QUIC
Internet-Draft
Intended status: Informational
Expires: July 19, 2020

A. Ferrieux, Ed.
I. Hamchaoui, Ed.
Orange Labs
I. Lubashev, Ed.
Akamai Technologies
D. Tikhonov, Ed.
LiteSpeed Technologies
January 16, 2020

Packet Loss Signaling for Encrypted Protocols draft-ferrieuxhamchaoui-quic-lossbits-03

Abstract

This document defines an extension to the QUIC transport protocol to allow endpoints to signal packet loss in a way that can be used by network devices to measure and locate the source of the loss.

Discussion of this work is encouraged to happen on the QUIC IETF mailing list quic@ietf.org [1] or on the GitHub repository which contains the draft: https://github.com/igorlord/draft-ferrieuxhamchaoui-lossbits [2].

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 July 19, 2020.

Copyright Notice

Table of Contents

1. Introduction

Packet loss is a hard and pervasive problem of day-to-day network operation. Proactively detecting, measuring, and locating it is crucial to maintaining high QoS and timely resolution of crippling end-to-end throughput issues. To this effect, in a TCP-dominated

world, network operators have been heavily relying on information present in the clear in TCP headers: sequence and acknowledgment numbers, and SACK when enabled. These allow for quantitative estimation of packet loss by passive on-path observation. Additionally, the lossy segment (upstream or downstream from the observation point) can be quickly identified by moving the passive observer around.

With QUIC, the equivalent transport headers are encrypted and passive packet loss observation is not possible, as described in [TRANSPORT-ENCRYPT].

Measuring TCP loss between similar endpoints cannot be relied upon to evaluate QUIC loss. QUIC could be routed by the network differently and the fraction of Internet traffic delivered using QUIC is increasing every year. It is imperative to measure packet loss experienced by QUIC users directly.

Since explicit path signals are preferred by [RFC8558], two explicit loss bits in the clear portion of short headers are used to signal packet loss to on-path network devices.

This draft adapts the general technique described in [LOSSBITS] for QUIC using reserved bits in QUIC v1 short header.

2. Notational Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Loss Bits

The draft introduces two bits that are to be present in packets with a short header. Therefore, only loss of short header packets is reported using loss bits. Whenever this specification refers to packets, it is referring only to packets with short headers.

- Q: The "sQuare signal" bit is toggled every N outgoing packets as explained below in Section 3.1.

- L: The "Loss event" bit is set to 0 or 1 according to the Unreported Loss counter, as explained below in Section 3.2.

Each endpoint maintains appropriate counters independently and separately for each connection 4-tuple and Destination Connection ID. Whenever this specification refers to connections, it is referring to packets sharing the same 4-tuple and Destination Connection ID. A

"QUIC connection", however, refers to connections in the traditional QUIC sense.

3.1. Setting the sQuare Signal Bit on Outgoing Packets

The sQuare Value is initialized to the Initial Q Value (0 or 1) and is reflected in the Q bit of every outgoing packet. The sQuare value is inverted after sending every N packets (a Q run). Hence, Q Period is 2*N. The Q bit represents "packet color" as defined by [RFC8321].

Observation points can estimate upstream losses by counting the number of packets during one period of the square signal, as described in Section 4.

3.1.1. Q Run Length Selection

The sender is expected to choose N (Q run length) based on the expected amount of loss and reordering on the path. The choice of N strikes a compromise - the observation could become too unreliable in case of packet reordering and/or severe loss if N is too small, while short connections may not yield a useful upstream loss measurement if N is too large (see Section 4.2).

The value of N MUST be at least 64 and be a power of 2. This requirement allows an Observer to infer the Q run length by observing one period of the square signal. It also allows the Observer to identify flows that set the loss bits to arbitrary values (see Section 6).

If the sender does not have sufficient information to make an informed decision about Q run length, the sender SHOULD use N=64, since this value has been extensively tried in large-scale field tests and yielded good results. Alternatively, the sender MAY also choose a random N for each connection, increasing the chances of using a Q run length that gives the best signal for some connections.

The sender MUST keep the value of N constant for a given connection. The sender can change the value of N during a QUIC connection by switching to a new Destination Connection ID, if one is available.

3.2. Setting the Loss Event Bit on Outgoing Packets

The Unreported Loss counter is initialized to 0, and the L bit of every outgoing packet indicates whether the Unreported Loss counter is positive (L=1 if the counter is positive, and L=0 otherwise). The value of the Unreported Loss counter is decremented every time a packet with L=1 is sent.

