

Network Working Group
Internet-Draft
Intended status: Experimental
Expires: January 4, 2016

M. Boucadair
C. Jacquenet
P. Seite
France Telecom
O. Bonaventure
Tessares
L. Deng
China Mobile
July 3, 2015

An MPTCP Profile for NAT- and Firewall-Free Networks: Network-Assisted
MPTCP Deployments
draft-boucadair-mptcp-natfwfree-profile-00

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 such resources, thereby providing better serviceability overall (including whenever the CPE fails to connect to one of the access networks). This document specifies a MPTCP profile for such deployments in network regions that are firewall- and NAT-free.

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 January 4, 2016.

Internet-Draft

Transport Profiles

July 2015

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	2
2.	Assumptions	4
3.	Target Use Cases	5
4.	Relaxing Some Base MPTCP Features	5
5.	IANA Considerations	8
6.	Security Considerations	8
7.	Acknowledgements	9
8.	References	9
8.1.	Normative References	9
8.2.	Informative References	9
	Authors' Addresses	10

[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 on both the CPE and the network sides (Figure 1). The latter plays the role of traffic concentrator. A concentrator terminates the MPTCP sessions, from a CPE, before relaying it into a legacy TCP session.

Internet-Draft

Transport Profiles

July 2015

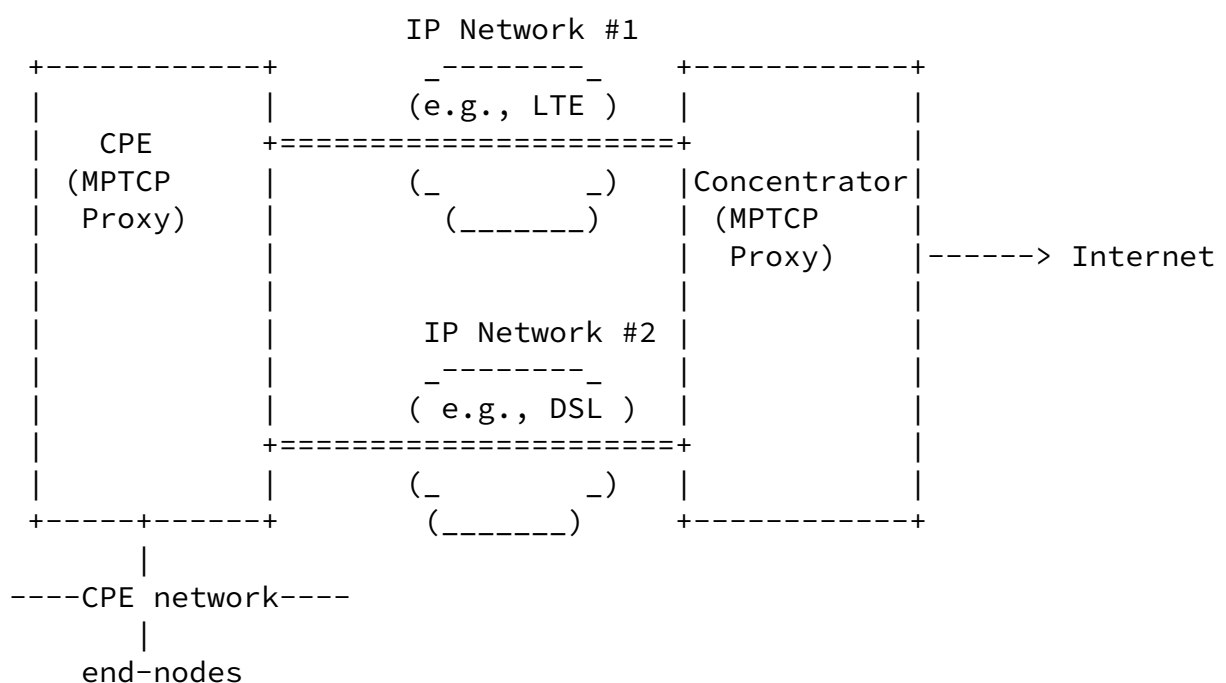


Figure 1: "Network-Assisted" MPTCP Design

Because the paths between the CPE and the concentrator are firewall- and NAT-free, the complexity of the MPTCP specification that was initially induced by the need to handle the presence of firewalls as well as routing asymmetry effects, is not justified anymore. Concretely, in the situations where the paths between the CPE and the concentrator are firewall- and NAT-free, the MPTCP stack is not required to support the dedicated features required to handle the presence of firewalls as well as routing asymmetry effects.

Such context encourages the specification of a dedicated MPTCP profile that would in turn foster the adoption of MPTCP. This document specifies such MPTCP profile that is adapted to network regions that are firewall- and NAT-free.

The constraint discussed in [[RFC6824](#)] does not apply for such

deployments:

"External Constraints: The protocol must function through the vast majority of existing middleboxes such as NATs, firewalls, and proxies, and as such must resemble existing TCP as far as possible on the wire. Furthermore, the protocol must not assume the segments it sends on the wire arrive unmodified at the destination: they may be split or coalesced; TCP options may be removed or duplicated."

NAT is used through this document to refer to any function that rewrites a source IP address/prefix to another IP address/prefix

