

IDR Workgroup
Internet-Draft
Intended status: Standards Track
Expires: April 22, 2022

A. Retana
Y. Qu
Futurewei Technologies, Inc.
October 19, 2021

Use of Streams in BGP over QUIC
draft-retana-idr-bgp-quic-stream-00

Abstract

This document specifies the use of QUIC Streams to support multiple BGP sessions over one connection in order to achieve high resiliency.

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 April 22, 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.

Internet-Draft

BGP QUIC Streams

October 2021

Table of Contents

1.	Introduction	2
1.1.	Requirements Language	3
2.	Multiple BGP Sessions	3
2.1.	Multiple QUIC Streams	3
2.2.	Multiple BGP Sessions Using QUIC Streams	3
3.	MultiStream Capability	4
4.	Error Handling	5
5.	BGP Session Establishment and Collision Avoidance	5
6.	Modifications to FSM	7
7.	Operational Considerations	7
7.1.	Backward Compatibility	7
7.2.	Session Prioritization	7
7.3.	Other Considerations	7
8.	Security Considerations	7
9.	IANA Considerations	8
10.	Acknowledgement	8
11.	References	8
11.1.	Normative References	8
11.2.	Informative References	9
	Authors' Addresses	10

[1.](#) Introduction

The Border Gateway Protocol (BGP) [[RFC4271](#)] uses TCP as its transport protocol. BGP establishes peer relationship between routers using a TCP session on port 179. TCP provides reliable packet communication.

Multiprotocol Extensions for BGP-4 (MP-BGP) [[RFC4760](#)] allow BGP to carry information for multiple Network Layer protocols. However, only a single TCP connection can reach the Established state between a pair of peers [[RFC4271](#)].

As pointed out by [[I-D.ietf-idr-bgp-multisession](#)], there are some disadvantages of using a single BGP session:

A common criticism of BGP is the fact that most malformed messages cause the session to be terminated. While this behavior is necessary for protocol correctness, one may observe that the protocol machinery of a given implementation may only be defective with respect to a given AFI/SAFI. Thus, it would be desirable to allow the session related to that family to be terminated while

leaving other AFI/SAFI unaffected. As BGP is commonly deployed, this is not possible.

A second criticism of BGP is that it is difficult or in some cases impossible to manage control plane resource contention when BGP is

used to support diverse services over a single session. In contrast, if a single BGP session carries only information for a single service (or related set of services) it may be easier to manage such contention.

QUIC [[RFC9000](#)] is a UDP-based multiplexed and secure transport protocol. QUIC can provide low latency and encrypted transport with resilient connections. [[I-D.chen-idr-bgp-over-quic](#)] specifies the procedure to use BGP over QUIC. Complementary to it, this document specifies a mechanism to support multiple BGP sessions using QUIC streams.

Each BGP session operates independently. An error on one session has no impact on any other session. Sessions are distinguished by the Network Layer protocol(s) negotiated in the BGP OPEN message.

[1.1](#). 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](#). Multiple BGP Sessions

[2.1](#). Multiple QUIC Streams

QUIC [[RFC9000](#)] is a UDP-based multiplexed and secure transport protocol. It provides connection-oriented and stateful interaction between a client and server. QUIC integrates TLS and allows the exchange of application data as soon as possible.

In QUIC, application protocols exchange information via streams, and multiple streams can be multiplexed onto an underlying connection. Each stream is a separate unidirectional or bidirectional channel of

"order stream of bytes". Each stream has flow control which limits bytes sent on a stream, together with flow control of the connection.

2.2. Multiple BGP Sessions Using QUIC Streams

BGP over QUIC [[I-D.chen-idr-bgp-over-quic](#)] proposes different options to map streams. This document specifies a complementary and backwards compatible mechanism to establish multiple BGP sessions using QUIC streams. An implementation can assign one or more Network Layer protocols to a BGP session.

A QUIC stream is created by sending a BGP OPEN message, and each stream MUST be bidirectional as described in [Section 2.1 of \[RFC9000\]](#). When a BGP session is terminated, the corresponding stream MUST end (clean termination) as described in [Section 2.4 of \[RFC9000\]](#).

The Connection Collision Detection procedure to be used with streams is described in [Section 5](#). Each BGP session operates independently, which means critical conditions (such as malformed message) in one session won't affect others.

3. MultiStream Capability

The MultiStream Capability (MSC) is defined to indicate that a BGP speaker supports multiple sessions as specified in this document. The capability [[RFC5492](#)] is defined as follows:

Capability code (1 octet): TBD1

Capability length (1 octet): 1

Capability value (1 octet): flag field reserved.

```
0 1 2 3 4 5 6 7
+---+---+---+---+
|   Reserved   |
+---+---+---+---+
```

Flags: bitfield - MUST be set to zero and ignored by the receiver.

