

Workgroup: QUIC Working Group  
Internet-Draft: draft-lmbdhk-quic-multipath-00  
Published: 25 October 2021  
Intended Status: Standards Track  
Expires: 28 April 2022

A            Y. Liu                    Y. Ma                    Q. De Coninck  
             uAlibaba Inc.    Alibaba Inc.    UCLouvain  
             t  
             h  
             o  
             r  
             s  
             :  
             O. Bonaventure    C. Huitema  
             UCLouvain            Private Octopus Inc.  
             M. Kuehlewind, Ed.  
             Ericsson

## Multipath Extension for QUIC

### Abstract

This document specifies a multipath extension for the QUIC protocol to enable the simultaneous usage of multiple paths for a single connection.

### Discussion Venues

This note is to be removed before publishing as an RFC.

Discussion of this document takes place on the QUIC Working Group mailing list ([quic@ietf.org](mailto:quic@ietf.org)), which is archived at <https://mailarchive.ietf.org/arch/browse/quic/>.

Source for this draft and an issue tracker can be found at <https://github.com/mirjak/draft-lmbdhk-quic-multipath>.

### 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 28 April 2022.

## Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

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

## Table of Contents

- [1. Introduction](#)
  - [1.1. Conventions and Definitions](#)
- [2. Handshake Negotiation and Transport Parameter](#)
- [3. Path Setup and Removal](#)
  - [3.1. Path Initiation](#)
  - [3.2. Path Close](#)
    - [3.2.1. Use PATH\\_ABANDON Frame to Close a Path](#)
    - [3.2.2. Effect of RETIRE\\_CONNECTION\\_ID Frame](#)
    - [3.2.3. Idle Timeout](#)
  - [3.3. Path States](#)
- [4. Congestion Control](#)
- [5. Computing Path RTT](#)
- [6. Packet Scheduling](#)
- [7. Packet Number Space and Use of Connection ID](#)
  - [7.1. Using One Packet Number Space](#)
    - [7.1.1. Sending Acknowledgements and Handling Ranges](#)
  - [7.2. Using Multiple Packet Number Spaces](#)
    - [7.2.1. Packet Protection for QUIC Multipath](#)
    - [7.2.2. Key Update for QUIC Multipath](#)
- [8. Examples](#)
  - [8.1. Path Establishment](#)
  - [8.2. Path Closure](#)
- [9. Implementation Considerations](#)
- [10. New Frames](#)
  - [10.1. PATH\\_ABANDON Frame](#)
  - [10.2. ACK\\_MP Frame](#)
- [11. Error Codes](#)
- [12. IANA Considerations](#)
- [13. Security Considerations](#)
- [14. Contributors](#)
- [15. Acknowledgments](#)
- [16. References](#)
  - [16.1. Normative References](#)
  - [16.2. Informative References](#)
- [Authors' Addresses](#)

## 1. Introduction

This document specifies an extension to QUIC v1 [[QUIC-TRANSPORT](#)] to enable the simultaneous usage of multiple paths for a single connection.

This proposal is based on several basic design points:

- \*Re-use as much as possible mechanisms of QUIC-v1. In particular this proposal uses path validation as specified for QUIC v1 and aims to re-use as much as possible of QUIC's connection migration.
- \*Use the same packet header formats as QUIC v1 to avoid the risk of packets being dropped by middleboxes (which may only support QUIC v1)
- \*Congestion Control, RTT measurements and PMTU discovery should be per-path (following [[QUIC-TRANSPORT](#)])
- \*A path is determined by the 4-tuple of source and destination IP address as well as source and destination port. Therefore there can be at most one active paths/connection ID per 4-tuple.

The path management specified in section 9 of [[QUIC-TRANSPORT](#)] fulfills multiple goals: it directs a peer to switch sending through a new preferred path, and it allows the peer to release resources associated with the old path. Multipath requires several changes to that mechanism:

- \*Allow simultaneous transmission of non probing frames on multiple paths.
- \*Continue using an existing path even if non-probing frames have been received on another path.
- \*Manage the removal of paths that have been abandoned.

As such this extension specifies a departure from the specification of path management in section 9 of [[QUIC-TRANSPORT](#)] and therefore requires negotiation between the two endpoints using a new transport parameter, as specified in [Section 2](#).

This proposal supports the negotiation of either the use of one packet number space for all paths or the use of separate packet number spaces per path. While separate packet number spaces allow for more efficient ACK encoding, especially when paths have highly different latencies, this approach requires the use of a connection ID. Therefore use of a single number space can be beneficial in highly constrained networks that do not benefit from exposing the connection ID in the header. While both approaches are supported by the specification in this version of the document, the intention for the final publication of a multipath extension for QUIC is to choose one option in order to avoid incompatibility. More evaluation and implementation experience is needed to select one approach before final publication. Some discussion about pros and cons can be found here: [https://github.com/mirjak/draft-lmbdtk-quic-multipath/blob/master/presentations/PacketNumberSpace\\_s.pdf](https://github.com/mirjak/draft-lmbdtk-quic-multipath/blob/master/presentations/PacketNumberSpace_s.pdf)

As currently defined in this version of the draft the use of multiple packet number spaces requires the use of connection IDs in both directions. Today's deployments often only use destination connection ID when sending packets from the client to the server as this addresses the most important use cases for migration, like NAT

rebinding or mobility events. Further discussion and work is required to evaluate if the use of multiple packet number spaces could be supported as well when the connection ID is only present in one direction.

