An MPTCP Option for Network-Assisted MPTCP Deployments: Plain Transport Mode
draft-boucadair-mptcp-plain-mode-06

Abstract

One of the promising deployment scenarios for Multipath TCP (MPTCP) is to enable a Customer Premises Equipment (CPE) that is connected to multiple networks (e.g., DSL, LTE, WLAN) to optimize the usage of its network attachments. Because of the lack of 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 Concentrator. This document focuses on a deployment scheme where the identity of the MPTCP Concentrator(s) is explicitly configured on connected hosts.

This document specifies an MPTCP option that is used to avoid an encapsulation scheme between the CPE and the MPTCP Concentrator. Also, this document specifies how UDP traffic can be distributed among available paths without requiring any encapsulation scheme.

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 June 10, 2016.

Copyright Notice

Copyright (c) 2015 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 .............................................. 4
3. Assumptions .............................................. 5
4. Introducing the MPTCP Plain Transport Mode ............... 5
   4.1. An Alternative to Encapsulation ..................... 5
   4.2. Plain Mode MPTCP Option ............................. 6
   4.3. Theory of Operation ................................. 7
   4.4. Flow Example ....................................... 8
5. UDP Traffic .............................................. 9
6. IANA Considerations ..................................... 11
7. Security Considerations ................................. 11
8. Acknowledgements ....................................... 12
9. References .............................................. 12
   9.1. Normative References ............................... 12
   9.2. Informative References ............................ 12
Authors’ Addresses ......................................... 13
1. Introduction

One of the promising deployment scenarios for Multipath TCP (MPTCP, [RFC6824]) is to enable a Customer Premises Equipment (CPE) that is connected to multiple networks (e.g., DSL, LTE, WLAN) to optimize the usage of such resources, see for example [I-D.deng-mptcp-proxy] or [RFC4908]. This deployment scenario relies on MPTCP proxies located on both the CPE and network sides (Figure 1). The latter plays the role of traffic concentrator. A concentrator terminates the MPTCP sessions established from a CPE, before redirecting traffic into a legacy TCP session.

Both implicit and explicit models are considered to steer traffic towards an MPTCP Concentrator. This document focuses on the explicit model that consists in configuring explicitly the reachability information of the MPTCP concentrator on a host (e.g., [I-D.boucadair-mptcp-dhc]).

This specification assumes an MPTCP Concentrator is reachable through one or multiple IP addresses. Also, it assumes the various network attachments provided to an MPTCP-enabled CPE are managed by the same administrative entity. Additional assumptions are listed in Section 3.

This document explains how a plain transport mode, where packets are exchanged between the CPE and the concentrator without requiring the
activation of any encapsulation scheme (e.g., IP-in-IP [RFC2473], GRE [RFC1701], SOCKS [RFC1928], etc.), can be enabled.

Also, this document investigates an alternate track where UDP flows can be distributed among available paths without requiring any encapsulation scheme.

The solution in this document does not require the modification of the binding information base (BIB) structure maintained by both the CPE and the Concentrator. Likewise, this approach does not infer any modification of the Network Address Translator (NAT) functions that may reside in both the CPE and the device that embeds the concentrator.

The solution also works properly when NATs are present in the network between the CPE and the Concentrator, unlike solutions that rely upon GRE tunneling. Likewise, the solution accommodates deployments that involve CGN (Carrier Grade NAT) upstream the Concentrator.

2. Terminology

This document makes use of the following terms:

- **Customer-facing interface**: is an interface of the MPTCP Concentrator that is visible to a CPE and which is used for communication purposes between a CPE and the MPTCP Concentrator.

- **MPTCP Proxy**: is a software module that is responsible for transforming a TCP connection into an MPTCP connection, and vice versa. Typically, an MPTCP proxy can be embedded in a CPE and/or a Concentrator.

- **MPTCP leg**: Refers to a network segment on which MPTCP is used to establish TCP connections.

- **MPTCP Concentrator (or concentrator)**: refers to a functional element that is responsible for aggregating the traffic of a group of CPEs. This element is located upstream in the network. One or multiple concentrators can be deployed in the network side to assist MPTCP-enabled CPEs to establish MPTCP connections via available network attachments.

  On the uplink path, the concentrator terminates the MPTCP connections received from its customer-facing interfaces and transforms these connections into legacy TCP connections towards upstream servers.
On the downlink path, the concentrator turns the legacy server’s TCP connection into MPTCP connections towards its customer-facing interfaces.

