and Authorization succeed, the HAAP sets the LTE WAN interface IP address (E) which is obtained from the GRE Tunnel Setup Request message (i.e., its outer source IP address) as the destination endpoint IP address of the GRE tunnel and replies to the HG's LTE WAN interface with the GRE Tunnel Setup Accept message in which an IP address (H) of the HAAP (e.g. an IP address of a Line Card in the HAAP) and a Session ID randomly generated by the HAAP are carried as attributes. The outer source IP address for this message is the IP address (H) or the anycast HAAP IP address (G) while the outer destination IP address is the LTE WAN interface IP address (E). Otherwise, the HAAP MUST send to the HG's LTE WAN interface the GRE Tunnel Setup Deny message and the HG MUST terminate the tunnel set up process once it receives the GRE Tunnel Setup Deny message.

After the LTE GRE tunnel is successfully set up, the HG will obtain the C address over the tunnel from the HAAP through Dynamic Host Configuration Protocol (DHCP). After that, the HG starts to set up the DSL GRE tunnel. It sends a GRE Tunnel Setup Request message via the DSL WAN interface, carrying the aforementioned Session ID received from the HAAP. The outer source IP address for this message is the DSL WAN interface IP address (D) while the outer destination IP address is the IP address (H) of the HAAP. The HAAP, which receives the GRE Tunnel Setup Request message, will initiate the Authentication and Authorization procedure in order to check whether the HG is trusted by the BRAS.

If the Authentication and Authorization succeed, the HAAP sets the DSL WAN interface IP address (D) which is obtained from the GRE Tunnel Setup Request message (i.e., its outer source IP address) as the destination endpoint IP address of the GRE tunnel and replies to the HG's DSL WAN interface with the GRE Tunnel Setup Accept message. The outer source IP address for this message is the IP address (H) of the HAAP while the outer destination IP address is the DSL WAN interface IP address (D). In this way, the two tunnels with the same Session ID can be used to carry traffic from the same user. That is to say, the two tunnels are "bonded" together. Otherwise, if the Authentication and Authorization fail, the HAAP MUST send to the HG's DSL WAN interface the GRE Tunnel Setup Deny message. Meanwhile, it MUST send to the HG's LTE WAN interface the GRE Tunnel Tear Down message. The HG MUST terminate the tunnel set up process once it receives the GRE Tunnel Setup Deny message and MUST tear down the LTE GRE tunnel that has been set up once it receives the GRE Tunnel Tear Down Message.

## 7. Security Considerations

Malicious devices controlled by attackers may intercept the control messages sent on the GRE tunnels. Later on, the rogue devices may fake control messages to disrupt the GRE tunnels or attract traffic of the target HG.

As a security feature, the Key field of the GRE header of the control messages and the data packets is generated as a 32-bit clear-text password, except the first GRE Setup Request message per bonding connection sent from HG to HAAP, whose Key field is filled with all zeros. HAAP and HG validate the Key value and the outer source IP address and discard packets with any invalid combination.

Moreover, GRE over IP Security (IPSec) could be used to enhance the security.

## 8. IANA Considerations

IANA need not to assign anything for the GRE Tunnel Bonding Protocol. The GRE Protocol Type for the GRE Channel is set to 0xB7EA which is under the control of IEEE Registration Authority. However, IANA may update the "IEEE 802 Numbers" IANA web page [802Type] which is of primarily historic interest.

## 9. Contributors

Li Xue  
Individual

E-Mail: [xueli\\_jas@163.com](mailto:xueli_jas@163.com)

Zhongwen Jiang  
Huawei Technologies

E-Mail: [jiangzhongwen@huawei.com](mailto:jiangzhongwen@huawei.com)

## 10. References

### 10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC2697] Heinanen, J. and R. Guerin, "A Single Rate Three Color Marker", RFC 2697, DOI 10.17487/RFC2697, September 1999, <<http://www.rfc-editor.org/info/rfc2697>>.
- [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>>.
- [RFC2890] Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, DOI 10.17487/RFC2890, September 2000, <<http://www.rfc-editor.org/info/rfc2890>>.
- [TR-069] Broadband Forum, "CPE WAN Management Protocol", Issue: 1 Amendment 5, Nov, 2013, <[https://www.broadband-forum.org/technical/download/TR-069\\_Amendment-5.pdf](https://www.broadband-forum.org/technical/download/TR-069_Amendment-5.pdf)>
- [TS23.401] "3GPP TS23.401, General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access", September 2013.
- [Port-NO] IANA, "Service Name and Transport Protocol Port Number Registry", <<http://www.iana.org/assignments/service-names-port-numbers>>
- [Pro-NO] IANA, "Assigned Internet Protocol Numbers", <<http://www.iana.org/assignments/protocol-numbers>>
- [RFC4632] Fuller, V. and T. Li, "Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan", BCP 122, RFC 4632, DOI 10.17487/RFC4632, August 2006, <<http://www.rfc-editor.org/info/rfc4632>>.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <<http://www.rfc-editor.org/info/rfc4291>>.

## 10.2. Informative References

[RFC1594] Marine, A., Reynolds, J., and G. Malkin, "FYI on Questions and Answers - Answers to Commonly asked "New Internet User" Questions", RFC 1594, March 1994.

[RFC2724] Handelman, S., Stibler, S., Brownlee, N., and G. Ruth, "RTFM: New Attributes for Traffic Flow Measurement", RFC 2724, DOI 10.17487/RFC2724, October 1999, <<http://www.rfc-editor.org/info/rfc2724>>.

[RFC6320] Wadhwa, S., Moisand, J., Haag, T., Voigt, N., and T. Taylor, Ed., "Protocol for Access Node Control Mechanism in Broadband Networks", RFC 6320, DOI 10.17487/RFC6320, October 2011, <<http://www.rfc-editor.org/info/rfc6320>>.

[RFC6733] Fajardo, V., Ed., Arkko, J., Loughney, J., and G. Zorn, Ed., "Diameter Base Protocol", RFC 6733, DOI 10.17487/RFC6733, October 2012, <<http://www.rfc-editor.org/info/rfc6733>>.

[RFC7676] Pignataro, C., Bonica, R., and S. Krishnan, "IPv6 Support for Generic Routing Encapsulation (GRE)", RFC 7676, DOI 10.17487/RFC7676, October 2015, <<http://www.rfc-editor.org/info/rfc7676>>.

[802Type] IANA, "IEEE 802 Numbers", <<http://www.iana.org/assignments/ieee-802-numbers>>.

## Author's Addresses

Nicolai Leymann  
Deutsche Telekom AG  
Winterfeldtstrasse 21-27  
Berlin 10781  
Germany

Phone: +49-170-2275345  
Email: n.leymann@telekom.de

Cornelius Heidemann  
Deutsche Telekom AG  
Heinrich-Hertz-Strasse 3-7  
Darmstadt 64295  
Germany

Phone: +4961515812721  
Email: heidemanc@telekom.de

Mingui Zhang  
Huawei Technologies  
No.156 Beiqing Rd. Haidian District,  
Beijing 100095 P.R. China

EMail: zhangmingui@huawei.com

Behcet Sarikaya  
Huawei USA  
5340 Legacy Dr. Building 3  
Plano, TX 75024

EMail: sarikaya@ieee.org

Margaret Cullen  
Painless Security  
14 Summer St. Suite 202  
Malden, MA 02148 USA

EMail: margaret@painless-security.com