This proposal does not cover address discovery and management. Addresses and the actual decision process to setup or tear down paths are assumed to be handled by the application that is using the QUIC multipath extension. Further, this proposal only specifies a simple basic packet scheduling algorithm in order to provide some basic implementation guidance. However, more advanced algorithms as well as potential extensions to enhance signaling of the current path state are expected as future work.

### 1.1. Conventions and Definitions

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.

We assume that the reader is familiar with the terminology used in [[QUIC-TRANSPORT](#)]. In addition, we define the following terms:

\*Path Identifier (Path ID): An identifier that is used to identify a path in a QUIC connection at an endpoint. Path Identifier is used in multi-path control frames (etc. PATH\_ABANDON frame) to identify a path. By default, it is defined as the sequence number of the destination Connection ID used for sending packets on that particular path, but alternative definitions can be used if the length of that connection ID is zero.

\*Packet Number Space Identifier (PN Space ID): An identifier that is used to distinguish packet number spaces for different paths. It is used in 1-RTT packets and ACK\_MP frames. Each node maintains a list of "Received Packets" for each of the CID that it provided to the peer, which is used for acknowledging packets received with that CID.

The difference between Path Identifier and Packet Number Space Identifier, is that the Path Identifier is used in multipath control frames to identify a path, and the Packet Number Space Identifier is used in 1-RTT packets and ACK\_MP frames to distinguish packet number spaces for different paths. Both identifiers have the same value, which is the sequence number of the connection ID, if a non-zero connection ID is used. If the connection ID is zero length, the Packet Number Space Identifier is 0, while the Path Identifier is selected on path establishment.

## 2. Handshake Negotiation and Transport Parameter

This extension defines a new transport parameter, used to negotiate the use of the multipath extension during the connection handshake, as specified in [[QUIC-TRANSPORT](#)]. The new transport parameter is defined as follow:

\*name: enable\_multipath (TBD - experiments use 0xbabf)

\*value: 0 (default) for disabled. Endpoints use 2-bits in the value field for negotiating one or more PN spaces, available option value for client and server are listed in [Table 1](#) :

Client Option	Definition	Allowed server responses
0x0	don't support multi-path	0x0
0x1	only support one PN space for multi-path	0x0 or 0x1
0x2	only support multiple PN spaces for multi-path	0x0 or 0x2
0x3	support both one PN space and multiple PN space	0x0, 0x1 or 0x2

Table 1: Available value for enable\_multipath

If the peer does not carry the enable\_multipath transport parameter, which means the peer does not support multipath, endpoint MUST fallback to [\[QUIC-TRANSPORT\]](#) with single path and MUST NOT use any frame or mechanism defined in this document. If endpoint receives unexpected value for the transport parameter "enable\_multipath", it MUST treat this as a connection error of type MP\_CONNECTION\_ERROR and close the connection.

Note that the transport parameter "active\_connection\_id\_limit" [\[QUIC-TRANSPORT\]](#) limits the number of usable Connection IDs, and also limits the number of concurrent paths. For the QUIC multipath extension this limit even applies when no connection ID is exposed in the QUIC header.

### 3. Path Setup and Removal

After completing the handshake, endpoints have agreed to enable multipath feature and can start using multiple paths. This document does not discuss when a client decides to initiate a new path. We delegate such discussion in separate documents.

This proposal adds one multi-path control frame for path management:

\*PATH\_ABANDON frame for the receiver side to abandon the path [Section 10.1](#)

All the new frames are sent in 1-RTT packets [\[QUIC-TRANSPORT\]](#).

#### 3.1. Path Initiation

When the multipath option is negotiated, clients that want to use an additional path MUST first initiate the Address Validation procedure with PATH\_CHALLENGE and PATH\_RESPONSE frames described in Section 8 of [\[QUIC-TRANSPORT\]](#). After receiving packets from the client on the new paths, the servers MAY in turn attempt to validate these paths using the same mechanisms.

If validation succeed, the client can send non-probing, 1-RTT packets on the new paths. In contrast with the specification in section 9 of [\[QUIC-TRANSPORT\]](#), the server MUST NOT assume that receiving non-probing packets on a new path indicates an attempt to

migrate to that path. Instead, servers SHOULD consider new paths over which non-probing packets have been received as available for transmission.

### **3.2. Path Close**

Each endpoint manages the set of paths that are available for transmission. At any time in the connection, each endpoint can decide to abandon one of these paths, following for example changes in local connectivity or changes in local preferences. After an endpoint abandons a path, the peer will not receive any more non-probing packets on that path.

An endpoint that wants to close a path SHOULD NOT rely on implicit signals like idle time or packet losses, but instead SHOULD use explicit request to terminate path by sending the PATH\_ABANDON frame (see [Section 10.1](#)).

#### **3.2.1. Use PATH\_ABANDON Frame to Close a Path**

Both endpoints, namely the client and the server, can close a path, by sending PATH\_ABANDON frame (see [Section 10.1](#)) which abandons the path with a corresponding Path Identifier. Once a path is marked as "abandoned", it means that the resources related to the path, such as the used connection IDs, can be released. However, information related to data delivered over that path SHOULD not be released immediately as acknowledgments can still be received or other frames that also may trigger retransmission of data on another path.

The endpoint sending the PATH\_ABANDON frame SHOULD consider a path as abandoned when the packet that contained the PATH\_ABANDON frame is acknowledged. When releasing resources of a path, the endpoint SHOULD send a RETIRE\_CONNECTION\_ID frame for the connection IDs used on the path, if any.