The value of the Unreported Loss counter is incremented for every packet that the protocol declares lost, using QUIC's existing loss detection machinery. If the implementation is able to rescind the loss determination later, a positive Unreported Loss counter MAY be decremented due to the rescission, but it SHOULD NOT become negative.

This loss signaling is similar to loss signaling in [RFC7713], except the Loss Event bit is reporting the exact number of lost packets, whereas the Echo Loss bit in [RFC7713] is reporting an approximate number of lost bytes.

Observation points can estimate the end-to-end loss, as determined by the upstream endpoint, by counting packets in this direction with the L bit equal to 1, as described in Section 4.

4. Using Loss Bits for Passive Loss Measurement

There are three sources of observable loss:

- _upstream loss_ - loss between the sender and the observation point (Section 4.2)

- _downstream loss_ - loss between the observation point and the destination (Section 4.4)

- _observer loss_ - loss by the observer itself that does not cause downstream loss (Section 4.5)

The upstream and downstream loss together constitute _end-to-end loss_ (Section 4.1).

The Q and L bits allow detection and measurement of all these types of loss.

4.1. End-To-End Loss

The Loss Event bit allows an observer to calculate the end-to-end loss rate by counting packets with the L bit value of 0 and 1 for a given connection. The end-to-end loss rate is the fraction of packets with L=1.

The assumption here is that upstream loss affects packets with L=0 and L=1 equally. If some loss is caused by tail-drop in a network device, this may be a simplification. If the sender congestion controller reduces the packet send rate after loss, there may be a sufficient delay before sending packets with L=1 that they have a greater chance of arriving at the observer.

4.2. Upstream Loss

Blocks of N (Q run length) consecutive packets are sent with the same value of the Q bit, followed by another block of N packets with an inverted value of the Q bit. Hence, knowing the value of N, an on-path observer can estimate the amount of loss after observing at least N packets. The upstream loss rate ("u") is one minus the average number of packets in a block of packets with the same Q value ("p") divided by N ("u=1-avg(p)/N").

The observer needs to be able to tolerate packet reordering that can blur the edges of the square signal.

The observer needs to differentiate packets as belonging to different connections, since they use independent counters.

4.3. Correlating End-to-End and Upstream Loss

Upstream loss is calculated by observing packets that did not suffer the upstream loss. End-to-end loss, however, is calculated by observing subsequent packets after the sender's protocol detected the loss. Hence, end-to-end loss is generally observed with a delay of between 1 RTT (loss declared due to multiple duplicate acknowledgments) and 1 RTO (loss declared due to a timeout) relative to the upstream loss.

The connection RTT can sometimes be estimated by timing protocol handshake messages. This RTT estimate can be greatly improved by observing a dedicated protocol mechanism for conveying RTT information, such as the Latency Spin bit of [QUIC-TRANSPORT].

Whenever the observer needs to perform a computation that uses both upstream and end-to-end loss rate measurements, it SHOULD use upstream loss rate leading the end-to-end loss rate by approximately 1 RTT. If the observer is unable to estimate RTT of the connection, it should accumulate loss measurements over time periods of at least 4 times the typical RTT for the observed connections.

If the calculated upstream loss rate exceeds the end-to-end loss rate calculated in Section 4.1, then either the Q run length is too short for the amount of packet reordering or there is observer loss, described in Section 4.5. If this happens, the observer SHOULD adjust the calculated upstream loss rate to match end-to-end loss rate.

4.4. Downstream Loss

Because downstream loss affects only those packets that did not suffer upstream loss, the end-to-end loss rate ("e") relates to the upstream loss rate ("u") and downstream loss rate ("d") as "(1-u)(1-d)=1-e". Hence, "d=(e-u)/(1-u)".

4.5. Observer Loss

A typical deployment of a passive observation system includes a network tap device that mirrors network packets of interest to a device that performs analysis and measurement on the mirrored packets. The observer loss is the loss that occurs on the mirror path.

Observer loss affects upstream loss rate measurement, since it causes the observer to account for fewer packets in a block of identical Q bit values (see {{upstreamloss}}). The end-to-end loss rate measurement, however, is unaffected by the observer loss, since it is a measurement of the fraction of packets with the set L bit value, and the observer loss would affect all packets equally (see Section 4.1).

