Workgroup: Transport Area Working Group
Internet-Draft:
draft-ietf-tsvwg-multipath-dccp-14
Published: 17 March 2024
Intended Status: Standards Track
Expires: 18 September 2024
Authors: M. Amend, Ed.
DT
A. Kassler
Karlstad University
S. Johnson
BT
A. Brunstrom
Karlstad University
V. Rakocevic
City, University of London

DCCP Extensions for Multipath Operation with Multiple Addresses

Abstract

DCCP communications as defined in [RFC4340] are restricted to a single path per connection, yet multiple paths often exist between peers. The simultaneous use of available multiple paths for a DCCP session could improve resource usage within the network and, thus, improve user experience through higher throughput and improved resilience to network failures. Use cases for Multipath DCCP (MP-DCCP) are mobile devices (e.g., handsets, vehicles) and residential home gateways simultaneously connected to distinct networks as, e.g., a cellular and a Wireless Local Area (WLAN) network or a cellular and a fixed access network. Compared to the existing multipath protocols, such as MPTCP, MP-DCCP provides specific support for non-TCP user traffic (e.g., UDP or plain IP).

This document specifies a set of extensions to DCCP to support multipath operations. Multipath DCCP provides the ability to simultaneously use multiple paths between peers. The protocol offers the same type of service to applications as DCCP and provides the components necessary to establish and use multiple DCCP flows across different paths simultaneously.

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 https://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 18 September 2024.

Copyright Notice

Table of Contents

1. Introduction

Datagram Congestion Control Protocol (DCCP) [RFC4340] is a transport protocol that provides bidirectional unicast connections of congestion-controlled unreliable datagrams. DCCP communications are restricted to one single path. This document specifies a set of protocol changes that add multipath support to DCCP; specifically, support for signaling and setting up multiple paths (a.k.a, "subflows"), managing these subflows, reordering of data, and termination of sessions.

Multipath DCCP (MP-DCCP) enables a DCCP connection to simultaneously establish a flow across multiple paths. This can be beneficial to applications that transfer large amounts of data, by utilizing the capacity/connectivity offered by multiple paths. In addition, the multipath extensions enable to tradeoff timeliness and reliability, which is important for low-latency applications that do not require guaranteed delivery services, such as Audio/Video streaming.

MP-DCCP was first suggested in the context of the 3GPP work on 5G multi-access solutions [I-D.amend-tsvwg-multipath-framework-mpdccp] and for hybrid access networks [I-D.lhwxz-hybrid-access-network-architecture] [I-D.muley-network-based-bonding-hybrid-access], where MP-DCCP can be applied for load-balancing, seamless session handover, and bandwidth aggregation (referred to as Access Traffic Steering, Switching, and Splitting (ATSSS) in the 3GPP terminology [TS23.501]). More details on potential use cases for MP-DCCP are provided in [multipath-dccp.org], [IETF115.Slides], and [MP-DCCP.Paper]. All these use cases profit from an Open Source Linux reference implementation provided under [multipath-dccp.org].

Encapsulation for DCCP in UDP is defined in [RFC6773]. [I-D.amend-tsvwg-multipath-framework-mpdccp] proposes a lightweight encapsulation for DCCP flow headers appropriate for unreliable IP traffic in terms of UDP and other non-TCP packets in comparison to MPTCP. This is not considered in the present specification.

Similar to MP-DCCP, MP-QUIC is designed to enable the simultaneous usage of multiple paths for a single connection [I-D.ietf-quic-multipath]. MP-QUIC is based on QUIC in a similar way as MP-DCCP is based on DCCP. MP-QUIC inherits the properties of QUIC with its various facets of encryption, multi-streaming and the STREAM and DATAGRAM transport characteristic. This makes a practical multipath implementation very complex. In contrast, MP-DCCP is based exclusively on the lean concept of DCCP. For traffic that is already encrypted, MP-DCCP is the more efficient choice as it does not apply its own encryption mechanisms. Also, the procedures defined by MP-DCCP, which allow subsequent reordering of traffic, improve performance, as shown in [MP-DCCP.Paper], and are not available in MP-QUIC.

1.1. Multipath DCCP in the Networking Stack

MP-DCCP provides a set of features to DCCP; Figure 1 illustrates this layering. It operates at the transport layer and can be used as a transport for both higher and lower layers. It is designed to be used by applications in the same way as DCCP with no changes to the application itself.

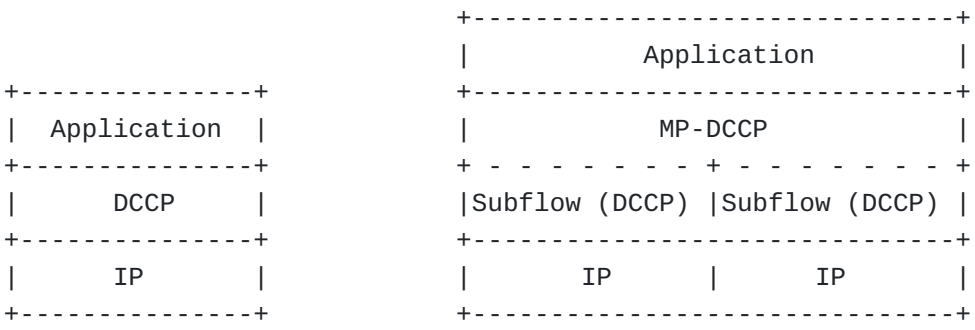

Figure 1: Comparison of standard DCCP and MP-DCCP protocol stacks

1.2. Terminology

This document uses terms that are either specific for multipath transport or are defined in the context of MP-DCCP, similar to [RFC8684], as follows:

Path: A sequence of links between a sender and a receiver, defined in this context by a 4-tuple of source and destination address/port pairs.

(MP-DCCP) Connection: A set of one or more subflows, over which an application can communicate between two hosts. The MP-DCCP connection is exposed as single DCCP socket to the application.

Connection Identifier (CI): A locally unique identifier given to a multipath connection by a host.

Host: An end host operating an MP-DCCP implementation, and either initiating or accepting an MP-DCCP connection.

Subflow: A subflow refers to a DCCP flow transmitted using a specific path (4-tuple of source and destination address/port pairs) that forms one of the multipath flows used by a single connection.

In addition to these terms, within the framework of MP-DCCP, the interpretation of, and effect on, regular single-path DCCP semantics is discussed in Section 3.

1.3. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Operation Overview

DCCP transmits congestion-controlled unreliable datagrams over a single path.
Various congestion control mechanisms have been specified to optimize DCCP performance for specific traffic types in terms of profiles denoted by a Congestion Control IDentifier (CCID). However, DCCP does not provide built-in support for managing multiple subflows within one DCCP connection.

The extension of DCCP for Multipath-DCCP (MP-DCCP) is described in detail in Section 3.

At a high level of the MP-DCCP operation, the data stream from a DCCP application is split by MP-DCCP operation into one or more subflows which can be transmitted via different - also physically isolated - paths. The corresponding control information allows the receiver to optionally re-assemble and deliver the received data in the originally transmitted order to the recipient application. This may be necessary because DCCP does not guarantee in-order delivery. The details of the transmission scheduling mechanism and optional reordering mechanism are up to the sender and receiver, respectively, and are outside the scope of this document.

A Multipath DCCP connection provides a bidirectional connection of datagrams between two hosts exchanging data using DCCP. It does not require any change to the applications. Multipath DCCP enables the hosts to use multiple paths with different IP addresses to transport the packets of an MP-DCCP connection. MP-DCCP manages the request, set-up, authentication, prioritization, modification, and removal of the DCCP subflows on different paths as well as the exchange of performance parameters.
The number of DCCP subflows can vary during the lifetime of a Multipath DCCP connection. The details of the path management decisions for when to add or remove subflows are outside the scope of this document.

The Multipath Capability for MP-DCCP is negotiated with a new DCCP feature, as specified in Section 3.1. Once negotiated, all subsequent MP-DCCP operations for that connection are signalled with a variable length multipath-related option, as described in Section 3. All MP-DCCP operations are signaled by Multipath options described in Section 3.2. Options that require confirmation from the remote peer are retransmitted by the sender until confirmed or until confirmation is no longer considered relevant.

The following sections define MP-DCCP behavior in detail.

2.1. MP-DCCP Concept

Figure 2 provides a general overview of the MP-DCCP working mode, whose main characteristics are summarized in this section.

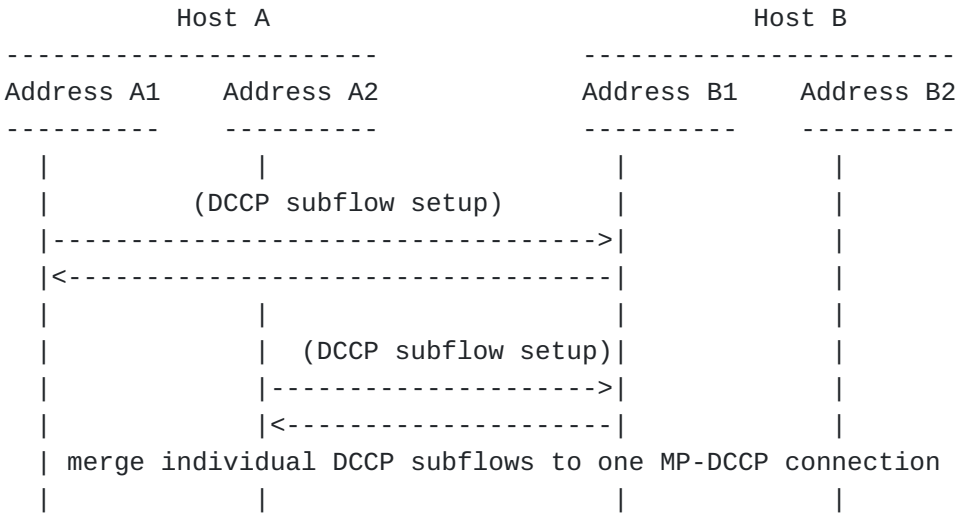

Figure 2: Example MP-DCCP usage scenario

*An MP-DCCP connection begins with a 4-way handshake, between two hosts. In Figure 2, an MP-DCCP connection is established between addresses A1 and B1 on Hosts A and B, respectively. In the

handshake, a Multipath Capable feature is used to negotiate multipath support for the connection. Host specific keys are also exchanged between Host A and Host B during the handshake. The details of the MP-DCCP handshaking procedure is described in Section 3.3. MP-DCCP does not require both peers to have more than one address.

*When additional paths and corresponding addresses/ports are available, additional DCCP subflows can be created on these paths and attached to the existing MP-DCCP connection. An MP_JOIN option is used to connect a new DCCP subflow to an existing MP-DCCP connection. It contains a Connection Identifier during the setup of the initial subflow and is exchanged in the 4-way handshake for the subflow together with the Multipath Capable feature. The example in Figure 2 illustrates creation of an additional DCCP subflow between Address A2 on Host A and Address B1 on Host B. The two subflows continues to provide a single connection to the applications at both endpoints.

*MP-DCCP identifies multiple paths by the presence of multiple addresses/ports at hosts. Combinations of these multiple addresses/ports indicate the additional paths. In the example, other potential paths that could be set up are A1<->B2 and A2<->B2. Although the additional subflow in the example is shown as being initiated from A2, an additional subflow could alternatively have been initiated from B1 or B2.

*The discovery and setup of additional subflows is achieved through a path management method including the logic and details of the procedures for adding/removing subflows; this document describes measures to allow a host to initiate new subflows and signal available addresses between peers. The definition of a path management method is, however, out of scope of this document and subject to a corresponding policy and the specifics of the implementation. If a MP-DCCP peer host limits the maximum number of paths that can be maintained (e.g., similar to what is discussed in Section 3.4 of [RFC8041], the creation of new subflows from that peer host needs to be avoided and incoming subflow requests terminated.

*Through the use of multipath options, MP-DCCP adds connection-level sequence numbers and exchange of Round-Trip Time (RTT) information to enable optional reordering features. As a hint for scheduling decisions, a multipath option that allows a peer to indicate its priorities for what path to use is also defined.

*Subflows are terminated in the same way as regular DCCP connections, as described in ([RFC4340], Section 8.3). MP-DCCP connections are closed by including an MP_CLOSE option in subflow

DCCP-CloseReq or DCCP-Close messages. An MP-DCCP connection may also be reset through the use of an MP_FAST_CLOSE option. Key data from the initial handshake is included in the MP_CLOSE and MP_FAST_CLOSE to protect from unauthorized shutdown of MP-DCCP connections.

3. MP-DCCP Protocol

The DCCP protocol feature list ([RFC4340], Section 6.4) is updated by adding a new Multipath feature with Feature number 10, as shown in Table 1.

Number	Meaning	Rec'n Rule	Initial Value	Req'd
10	Multipath Capable	SP	0	N

Table 1: Multipath feature

Rec'n Rule: The reconciliation rule used for the feature. SP indicates the server-priority.

Initial Value: The initial value for the feature. Every feature has a known initial value.

Req'd: This column is "Y" if and only if every DCCP implementation MUST understand the feature. If it is "N", then the feature behaves like an extension, and it is safe to respond to Change options for the feature with empty Confirm options.

This specification adds a DCCP protocol option as defined in ([RFC4340], Section 5.8) providing a new Multipath related variable-length option with option type 46, as shown in Table 2.