The receiver of a PATH\_ABANDON frame SHOULD NOT release its resources immediately but SHOULD wait for the receive of the RETIRE\_CONNECTION\_ID frame for the used connection IDs or 3 RTOs.

Usually it is expected that the PATH\_ABANDON frame is used by the client to indicate to the server that path conditions have changed such that the path is or will be not usable anymore, e.g. in case of an mobility event. The PATH\_ABANDON frame therefore indicates to the receiving peer that the sender does not intend to send any packets on that path anymore but also recommends to the receiver that no packets should be sent in either direction. The receiver of an PATH\_ABANDON frame MAY also send an PATH\_ABANDON frame to signal its own willingness to not send any packet on this path anymore.

If connection IDs are used, PATH\_ABANDON frames can be sent on any path, not only the path that is intended to be closed. Thus a connection can be abandoned even if connectivity on that path is already broken. If no connection IDs are used and the PATH\_ABANDON frame has to be sent on the path that is intended to be closed, it is possible that the packet containing the PATH\_ABANDON frame or the packet containing the ACK for the PATH\_ABANDON frame cannot be received anymore and the endpoint might need to rely on an idle time out to close the path, as described in [Section 3.2.3](#).

Retransmittable frames, that have previously been send on the abandoned path and are considered lost, SHOULD be retransmitted on a different path.

If a PATH\_ABANDON frame is received for the only active path of a QUIC connection, the receiving peer SHOULD send a CONNECTION\_CLOSE frame and enters the closing state. If the client received a PATH\_ABANDON frame for the last open path, it MAY instead try to open a new path, if available, and only initiate connection closure if path validation fails or a CONNECTION\_CLOSE frame is received from the server. Similarly the server MAY wait for a short, limited time such as one RTO if a path probing packet is received on a new path before sending the CONNECTION\_CLOSE frame.

### 3.2.2. Effect of RETIRE\_CONNECTION\_ID Frame

Receiving a RETIRE\_CONNECTION\_ID frame causes the endpoint to discard the resources associated with that connection ID. If the connection ID was used by the peer to identify a path from the peer to this endpoint, the resources include the list of received packets used to send acknowledgements. The peer MAY decide to keep sending data using the same IP addresses and UDP ports previously associated with the connection ID, but MUST use a different connection ID when doing so.

### 3.2.3. Idle Timeout

[[QUIC-TRANSPORT](#)] allows for closing of connections if they stay idle for too long. The connection idle timeout in multipath QUIC is defined as "no packet received on any path for the duration of the idle timeout". When only one path is available, servers MUST follow the specifications in [[QUIC-TRANSPORT](#)].

When more than one path is available, servers shall monitor the arrival of non-probing packets on the available paths. Servers SHOULD stop sending traffic on paths through where no non-probing packet was received in the last 3 path RTTs, but MAY ignore that rule if it would disqualify all available paths. Server MAY release the resource associated with paths for which no non-probing packet was received for a sufficiently long path-idle delay, but SHOULD only release resource for the last available path if no traffic is received for the duration of the idle timeout, as specified in section 10.1 of [[QUIC-TRANSPORT](#)]. This means if all paths remain idle for the idle timeout, the connection is implicitly closed.

Server implementations need to select the sub-path idle timeout as a trade-off between keeping resources, such as connection IDs, in use for an excessive time or having to promptly reestablish a path after a spurious estimate of path abandonment by the client.

## 3.3. Path States

[Figure 1](#) shows the states that an endpoint's path can have.

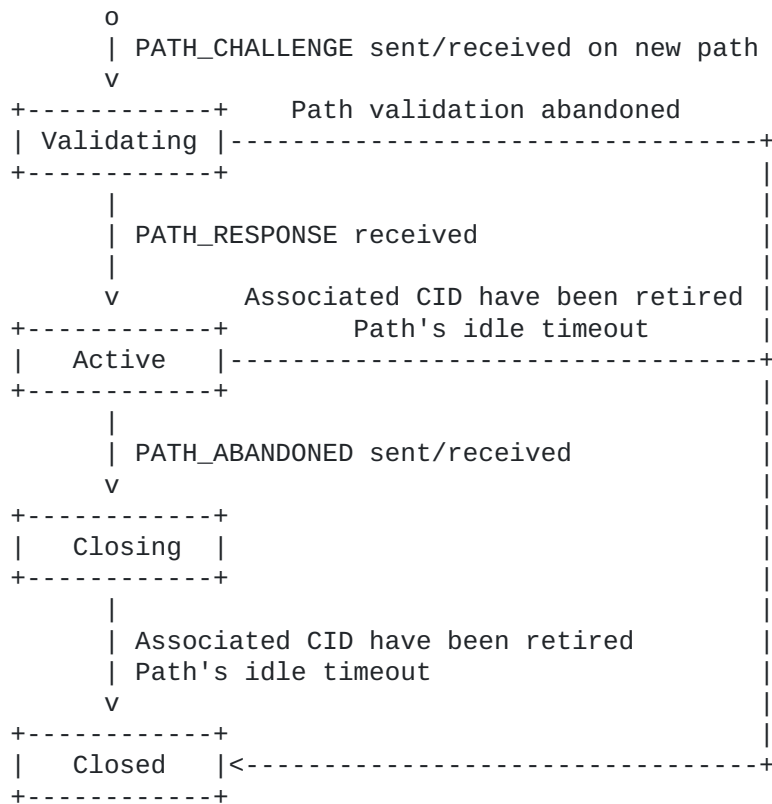


Figure 1: States of a path

In non-final states, hosts have to track the following information.