within the same or distinct address family. NAT is also to refer to any function that rewrites port numbers. Typical examples are traditional IPv4-IPv4 NAT ([RFC3022]), NPTv6 ([RFC6296]), NAT64 ([RFC6146]), DS-Lite AFTR ([RFC6333]), or Carrier Grade NAT (CGN, [RFC6888]).

Lawful intercept and data retention implications due to the use of MPTCP are out of the scope of this document.

2. Assumptions

The following assumptions are made:

- o One or multiple concentrators are deployed on the network side to assist MPTCP-enabled hosts to establish MPTCP connections via available network attachments.
- o 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.
- o Various network attachments are provided to an MPTCP-enabled host/CPE; all these network attachments are managed by the same administrative entity.
- o The CPE implements an MPTCP proxy as well. This MPTCP proxy acts as a traffic concentrator from the end-nodes (i.e., hosts attached to the CPE) standpoint.
- o The network legs between the hosts and a concentrator instance are

NAT- and firewall-free.

- o The logic for mounting network attachments by a host is deployment- and implementation-specific that are out of scope of this document .
- o 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.
- o 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.
- o 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 probing operations are executed.
- o Ingress filtering ([\[RFC2827\]](#)) is implemented at the boundaries of the networks to provide anti-spoofing.

- o An MPTCP-enabled multiple Interfaces host, that is directly connected to one or multiple access networks obtains address/ prefixes via legacy mechanisms of the various available network attachments. The host may be assigned the same or distinct IP address/prefixes via the various available network attachments.
- o The CPE does not alter or strip MPTCP signals received from end-nodes.
- o The concentrator may behave in a transparent mode (that is, hosts are unaware of the presence of the concentrator in the communication path(s)) or in a non-transparent mode (i.e., the identity of the concentrator is explicitly configured to the hosts).
- o A mechanism should be used to make sure the same IP address is assigned to the host when transforming an MPTCP connection into a TCP connection. No assumption is made about how such mechanism is implemented. Network Providers should be aware of the complications that may arise if a same IP address/prefix is shared among multiple hosts (see [\[RFC6967\]](#)). Whether these complications apply or not is deployment-specific.
- o 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.

3. Target Use Cases

Two main use cases are targeted by this profile:

1. Multi-homed CPEs (e.g., [[RFC4908](#)]).
2. MPTCP within core networks to achieve load-balancing or bandwidth aggregation. This use case assumes that MPTCP connections are established between nodes that are managed by the same administrative entity.