Type	Option Length	Meaning	DCCP-Data?
46	variable	Multipath	Y

Table 2: Multipath option set

3.1. Multipath Capable Feature

A DCCP endpoint negotiates the Multipath Capable Feature to determine whether multipath extensions can be enabled for a DCCP connection.

The Multipath Capable feature (MP_CAPABLE) has feature number 10 and follows the structure for features given in [RFC4340] Section 6. Beside the negotiation of the feature itself, also one or several values can be exchanged. The value field specified here for the Multipath Capable feature has a length of one-byte and can be repeated several times within the DCCP option for feature negotiation if required for example to announce support of different

versions of the protocol. For that, the leftmost four bits in Figure 3 specify the compatible version of the MP-DCCP implementation and MUST be set to 0 following this specification. The four bits following the Version field are unassigned in version 0 and MUST be set to zero by the sender and MUST be ignored by the receiver.

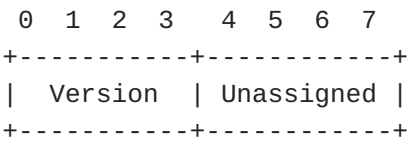

Figure 3: Format of the Multipath Capable feature value field

The setting of the MP_CAPABLE feature MUST follow the server-priority reconciliation rule described in ([RFC4340], Section 6.3.1). This allows multiple versions to be specified in order of priority.

The negotiation MUST be a part of the initial handshake procedure described in Section 3.3. No subsequent re-negotiation of the MP_CAPABLE feature is allowed for the same MP-DCCP connection.

Clients MUST include a Change R option during the initial handshake request to supply a list of supported MP-DCCP protocol versions, ordered by preference.

Servers MUST include a Confirm L option in the subsequent response to agree on an MP-DCCP version to be used from the Client list, followed by its own supported version(s), ordered by preference. Any subflow added to an existing MP-DCCP connection MUST use the version negotiated for the first subflow.

If no agreement is found, the Server MUST reply with an empty Confirm L option with feature number 10 and no values.

An example of successful version negotiation is shown hereafter and follows the negotiation example shown in [RFC4340] Section 6.5. For better understanding, this example uses the unspecified MP-DCCP versions 1 and 2 in addition to the MP-DCCP version 0 specified in this document:

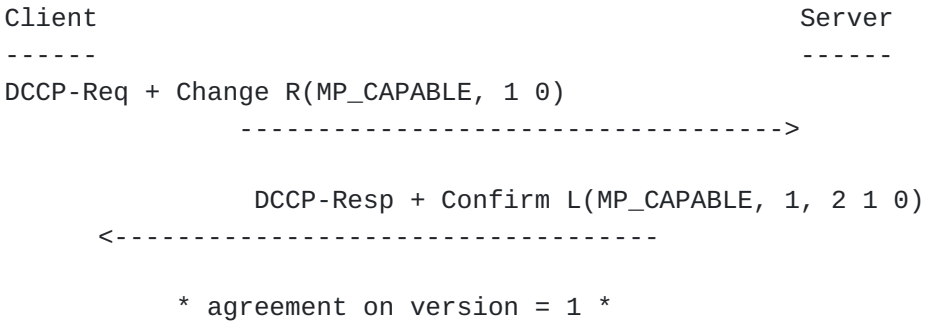

Figure 4: Example of MP-DCCP support negotiation using MP_CAPABLE

1. The Client indicates support for both MP-DCCP versions 1 and 0, with a preference for version 1.

2. Server agrees on using MP-DCCP version 1 indicated by the first value, and supplies its own preference list with the following values.

3. MP-DCCP is then enabled between the Client and Server with version 1.

Unlike the example in Figure 4, this document only allows the negotiation of MP-DCCP version 0, which means that client and server must support it.

If the version negotiation fails or the MP_CAPABLE feature is not present in the DCCP-Request or DCCP-Response packets of the initial handshake procedure, the MP-DCCP connection SHOULD fallback to regular DCCP or MUST close the connection. Further details are specified in Section 3.6

3.2. Multipath Option

MP-DCCP uses one single option to signal various multipath-related operations. The format of this multipath option is shown in Figure 5.

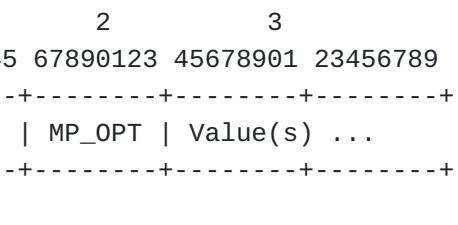

Figure 5: Multipath option format

The fields used by the the multipath option are described in Table 3. MP_OPT refers to an Multipath option.

Type	Option Length	MP_OPT	Meaning
46	var	0 =MP_CONFIRM	Confirm reception and processing of an MP_OPT option
46	12	1 =MP_JOIN	Join path to an existing MP-DCCP connection
46	var	2 =MP_FAST_CLOSE	Close an MP-DCCP connection unconditionally
46	var	3 =MP_KEY	Exchange key material for MP_HMAC
46	9	4 =MP_SEQ	Multipath Sequence Number
46	23	5 =MP_HMAC	HMA Code for authentication
46	12	6 =MP_RTT	Transmit RTT values
46	var	7 =MP_ADDADDR	Advertise additional Address
46	4	8 =MP_REMOVEADDR	Remove Address
46	4	9 =MP_PRIO	Change subflow Priority
46	var	10 =MP_CLOSE	Close an MP-DCCP subflow
46	var	11 =MP_EXP	Experimental for private use
46	TBD	>11	Reserved for future MP options.

Table 3: MP_OPT option types

Future MP options could be defined in a later version or extension to this specification.

These operations are largely inspired by the signals defined in [RFC8684].

3.2.1. MP_CONFIRM

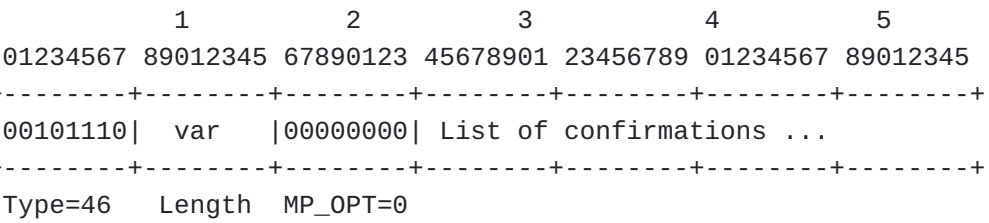

Figure 6: Format of the MP_CONFIRM option

Some multipath options require confirmation from the remote peer (see Table 4). Such options will be retransmitted by the sender until an MP_CONFIRM is received or confirmation of options is identified outdated. The further processing of the multipath options in the receiving host is not the subject of MP_CONFIRM.

Multipath options could arrive out-of-order, therefore multipath options defined in Table 4 MUST be sent in a DCCP datagram with MP_SEQ Section 3.2.5. This allows a receiver to identify whether

multipath options are associated with obsolete datasets (information carried in the option header) that would otherwise conflict with newer datasets. In the case of MP_ADDADDR or MP_REMOVEADDR the same dataset is identified based on AddressID, whereas the same dataset for MP_PRIO is identified by the subflow in use. An outdated multipath option is detected at the receiver if a previous multipath option referring to the same dataset contained a higher sequence number in the MP_SEQ. An MP_CONFIRM MAY be generated for multipath options that are identified as outdated.

Similarly an MP_CONFIRM could arrive out of order. The associated MP_SEQ received MUST be echoed to ensure that the most recent multipath option is confirmed. This protects from inconsistencies that could occur, e.g. if three MP_PRIO options are sent one after the other on one path in order to first set the path priority to 0, then to 1 and finally to 0 again. Without an associated MP_SEQ, a loss of the third MP_PRIO option and a loss of the MP_CONFIRM of the second update and the third update would cause the sender to incorrectly interpret that the priority value was set to 0 without recognizing that the receiver has applied priority value 1.

The length of the MP_CONFIRM option and the path over which the option is sent depend on the confirmed multipath options and the received MP_SEQ, which are both copied verbatim and appended as a list of confirmations. The list is structured by first listing the received MP_SEQ followed by the related multipath option or options to confirm. The same rules apply when multipath options with different MP_SEQs are confirmed at once. This could happen if a datagram with MP_PRIO and a first MP_SEQ_1 and another datagram with MP_ADDADDR and a second MP_SEQ_2 are received in short succession. In this case, the structure described above is concatenated resulting in MP_SEQ_2 + MP_ADDADDR + MP_SEQ_1 + MP_PRIO. The order of the confirmed multipath options in the list of confirmations MUST reflect the incoming order at the host who sends the MP_CONFIRM, with the most recent suboption received listed first. This could allow the host receiving the MP_CONFIRM to verify that the options were applied in the correct order and to take countermeasures if they were not, e.g., if an MP_REMOVEADDR overtakes an MP_ADDADDR that refers to the same dataset.

Type	Option Length	MP_OPT	MP_CONFIRM Sending path
46	var	7 =MP_ADDADDR	Any available
46	4	8 =MP_REMOVEADDR	Any available
46	4	9 =MP_PRIO	Any available

Table 4: Multipath options requiring confirmation

An example to illustrate the MP-DCCP confirm procedure for the MP_PRIO option is shown in Figure 7. The host A sends a DCCP-Request

on path A2-B2 with an MP_PRIO option with value 1 and associated sequence number of 1. Host B replies on the same path in this instance (any path can be used) with a DCCP-Response containing the MP_CONFIRM option and a list containing the original sequence number (1) together with the associated option (MP_PRIO).

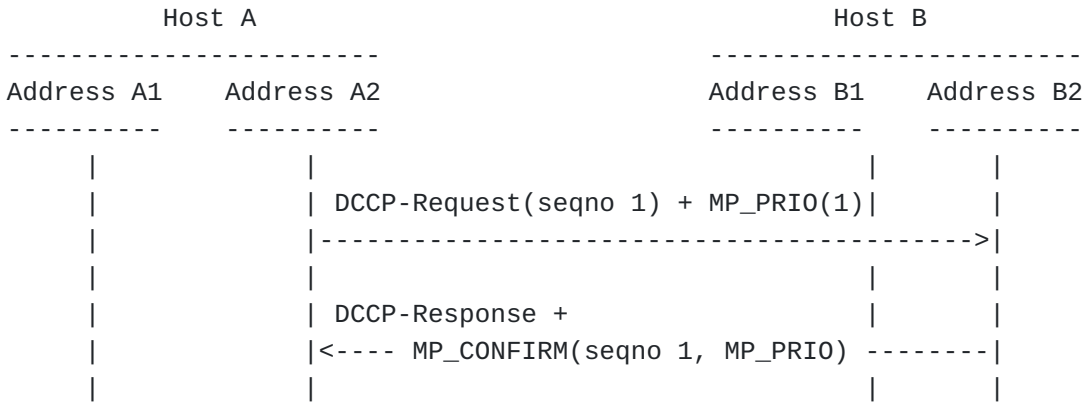

Figure 7: Example MP-DCCP CONFIRM procedure

A second example to illustrate the same MP-DCCP confirm procedure but where an out of date option is also delivered is shown in (Figure 8. Here, the first DCCP-Data is sent from Host A to Host B with option MP_PRIO set to 4. Host A subsequently sends the second DCCP-Data with option MP_PRIO set to 1. In this case, the delivery of the first MP_PRIO is delayed in the network between Host A and Host B and arrives after the second MP_PRIO. Host B ignores this second MP_PRIO as the associated sequence number is earlier than the first. Host B sends a DCCP-Ack confirming receipt of the MP_PRIO(1) with sequence number 2.