- \*Associated 4-tuple: The tuple (source IP, source port, destination IP, destination port) used by the endhost to send packets over the path.

- \*Associated Destination Connection ID: The Connection ID used to send packets over the path.

If multiple packet number spaces are used over the connection, hosts MUST also track the following information.

- \*Path Packet Number Space: The endpoint maintains a separate packet number for sending and receiving packets over this path. Packet number considerations described in [\[QUIC-TRANSPORT\]](#) apply within the given path.

In the "Active" state, hosts MUST also track the following information.

- \*Associated Source Connection ID: The Connection ID used to receive packets over the path.

A path in the "Validating" state performs path validation as described in Section 8.2 of [\[QUIC-TRANSPORT\]](#). An endhost should not send non-probing frames on a path in "Validating" state, as it has no guarantee that packets will actually reach the peer.



The endhost can use all the paths in the "Active" state, provided that the congestion control and flow control currently allow sending of new data on a path.

In the "Closing" state, the endhost SHOULD NOT send packets on this path anymore, as there is no guarantee that the peer can still map the packets to the connection. The endhost SHOULD wait for the acknowledgment of the PATH\_ABANDONED frame before moving the path to the "Closed" state to ensure a graceful termination of the path.

When a path reaches the "Closed" state, the endhost releases all the path's associated resources. Consequently, the endhost is not able to send nor receive packets on this path anymore.

#### 4. Congestion Control

Senders MUST manage per-path congestion status, and MUST NOT send more data on a given path than congestion control on that path allows. This is already a requirement of [\[QUIC-TRANSPORT\]](#).

When a Multipath QUIC connection uses two or more paths, there is no guarantee that these paths are fully disjoint. When two (or more paths) share the same bottleneck, using a standard congestion control scheme could result in an unfair distribution of the bandwidth with the multipath connection getting more bandwidth than competing single paths connections. Multipath TCP uses the LIA congestion control scheme specified in [\[RFC6356\]](#) to solve this problem. This scheme can immediately be adapted to Multipath QUIC. Other coupled congestion control schemes have been proposed for Multipath TCP such as [\[OLIA\]](#).

#### 5. Computing Path RTT

Acknowledgement delays are the sum of two one-way delays, the delay on the packet sending path and the delay on the return path chosen for the acknowledgements. When different paths have different characteristics, this can cause acknowledgement delays to vary widely. Consider for example multipath transmission using both a terrestrial path, with a latency of 50ms in each direction, and a geostationary satellite path, with a latency of 300ms in both directions. The acknowledgement delay will depend on the combination of paths used for the packet transmission and the ACK transmission, as shown in [Table 2](#).

ACK Path \ Data path	Terrestrial	Satellite
Terrestrial	100ms	350ms
Satellite	350ms	600ms

Table 2: Example of ACK delays using multiple paths

Using the default algorithm specified in [\[QUIC-RECOVERY\]](#) would result in suboptimal performance, computing average RTT and standard deviation from series of different delay measurements of different combined paths. At the same time, early tests showed that it is desirable to send ACKs through the shortest path, because a shorter ACK delay results in a tighter control loop and better performances.

The tests also showed that it is desirable to send copies of the ACKs on multiple paths, for robustness if a path experiences sudden losses.

An early implementation mitigated the delay variation issue by using time stamps, as specified in [[QUIC-Timestamp](#)]. When the timestamps are present, the implementation can estimate the transmission delay on each one-way path, and can then use these one way delays for more efficient implementations of recovery and congestion control algorithms.

If timestamps are not available, implementations could estimate one way delays using statistical techniques. For example, in the example shown in Table 1, implementations can use "same path" measurements to estimate the one way delay of the terrestrial path to about 50ms in each direction, and that of the satellite path to about 300ms. Further measurements can then be used to maintain estimates of one way delay variations, using logical similar to Kalman filters. But statistical processing is error-prone, and using time stamps provides more robust measurements.

## 6. Packet Scheduling

The transmission of QUIC packets on a regular QUIC connection is regulated by the arrival of data from the application and the congestion control scheme. QUIC packets can only be sent when the congestion window of at least one path is open.

Multipath QUIC implementations also need to include a packet scheduler that decides, among the paths whose congestion window is open, the path over which the next QUIC packet will be sent. Many factors can influence the definition of these algorithms and their precise definition is outside the scope of this document. Various packet schedulers have been proposed and implemented, notably for Multipath TCP. A companion draft [[I-D.bonaventure-iccr-g-schedulers](#)] provides several general-purpose packet schedulers depending on the application goals.

## 7. Packet Number Space and Use of Connection ID

If the connection ID is present (non-zero length) in the packet header, the connection ID is used to identify the path. If no connection ID is present, the 4 tuple identifies the path. The initial path that is used during the handshake (and multipath negotiation) has the path ID 0 and therefore all 0-RTT packets are also tracked and processed with the path ID 0. For 1-RTT packets the path ID is the sequence number of the Destination Connection ID present in the packet header, as defined in Section 5.1.1 of [[QUIC-TRANSPORT](#)], or also 0 if the Connection ID is zero-length.