The MSC applies only when BGP over QUIC [[I-D.chen-idr-bgp-over-quic](#)] is used. It MUST be included in all OPEN messages. It MUST be ignored otherwise.

The specification in this document is used only if both peers advertise the MSC during the establishment of the "initial session". Otherwise the processes specified in [[I-D.chen-idr-bgp-over-quic](#)] MUST be followed. In particular, if a peer that advertises the MSC doesn't receive an OPEN message with the MSC from its peer, it SHOULD NOT terminate the session.

Using the MSC allows peers to establish multiple BGP sessions, one per QUIC stream. Each new BGP session is established using its own OPEN message [[RFC4271](#)] and MUST include the MSC. If the MSC was exchanged by both peers in the "initial session" but is not present when establishing a new session, the session is terminated and the

Error Subcode MUST be set to MultiStream Conflict (TBD2), defined in [Section 4](#).

Once a BGP session is established, it follows the procedures specified in [[RFC4271](#)].

[4](#). Error Handling

OPEN message error handling is defined in [section 6.2 of \[RFC4271\]](#). This document introduces the following OPEN Message Error subcodes:

TBD2 - MultiSession Conflict - Used if the MSC was exchanged by both peers in the "initial session" but is not present when establishing a new session.

TBD3 - Session Capability Mismatch - Used if a BGP speaker terminates a session in the case where it sends an OPEN message with the MSC, but receives an OPEN message without it.

TBD4 - Network Layer Protocol Mismatch - Used if a BGP session has already been established for a signaled Network Layer Protocol, either individually or as part of a set.

[Section 3](#) recommends not terminating a session when only one peer supports the MSC. If such a BGP speaker does terminate the session, the Error Subcode MUST be set to Session Capability Mismatch (TBD3).

Any individual BGP session can be terminated as specified in [\[RFC4486\]](#). If multiple sessions are to be terminated, then the procedure MUST be followed for each one.

5. BGP Session Establishment and Collision Avoidance

Before creating a new session, a BGP speaker should check that no session exists for the same Network Layer protocol(s). If a session already exists, the BGP speaker SHOULD NOT attempt to create a new one.

If a pair of BGP speakers try to establish a BGP session with each other simultaneously, then two parallel sessions will be formed. In the case of BGP over QUIC, the IP addressees of the connection cannot be used to resolve collisions when using multiple streams.

For the purpose of avoiding connection collision, a session is identified by the My Autonomous System and BGP Identifier fields pair in the OPEN message. In this context, connection collision is described as the attempt to open a BGP session for which the set of

Network Layer protocols is the same. One of the connections MUST be closed.

The connection collision is resolved using the extension specified in [\[RFC6286\]](#). In other words, the session with the higher-valued BGP Identifier is preserved [\[RFC4271\]](#). If the BGP Identifiers are identical, then the session with the larger ASN is preserved [\[RFC6286\]](#).

Upon receipt of an OPEN message, the local system MUST examine all of its sessions that are in the OpenConfirm state. A BGP speaker MAY also examine sessions in an OpenSent state if it knows the BGP Identifier of the peer by means outside of the protocol. If, among these sessions, there is one to a remote BGP speaker whose BGP Identifier and ASN pair equals the one in the OPEN message, and this

session collides with the connection over which the OPEN message is received, then the local system performs the following collision resolution procedure:

- 1) The BGP Identifier of the local system is compared to the BGP Identifier of the remote system (as specified in the OPEN message). Comparing BGP Identifiers is done by converting them to host byte order and treating them as 4-octet unsigned integers.
- 2) If the value of the local BGP Identifier is less than the remote one, the local system closes the BGP connection that already exists (the one that is already in the OpenConfirm state), and accepts the BGP connection initiated by the remote system.
 - 2a) Otherwise, the local system closes the newly created BGP connection (the one associated with the newly received OPEN message), and continues to use the existing one (the one that is already in the OpenConfirm state).
- 3) If the BGP Identifiers of the peers involved in the connection collision are identical, then the session initiated by the BGP speaker with the larger AS number is preserved.

Unless allowed via configuration, a connection collision with an existing BGP session that is in the Established state causes closing of the newly created session.

Closing the BGP session (that results from the collision resolution procedure) is accomplished by sending the NOTIFICATION message with the Error Code Cease, Subcode Connection Collision Resolution (7) [[RFC4486](#)].

The remainder of the process is as specified in [[RFC4271](#)].

[6.](#) Modifications to FSM

[To be completed.]

[7.](#) Operational Considerations

[7.1.](#) Backward Compatibility

A BGP Speaker that doesn't understand the MSC will simply ignore it [RFC5492] . [Section 3](#) recommends not terminating a session when only one peer supports the MSC. Instead, the operation will continue as specified in [[I-D.chen-idr-bgp-over-quic](#)].