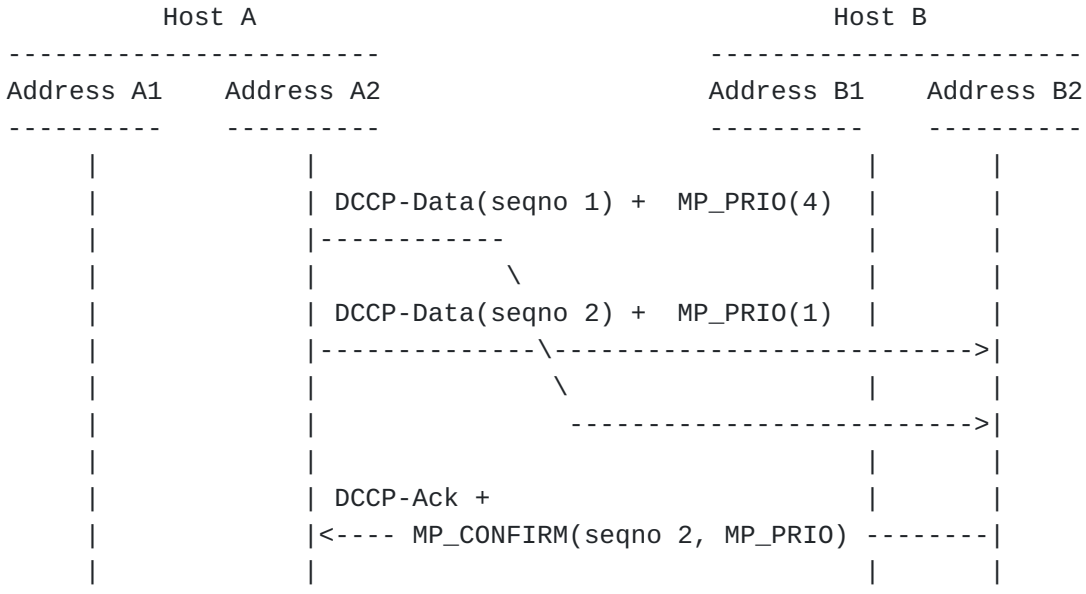

Figure 8: Example MP-DCCP CONFIRM procedure with outdated suboption

3.2.2. MP_JOIN

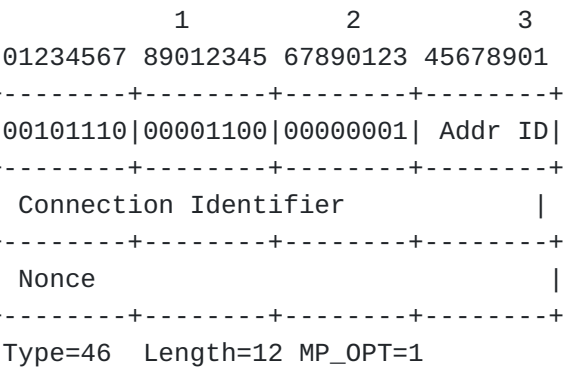

Figure 9: Format of the MP_JOIN suboption

The MP_JOIN option is used to add a new subflow to an existing MP-DCCP connection and REQUIRES a successful establishment of the first subflow using MP_KEY. The Connection Identifier (CI) is the one from the peer host, which was previously exchanged with the MP_KEY option. MP_HMAC MUST be set when using MP_JOIN within a DCCP-Response packet (See Section 3.2.6 for details).

The MP_JOIN option includes an "Addr ID" (Address ID) generated by the sender of the option, used to identify the source address of this packet, even if the IP header was changed in transit by a middlebox. The value of this field is generated by the sender and MUST map uniquely to a source IP address for the sending host. The Address ID allows address removal (Section 3.2.9) without needing to know what the source address at the receiver is, thus allowing address removal through NATs. The Address ID also allows correlation between new subflow setup attempts and address signaling (Section 3.2.8), to prevent setting up duplicate subflows on the same path, if an MP_JOIN and MP_ADDADDR are sent at the same time.

The Address IDs of the subflow used in the initial DCCP Request/Response exchange of the first subflow in the connection are implicit, and have the value zero. A host MUST store the mappings between Address IDs and addresses both for itself and the remote host. An implementation will also need to know which local and remote Address IDs are associated with which established subflows, for when addresses are removed from a local or remote host. An Address ID always MUST be unique over the lifetime of a subflow and can only be re-assigned if sender and receiver no longer have them in use.

The Nonce is a 32-bit random value locally generated for every MP_JOIN option. Together with the CI, the Nonce value builds the

basis to calculate the HMAC used in the handshaking process as described in Section 3.3.

If the CI cannot be verified by the receiving host during a handshake negotiation, the new subflow MUST be closed, as specified in Section 3.6.

3.2.3. MP_FAST_CLOSE

DCCP can send a Close or Reset signal to abruptly close a connection. Using MP-DCCP, a regular Close or Reset only has the scope of the subflow over which a signal was received. As such, it will only close the subflow and does not affect other remaining subflows or the MP-DCCP connection (unless it is the last subflow). This permits break-before-make handover between subflows.

In order to provide an MP-DCCP-level "reset" and thus allow the abrupt closure of the MP-DCCP connection, the MP_FAST_CLOSE suboption can be used.

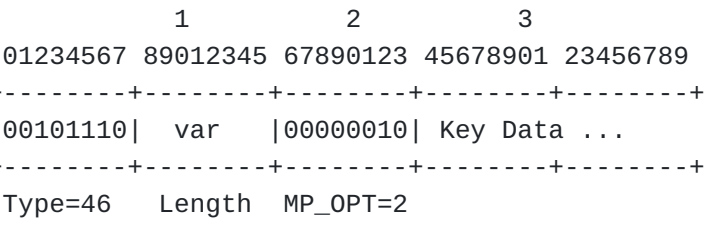

Figure 10: Format of the MP_FAST_CLOSE suboption

When host A wants to abruptly close an MP-DCCP connection with host B, it will send out the MP_FAST_CLOSE. The MP_FAST_CLOSE suboption MUST be sent from host A on all subflows using a DCCP-Reset packet with Reset Code 13. The requirement to send the MP_FAST_CLOSE on all subflows increases the probability that host B will receive the MP_FAST_CLOSE to take the same action. To protect from unauthorized shutdown of a MP-DCCP connection, the selected Key Data of the peer host during the handshaking procedure is carried by the MP_FAST_CLOSE option.

After sending the MP_FAST_CLOSE on all subflows, host A will tear down all subflows and the multipath DCCP connection immediately terminates.

Upon reception of the first MP_FAST_CLOSE with successfully validated Key Data, host B will send a DCCP Reset packet response on all subflows to host A with Reset Code 13 to clean potential middlebox states. Host B will then tear down all subflows and terminate the MP-DCCP connection.

3.2.4. MP_KEY

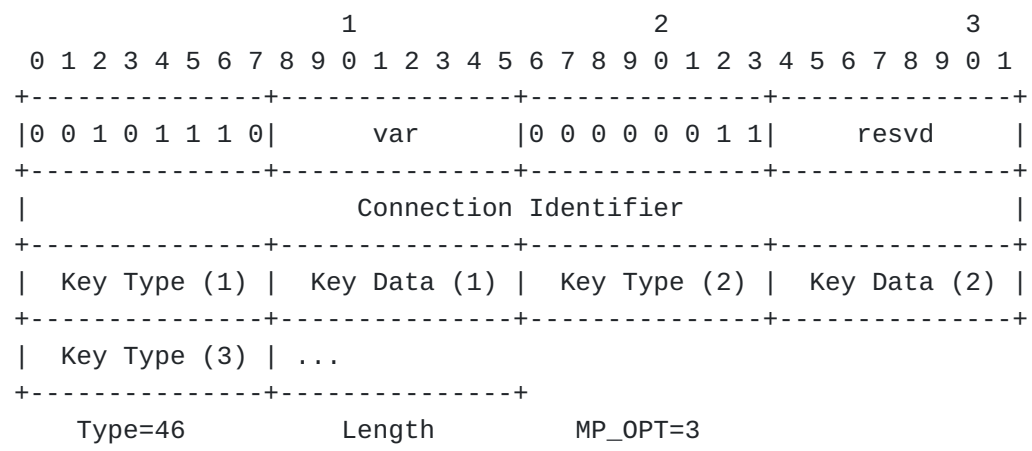

Figure 11: Format of the MP_KEY suboption

The MP_KEY suboption is used to exchange a Connection Identifier (CI) and key material between hosts for a given connection. The CI is a unique number that is configured per host during the initial exchange of a connection with MP_KEY and is necessary to connect other DCCP subflows to an MP-DCCP connection with MP_JOIN (Section 3.2.2). Its size of 32-bits also defines the maximum number of simultaneous MP-DCCP connections in a host to 2^32. According to the Key related elements of the MP_KEY suboption, the Length varies between 17 and 73 Bytes for a single-key message, and up to 115 Bytes when all specified Key Types 0-2 are provided. The Key Type field specifies the type of the following key data. The set of key types are shown in Table 5.

Key Type	Key Length (Bytes)	Meaning
0 =Plain Text	8	Plain Text Key
1 =ECDHE-C25519-SHA256	32	ECDHE with SHA256 and Curve25519
2 =ECDHE-C25519-SHA512	64	ECDHE with SHA512 and Curve25519
3-255		Unassigned

Table 5: MP_KEY key types

Plain Text

Key Material is exchanged in plain text between hosts, and the key parts (key-a, key-b) are used by each host to generate the derived key (d-key) by concatenating the two parts with the local key in front (e.g. hostA d-key(A)=(key-a+key-b), hostB d-key(B)=(key-b+key-a)).

ECDHE-SHA256-C25519

Public Key Material is exchanged via ECDHE key exchange with SHA256 and Curve 25519 to generate the derived key (d-key) from the shared secret. The full potential of ECDHE use is realized when it is combined with peer authentication technologies to protect against men-in-the-middle attacks. This can be achieved, for example, with separate use and verification of certificates issued by a certificate authority.

ECDHE-SHA512-C25519

Public Key Material is exchanged via ECDHE key exchange with SHA512 and Curve 25519 to generate the derived key (d-key) from the shared secret.

Multiple keys are only permitted in the DCCP-Request message of the handshake procedure for the first subflow. This allows the hosts to agree on a single key type to be used, as described in Section 3.3

It is possible that not all hosts will have all key types. If the key type cannot be agreed in the handshake procedure, the MP-DCCP connection MUST fall back to not using MP-DCCP, as indicated in Section 3.6.

3.2.5. MP_SEQ

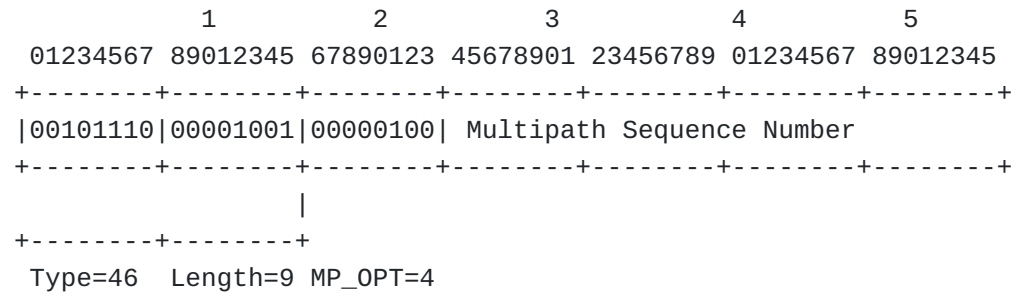

Figure 12: Format of the MP_SEQ suboption

The MP_SEQ suboption is used for end-to-end 48-bit datagram-based sequence numbers of an MP-DCCP connection. The initial data sequence number (IDSN) SHOULD be set randomly [RFC4086].

The MP_SEQ number space is independent from the path individual sequence number space and MUST be sent with all DCCP-Data and DCCP-DataACK packets.

When the sequence number space is exhausted, the sequence number MUST be wrapped. [RFC7323] provides guidance on selecting an appropriately sized sequence number space according to the maximum segment lifetime of TCP. 64 bits is the recommended size for TCP to avoid the sequence number space going through within the segment lifetime. For DCCP, the Maximum Segment Lifetime is the same as that

of TCP as specified in [RFC4340], Section 3.4. Compared to TCP, the sequence number for DCCP is incremented per packet rather than per byte transmitted. For this reason, the 48 bits chosen in MP_SEQ are considered sufficiently large.

3.2.6. MP_HMAC

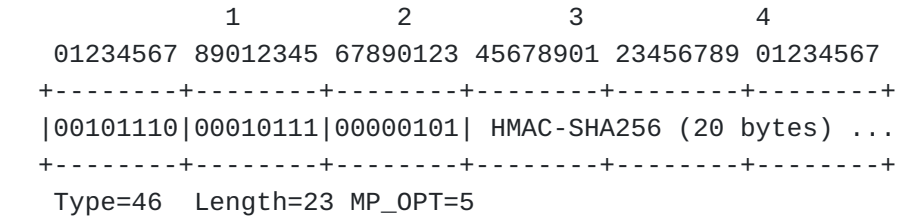

Figure 13: Format of the MP_HMAC suboption

The MP_HMAC suboption is used to provide authentication for the MP_ADDADDR, and MP_REMOVEADDR suboptions. In addition, it provides authentication for subflows joining an existing MP_DCCP connection, as described in the second and third step of the handshake of a subsequent subflow in Section 3.3. For this specification of MP-DCCP, the HMAC code is generated according to [RFC2104] in combination with the SHA256 hash algorithm described in [RFC6234], with the output truncated to the leftmost 160 bits (20 bytes).

The "Key" used for the HMAC computation is the derived key (d-key) described in Section 3.2.4, while the HMAC "Message" for MP_JOIN, MP_ADDADDR and MP_REMOVEADDR is a concatenation of:

- for MP_JOIN: The nonces of the MP_JOIN messages for which authentication shall be performed. Depending on whether Host A or Host B performs the HMAC-SHA256 calculation, it is carried out as follows: MP_HMAC(A) = HMAC-SHA256(Key=d-key(A), Msg=RA+RB) MP_HMAC(B) = HMAC-SHA256(Key=d-key(B), Msg=RB+RA) An usage example is shown in Figure 21.