If non-zero-length Connection IDs are used, an endpoint MUST use different Connection IDs on different paths. Still, the receiver may observe the same Connection ID used on different 4-tuples due to, e.g., NAT rebinding. In such case, the receiver reacts as specified in Section 9.3 of [[QUIC-TRANSPORT](#)].

Acknowledgements of Initial and Handshake packets MUST be carried using ACK frames, as specified in [[QUIC-TRANSPORT](#)]. The ACK frames,

as defined in [[QUIC-TRANSPORT](#)], do not carry path identifiers. If for any reason ACK frames are received in 1-RTT packets while the state of multipath negotiation is ambiguous, they MUST be interpreted as acknowledging packets sent on path 0.

### 7.1. Using One Packet Number Space

If the multipath option is negotiated to use one packet number space for all paths, the packet sequence numbers are allocated from the common number space, so that, for example, packet number N could be sent on one path and packet number N+1 on another.

ACK frames report the numbers of packets that have been received so far, regardless of the path on which they have been received. That means the senders needs to maintain an association between sent packet numbers and the path over which these packets were sent. This is necessary to implement per path congestion control.

When a packet is acknowledged, the state of the congestion control MUST be updated for the path where the acknowledged packet was originally sent. The RTT is calculated based on the delay between the transmission of that packet and its first acknowledgement (see [Section 5](#)) and is used to update the RTT statistics for the sending path.

Also loss detection MUST be adapted to allow for different RTTs on different paths. For example, timer computations should take into account the RTT of the path on which a packet was sent. Detections based on packet numbers shall compare a given packet number to the highest packet number received for that path.

#### 7.1.1. Sending Acknowledgements and Handling Ranges

If senders decide to send packets on paths with different transmission delays, some packets will very likely be received out of order. This will cause the ACK frames to carry multiple ranges of received packets. The large number of range increases the size of ACK frames, causing transmission and processing overhead.

The size and overhead of the ACK frames can be controlled by the combination of one or several of the following:

- \*Not transmitting again ACK ranges that were present in an ACK frame acknowledged by the peer.
- \*Delay acknowledgements to allow for arrival of "hole filling" packets.
- \*Limit the total number of ranges sent in an ACK frame.
- \*Limiting the number of transmissions of a specific ACK range, on the assumption that a sufficient number of transmissions almost certainly ensures reception by the peer.
- \*Send multiple messages for a given path in a single socket operation, so that a series of packets sent from a single path uses a series of consecutive sequence numbers without creating holes.

## 7.2. Using Multiple Packet Number Spaces

If the multipath option is enabled with a value of 2, each path has its own packet number space for transmitting 1-RTT packets and a new ACK frame format is used as specified in [Section 10.2](#). Compared to the QUIC v1 ACK frame, the MP\_ACK frames additionally contains a Packet Number Space Identifier (PN Space ID). The PN Space ID used to distinguish packet number spaces for different paths and is simply derived from the sequence number of Destination Connection ID. Therefore, the packet number space for 1-RTT packets can be identified based on the Destination Connection ID in each packets.

As soon as the negotiation of multipath support with value 2 is completed, endpoints SHOULD use ACK\_MP frames instead of ACK frames for acknowledgements of 1-RTT packets on path 0, as well as for 0-RTT packets that are acknowledged after the handshake concluded.

Following [\[QUIC-TRANSPORT\]](#), each endpoint uses NEW\_CONNECTION\_ID frames to issue usable connections IDs to reach it. Before an endpoint adds a new path by initiating path validation, it MUST check whether at least one unused Connection ID is available for each side.

If the transport parameter "active\_connection\_id\_limit" is negotiated as N, the server provided N Connection IDs, and the client is already actively using N paths, the limit is reached. If the client wants to start a new path, it has to retire one of the established paths.

ACK\_MP frame [Section 10.2](#) can be returned via either a different path, or the same path identified by the Path Identifier, based on different strategies of sending ACK\_MP frames.

Using multiple packet number spaces requires changes in the way AEAD is applied for packet protection, as explained in [Section 7.2.1](#), and tighter constraints for key updates, as explained in [Section 7.2.2](#).

### 7.2.1. Packet Protection for QUIC Multipath

Packet protection for QUIC v1 is specified in Section 5 of [\[QUIC-TLS\]](#). The general principles of packet protection are not changed for QUIC Multipath. No changes are needed for setting packet protection keys, initial secrets, header protection, use of 0-RTT keys, receiving out-of-order protected packets, receiving protected packets, or retry packet integrity. However, the use of multiple number spaces for 1-RTT packets requires changes in AEAD usage.

Section 5.3 of [\[QUIC-TLS\]](#) specifies AEAD usage, and in particular the use of a nonce, N, formed by combining the packet protection IV with the packet number. If multiple packet number spaces are used, the packet number alone would not guarantee the uniqueness of the nonce.

In order to guarantee the uniqueness of the None, the nonce N is calculated by combining the packet protection IV with the packet number and with the path identifier.

The path ID for 1-RTT packets is the sequence number of of [\[QUIC-TRANSPORT\]](#), or zero if the Connection ID is zero-length. Section 19 of [\[QUIC-TRANSPORT\]](#) encodes the Connection ID Sequence Number as a variable-length integer, allowing values up to  $2^{62}-1$ ; in this specification a range of less than  $2^{32}-1$  values MUST be used before updating the packet protection key.