The need to adjust the upstream loss rate down to match end-to-end loss rate as described in Section 4.3 is a strong indication of the observer loss, whose magnitude is between the amount of such adjustment and the entirety of the upstream loss measured in Section 4.2. Alternatively, a high apparent upstream loss rate could be an indication of significant reordering, possibly due to packets belonging to a single connection being multiplexed over several upstream paths with different latency characteristics.

5. QUIC v1 Implementation

5.1. Transport Parameter

The use of the loss bits is negotiated using a transport parameter:

loss_bits (0x1057): The loss bits transport parameter is an integer value, encoded as a variable-length integer, that can be set to 0 or 1 indicating the level of loss bits support.

When loss_bits parameter is present, the peer is allowed to use reserved bits in the short packet header as loss bits if the peer sends loss_bits=1.

When loss_bits is set to 1, the sender will use reserved bits as loss bits if the peer includes the loss_bits transport parameter.

A client MUST NOT use remembered value of loss_bits for 0-RTT connections.

5.2. Short Packet Header

When sending loss bits has been negotiated, the reserved (R) bits are replaced by the loss (Q and L) bits in the short packet header (see Section 17.3 of [QUIC-TRANSPORT]).

sQuare Signal Bit (Q): The fourth most significant bit (0x10) is the sQuare signal bit, set as described in Section 3.1.

Loss Event Bit (L): The fifth most significant bit (0x08) is the Loss event bit, set as described in Section 3.2.

5.3. Header Protection

Unlike the reserved (R) bits, the loss (Q and L) bits are not protected. When sending loss bits has been negotiated, the first byte of the header protection mask used to protect short packet headers has its five most significant bits masked out instead of three.

The algorithm specified in Section 5.4.1 of [QUIC-TLS] changes as follows:

```
else:
   # Short header: 3 bits masked
   packet[0] ^= mask[0] & 0x07
```

6. Ossification Considerations

Accurate loss reporting signal is not critical for the operation QUIC protocol, though its presence in a sufficient number of connections is important for the operation of networks.

The loss bits are amenable to "greasing" described in [GREASE] and MUST be greased. The greasing should be accomplished similarly to the Latency Spin bit greasing in [QUIC-TRANSPORT]. Namely, implementations MUST NOT include loss_bits transport parameter for a random selection of at least one in every 16 QUIC connections.

It is possible to observe packet reordering near the edge of the square signal. A middle box might observe the signal and try to fix packet reordering that it can identify, though only a small fraction of reordering can be fixed using this method. Latency spin bit signal edge can be used for the same purpose.

7. Security Considerations

In the absence of packet loss, the Q bit signal does not provide any information that cannot be observed by simply counting packets transiting a network path. The L bit signal discloses internal state of the protocol's loss detection machinery, but this state can often be gleamed by timing packets and observing congestion controller response. Hence, loss bits do not provide a viable new mechanism to attack QUIC data integrity and secrecy.

7.1. Optimistic ACK Attack

A defense against an Optimistic ACK Attack [QUIC-TRANSPORT] involves a sender randomly skipping packet numbers to detect a receiver acknowledging packet numbers that have never been received. The Q bit signal may inform the attacker which packet numbers were skipped on purpose and which had been actually lost (and are, therefore, safe for the attacker to acknowledge). To use the Q bit for this purpose, the attacker must first receive at least an entire Q run of packets, which renders the attack ineffective against a delay-sensitive congestion controller.

For QUIC v1 connections, if the attacker can make its peer transmit data using a single large stream, examining offsets in STREAM frames can reveal whether packet number skips are deliberate. In that case, the Q bit signal provides no new information (but it does save the attacker the need to remove packet protection). However, an endpoint that communicates using [DATAGRAM] and uses a loss-based congestion controller MAY shorten the current Q run by the number of skipped packets. For example, skipping a single packet number will invert the sQuare signal one outgoing packet sooner.

8. Privacy Considerations

To minimize unintentional exposure of information, loss bits provide an explicit loss signal - a preferred way to share information per [RFC8558].

[QUIC-TRANSPORT] allows changing connection IDs in the middle of a QUIC connection to reduce the likelihood of a passive observer linking old and new subflows to the same device. Hence, a QUIC implementation would need to reset all counters when it changes

connection ID used for outgoing packets. It would also need to avoid incrementing Unreported Loss counter for loss of packets sent with a different connection ID.