- for MP_ADDADDR: The Address ID with associated IP address and if defined port, otherwise two octets of value 0. IP address and port MUST be used in network byte order (NBO). Depending on whether Host A or Host B performs the HMAC-SHA256 calculation, it is carried out as follows: MP_HMAC(A) = HMAC-SHA256(Key=d-key(A), Msg=Address ID+NBO(IP)+NBO(Port)) MP_HMAC(B) = HMAC-SHA256(Key=d-key(B), Msg=Address ID+NBO(IP)+NBO(Port))

- for MP_REMOVEADDR: Solely the Address ID. Depending on whether Host A or Host B performs the HMAC-SHA256 calculation, it is carried out as follows: MP_HMAC(A) = HMAC-SHA256(Key=d-key(A), Msg=Address ID) MP_HMAC(B) = HMAC-SHA256(Key=d-key(B), Msg=Address ID)

MP_JOIN, MP_ADDADDR and MP_REMOVEADDR can co-exist or be used multiple times within a single DCCP packet. All these multipath options require an individual MP_HMAC option. This ensures that the MP_HMAC is correctly associated. Otherwise, the receiver cannot validate multiple MP_JOIN, MP_ADDADDR or MP_REMOVEADDR. Therefore, a MP_HMAC MUST directly follow its associated multipath option. In the likely case of sending a MP_JOIN together with a MP_ADDADDR, this results in concatenating MP_JOIN + MP_HMAC_1 + MP_ADDADDR + MP_HMAC_2, whereas the first MP_HMAC_1 is associated with the MP_JOIN and the second MP_HMAC_2 is associated with the MP_ADDADDR suboption.

On the receiver side, the HMAC validation of the suboptions MUST be carried out according to the sending sequence in which the associated MP_HMAC follows a suboption. If the suboption cannot be validated by a receiving host because the HMAC validation fails, the subsequent handling depends on which suboption was being verified. If the suboption to be authenticated was either MP_ADDADDR or MP_REMOVEADDR, the receiving host MUST silently ignore it (see Section 3.2.8 and Section 3.2.9). If the suboption to be authenticated was MP_JOIN, the subflow MUST be closed (see Section 3.6). In the event that an MP_HMAC cannot be associated with a suboption, unless it is an MP_HMAC sent in DCCP-Ack in response to a DCCP-Response packet containing an MP_JOIN option, this MP_HMAC MUST be ignored.

3.2.7. MP_RTT

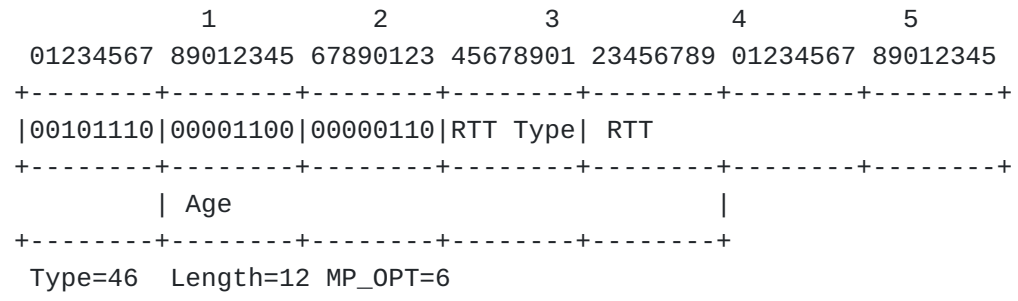

Figure 14: Format of the MP_RTT suboption

The MP_RTT suboption is used to transmit RTT values and age (represented in milliseconds) that belong to the path over which this information is transmitted. This information is useful for the receiving host to calculate the RTT difference between the subflows and to estimate whether missing data has been lost.

The RTT and Age information is a 32-bit integer. This covers a period of approximately 1193 hours.

The Field RTT type indicates the type of RTT estimation, according to the following description:

Raw RTT (=0)
Raw RTT value of the last Datagram Round-Trip

Min RTT (=1)
Min RTT value over a given period

Max RTT (=2)
Max RTT value over a given period

Smooth RTT (=3)
Averaged RTT value over a given period

Each CCID specifies the algorithms and period applied for their corresponding RTT estimations.The availability of the above described types, to be used in the MP_RTT option, depends on the CCID implementation in place.

Age
The Age parameter defines the time difference between now - creation of the MP_RTT option - and the conducted RTT measurement in milliseconds. If no previous measurement exists, e.g., when initialized, the value is 0.

An example of a flow showing the exchange of path individual RTT information is provided in Figure 15. RTT1 refers to the first path and RTT2 to the second path. The RTT values could be extracted from the sender's Congestion Control procedure and are conveyed to the receiving host using the MP_RTT suboption. With the reception of RTT1 and RTT2, the receiver is able to calculate the path_delta which corresponds to the absolute difference of both values. In the case that the path individual RTTs are symmetric in the down- and uplink directions and there is no jitter, packets with missing sequence number MP_SEQ, e.g., in a reordering process, can be assumed lost after path_delta/2.

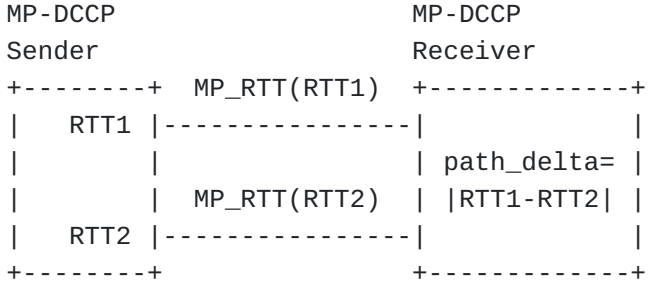

Figure 15: Exemplary flow of MP_RTT exchange and usage

3.2.8. MP_ADDADDR

The MP_ADDADDR suboption announces additional addresses (and, optionally, port numbers) by which a host can be reached. This can be sent at any time during an existing MP-DCCP connection, when the sender wishes to enable multiple paths and/or when additional paths become available. Multiple instances of this suboption within a packet can simultaneously advertise new addresses.

The Length is variable depending on the address family (IPv4 or IPv6) and whether a port number is used. This field is in range between 8 and 22 bytes.

The final 2 octets, optionally specify the DCCP port number to use, and their presence can be inferred from the length of the option. Although it is expected that the majority of use cases will use the same port pairs as used for the initial subflow (e.g., port 80 remains port 80 on all subflows, as does the ephemeral port at the client), there could be cases (such as port-based load balancing) where the explicit specification of a different port is required. If no port is specified, the receiving host MUST assume that any attempt to connect to the specified address uses the port already used by the subflow on which the MP_ADDADDR signal was sent.

Along with the MP_ADDADDR option an MP_HMAC option MUST be sent for authentication. The truncated HMAC parameter present in this MP_HMAC option is the leftmost 20 bytes of an HMAC, negotiated and calculated as described in Section 3.2.6. In the same way as for MP_JOIN, the key for the HMAC algorithm, in the case of the message transmitted by Host A, will be Key-A followed by Key-B, and in the case of Host B, Key-B followed by Key-A. These are the keys that were exchanged and selected in the original MP_KEY handshake. The message for the HMAC is the Address ID, IP address, and port number that precede the HMAC in the MP_ADDADDR option. If the port number is not present in the MP_ADDADDR option, the HMAC message will include 2 octets of value zero. The rationale for the HMAC is to prevent unauthorized entities from injecting MP_ADDADDR signals in an attempt to hijack a connection. Note that, additionally, the presence of this HMAC prevents the address from being changed in flight unless the key is known by an intermediary. If a host receives an MP_ADDADDR option for which it cannot validate the HMAC, it SHOULD silently ignore the option.

The presence of an MP_SEQ Section 3.2.5 MUST be ensured in a DCCP datagram in which MP_ADDADDR is sent, as described in Section 3.2.1.

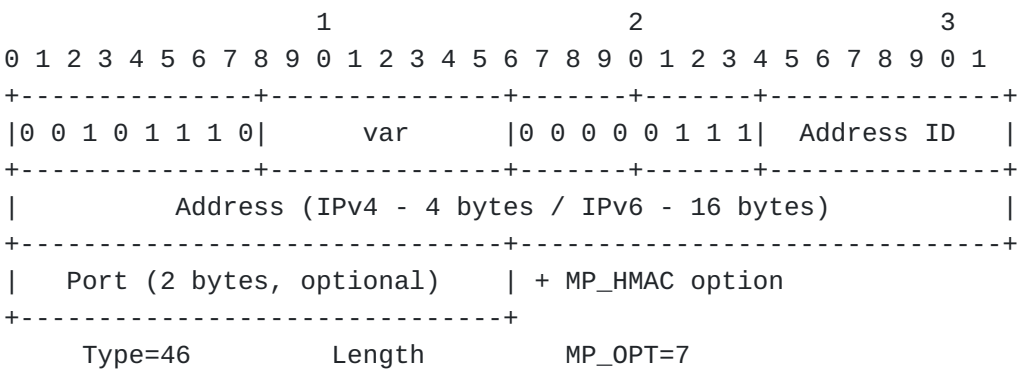

Figure 16: Format of the MP_ADDADDR suboption

Each address has an Address ID that could be used for uniquely identifying the address within a connection for address removal. Each host maintains a list of unique Address IDs and it manages these as it wishes. The Address ID is also used to identify MP_JOIN options (see Section 3.2.2) relating to the same address, even when address translators are in use. The Address ID MUST uniquely identify the address for the sender of the option (within the scope of the connection); the mechanism for allocating such IDs is implementation specific.

All Address IDs learned via either MP_JOIN or MP_ADDADDR can be stored by the receiver in a data structure that gathers all the Address-ID-to-address mappings for a connection (identified by a CI pair). In this way, there is a stored mapping between the Address ID, the observed source address, and the CI pair for future processing of control information for a connection. Note that an implementation MAY discard incoming address advertisements - for example, to avoid the required mapping state, or because advertised addresses are of no use to it (for example, IPv6 addresses when it has IPv4 only). Therefore, a host MUST treat address advertisements as soft state, and the sender MAY choose to refresh advertisements periodically.

A host MAY advertise private addresses, e.g., because there is a NAT on the path. It is desirable to allow this, since there could be cases where both hosts have additional interfaces on the same private network.

The MP_JOIN handshake to create a new subflow (Section 3.2.2) provides mechanisms to minimize security risks. The MP_JOIN message contains a 32-bit CI that uniquely identifies a connection to the receiving host. If the CI is unknown, the host MUST send a DCCP-Reset.

Further security considerations around the issue of MP_ADDADDR messages that accidentally misdirect, or maliciously direct, new

MP_JOIN attempts are discussed in Section 4. If a sending host of an MP_ADDADDR knows that no incoming subflows can be established at a particular address, an MP_ADDADDR SHOULD NOT announce that address unless the sending host has new knowledge about the possibility to do so. This information can be obtained from local firewall or routing settings, knowledge about availability of external NAT or firewall, or from connectivity checks performed by the host/ application.

The reception of an MP_ADDADDR message is acknowledged using MP_CONFIRM (Section 3.2.1). This ensures reliable exchange of address information.

A host MAY send an MP_ADDADDR message with an already assigned Address ID, but the Address MUST be the same as previously assigned to this Address ID, and the Port MUST be different from one already in use for this Address ID. If these conditions are not met, the receiver SHOULD silently ignore the MP_ADDADDR. A host wishing to replace an existing Address ID MUST first remove the existing one (Section 3.2.9).

A host that receives an MP_ADDADDR, but finds at connection set up that the IP address and port number is unsuccessful, SHOULD NOT perform further connection attempts to this address/port combination for this connection. However, a sender that wishes to trigger a new incoming connection attempt on a previously advertised address/port combination can therefore refresh the MP_ADDADDR information by sending the option again.

3.2.9. MP_REMOVEADDR

If, during the lifetime of an MP-DCCP connection, a previously announced address becomes invalid (e.g., if an interface disappears), the affected host SHOULD announce this. The peer can remove a previously added address with an Address ID from a connection using the Remove Address (MP_REMOVEADDR) suboption. This will terminate any subflows currently using that address.

Along with the MP_REMOVEADDR suboption a MP_HMAC option MUST be sent for authentication. The truncated HMAC parameter present in this MP_HMAC option is the leftmost 20 bytes of an HMAC, negotiated and calculated as described in Section 3.2.6. In the same way as for MP_JOIN, the key for the HMAC algorithm, in the case of the message transmitted by Host A, will be Key-A followed by Key-B, and in the case of Host B, Key-B followed by Key-A. These are the keys that were exchanged and selected in the original MP_KEY handshake. The message for the HMAC is the Address ID.

The rationale for using a HMAC is to prevent unauthorized entities from injecting MP_REMOVEADDR signals in an attempt to hijack a connection. Note that, additionally, the presence of this HMAC prevents the address from being modified in flight unless the key is known by an intermediary. If a host receives an MP_REMOVEADDR option for which it cannot validate the HMAC, it SHOULD silently ignore the option.

A receiver MUST include a MP_SEQ Section 3.2.5 in a DCCP datagram that sends an MP_REMOVEADDR. Further details are given in Section 3.2.1.

The reception of an MP_REMOVEADDR message is acknowledged using MP_CONFIRM (Section 3.2.1). This ensures reliable exchange of address information. To avoid inconsistent states, the sender releases the address ID only after MP_REMOVEADDR has been confirmed.

The sending and receiving of this message SHOULD trigger the closing procedure described in [RFC4340] between the client and the server, respectively on the affected subflow(s) (if possible). This helps remove middlebox state, before removing any local state.

Address removal is done by Address ID to allow the use of NATs and other middleboxes that rewrite source addresses. If there is no address at the requested Address ID, the receiver will silently ignore the request.