To calculate the nonce, a 96 bit path-and-packet-number is composed of the 32 bit Connection ID Sequence Number in byte order, two zero bits, and the 62 bits of the reconstructed QUIC packet number in network byte order. If the IV is larger than 96 bits, the path-and-packet-number is left-padded with zeros to the size of the IV. The exclusive OR of the padded packet number and the IV forms the AEAD nonce.

For example, assuming the IV value is 6b26114b9cba2b63a9e8dd4f, the connection ID sequence number is 3, and the packet number is aead, the nonce will be set to 6b2611489cba2b63a9e873e2.

### 7.2.2. Key Update for QUIC Multipath

The Key Phase bit update process for QUIC v1 is specified in Section 6 of [\[QUIC-TLS\]](#). The general principles of key update are not changed in this specification. Following QUIC v1, the Key Phase bit is used to indicate which packet protection keys are used to protect the packet. The Key Phase bit is toggled to signal each subsequent key update. Because of network delays, packets protected with the older key might arrive later than the packets protected with the new key. Therefore, the endpoint needs to retain old packet keys to allow these delayed packets to be processed and it must distinguish between the new key and the old key. In QUIC V1, this is done using packet numbers so that the rule is made simple: Use the older key if packet number is lower than any packet number frame the current key phase.

When using multiple packet number spaces on different paths, some care is needed when initiating the Key Update process, as different paths use different packet number spaces but share a single key. When a key update is initiated on one path, packets sent to another path needs to know when the transition is complete. Otherwise, it is possible that the other paths send packets with the old keys, but skip sending any packets in the current key phase and directly jump to sending packet in the next key phase. When that happens, as the endpoint can only retain two sets of packet protection keys with the 1-bit Key Phase bit, the other paths cannot distinguish which key should be used to decode received packets, which results in a key rotation synchronization problem.

To address such a synchronization issue, if key update is initialized on one path, the sender SHOULD send at least one packet with the new key on all active paths. Further, an endpoint MUST NOT initiate a subsequent key update until a packet with the current key has been acknowledged on each path.

Following Section 5.4 of [\[QUIC-TLS\]](#), the Key Phase bit is protected, so sending multiple packets with Key Phase bit flipping at the same time should not cause linkability issue.

## 8. Examples

### 8.1. Path Establishment

[Figure 2](#) illustrates an example of new path establishment using multiple packet number spaces.

```
Client                                                    Server

(Exchanges start on default path)
1-RTT[]: NEW_CONNECTION_ID[C1, Seq=1] -->
      <-- 1-RTT[]: NEW_CONNECTION_ID[S1, Seq=1]
      <-- 1-RTT[]: NEW_CONNECTION_ID[S2, Seq=2]
...
(starts new path)
1-RTT[0]: DCID=S2, PATH_CHALLENGE[X] -->
      Checks AEAD using nonce(CID sequence 2, PN 0)
      <-- 1-RTT[0]: DCID=C1, PATH_RESPONSE[X], PATH_CHALLENGE[Y],
      ACK_MP[Seq=2,PN=0]
Checks AEAD using nonce(CID sequence 1, PN 0)
1-RTT[1]: DCID=S2, PATH_RESPONSE[Y],
      ACK_MP[Seq=1, PN=0], ... -->
```

Figure 2: Example of new path establishment

In [Figure 2](#), the endpoints first exchange new available Connection IDs with the `NEW_CONNECTION_ID` frame. In this example the client provides one Connection ID (C1 with sequence number 1), and server provides two Connection IDs (S1 with sequence number 1, and S2 with sequence number 2).

Before the client opens a new path by sending a packet on that path with a `PATH_CHALLENGE` frame, it has to check whether there are unused Connection IDs available for each side. In this example the client chooses the Connection ID S2 as the Destination Connection ID in the new path.

If the client has used all the allocated CID, it is supposed to retire those that are not used anymore, and the server is supposed to provide replacements, as specified in [\[QUIC-TRANSPORT\]](#). Usually it is desired to provide one more connection ID as currently in use, to allow for new paths or migration.

### 8.2. Path Closure

In this example the client detects the network environment change (client's 4G/Wi-Fi is turned off, Wi-Fi signal is fading to a threshold, or the quality of RTT or loss rate is becoming worse) and wants to close the initial path.

In [Figure 3](#) the server's 1-RTT packets use DCID C1, which has a sequence number of 1, for the first path; the client's 1-RTT packets use DCID S2, which has a sequence number of 2. For the second path, the server's 1-RTT packets use DCID C2, which has a sequence number of 2; the client's 1-RTT packets use CID S3, which

has a sequence number of 3. Note that two paths use different packet number space.

The client initiates the path closure for the path with ID 1 by sending a packet with an PATH\_ABANDON frame. When the server received the PATH\_ABANDON frame, it also sends an PATH\_ABANDON frame in the next packet. Afterwards the connection IDs in both directions can be retired using the RETIRE\_CONNECTION\_ID frame.

```
Client                                                    Server

(client tells server to abandon a path)
1-RTT[X]: DCID=S2 PATH_ABANDON[path_id=1]->
          (server tells client to abandon a path)
  <-1-RTT[Y]: DCID=C1 PATH_ABANDON[path_id=2], ACK_MP[Seq=2, PN=X]
(client abandons the path that it is using)
1-RTT[U]: DCID=S3 RETIRE_CONNECTION_ID[2], ACK_MP[Seq=1, PN=Y] ->
          (server abandons the path that it is using)
  <- 1-RTT[V]: DCID=C2 RETIRE_CONNECTION_ID[1], ACK_MP[Seq=3, PN=U]
```

Figure 3: Example of closing a path (path id type=0x00)

## 9. Implementation Considerations

TDB

## 10. New Frames

All the new frames MUST only be sent in 1-RTT packet, and MUST NOT use other encryption levels.