3. Assumptions

The following assumptions are made:

- The logic for mounting network attachments by a host is deployment- and implementation-specific and is out of scope of this document.
- The Network Provider that manages the various network attachments (including the concentrators) can enforce authentication and authorization policies using appropriate mechanisms that are out of scope of this document.
- Policies can be enforced by a concentrator instance operated by the Network Provider to manage both upstream and downstream traffic. These policies may be subscriber-specific, connection-specific or system-wide.
- The concentrator may be notified about the results of monitoring (including probing) the various network legs to service a customer, a group of customers, a given region, etc. No assumption is made by this document about how these monitoring (including probing) operations are executed.
- An MPTCP-enabled, multi-interfaced host that is directly connected to one or multiple access networks is allocated addresses/prefixes via legacy mechanisms (e.g., DHCP) supported by the various available network attachments. The host may be assigned the same or distinct IP address/prefix via the various available network attachments.
- The location of the concentrator(s) is deployment-specific. Network Providers may choose to adopt centralized or distributed (even if they may not be present on the different network accesses) designs, etc. Nevertheless, in order to take advantage of MPTCP, the location of the concentrator should not jeopardize packet forwarding performance for traffic sent from or directed to connected hosts.

4. Introducing the MPTCP Plain Transport Mode

4.1. An Alternative to Encapsulation

The design option for aggregating various network accesses often relies upon the use of an encapsulation scheme (such as GRE) between the CPE and the Concentrator. The use of encapsulation is motivated by the need to steer traffic through the concentrator and also to allow the distribution of UDP flows among the available paths without requiring any advanced traffic engineering tweaking technique in the
network side to intercept traffic and redirect it towards the appropriate concentrator.

This document specifies another approach that relies upon plain transport mode between the CPE and the Concentrator.

The use of a plain transport mode does not require the upgrade of any intermediate function (security, TCP optimizer, etc.) that may be located on-path. Thus, the introduction of MPTCP concentrators in operational networks to operate plain mode does not add any extra complexity as far as the operation of possible intermediate functions is concerned.

4.2. Plain Mode MPTCP Option

The format of the Plain Mode MPTCP option is shown in Figure 2.