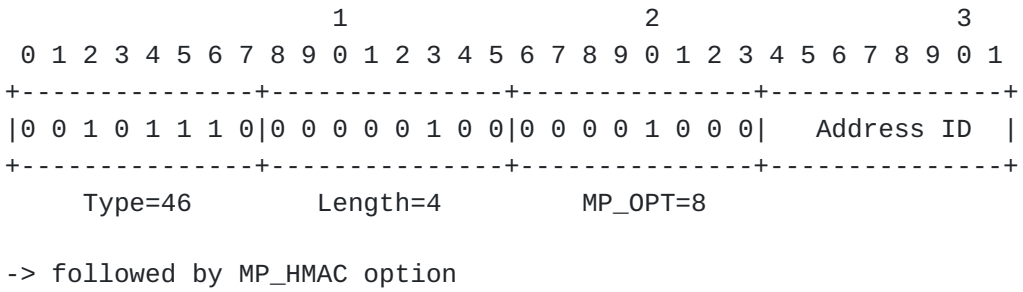

Figure 17: Format of the MP_REMOVEADDR suboption

A subflow that is still functioning MUST be closed with a DCCP-Close exchange as in regular DCCP, rather than using this option. For more information, see Section 3.5.

3.2.10. MP_PRIO

The path priority SHOULD be considered as hints for the packet scheduler when making decisions which path to use for payload traffic. When a single specific path from the set of available paths is treated with higher priority compared to the others when making scheduling decisions for payload traffic, a host can signal such

change in priority to the peer. This could be used when there are different costs for using different paths (e.g., WiFi is free while cellular has limit on volume, 5G has higher energy consumption). The priority of a path could also change, for example, when a mobile host runs out of battery, the usage of only a single path may be the preferred choice of the user.

The MP_PRIO suboption, shown below, can be used to set a priority flag for the subflow over which the suboption is received.

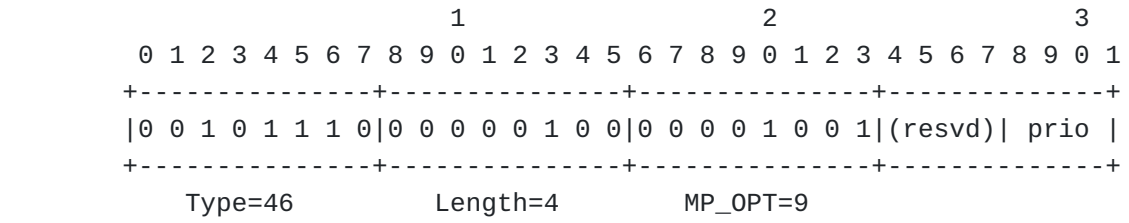

Figure 18: Format of the MP_PRIO suboption

The following values are available for the Prio field:

- *0: Do not use. The path is not available.

- *1: Standby: do not use this path for traffic scheduling, if another path (secondary or primary) is available. The path will only be used if other secondary or primary paths are not established.

- *2: Secondary: do not use this path for traffic scheduling, if the other paths are good enough. The path will be used occasionally for increasing temporarily the available capacity, e.g. when primary paths are congested or are not available. This is the recommended setting for paths that have costs or data caps as these paths will be used less frequently then primary paths.

- *3 - 15: Primary: The path can be used for packet scheduling decisions. The priority number indicates the relative priority of one path over the other for primary paths. Higher numbers indicate higher priority. The peer should consider sending traffic first over higher priority paths. This is the recommended setting for paths that do not have a cost or data caps associated with them as these paths will be frequently used.

Example use cases include: 1) Setting Wi-Fi path to Primary and Cellular paths to Secondary. In this case Wi-Fi will be used and Cellular will be used only if the Wi-Fi path is congested or not available. Such setting results in using the Cellular path only temporally, if more capacity is needed than the WiFi path can provide, indicating a clear priority of the Wi-Fi path over the

Cellular due to e.g. cost reasons. 2) Setting Wi-Fi path to Primary and Cellular to Standby. In this case Wi-Fi will be used and Cellular will be used only if the Wi-Fi path is not available. 3) Setting Wi-Fi path to Primary and Cellular path to Primary. In this case, both paths can be used when making packet scheduling decisions.

If not specified, the default behavior is to always use a path for packet scheduling decisions (MP_PRIO=3), when the path has been established and added to an existing MP-DCCP connection. At least one path ought to have a MP_PRIO value greater or equal to one for it to be allowed to send on the connection. It is RECOMMENDED to update at least one path to a non-zero MP_PRIO value when an MP-DCCP connection enters a state where all paths remain with an MP_PRIO value of zero. This helps an MP-DCCP connection to schedule when the multipath scheduler strictly respects MP_PRIO value 0. MP_PRIO is assumed to be exchanged reliably using the MP_CONFIRM mechanisms (see Table 4).

The relative ratio of the primary path values 3-15 depends on the path usage strategy, which is described in more detail in Section 3.11. In the case of path mobility Section 3.11.1, only one path can be used at a time and MUST be the appropriate one that has the highest available priority value including also the prio numbers 1 and 2. In the other case of concurrent path usage (Section 3.11.2), the definition is up to the multipath scheduler logic.

A MP_SEQ Section 3.2.5 MUST be present in a DCCP datagram in which MP_PRIO is sent. Further details are given in Section 3.2.1.

3.2.11. MP_CLOSE

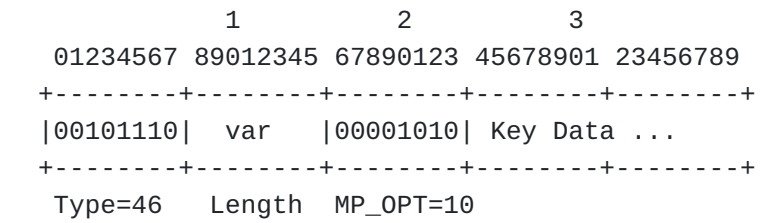

Figure 19: Format of the MP_CLOSE suboption

An MP-DCCP connection can be gracefully closed by sending and MP_CLOSE to the peer host. On all subflows, the regular termination procedure as described in [RFC4340] MUST be initiated using MP_CLOSE in the initial packet (either a DCCP-CloseReq or a DCCP-Close). When a DCCP-CloseReq is used, the following DCCP-Close MUST also carry the MP_CLOSE to avoid keeping a state in the sender of the DCCP-CloseReq. At the initiator of the DCCP-CloseReq, all sockets

including the MP-DCCP connection socket, transition to CLOSEREQ state. To protect from unauthorized shutdown of a multi-path connection, the selected Key Data of the peer host during the handshaking procedure MUST be included in by the MP_CLOSE option and must be validated by the peer host. Note, the Key Data is different between MP_CLOSE option carried by DCCP-CloseReq or DCCP-Close.

On reception of the first DCCP-CloseReq carrying a MP_CLOSE with valid Key Data, or due to a local decision, all subflows transition to the CLOSING state before transmitting a DCCP-Close carrying MP_CLOSE. The MP-DCCP connection socket on the host sending the DCCP-Close reflects the state of the initial subflow during handshake with MP_KEY option. If the initial subflow no longer exists, the state moves immediately to CLOSED.

Upon reception of the first DCCP-Close carrying a MP_CLOSE with valid Key Data at the peer host, all subflows, as well as the MP-DCCP connection socket, move to the CLOSED state. After this, a DCCP-Reset with Reset Code 1 MUST be sent on any subflow in response to a received DCCP-Close containing a valid MP_CLOSE option.

When the MP-DCCP connection socket is in CLOSEREQ or CLOSE state, new subflow requests using MP_JOIN MUST be ignored.

Contrary to a MP_FAST_CLOSE Section 3.2.3, no single-sided abrupt termination is applied.

3.2.12. Experimental Multipath option MP_EXP for private use

This section reserves a Multipath option to define and specify any experimental additional feature for improving and optimization of the MP-DCCP protocol. This option could be applicable to specific environments or scenarios according to potential new requirements and is meant for private use only. MP_OPT feature number 11 is specified with an exemplary description as below:

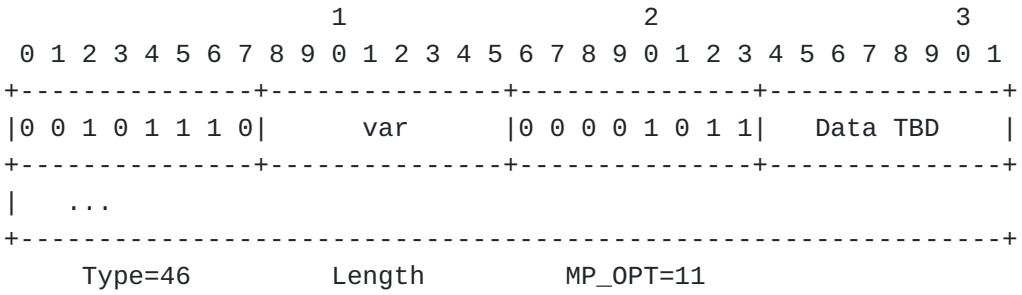

Figure 20: Format of the MP_EXP suboption

The Data field can carry any data according to the foreseen use by the experimenters with a maximum length of 252 Bytes.

3.3. MP-DCCP Handshaking Procedure

An example to illustrate the MP-DCCP handshake procedure is shown in Figure 21.

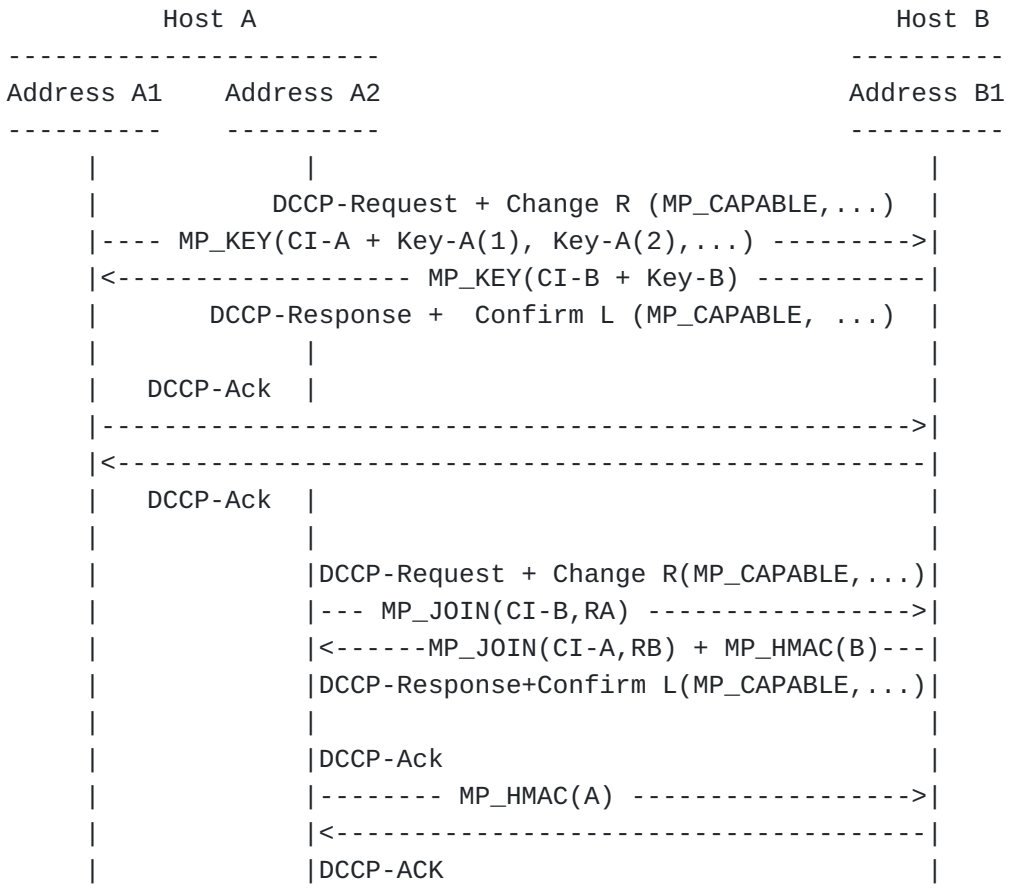

Figure 21: Example MP-DCCP handshake

The basic initial handshake for the first subflow is as follows:

- Host A sends a DCCP-Request with the MP-Capable feature Change request and the MP_KEY option with an Host-specific CI-A and a Key-A for each of the supported key types as described in Section 3.2.4. CI-A is a unique identifier during the lifetime of a MP-DCCP connection.

- Host B sends a DCCP-Response with Confirm feature for MP-Capable and the MP_Key option with a unique Host-specific CI-B and a single Host-specific Key-B. The type of the key is chosen from the list of supported types from the previous request.

- Host A sends a DCCP-Ack to confirm the proper key exchange.

- Host B sends a DCCP-Ack to complete the handshake and set both connection ends to the OPEN state.

It should be noted that DCCP is protected against corruption of DCCP header data (section 9 of [RFC4340]), so no additional mechanisms beyond the general confirmation are required to ensure that the header data has been properly received.

Host A waits for the final DCCP-Ack from host B before starting any establishment of additional subflow connections.

The handshake for subsequent subflows based on a successful initial handshake is as follows:

*Host A sends a DCCP-Request with the MP-Capable feature Change request and the MP_JOIN option with Host B's CI-B, obtained during the initial handshake. Additionally, an own random nonce RA is transmitted with the MP_JOIN.