If an endpoint receives multipath-specific frames from packets of other encryption levels, it MUST return MP\_PROTOCOL\_VIOLATION as a connection error and close the connection.

### 10.1. PATH\_ABANDON Frame

The PATH\_ABANDON frame informs the peer to abandon a path. More complex path management can be made possible with additional extensions (e.g., PATH\_STATUS frame in [\[I-D.liu-multipath-quick\]](#) ).

PATH\_ABANDON frames are formatted as shown in [Figure 4](#).

```
PATH_ABANDON Frame {
  Type (i) = TBD-03 (experiments use 0xbaba05),
  Path Identifier (..),
  Error Code (i),
  Reason Phrase Length (i),
  Reason Phrase (..),
}
```

Figure 4: PATH\_ABANDON Frame Format

PATH\_ABANDON frames contain the following fields:



Path Identifier: An identifier of the path, which is formatted as shown in [Figure 5](#).

\*Identifier Type: Identifier Type field is set to indicate the type of path identifier.

-Type 0: Refer to the connection identifier used by the sender of the control frame when sending data over the specified path. This method SHOULD be used if this connection identifier is non-zero length. This method MUST NOT be used if this connection identifier is zero-length.

-Type 1: Refer to the connection identifier used by the receiver of the control frame when sending data over the specified path. This method MUST NOT be used if this connection identifier is zero-length.

-Type 2: Refer to the path over which the control frame is sent or received.

\*Path Identifier Content: A variable-length integer specifying the path identifier. If Identifier Type is 2, the Path Identifier Content MUST be empty.

```
Path Identifier {
  Identifier Type (i) = 0x00..0x02,
  [Path Identifier Content (i)],
}
```

Figure 5: Path Identifier Format

Note: If the receiver of the PATH\_ABANDON frame is using non-zero length Connection ID on that path, endpoint SHOULD use type 0x00 for path identifier in the control frame. If the receiver of the PATH\_ABANDON frame is using zero-length Connection ID, but the peer is using non-zero length Connection ID on that path, endpoints SHOULD use type 0x01 for path identifier. If both endpoints are using 0-length Connection IDs on that path, endpoints SHOULD only use type 0x02 for path identifier.

**Error Code:** A variable-length integer that indicates the reason for abandoning this path.

**Reason Phrase Length:** A variable-length integer specifying the length of the reason phrase in bytes. Because an PATH\_ABANDON frame cannot be split between packets, any limits on packet size will also limit the space available for a reason phrase.

**Reason Phrase:** Additional diagnostic information for the closure. This can be zero length if the sender chooses not to give details beyond the Error Code value. This SHOULD be a UTF-8 encoded string [[RFC3629](#)], though the frame does not carry information, such as language tags, that would aid comprehension by any entity other than the one that created the text.

PATH\_ABANDON frames SHOULD be acknowledged. If a packet containing a PATH\_ABANDON frame is considered lost, the peer SHOULD repeat it.



If the Identifier Type is 0x00 or 0x01, PATH\_ABANDON frames MAY be sent on any path, not only the path identified by the Path Identifier Content field. If the Identifier Type is 0x02, the PATH\_ABANDON frame MUST only be sent on the path that is intended to be abandoned.

## 10.2. ACK\_MP Frame

The ACK\_MP frame (types TBD-00 and TBD-01; experiments use 0xbaba00..0xbaba01) is an extension of the ACK frame defined by [\[QUIC-TRANSPORT\]](#). It is used to acknowledge packets that were sent on different paths when using multiple packet number spaces. If the frame type is TBD-01, ACK\_MP frames also contain the sum of QUIC packets with associated ECN marks received on the connection up to this point.

ACK\_MP frame is formatted as shown in [Figure 6](#).

```
ACK_MP Frame {
  Type (i) = TBD-00..TBD-01 (experiments use 0xbaba00..0xbaba01),
  Packet Number Space Identifier (i),
  Largest Acknowledged (i),
  ACK Delay (i),
  ACK Range Count (i),
  First ACK Range (i),
  ACK Range (..) ...,
  [ECN Counts (..)],
}
```

Figure 6: ACK\_MP Frame Format

Compared to the ACK frame specified in [\[QUIC-TRANSPORT\]](#), the following field is added.

**Packet Number Space Identifier:** An identifier of the path packet number space, which is the sequence number of Destination Connection ID of the 1-RTT packets which are acknowledged by the ACK\_MP frame. If the endpoint receives 1-RTT packets with zero-length Connection ID, it SHOULD use Packet Number Space Identifier 0 in ACK\_MP frames. If an endpoint receives a ACK\_MP frame with a non-existing packet number space ID, it MUST treat this as a connection error of type MP\_PROTOCOL\_VIOLATION and close the connection.

When using a single packet number space, endhosts MUST NOT send ACK\_MP frames. If an endhost receives an ACK\_MP frame while a single packet number space was negotiated, it MUST treat this as a connection error of type MP\_PROTOCOL\_VIOLATION and close the connection.

## 11. Error Codes

Multi-path QUIC transport error codes are 62-bit unsigned integers following [\[QUIC-TRANSPORT\]](#).