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+---------------+---------------+-------+-------+---------------+
|     Kind      |     Length    |SubType|D|U|       Flag Bits   |
+---------------+---------------+-------+-------+---------------+
|   Address (IPv4 - 4 octets / IPv6 - 16 octets)   |
|---------------------------------------------------+---------------------+
|   Port (2 octets, optional)                       |
+---------------------------------------------------+
```

Figure 2: Plain Mode MPTCP Option

The description of the fields is as follows:

- Kind and Length: are the same as in [RFC6824].
- Subtype: to be defined by IANA (Section 6).
- D-bit (direction bit): This flag indicates whether the enclosed IP address (and a port number) reflects the source or destination IP address (and port). 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.
- U-bit (UDP bit): The use of this flag is detailed in Section 5.
- The "Flag" 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.
4.3. Theory of Operation

Plain mode operation is as follows:

1. The CPE is provisioned with the reachability information of one or several Concentrators (e.g., [I-D.boucadair-mptcp-dhc]).

2. Outgoing TCP packets that can be forwarded by a CPE along MPTCP subflows are transformed into TCP packets carried over a MPTCP connection. The decision-making process to decide whether a flow should be MPTCP-tagged or not is local to the Concentrator and the CPE. It depends on the policies provisioned by the network provider. As such, the decision-making process is policy-driven, implementation- and deployment-specific.

3. MPTCP packets are sent using a plain transport mode (i.e., without any encapsulation header).

The source IP address and source port number are those assigned locally by the CPE. Because multiple IP addresses may be available to the CPE, the address used to rewrite the source IP address for an outgoing packet forwarded through a given network attachment (typically, a WAN interface) MUST be associated with that network attachment. It is assumed that ingress filtering ([RFC2827]) is implemented at the boundaries of the networks to prevent any spoofing attack.

The destination IP address is replaced by the CPE with one of the IP addresses of the Concentrator.

The destination port number may be maintained as initially set by the host or altered by the CPE.

The original destination IP address is copied into a dedicated MPTCP option called Plain Mode MPTCP option (see Section 4.2). Because of the limited TCP option space, it is RECOMMENDED to implement the solution specified in [I-D.ietf-tcpm-tcp-edo]. As a reminder, [I-D.touch-tcpm-tcp-syn-ext-opt] specifies a proposal for TCP SYN extended option space.

A binding entry must be maintained by the CPE for that outgoing packet. This binding entry is instantiated by the NAT and/or the firewall functions embedded in the CPE.
(4) Upon receipt of the packet on the MPTCP leg, the Concentrator extracts the IP address included in the Plain Mode MPTCP Option that it uses as the destination IP address of the packet generated in the TCP leg towards its ultimate destination.

The source IP address and port are those of the Concentrators. A binding entry is instantiated by the Concentrator to record the state.

The concentrator may be configured to behave as either a 1:1 address translator or a N:1 translator where the same address is shared among multiple CPEs. Network Providers should be aware of the complications that may arise if a given IP address/prefix is shared among multiple hosts (see [RFC6967]). Whether these complications apply or not is deployment-specific.

The Concentrator should preserve the same IP address that was assigned to a given CPE for all its outgoing connections when transforming an MPTCP connection into a TCP connection.

(5) For incoming TCP packets that need to be forwarded to a CPE, the Concentrator records the source IP address in a Plain Mode MPTCP Option.

The source IP address is replaced with one of the IP addresses listed in the aforementioned binding information base maintained by the Concentrator (if such a state entry exists) or with one of the Concentrator’s IP addresses.

The destination IP address is replaced with the CPE’s IP address (if the corresponding state entry is found in the Concentrator’s binding table) or with one of the CPE’s IP addresses (that are known by the concentrator using some means that are out of the scope of the document).

4.4. Flow Example

A typical flow exchange is shown in Figure 3.

This example assumes no NAT is located between the CPE and the concentrator.

Because the remote server is not MPTCP-aware, the Concentrator is responsible for preserving the same IP address (conc_@, in the example) for the same CPE even if distinct IP addresses (cpe_@1 and cpe_@2, in the example) are used by the CPE to establish subflows with the Concentrator.
From an application standpoint, there may be a value to distribute UDP datagrams among available network attachments for the sake of network resource optimisation, for example.

Unlike existing proposals that rely upon encapsulation schemes such as IP-in-IP or GRE, this document suggests the use of MPTCP features to control how UDP datagrams are distributed among existing network attachments. UDP datagrams are therefore transformed into TCP-formatted packets.

Figure 3: Flow Example (No NAT between the CPE and the Concentrator)

5. UDP Traffic

Legend:
* "--Plain Mode MPTCP Option()--" indicates the packet is sent in a plain mode, i.e., without any encapsulation header, and that "Plain Mode MPTCP Option" is carried in the packet.
The CPE and the Concentrator establish a set of MPTCP subflows. These subflows are used to transport UDP datagrams that are distributed among existent subflows. TCP session tracking may not be enabled for the set of subflows that are dedicated to transport UDP traffic. The establishment of these subflows is not conditioned by the receipt of UDP packets; instead, these subflows are initiated upon CPE reboot or when network conditions change (e.g., whenever a new Concentrator is discovered or a new IP address is assigned to the Concentrator). Additional MPTCP connections may be established to anticipate UDP traffic to be distributed among several paths. The maximum number of MPTCP connections that can be dedicated to UDP traffic may be configured locally to the CPE and the Concentrator. How this parameter is configured is implementation and deployment-specific.

When the CPE (or the Concentrator) transforms a UDP packet into a TCP one, it must insert the Plain Mode MPTCP Option with the U-bit set. When setting the source IP address, the destination IP address, and the IP address enclosed in the Plain Mode MPTCP Option, the same considerations specified in Section 4.3 must be followed.

In addition, the CPE (or the Concentrator) must replace the UDP header with a TCP header. Upon receipt of the packet with the U-bit set, the Concentrator (or the CPE) transforms the packet into a UDP packet and follows the same considerations specified in Section 4.3. Both the CPE and the Concentrator may be configured to disable some features (e.g., reordering). Enabling these features is deployment and implementation-specific.

Relaying UDP packets is not conditioned by TCP session establishment because the required subflows that are dedicated to transport UDP traffic are already in place (either at the CPE or the Concentrator).

A flow example is shown in Figure 4.
6. IANA Considerations

This document requests an MPTCP subtype code for this option:

- Plain Mode MPTCP Option

7. Security Considerations

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

Means to protect the MPTCP concentrator against Denial-of-Service (DoS) attacks must be enabled. Such means include the enforcement of ingress filtering policies at the boundaries of the network. In order to prevent exhausting the resources of the concentrator by creating an aggressive number of simultaneous subflows for each MPTCP connection, the administrator should limit the number of allowed subflows per host for a given connection.
Attacks outside the domain can be prevented if ingress filtering is enforced. Nevertheless, attacks from within the network between a host and a concentrator instance are yet another actual threat. Means to ensure that illegitimate nodes cannot connect to a network should be implemented.

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

8. Acknowledgements

Many thanks to Chi Dung Phung, Mingui Zhang, and Christoph Paasch for the comments.

9. References

9.1. Normative References


9.2. Informative References


[I-D.ietf-tcpm-tcp-edo]
Touch, J. and W. Eddy, "TCP Extended Data Offset Option",
draft-ietf-tcpm-tcp-edo-04 (work in progress), November 2015.

[I-D.touch-tcpm-tcp-syn-ext-opt]
Touch, J. and T. Faber, "TCP SYN Extended Option Space Using an Out-of-Band Segment",
draft-touch-tcpm-tcp-syn-ext-opt-03 (work in progress), October 2015.

[RFC1701] Hanks, S., Li, T., Farinacci, D., and P. Traina, "Generic Routing Encapsulation (GRE)",
RFC 1701, DOI 10.17487/RFC1701, October 1994,

RFC 1928, DOI 10.17487/RFC1928, March 1996,

RFC 2473, DOI 10.17487/RFC2473, December 1998,

[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,

network Using Mobile IP and NEMO",
RFC 4908, DOI 10.17487/RFC4908, June 2007,

RFC 6967, DOI 10.17487/RFC6967, June 2013,

Authors’ Addresses