*Host B computes the HMAC of the DCCP-Request and sends a DCCP-Response with Confirm feature option for MP-Capable and the MP_JOIN option with the CI-A and a random nonce RB together with the computed MP_HMAC. The HMAC is calculated by taking the leftmost 20 bytes from the SHA256 hash of a HMAC code created by using the nonce received with MP_JOIN(A) and the local nonce RB as message and the derived key described in Section 3.2.4 as key:

MP_HMAC(B) = HMAC-SHA256(Key=d-key(B), Msg=RB+RA)

*Host A sends a DCCP-Ack with the HMAC computed for the DCCP-Response. The HMAC is calculated by taking the leftmost 20 bytes from the SHA256 hash of a HMAC code created by using the local nonce RA and the nonce received with MP_JOIN(B) as message and the derived key described in Section 3.2.4 as key:

MP_HMAC(A) = HMAC-SHA256(Key=d-key(A), Msg=RA+RB)

*Host B sends a DCCP-Ack to confirm the HMAC and to conclude the handshaking.

3.4. Address knowledge exchange

3.4.1. Advertising a new path (MP_ADDADDR)

When a host (Host A) wants to advertise the availability of a new path, it should use the MP_ADDADDR option (Section 3.2.8) as shown in the example in Figure 22. The MP_ADDADDR option passed in the DCCP-Data contains the following parameters: * an identifier (id 2) for the new IP address which is used as a reference in subsequent control exchanges. * the IP address of the new path (A2_IP) * A pair of octets specifying the port number associated with this IP address. The value of 00 here indicates that the port number is the same as that used for the initial subflow address A1_IP

The following options MUST be included in a packet carrying MP_ADDADDR: * the leftmost 20 bytes of the HMAC(A) generated during the initial handshaking procedure described in Section 3.3 and Section 3.2.6 * the MP_SEQ option with the sequence number (seqno 12) for this message according to Section 3.2.5.

Host B acknowledges receipt of the MP_ADDADDR message with a DCCP-Ack containing the MP_CONFIRM option. The parameters supplied in this response are as follows: * an MP_CONFIRM containing the MP_SEQ number (seqno 12) of the packet carrying the option that we are confirming together with the MP_ADDADDR option * the leftmost 20 bytes of the HMAC(B) generated during the initial handshaking procedure Section 3.3

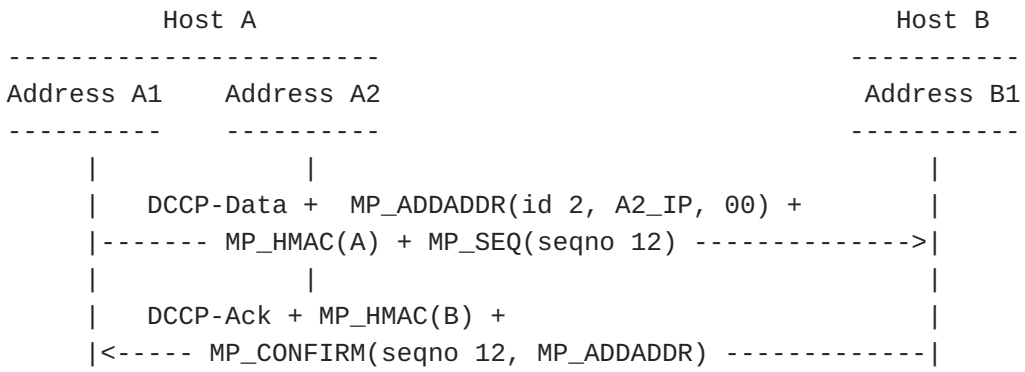

Figure 22: Example MP-DCCP ADDADDR procedure

3.4.2. Removing a path (MP_REMOVEADDR)

When a host (Host A) wants to indicate that a path is no longer available, it should use the MP_REMOVEADDR option (Section 3.2.9) as shown in the example in Figure 23. The MP_REMOVEADDR option passed in the DCCP-Data contains the following parameters: * an identifier (id 2) for the IP address to remove (A2_IP) and which was specified in a previous MP_ADDADDR message.

The following options must be included in a packet carrying MP_REMOVEADDR * the leftmost 20 bytes of the HMAC(A) generated during the initial handshaking procedure described in Section 3.3 and Section 3.2.6 * the MP_SEQ option with the sequence number (seqno 33) for this message according to Section 3.2.5.

Host B acknowledges receipt of the MP_REMOVEADDR message with a DCCP-Ack containing the MP_CONFIRM option. The parameters supplied in this response are as follows: * an MP_CONFIRM containing the MP_SEQ number (seqno 33) of the packet carrying the option that we are confirming, together with the MP_REMOVEADDR option * the leftmost 20 bytes of the HMAC(B) generated during the initial handshaking procedure Section 3.3

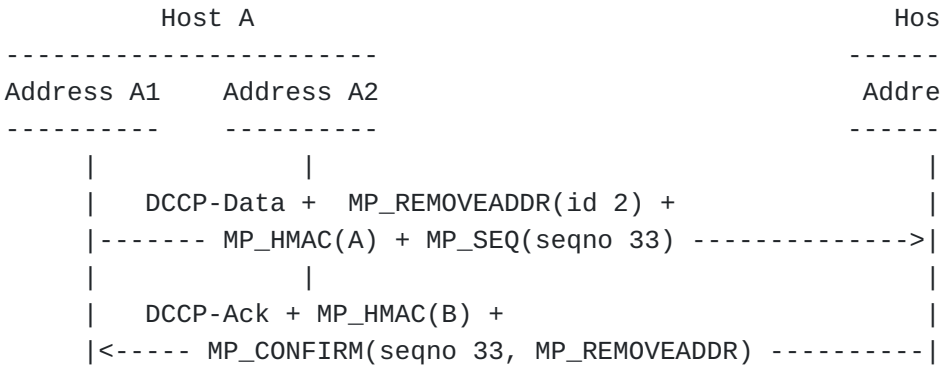

Figure 23: Example MP-DCCP REMOVEADDR procedure

3.5. Closing an MP-DCCP connection

When a host wants to close an existing subflow but not the whole MP-DCCP connection, it MUST initiate the regular DCCP connection termination procedure as described in Section 5.6 of [RFC4340], i.e., it sends a DCCP-Close/DCCP-Reset on the subflow. This may be preceded by a DCCP-CloseReq. In the event of an irregular termination of a subflow, e.g., during subflow establishment, it MUST use an appropriate DCCP reset code as specified in IANA [DCCP.Parameter] for DCCP operations. This could be, for example, sending reset code 5 (Option Error) when an MP-DCCP option provides invalid data or reset code 9 (Too Busy) when the maximum number of maintainable paths is reached. Note that receiving a reset code 9 for secondary subflows SHOULD NOT impact already existing active subflows. If necessary, these subflows are terminated in a subsequent step using the procedures described in this section.

A host terminates an MP-DCCP connection using the DCCP connection termination specified in section 5.5 of [RFC4340] on each subflow with the first packet on each subflow carrying MP_CLOSE (see Section 3.2.11).

```
Host A                                  Host B
------                                  ------
                                <-      Optional DCCP-CloseReq +
                                        MP_CLOSE [A's key]
                                        [on all subflows]
DCCP-Close + MP_CLOSE           ->
[B's key] [on all subflows]
                                <-      DCCP-Reset
                                        [on all subflows]
```

Additionally, an MP-DCCP connection may be closed abruptly using the "Fast Close" procedure described in Section 3.2.3, where a DCCP-

Reset is sent on all subflows, each carrying the MP_FAST_CLOSE option.

```
Host A                                  Host B
------                                  ------
DCCP-Reset + MP_FAST_CLOSE       ->
[B's key] [on all subflows]

                                 <-     DCCP-Reset
                                        [on all subflows]
```

3.6. Fallback

When a subflow fails to operate following MP-DCCP intended behavior, it is necessary to proceed with a fallback. This may be either falling back to regular DCCP [RFC4340] or removing a problematic subflow. The main reasons for subflow failing include: no MP support at peer host, failure to negotiate protocol version, loss of Multipath options, faulty/non-supported MP-DCCP options or modification of payload data.

At the start of an MP-DCCP connection, the handshake ensures exchange of MP-DCCP feature and options and thus ensures that the path is fully MP-DCCP capable. If during the handshake procedure it appears that DCCP-Request or DCCP-Response messages do not carry the MP_CAPABLE feature, the MP-DCCP connection will not be established and the handshake SHOULD fallback to regular DCCP (if this is not possible it MUST be closed).

A connection SHOULD fallback to regular DCCP if the endpoints fail to agree on a protocol version to use during the Multipath Capable feature negotiation. This is described in Section 3.1. The protocol version negotiation distinguishes between negotiation for the initial connection establishment, and addition of subsequent subflows. If protocol version negotiation is not successful during the initial connection establishment, MP-DCCP connection will fallback to regular DCCP.

The fallback procedure to regular DCCP MUST be also applied if the MP_KEY Section 3.2.4 Key Type cannot be negotiated.

If a subflow attempts to join an existing MP-DCCP connection, but MP-DCCP options or MP_CAPABLE feature are not present or are faulty in the handshake procedure, that subflow MUST be closed. This is especially the case if a different MP_CAPABLE version than the originally negotiated version is used. Reception of a non-verifiable MP_HMAC (Section 3.2.6) or an invalid CI used in MP_JOIN (Section 3.2.2) during flow establishment MUST cause the subflow to be closed.

The subflow closing procedure MUST be also applied if a final ACK carrying MP_KEY with wrong Key-A/Key-B is received or MP_KEY option is malformed.

Another relevant case is when payload data is modified by middleboxes. DCCP uses checksum to protect the data, as described in section 9 of [RFC4340]. A checksum will fail if the data has been changed in any way. All data from the start of the segment that failed the checksum onwards cannot be considered trustworthy. DCCP defines that if the checksum fails, the receiving endpoint MUST drop the application data and report that data as dropped due to corruption using a Data Dropped option (Drop Code 3, Corrupt). If data is dropped due to corruption for an MP-DCCP connection, the affected subflow MAY be closed.

3.7. State Diagram

The MP-DCCP per subflow state transitions to a large extent follow the state transitions defined for DCCP in [RFC4340], with some modifications due to the MP-DCCP four-way handshake and fast close procedures. The state diagram below illustrates the most common state transitions. The diagram is illustrative. For example, there are arcs (not shown) from several additional states to TIMEWAIT, contingent on the receipt of a valid DCCP-Reset.

The states transitioned when moving from the CLOSED to OPEN state during the four-way handshake remain the same as for DCCP, but it is no longer possible to transmit application data while in the REQUEST state. The fast close procedure can be triggered by either the client or the server and results in the transmission of a Reset packet. The fast close procedure moves the state of the client and server directly to TIMEWAIT and CLOSED, respectively.

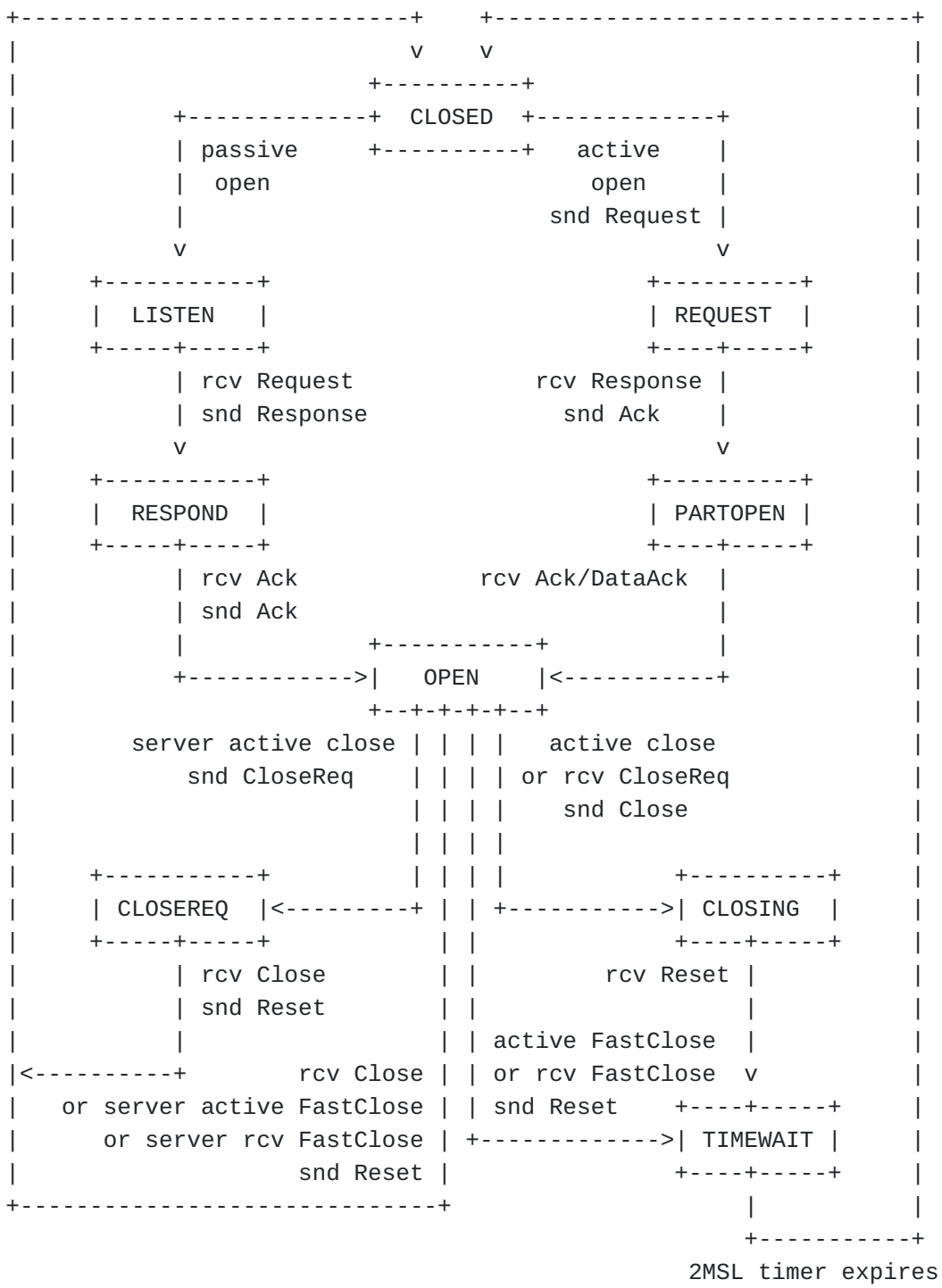

Figure 24: Most common state transitions of a MP-DCCP subflow

3.8. Congestion Control Considerations

Senders MUST manage per-path congestion status, and avoid to sending more data on a given path than congestion control for each path allows.