Accurate loss information allows identification and correlation of network conditions upstream and downstream of the observer. This could be a powerful tool to identify connections that attempt to hide their origin networks, if the adversary is able to affect network conditions in those origin networks. Similar information can be obtained by packet timing and inferring congestion controller response to network events, but loss information provides a clearer signal.

Implementations MUST allow administrators of clients and servers to disable loss reporting either globally or per QUIC connection. Additionally, as described in Section 6, loss reporting MUST be disabled for a certain fraction of all QUIC connections.

9. IANA Considerations

This document registers a new value in the QUIC Transport Parameter Registry:

Value: 0x1057 (if this document is approved)

Parameter Name: loss_bits

Specification: Indicates that the endpoint supports loss bits. An endpoint that advertises this transport parameter can receive loss bits. An endpoint that advertises this transport parameter with value 1 can also send loss bits.

10. Change Log

10.1. Since version 02

- Add QUIC v1 negotiation using transport parameter, use short header reserved bits as loss bits, header protection change, IANA Considerations

- Add Optimistic ACK Attack Defense to Security Considerations

- Expand Privacy Considerations

- Clarify Q run length selection

10.2. Since version 01

- Add reference to RFC7713

10.3. Since version 00

- Rewrote to base this draft on [LOSSBITS]

11. Acknowledgments

The sQuare signal bit was originally specified by Kazuho Oku in early proposals for loss measurement and is an instance of the "alternate marking" as defined in [RFC8321].

Many thanks to Christian Huitema for pointing out the interaction of Q bit and Optimistic ACK Attack defence.

12. References

12.1. Normative References

[QUIC-TLS] Thomson, M. and S. Turner, "Using TLS to Secure QUIC", draft-ietf-quic-tls-24 (work in progress), November 2019.

[QUIC-TRANSPORT] Iyengar, J. and M. Thomson, "QUIC: A UDP-Based Multiplexed and Secure Transport", draft-ietf-quic-transport-24 (work in progress), November 2019.

[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>.

[RFC8558] Hardie, T., Ed., "Transport Protocol Path Signals", RFC 8558, DOI 10.17487/RFC8558, April 2019, <https://www.rfc-editor.org/info/rfc8558>.

[TRANSPORT-ENCRYPT] Fairhurst, G. and C. Perkins, "Considerations around Transport Header Confidentiality, Network Operations, and the Evolution of Internet Transport Protocols", draft-ietf-tsvwg-transport-encrypt-10 (work in progress), January 2020.

12.2. Informative References

[DATAGRAM] Pauly, T., Kinnear, E., and D. Schinazi, "An Unreliable Datagram Extension to QUIC", draft-pauly-quic-datagram-05 (work in progress), November 2019.

[GREASE] Benjamin, D., "Applying GREASE to TLS Extensibility", draft-ietf-tls-grease-04 (work in progress), August 2019.

[LOSSBITS] Ferrieux, A., Hamchaoui, I., and I. Lubashev, "Packet Loss Signaling for Encrypted Protocols", draft-ferrieuxhamchaoui-tsvwg-lossbits-02 (work in progress), November 2019.

[RFC7713] Mathis, M. and B. Briscoe, "Congestion Exposure (ConEx) Concepts, Abstract Mechanism, and Requirements", RFC 7713, DOI 10.17487/RFC7713, December 2015, <https://www.rfc-editor.org/info/rfc7713>.

[RFC8321] Fioccola, G., Ed., Capello, A., Cociglio, M., Castaldelli, L., Chen, M., Zheng, L., Mirsky, G., and T. Mizrahi, "Alternate-Marking Method for Passive and Hybrid Performance Monitoring", RFC 8321, DOI 10.17487/RFC8321, January 2018, <https://www.rfc-editor.org/info/rfc8321>.

12.3. URIs

[1] mailto:quic@ietf.org

[2] https://github.com/igorlord/draft-ferrieuxhamchaoui-lossbits

Authors' Addresses

Alexandre Ferrieux (editor)
Orange Labs

EMail: alexandre.ferrieux@orange.com

Isabelle Hamchaoui (editor)
Orange Labs

EMail: isabelle.hamchaoui@orange.com

Igor Lubashev (editor)
Akamai Technologies

EMail: ilubashe@akamai.com

Dmitri Tikhonov (editor)
LiteSpeed Technologies

EMail: dtikhonov@litespeedtech.com