[7.2.](#) Session Prioritization

One of the major drawbacks of a single BGP session is that control plane messages for all supported Network Layer protocols use the same connection, which may cause resource contention.

QUIC [[RFC9000](#)] does not provide a mechanism for exchanging prioritization information and it recommends that implementations provide ways for an application to indicate the relative priority of streams, in this case mapped to BGP sessions. If the functionality is available, an operator should prioritize BGP sessions (streams) that carry critical control plane information. The definition of this functionality and the determination of the importance of a BGP session are both outside the scope of this document.

[7.3.](#) Other Considerations

A configuration command SHOULD be implemented to allow grouping of some AFI/SAFIs into one session.

[8.](#) Security Considerations

This document specifies how to establish multiple BGP sessions over a single QUIC connection. The general operation of BGP is not changed nor is its security model. The security considerations of [[I-D.chen-idr-bgp-over-quic](#)] apply. Also, the non-TCP-related considerations of [[RFC4271](#)], [[RFC4272](#)], and [[RFC7454](#)] are applicable to the specification in this document.

By separating the control plane traffic over multiple sessions, the effect of a session-based vulnerability is reduced; only a single session is affected and not the whole connection. The result is increased resiliency.

On the other hand, a high number of BGP sessions may result in higher

resource utilization and the risk of depletion. Also, more sessions may imply additional configuration and operational complexity. This risk is mitigated by the fact that BGP sessions typically require explicit configuration by the operator.

9. IANA Considerations

IANA is asked to assign a new Capability Code for the MultiStream Capability ([Section 3](#)) as follows:

Value	Description	Reference	Change Controller
TBD1	MultiStream Capability	[This Document]	IETF

MultiStream Capability

IANA is asked to assign three values from the OPEN Message Error subcodes registry as follows:

Value	Name	Reference
TBD2	MultiSession Conflicty	[This Document]
TBD3	Session Capability Mismatch	[This Document]
TBD4	Network Layer Protocol Mismatch	[This Document]

10. Acknowledgement

This document references the text and procedures defined in [\[I-D.ietf-idr-bgp-multisession\]](#), and we are grateful for their contributions.

The authors would like to thank xx for review and comments.

11. References

11.1. Normative References

[I-D.chen-idr-bgp-over-quic]

Chen, S., Zhang, Y., Wang, H., and Z. Li, "BGP Over QUIC", [draft-chen-idr-bgp-over-quic-00](#) (work in progress), June 2021.

- [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>>.
- [RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", [RFC 4271](#), DOI 10.17487/RFC4271, January 2006, <<https://www.rfc-editor.org/info/rfc4271>>.
- [RFC4486] Chen, E. and V. Gillet, "Subcodes for BGP Cease Notification Message", [RFC 4486](#), DOI 10.17487/RFC4486, April 2006, <<https://www.rfc-editor.org/info/rfc4486>>.
- [RFC5492] Scudder, J. and R. Chandra, "Capabilities Advertisement with BGP-4", [RFC 5492](#), DOI 10.17487/RFC5492, February 2009, <<https://www.rfc-editor.org/info/rfc5492>>.
- [RFC6286] Chen, E. and J. Yuan, "Autonomous-System-Wide Unique BGP Identifier for BGP-4", [RFC 6286](#), DOI 10.17487/RFC6286, June 2011, <<https://www.rfc-editor.org/info/rfc6286>>.
- [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>>.
- [RFC9000] 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>>.

11.2. Informative References

- [I-D.ietf-idr-bgp-multisession]
Scudder, J., Appanna, C., and I. Varlashkin, "Multisession BGP", [draft-ietf-idr-bgp-multisession-07](#) (work in progress), September 2012.
- [RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", [RFC 4272](#), DOI 10.17487/RFC4272, January 2006, <<https://www.rfc-editor.org/info/rfc4272>>.
- [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", [RFC 4760](#), DOI 10.17487/RFC4760, January 2007, <<https://www.rfc-editor.org/info/rfc4760>>.

Internet-Draft

BGP QUIC Streams

October 2021

[RFC7454] Durand, J., Pepelnjak, I., and G. Doering, "BGP Operations and Security", [BCP 194](#), [RFC 7454](#), DOI 10.17487/RFC7454, February 2015, <<https://www.rfc-editor.org/info/rfc7454>>.

Authors' Addresses

Alvaro Retana
Futurewei Technologies, Inc.
2330 Central Expressway
Santa Clara, CA 95050
USA

Email: aretana@futurewei.com

Yingzhen Qu
Futurewei Technologies, Inc.
2330 Central Expressway
Santa Clara, CA 95050
USA

Email: yingzhen.qu@futurewei.com