4. Relaxing Some Base MPTCP Features

The following list is a set of items of the MPTCP specification that can be relaxed to facilitate and improve MPTCP operation in firewall- and NAT-free regions of the network. A technical justification is provided to relax each of these items. The rest of the MPTCP specifications that are not mentioned in this section MUST be followed even in firewall and NAT-free networks.

Item 1: Checksum SHOULD be disabled. This behavior implies in particular that "A" flag bit must always be set to 0.

Justification: This is a direct consequence of the absence of NATs and firewalls in the network leg between the host and a concentrator.

Item 2: [Section 3.6 of \[RFC6824\]](#) does not apply.

Justification: The target deployments assume that all paths are MPTCP-compliant; once the first subflow is established, it is safe to assume that any additional subflow will be successfully established over an MPTCP-compliant path. There is no need to envision a TCP fallback mechanism except for the first subflow.

Operators may run tests to assess whether available paths are MPTCP-compliant. For example, Operators can perform tests with tools like tracebox to validate the absence of middleboxes on the network legs that are used. It is out of scope of this document to define those tests.

- Item 3: Endpoints may rely on the source address of a sub-flow established by an initiating peer to establish new subflows or enforce policies (e.g., rate-limit at the concentrator side).

Justification: The point about private IP addresses discussed in Section of [[RFC6824](#)] does not apply, since there is no NAT in the path between the involved MPTCP endpoints.

- Item 4: Given that the network legs that are used are trusted, there is no need to authenticate the establishment of the additional subflows with a HMAC in the MP_JOIN. MP_JOIN options are still used, but they neither contain random numbers nor truncated HMACs. The MP_JOIN option in the SYN has a length of 8 bytes and contains the receiver's token. In the SYN+ACK and the third ACK, the MP_JOIN options have a length of 4 bytes.

Justification: The network leg between the directly-connected host and a concentrator instance is trusted. There is thus no risk of attack on this part of the network.

- Item 5: If the concentrator's reachability information is explicitly configured on the MPTCP host, and the concentrator is aware of addresses assigned to the MPTCP host, then the ADD_ADDR

option SHOULD NOT be supported. In such case, the host MUST rely upon the provided configuration information to manage an MPTCP connection.

Justification: A host that is configured with the addresses of a concentrator can use these addresses to establish one or multiple subflows for a given connection; each connection is then bound to an IP address of the concentrator that has

been "assigned" to the host, as per the concentrator's reachability information (a name, an IP address, etc.) provided to the host. The locators of the concentrators are likely to be stable. A locator can be a name, IPv4/v6 addresses, etc.

- Item 6: If the MPTCP endpoint is explicitly configured so that it behaves in a network-assisted mode, subflows can be created to the remote MPTCP concentrator, using a locally configured set of addresses, without advertising the available set of addresses. Triggers to decide how many sub-flows are to be initiated or when to establish additional ones are application-specific.

Justification: Multiple addresses are configured out of band (e.g., using DHCP [[I-D.boucadair-mptcp-dhc](#)], TR-69, etc.). The use of out-of-band configuration mechanism is justified in some deployments for engineering purpose (e.g., assign the concentrator to service a host based on criteria such as the load of the concentrator, geo-proximity, etc.).

- Item 7: If the concentrator has access to the information about address(es) assigned to a directly-connected host and their associated leases (e.g., because the concentrator is collocated with a DHCP relay or acquires such information by means of an out-of-band mechanism), the concentrator SHOULD undertake actions for a better quality of experience of MPTCP connections such as: add a new sub-flow even if the concentrator is not the initiator of the MPTCP connection, migrate flows to another alternate address if the remote address is not valid anymore or because its validity timeout will expire soon, etc.

Justification: This approach is meant to anticipate retransmissions that may be induced by the invalidity of an IP address. Also, this allows to reduce the delay from waiting for a notification from a remote peer. The concentrator can also decide to add new subflows for better quality of experience based, for example, on local policies.

- Item 8: Address management MAY not be specific to each active MPTCP

connection, but MAY be on a per host basis.