This section lists the defined multipath QUIC transport error codes that can be used in a CONNECTION\_CLOSE frame with a type of 0x1c. These errors apply to the entire connection.

MP\_PROTOCOL\_VIOLATION (experiments use 0xba01): An endpoint detected an error with protocol compliance that was not covered by more specific error codes.

## 12. IANA Considerations

This document defines a new transport parameter for the negotiation of enable multiple paths for QUIC, and two new frame types. The draft defines provisional values for experiments, but we expect IANA to allocate short values if the draft is approved.

The following entry in [Table 3](#) should be added to the "QUIC Transport Parameters" registry under the "QUIC Protocol" heading.

Value	Parameter Name.	Specification
TBD (experiments use 0xbabf)	enable_multipath	<a href="#">Section 2</a>

Table 3: Addition to QUIC Transport Parameters Entries

The following frame types defined in [Table 4](#) should be added to the "QUIC Frame Types" registry under the "QUIC Protocol" heading.

Value	Frame Name	Specification
TBD-00 - TBD-01 (experiments use 0xbaba00-0xbaba01)	ACK_MP	<a href="#">Section 10.2</a>
TBD-02 (experiments use 0xbaba05)	PATH_ABANDON	<a href="#">Section 10.1</a>

Table 4: Addition to QUIC Frame Types Entries

The following transport error code defined in [Table 5](#) should be added to the "QUIC Transport Error Codes" registry under the "QUIC Protocol" heading.

Value	Code	Description	Specification
TBD (experiments use 0xba01)	MP_PROTOCOL_VIOLATION	Multi-path protocol violation	<a href="#">Section 11</a>

Table 5: Error Code for Multi-path QUIC

## 13. Security Considerations

TBD

## 14. Contributors

This document is a collaboration of authors that combines work from three proposals. Further contributors that were also involved one of the original proposals are:

\*Qing An

\*Zhenyu Li

## 15. Acknowledgments

TBD

## 16. References

### 16.1. Normative References

- [QUIC-TLS] Thomson, M., Ed. and S. Turner, Ed., "Using TLS to Secure QUIC", RFC 9001, DOI 10.17487/RFC9001, May 2021, <<https://www.rfc-editor.org/info/rfc9001>>.
- [QUIC-TRANSPORT] Iyengar, J., Ed. and M. Thomson, Ed., "QUIC: A UDP-Based Multiplexed and Secure Transport", RFC 9000, DOI 10.17487/RFC9000, May 2021, <<https://www.rfc-editor.org/info/rfc9000>>.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, DOI 10.17487/RFC3629, November 2003, <<https://www.rfc-editor.org/info/rfc3629>>.
- [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/info/rfc8174>>.

### 16.2. Informative References

- [I-D.bonaventure-iccr-g-schedulers] Bonaventure, O., Piraux, M., Coninck, Q. D., Baerts, M., Paasch, C., and M. Amend, "Multipath schedulers", Work in Progress, Internet-Draft, draft-bonaventure-iccr-g-schedulers-02, 25 October 2021, <<https://www.ietf.org/archive/id/draft-bonaventure-iccr-g-schedulers-02.txt>>.
- [I-D.liu-multipath-quic] Liu, Y., Ma, Y., Huitema, C., An, Q., and Z. Li, "Multipath Extension for QUIC", Work in Progress, Internet-Draft, draft-liu-multipath-quic-04, 5 September 2021, <<https://www.ietf.org/archive/id/draft-liu-multipath-quic-04.txt>>.
- [OLIA] Khalili, R., Gast, N., Popovic, M., Upadhyay, U., and J.-Y. 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.
- [QUIC-Invariants] Thomson, M., "Version-Independent Properties of QUIC", RFC 8999, DOI 10.17487/RFC8999, May 2021, <<https://www.rfc-editor.org/info/rfc8999>>.
- [QUIC-RECOVERY] Iyengar, J., Ed. and I. Swett, Ed., "QUIC Loss Detection and Congestion Control", RFC 9002, DOI

10.17487/RFC9002, May 2021, <<https://www.rfc-editor.org/info/rfc9002>>.

[**QUIC-Timestamp**] Huitema, C., "Quic Timestamps For Measuring One-Way Delays", Work in Progress, Internet-Draft, draft-huitema-  
quic-ts-06, 12 September 2021, <<https://www.ietf.org/archive/id/draft-huitema-quic-ts-06.txt>>.

[**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/info/rfc6356>>.

## Authors' Addresses

Yanmei Liu  
Alibaba Inc.

Email: [miaoji.lym@alibaba-inc.com](mailto:miaoji.lym@alibaba-inc.com)

Yunfei Ma  
Alibaba Inc.

Email: [yunfei.ma@alibaba-inc.com](mailto:yunfei.ma@alibaba-inc.com)

Quentin De Coninck  
UCLouvain

Email: [quentin.deconinck@uclouvain.be](mailto:quentin.deconinck@uclouvain.be)

Olivier Bonaventure  
UCLouvain

Email: [olivier.bonaventure@uclouvain.be](mailto:olivier.bonaventure@uclouvain.be)

Christian Huitema  
Private Octopus Inc.

Email: [huitema@huitema.net](mailto:huitema@huitema.net)

Mirja Kuehlewind (editor)  
Ericsson

Email: [mirja.kuehlewind@ericsson.com](mailto:mirja.kuehlewind@ericsson.com)