3.9. Maximum Packet Size Considerations

A DCCP implementation maintains the maximum packet size (MPS) during operation of a DCCP session. This procedure is specified for single-path DCCP in [RFC4340], Section 14. Without any restrictions, this is adopted for MP-DCCP operations, in particular the PMTU measurement and the Sender Behaviour. The DCCP application interface SHOULD allow the application to discover the current MPS. This reflects the current supported largest size for the data stream that can be used across the set of all active MP-DCCP subflows.

3.10. Maximum number of Subflows

In theory, an infinite number of subflows can be created within an MP-DCCP connection, as there is no element in the protocol that represents a restriction. In practical scenarios, however, there will be resource limitations on the host or use cases that do not benefit from additional subflows.

It is RECOMMENDED to limit the number of subflows in implementations and to reject incoming subflow requests with a DCCP-Reset using the Reset Code "too busy" as specified in [RFC4340] if this limit is exceeded.

3.11. Path usage strategies

MP-DCCP can be configured to realize one of several strategies for path usage, via selecting one DCCP subflow of the multiple DCCP subflows within a MP-DCCP connection for data transmission. This can be a dynamic process further facilitated by the means of DCCP and MP-DCCP defined options such as path preference using MP-PRIO, adding or removing DCCP subflows using MP_REMOVEADDR, MP_ADDADDR or DCCP-Close/DCCP-Reset and also path metrics such as packet-loss-rate, CWND or RTT provided by the Congestion Control Algorithm. Selecting an appropriate method can allow MP-DCCP to realize different path utilization strategies that make MP-DCCP suitable for end-to-end implementation over the Internet or in controlled environments such as Hybrid Access or 5G ATSSS.

3.11.1. Path mobility

The path mobility strategy provides the use of a single path with a seamless handover function to continue the connection when the currently used path is deemed unsuitable for service delivery. Some of the DCCP subflows of a MP-DCCP connection might become inactive due to either the occurrence of certain error conditions (e.g., DCCP timeout, packet loss threshold, RTT threshold, closed/removed) or adjustments from the MP-DCCP user. When there is outbound data to send and the primary path becomes inactive (e.g., due to failures) or de-prioritized, the MP-DCCP endpoint SHOULD try to send the data

through an alternate path with a different source or destination address (depending on the point of failure), if one exists. This process SHOULD respect the path priority configured by MP_PRIO or if not available pick the most divergent source-destination pair from the original used source-destination pair. Note: Rules for picking the most appropriate source-destination pair are an implementation decision and are not specified within this document. Path mobility is supported in the current Linux reference implementation [multipath-dccp.org].

3.11.2. Concurrent path usage

Different to a path mobility strategy, the selection between MP-DCCP subflows is a per-packet decision that is a part of the multipath scheduling process. This method would allow multiple subflows to be simultaneously used to aggregate the path resources to obtain higher connection throughput.

In this scenario, the selection of congestion control, per-packet scheduling and potential re-ordering method determines a concurrent path utilization strategy and result in a particular transport characteristic. A concurrent path usage method uses a scheduling design that could seek to maximize reliability, throughput, minimizing latency, etc.

Concurrent path usage over the Internet can have implications. When a Multipath DCCP connection uses two or more paths, there is no guarantee that these paths are fully disjoint. When two (or more) subflows share the same bottleneck, using a standard congestion control scheme could result in an unfair distribution of the capacity with the multipath connection using more capacity than competing single path connections.
Multipath TCP uses the coupled congestion control Linked Increases Algorithm (LIA) specified in the experimental specification [RFC6356] to solve this problem. This scheme could also be specified for Multipath DCCP. The same applies to other coupled congestion control schemes that have been proposed for Multipath TCP such as Opportunistic Linked Increases Algorithm [OLIA].

The specification of scheduling for concurrent multipath and related the congestion control algorithms and re-ordering methods for use in the general Internet are outside the scope of this document. If, and when, the IETF specifies a method for concurrent usage of multiple paths for the general Internet, the framework specified in this document could be used to provide an IETF recommended method for MP-DCCP.

4. Security Considerations

Similar to DCCP, MP-DCCP does not provide cryptographic security guarantees inherently. Thus, if applications need cryptographic security (integrity, authentication, confidentiality, access control, and anti-replay protection) the use of IPsec, DTLS over DCCP [RFC5238] or other end-to-end security is recommended; Secure Real-time Transport Protocol (SRTP) [RFC3711] is one candidate protocol for authentication. Together with Encryption of Header Extensions in SRTP, as provided by [RFC6904], also integrity would be provided.

DCCP [RFC4340] provides protection against hijacking and limits the potential impact of some denial-of-service attacks, but DCCP provides no inherent protection against an on-path attacker snooping on data packets. Regarding the security of MP-DCCP no additional risks should be introduced compared to regular DCCP. Thereof derived are the following key security requirements to be fulfilled by MP-DCCP:

- Provide a mechanism to confirm that parties involved in a subflow handshake are identical to those in the original connection setup.

- Provide verification that the new address to be included in a MP connection is valid for a peer to receive traffic at before using it.

- Provide replay protection, i.e., ensure that a request to add/remove a subflow is 'fresh'.

To achieve these goals, MP-DCCP includes a hash-based handshake algorithm documented in Sections Section 3.2.4, Section 3.2.6 and Section 3.3. The security of the MP-DCCP connection depends on the use of keys that are shared once at the start of the first subflow and are never sent again over the network. To ease demultiplexing while not revealing cryptographic material, subsequent subflows use the initially exchanged CI information. The keys exchanged once at the beginning are concatenated and used as keys for creating Hash-based Message Authentication Codes (HMACs) used on subflow setup, in order to verify that the parties in the handshake of subsequent subflows are the same as in the original connection setup. It also provides verification that the peer can receive traffic at this new address. Replay attacks would still be possible when only keys are used; therefore, the handshakes use single-use random numbers (nonces) at both ends -- this ensures that the HMAC will never be the same on two handshakes. Guidance on generating random numbers suitable for use as keys is given in [RFC4086]. During normal operation, regular DCCP protection mechanisms (such as header

checksum to protect DCCP headers against corruption) is designed to provide the same level of protection against attacks on individual DCCP subflows as exists for regular DCCP.

As discussed in Section 3.2.8, a host may advertise its private addresses, but these might point to different hosts in the receiver's network. The MP_JOIN handshake (Section 3.2.2) is designed to ensure that this does not set up a subflow to the incorrect host. However, it could still create unwanted DCCP handshake traffic. This feature of MP-DCCP could be a target for denial-of-service exploits, with malicious participants in MP-DCCP connections encouraging the recipient to target other hosts in the network. Therefore, implementations should consider heuristics at both the sender and receiver to reduce the impact of this.

As described in Section 3.9, a Maximum Packet Size (MPS) is maintained for a MP-DCCP connection. If MP-DCCP exposes a minimum MPS across all paths, any change to one path impacts the sender for all paths. To mitigate attacks that seek to force a low MPS, MP-DCCP could detect an attempt to reduce the MPS less than a minimum MPS, and then stop using these paths.

5. Interactions with Middleboxes

Issues from interaction with on-path middleboxes such as NATs, firewalls, proxies, intrusion detection systems (IDSs), and others have to be considered for all extensions to standard protocols since otherwise unexpected reactions of middleboxes may hinder its deployment. DCCP already provides means to mitigate the potential impact of middleboxes, also in comparison to TCP (see [RFC4043], Section 16). When both hosts are located behind a NAT or firewall entity, specific measures have to be applied such as the [RFC5596]-specified simultaneous-open technique that update the (traditionally asymmetric) connection-establishment procedures for DCCP. Further standardized technologies addressing middleboxes operating as NATs are provided in [RFC5597].

[RFC6773] specifies UDP Encapsulation for NAT Traversal of DCCP sessions, similar to other UDP encapsulations such as for SCTP [RFC6951]. Future specifications by the IETF could specify other methods for DCCP encapsulation.

The security impact of MP-DCCP aware middleboxes is discussed in Section 4

6. Implementation

The approach described above has been implemented in open source across different testbeds and a new scheduling algorithm has been

extensively tested. Also demonstrations of a laboratory setup have been executed and have been published at [multipath-dccp.org].

7. Acknowledgments

[RFC6824] and [RFC8684] defined Multipath TCP and provided important inputs for this specification.

The authors gratefully acknowledge significant input into this document from Dirk von Hugo, Nathalie Romo Moreno, Omar Nassef, Mohamed Boucadair, Simone Ferlin, Olivier Bonaventure, Gorry Fairhurst and Behcet Sarikaya.

8. IANA Considerations

This section provides guidance to the Internet Assigned Numbers Authority (IANA) regarding registration of values related to the MP extension of the DCCP protocol in accordance with [RFC8126]. This document defines one new value which is requested to be allocated in the IANA DCCP Feature Numbers registry and three new registries to be allocated in the DCCP registry group.

This document requests IANA to assign a new DCCP feature parameter for negotiating the support of multipath capability for DCCP sessions between hosts as described in Section 3. The following entry in Table 6 should be added to the Feature Numbers registry in the DCCP registry group according to [RFC4340], Section 19.4. under the "DCCP Protocol" heading.

Value	Feature Name	Specification
10 suggested	Multipath Capable	[ThisDocument]

Table 6: Addition to DCCP Feature Numbers registry

Sect. Section 3.1 specifies the new 1-Byte entry above includes a 4-bit part to specify the version of the used MP-DCCP implementation. This document requests IANA to create a new 'MP-DCCP Versions' registry within the DCCP registry group to track the MP-DCCP version. The initial content of this registry is as follows:

Version	Value	Specification
0	0000 suggested	[ThisDocument]
Unassigned	0001 - 1111	

Table 7: MP-DCCP Versions Registry

Future MP-DCCP versions 1 to 15 are assigned from this registry using the Specification Required policy (Section 4.6 of [RFC8126]).

This document requests IANA to assign value 46 in the DCCP "Option Types" registry to "Multipath Options", as described in Section 3.2.

IANA is requested to create a new 'Multipath Options' registry within the DCCP registry group. The following entries in Table 8 should be added to the new 'Multipath Options' registry. The registry in Table 8 has an upper boundary of 255 in the numeric value field.

Multipath Option	Name	Description	Reference
MP_OPT=0	MP_CONFIRM	Confirm reception/ processing of an MP_OPT option	Section 3.2.1
MP_OPT=1	MP_JOIN	Join subflow to existing MP-DCCP connection	Section 3.2.2
MP_OPT=2	MP_FAST_CLOSE	Close MP-DCCP connection	Section 3.2.3
MP_OPT=3	MP_KEY	Exchange key material for MP_HMAC	Section 3.2.4
MP_OPT=4	MP_SEQ	Multipath sequence number	Section 3.2.5
MP_OPT=5	MP_HMAC	Hash-based message auth. code for MP-DCCP	Section 3.2.6
MP_OPT=6	MP_RTT	Transmit RTT values and calculation parameters	Section 3.2.7
MP_OPT=7	MP_ADDADDR	Advertise additional address(es)/port(s)	Section 3.2.8
MP_OPT=8	MP_REMOVEADDR	Remove address(es)/ port(s)	Section 3.2.9
MP_OPT=9	MP_PRIO	Change subflow priority	Section 3.2.10
MP_OPT=10	MP_CLOSE	Close MP-DCCP subflow	Section 3.2.11
MP_OPT=11	MP_EXP	Experimental suboption for private use	Section 3.2.12
MP_OPT>11	Unassigned	Reserved for future Multipath Options	

Table 8: Multipath Options registry

Future Multipath options with MP_OPT>11 are assigned from this registry using the Specification Required policy (Section 4.6 of [RFC8126]).

In addition IANA is requested to assign a new DCCP Reset Code value 13 suggested in the DCCP Reset Codes Registry, with the short description "Abrupt MP termination". Use of this reset code is defined in section Section 3.2.3.