Justification: This is because all the MPTCP connections initiated by a host (resp. a concentrator) involve the same MPTCP endpoint (concentrator).

How such address management is actually achieved is implementation-specific. Nevertheless, for illustration purposes, a dedicated session can be enabled between an MPTCP-enabled host and a concentrator for control purposes. This information exchanged in this dedicated connection will be used to adjust other (data) connections.

Item 9: The maximum number of subflows for a given connection SHOULD be set by default to 4.

Justification: For dimensioning purposes, an operator needs to control the number of flows to be handled by a concentrator.

4 corresponds to a dual-homed host that is assigned both an IPv4 address and an IPv6 prefix for each network attachment.

An MPTCP endpoint that is dimensioned to maintain a maximum number of subflows per MPTCP connection may accept to maintain more subflows for some connections.

[5.](#) IANA Considerations

This document makes no request of IANA.

[6.](#) 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. This profile recommends this value to be set to 4.

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 another yet actual threat. Means to ensure that illegitimate nodes cannot connect to a network should be implemented.

Traffic theft can be achieved if an illegitimate concentrator is inserted in the path. Indeed, inserting an illegitimate concentrator in the forwarding path allows to intercept traffic and therefore have 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.

This document relax checksum validations (Item (1), [Section 4](#)) and MP_JOIN authentication constraints (Item (4), [Section 4](#)) because the networks between two MPTCP endpoints is trusted. Furthermore, ingress filtering is enforced at these networks for source address validation.

[7.](#) Acknowledgements

TBC.

[8.](#) References

[8.1.](#) Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC6824] Ford, A., Raiciu, C., Handley, M., and O. Bonaventure, "TCP Extensions for Multipath Operation with Multiple Addresses", [RFC 6824](#), January 2013.

[8.2.](#) Informative References

- [I-D.boucadair-mptcp-dhc]
Boucadair, M., Jacquenet, C., and T. Reddy, "IP Flow Information Export (IPFIX) Entities", July 2015, <<https://datatracker.ietf.org/doc/draft-boucadair-mptcp-dhc/>>.
- [I-D.deng-mptcp-proxy]
Lingli, D., Liu, D., Sun, T., Boucadair, M., and G. Cauchie, "Use-cases and Requirements for MPTCP Proxy in ISP Networks", [draft-deng-mptcp-proxy-01](#) (work in

progress), October 2014.

Boucadair, et al.

Expires January 4, 2016

[Page 9]

Internet-Draft

Transport Profiles

July 2015

- [RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", [BCP 38](#), [RFC 2827](#), May 2000.
- [RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", [RFC 3022](#), January 2001.
- [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](#), June 2007.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", [RFC 6146](#), April 2011.
- [RFC6296] Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Prefix Translation", [RFC 6296](#), June 2011.
- [RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", [RFC 6333](#), August 2011.
- [RFC6888] Perreault, S., Yamagata, I., Miyakawa, S., Nakagawa, A., and H. Ashida, "Common Requirements for Carrier-Grade NATs (CGNs)", [BCP 127](#), [RFC 6888](#), April 2013.
- [RFC6967] Boucadair, M., Touch, J., Levis, P., and R. Penno, "Analysis of Potential Solutions for Revealing a Host Identifier (HOST_ID) in Shared Address Deployments", [RFC 6967](#), June 2013.

Authors' Addresses

Mohamed Boucadair
France Telecom
Rennes 35000
France

Email: mohamed.boucadair@orange.com

Christian Jacquenet
France Telecom
Rennes
France

Email: christian.jacquenet@orange.com

Boucadair, et al.

Expires January 4, 2016

[Page 10]

Internet-Draft

Transport Profiles

July 2015

Pierrick Seite
France Telecom
Rennes
France

Email: pierrick.seite@orange.com

Olivier Bonaventure
Tessares

Email: Olivier.Bonaventure@tessares.net

URI: <http://www.tessares.net>

Lingli Deng
China Mobile

Email: denglingli@chinamobile.com