In addition IANA is requested to assign for this version of the MP-DCCP protocol a new 'Multipath Key Type' registry containing three

different suboptions to the MP_KEY option to identify the MP_KEY Key types in terms of 8-bit values as specified in Section 3.2.4 according to the entries in Table 9 below. Values in range 3-255 (decimal) inclusive remain unassigned in this here specified version 0 of the protocol and are assigned via Specification Required [RFC8126] in potential future versions of the MP-DCCP protocol.

Type	Name	Meaning	Reference
0	Plain Text	Plain text key	Section 3.2.4
1	ECDHE-C25519-SHA256	ECDHE with SHA256 and Curve25519	Section 3.2.4
2	ECDHE-C25519-SHA512	ECDHE with SHA512 and Curve25519	Section 3.2.4
3-255	Unassigned	Reserved for future use	Section 3.2.4

Table 9: Multipath Key Type registry with the MP_KEY Key Types for key data exchange on different paths

9. References

9.1. Normative References

[DCCP.Parameter] "IANA Datagram Congestion Control Protocol (DCCP) Parameters", n.d., <https://www.iana.org/assignments/dccp-parameters/dccp-parameters.xhtml>.

[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/rfc/rfc2119>.

[RFC4340] Kohler, E., Handley, M., and S. Floyd, "Datagram Congestion Control Protocol (DCCP)", RFC 4340, DOI 10.17487/RFC4340, March 2006, <https://www.rfc-editor.org/rfc/rfc4340>.

9.2. Informative References

[I-D.amend-iccrg-multipath-reordering] Amend, M. and D. Von Hugo, "Multipath sequence maintenance", Work in Progress, Internet-Draft, draft-amend-iccrg-multipath-reordering-03, 25 October 2021, <https://datatracker.ietf.org/doc/html/draft-amend-iccrg-multipath-reordering-03>.

[I-D.amend-tsvwg-multipath-framework-mpdccp] Amend, M., Bogenfeld, E., Brunstrom, A., Kassler, A., and V. Rakocevic, "A multipath framework for UDP traffic over heterogeneous access networks", Work in Progress,

Internet-Draft, draft-amend-tsvwg-multipath-framework-mpdccp-01, 8 July 2019, <https://datatracker.ietf.org/doc/html/draft-amend-tsvwg-multipath-framework-mpdccp-01>.

[I-D.ietf-quic-multipath] Liu, Y., Ma, Y., De Coninck, Q., Bonaventure, O., Huitema, C., and M. Kuehlewind, "Multipath Extension for QUIC", Work in Progress, Internet-Draft, draft-ietf-quic-multipath-06, 23 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-quic-multipath-06>.

[I-D.lhwxz-hybrid-access-network-architecture] Leymann, N., Heidemann, C., Cullen, M., Xue, L., and M. Zhang, "Hybrid Access Network Architecture", Work in Progress, Internet-Draft, draft-lhwxz-hybrid-access-network-architecture-02, 13 January 2015, <https://datatracker.ietf.org/doc/html/draft-lhwxz-hybrid-access-network-architecture-02>.

[I-D.muley-network-based-bonding-hybrid-access] Muley, P., Henderickx, W., Geng, L., Liu, H., Cardullo, L., Newton, J., Seo, S., Draznin, S., and B. Patil, "Network based Bonding solution for Hybrid Access", Work in Progress, Internet-Draft, draft-muley-network-based-bonding-hybrid-access-03, 22 October 2018, <https://datatracker.ietf.org/doc/html/draft-muley-network-based-bonding-hybrid-access-03>.

[IETF115.Slides] Amend, M., "MP-DCCP for enabling transfer of UDP/IP traffic over multiple data paths in multi-connectivity networks", IETF105 , n.d., <https://datatracker.ietf.org/meeting/105/materials/slides-105-tsvwg-sessa-62-dccp-extensions-for-multipath-operation-00>.

[MP-DCCP.Paper] Amend, M., Bogenfeld, E., Cvjetkovic, M., Rakocevic, V., Pieska, M., Kassler, A., and A. Brunstrom, "A Framework for Multiaccess Support for Unreliable Internet Traffic using Multipath DCCP", DOI 10.1109/LCN44214.2019.8990746, October 2019, <https://doi.org/10.1109/LCN44214.2019.8990746>.

[multipath-dccp.org] "Multipath extension for DCCP", n.d., <https://multipath-dccp.org/>.

[OLIA] Khalili, R., Gast, N., Popovic, M., Upadhyay, U., and J. Le Boudec, "MPTCP is not pareto-optimal: performance issues and a possible solution", Proceedings of the 8th

international conference on Emerging networking experiments and technologies, ACM , 2012.

[RFC0793] Postel, J., "Transmission Control Protocol", RFC 793, DOI 10.17487/RFC0793, September 1981, <https://www.rfc-editor.org/rfc/rfc793>.

[RFC2104] Krawczyk, H., Bellare, M., and R. Canetti, "HMAC: Keyed-Hashing for Message Authentication", RFC 2104, DOI 10.17487/RFC2104, February 1997, <https://www.rfc-editor.org/rfc/rfc2104>.

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, DOI 10.17487/RFC3711, March 2004, <https://www.rfc-editor.org/rfc/rfc3711>.

[RFC4043] Pinkas, D. and T. Gindin, "Internet X.509 Public Key Infrastructure Permanent Identifier", RFC 4043, DOI 10.17487/RFC4043, May 2005, <https://www.rfc-editor.org/rfc/rfc4043>.

[RFC4086] Eastlake 3rd, D., Schiller, J., and S. Crocker, "Randomness Requirements for Security", BCP 106, RFC 4086, DOI 10.17487/RFC4086, June 2005, <https://www.rfc-editor.org/rfc/rfc4086>.

[RFC5238] Phelan, T., "Datagram Transport Layer Security (DTLS) over the Datagram Congestion Control Protocol (DCCP)", RFC 5238, DOI 10.17487/RFC5238, May 2008, <https://www.rfc-editor.org/rfc/rfc5238>.

[RFC5595] Fairhurst, G., "The Datagram Congestion Control Protocol (DCCP) Service Codes", RFC 5595, DOI 10.17487/RFC5595, September 2009, <https://www.rfc-editor.org/rfc/rfc5595>.

[RFC5596] Fairhurst, G., "Datagram Congestion Control Protocol (DCCP) Simultaneous-Open Technique to Facilitate NAT/Middlebox Traversal", RFC 5596, DOI 10.17487/RFC5596, September 2009, <https://www.rfc-editor.org/rfc/rfc5596>.

[RFC5597] Denis-Courmont, R., "Network Address Translation (NAT) Behavioral Requirements for the Datagram Congestion Control Protocol", BCP 150, RFC 5597, DOI 10.17487/RFC5597, September 2009, <https://www.rfc-editor.org/rfc/rfc5597>.

[RFC6234] Eastlake 3rd, D. and T. Hansen, "US Secure Hash Algorithms (SHA and SHA-based HMAC and HKDF)", RFC 6234,

DOI 10.17487/RFC6234, May 2011, <https://www.rfc-editor.org/rfc/rfc6234>.

[RFC6356] Raiciu, C., Handley, M., and D. Wischik, "Coupled Congestion Control for Multipath Transport Protocols", RFC 6356, DOI 10.17487/RFC6356, October 2011, <https://www.rfc-editor.org/rfc/rfc6356>.

[RFC6773] Phelan, T., Fairhurst, G., and C. Perkins, "DCCP-UDP: A Datagram Congestion Control Protocol UDP Encapsulation for NAT Traversal", RFC 6773, DOI 10.17487/RFC6773, November 2012, <https://www.rfc-editor.org/rfc/rfc6773>.

[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, <https://www.rfc-editor.org/rfc/rfc6824>.

[RFC6904] Lennox, J., "Encryption of Header Extensions in the Secure Real-time Transport Protocol (SRTP)", RFC 6904, DOI 10.17487/RFC6904, April 2013, <https://www.rfc-editor.org/rfc/rfc6904>.

[RFC6951] Tuexen, M. and R. Stewart, "UDP Encapsulation of Stream Control Transmission Protocol (SCTP) Packets for End-Host to End-Host Communication", RFC 6951, DOI 10.17487/RFC6951, May 2013, <https://www.rfc-editor.org/rfc/rfc6951>.

[RFC7323] Borman, D., Braden, B., Jacobson, V., and R. Scheffenegger, Ed., "TCP Extensions for High Performance", RFC 7323, DOI 10.17487/RFC7323, September 2014, <https://www.rfc-editor.org/rfc/rfc7323>.

[RFC8041] Bonaventure, O., Paasch, C., and G. Detal, "Use Cases and Operational Experience with Multipath TCP", RFC 8041, DOI 10.17487/RFC8041, January 2017, <https://www.rfc-editor.org/rfc/rfc8041>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/rfc/rfc8126>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/rfc/rfc8174>.

[RFC8684] Ford, A., Raiciu, C., Handley, M., Bonaventure, O., and C. Paasch, "TCP Extensions for Multipath Operation with

Multiple Addresses", RFC 8684, DOI 10.17487/RFC8684, March 2020, <https://www.rfc-editor.org/rfc/rfc8684>.

[TS23.501] 3GPP, "System architecture for the 5G System; Stage 2; Release 16", December 2020, <https://www.3gpp.org/ftp//Specs/archive/23_series/23.501/23501-g70.zip>.

Appendix A. Differences from Multipath TCP

This appendix is Informative.

Multipath DCCP is similar to Multipath TCP [RFC8684], in that it extends the related basic DCCP transport protocol [RFC4340] with multipath capabilities in the same way as Multipath TCP extends TCP [RFC0793]. However, because of the differences between the underlying TCP and DCCP protocols, the transport characteristics of MPTCP and MP-DCCP are different.

Table 10 compares the protocol characteristics of TCP and DCCP, which are by nature inherited by their respective multipath extensions. A major difference lies in the delivery of payload, which is for TCP an exact copy of the generated byte-stream. DCCP behaves in a different way and does not guarantee to deliver any payload nor the order of delivery. Since this is mainly affecting the receiving endpoint of a TCP or DCCP communication, many similarities on the sender side can be identified. Both transport protocols share the 3-way initiation of a communication and both employ congestion control to adapt the sending rate to the path characteristics.

Feature	TCP	DCCP
Full-Duplex	yes	yes
Connection-Oriented	yes	yes
Header option space	40 bytes	< 1008 bytes or PMTU
Data transfer	reliable	unreliable
Packet-loss handling	re-transmission	report only
Ordered data delivery	yes	no
Sequence numbers	one per byte	one per PDU
Flow control	yes	no
Congestion control	yes	yes
ECN support	yes	yes
Selective ACK	yes	depends on congestion control
Fix message boundaries	no	yes
Path MTU discovery	yes	yes
Fragmentation	yes	no

Feature	TCP	DCCP
SYN flood protection	yes	no
Half-open connections	yes	no

Table 10: TCP and DCCP protocol comparison

Consequently, the multipath features, shown in Table 11, are the same, supporting volatile paths having varying capacity and latency, session handover and path aggregation capabilities. All of them profit by the existence of congestion control.

Feature	MPTCP	MP-DCCP
Volatile paths	yes	yes
Session handover	yes	yes
Path aggregation	yes	yes
Data reordering	yes	optional
Expandability	limited by TCP header	flexible

Table 11: MPTCP and MP-DCCP protocol comparison

Therefore, the sender logic is not much different between MP-DCCP and MPTCP.

The receiver side for MP-DCCP has to deal with the unreliable delivery provided by DCCP. The multipath sequence numbers included in MP-DCCP (see Section 3.2.5) facilitates adding optional mechanisms for data stream packet reordering at the receiver. Information from the MP_RTT multipath option (Section 3.2.7), DCCP path sequencing and the DCCP Timestamp Option provide further means for advanced reordering approaches, e.g., as proposed in [I-D.amend-iccrg-multipath-reordering]. Such mechanisms do, however, not affect interoperability and are not part of the MP-DCCP protocol. Many applications that use unreliable transport protocols can also inherently process out-of-sequence data (e.g., through adaptive audio and video buffers), and so additional reordering support might not be necessary. The addition of optional reordering mechanisms are likely to be needed when the different DCCP subflows are routed across paths with different latencies. In theory, applications using DCCP are aware that packet reordering could occur, because DCCP does not provide mechanisms to restore the original packet order.

In contrast to TCP, the receiver processing for MPTCP adopted a rigid "just wait" approach, because TCP guarantees reliable in-order delivery.

Authors' Addresses

Markus Amend (editor)
Deutsche Telekom

Deutsche-Telekom-Allee 9
64295 Darmstadt
Germany

Email: Markus.Amend@telekom.de

Anna Brunstrom
Karlstad University
Universitetsgatan 2
SE-651 88 Karlstad
Sweden

Email: anna.brunstrom@kau.se

Andreas Kassler
Karlstad University
Universitetsgatan 2
SE-651 88 Karlstad
Sweden

Email: andreas.kassler@kau.se

Veselin Rakocevic
City, University of London
Northampton Square
London
United Kingdom

Email: veselin.rakocevic.1@city.ac.uk

Stephen Johnson
BT
Adastral Park
Martlesham Heath
IP5 3RE
United Kingdom

Email: stephen.h.johnson@bt.com