TSVWG Working Group
Internet-Draft
Intended status: Standards Track
Expires: October 14, 2014

G. Fairhurst
University of Aberdeen

April 12, 2014

Network Transport Circuit Breakers
draft-fairhurst-tsvwg-00

Abstract

This note explains what is meant by the term "transport circuit breaker" in the context of an Internet tunnel service.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on October 14, 2014.

Table of Contents

1. Introduction

A transport Circuit Breaker (CB) is an automatic mechanism that is used to estimate congestion caused by a flow, and to terminate (or significantly reduce the rate of) the flow when excessive congestion is detected. This is a safety measure to prevent congestion collapse (starvation of resources available to other flows), essential for an Internet that is heterogeneous and for traffic that is hard to predict in advance.

A CB is intended as a protection mechanism of last resort. Under normal circumstances, a CB should not be triggered; It is designed to protect things when there is overload. Just as people do not expect the electrical circuit-breaker (or fuse) in their home to be triggered, except when there is a wiring fault or a problem with an electrical appliance.

Persistent congestion (also known as "congestion collapse") was a feature of the early Internet of the 1980s. This resulted in excess traffic starving other connection from access to the Internet. It was countered by the requirement to use congestion control (CC) by

the TCP transport protocol[Jacobsen88] [RFC1112]. These mechanisms operate in Internet hosts to cause TCP connections to "back off" during congestion. The introduction of CC in TCP (currently documented in [RFC5681] ensured the stability of the Internet, because it was able to detect congestion and promptly react. This worked well while TCP was by far the dominant traffic in the Internet, and most TCP flows were long-lived (ensuring that they could detect and respond to congestion before the flows terminated). This is no longer the case, and non-congestion controlled traffic, such as UDP can form a significant proportion of the total traffic traversing a link. The current Internet therefore requires that non-congestion controlled traffic needs to be considered to avoid congestion collapse.

There are important differences between a transport circuit-breaker and a congestion-control method. Specifically, congestion control (as implemented in TCP, SCTP, and DCCP) needs to operate on the timescale on the order of a packet round-trip-time (RTT), the time from sender to destination and return. Congestion control methods may react to a single packet loss/marking and reduce the transmission rate for each loss or congestion event. The goal is usually to limit the maximum transmission rate that reflects the available capacity of a network path. These methods typically operate on individual traffic flows (e.g. a 5-tuple).

In contrast, CBs are recommended for traffic aggregates, e.g.traffic sent using a network tunnel. Later sections provide examples of cases where circuit-breakers may or may not be desirable.

A CB needs to be designed to trigger robustly when there is persistent congestion. It will often operate on a much longer timescale: many RTTs, possibly many 10s of seconds. This longer period is needed to provide sufficient time for transports (or applications) to adjust their rate following congestion, and for the network load to stabilise after adjustment. A CB also needs to decide if a reaction is required based on a series of successive samples taken over a reasonably long period of time. This is to ensure that a CB does not accidentally trigger following a single (or even successive) congestion events (congestion events are what triggers congestion control, and are to be regarded as normal on a network link operating near its capacity).

1.1. Types of Circuit-Breaker

There are various forms of circuit breaker, which are differentiated mainly on the timescale over which they are triggered, but also in the intended protection they offer:

- Fast-Trip Circuit Breakers: The relatively short timescale used by this form of circuit breaker is intended to protect a flow or related group of flows.

- Slow-Trip Circuit Breakers: This circuit breaker utilises a longer timescale and is designed to protect traffic aggregates.

- Managed Circuit Breakers: Utilise the operations and management functions that may be present in a managed service to implement a circuit breaker.

Examples of each type of circuit breaker are provided in section 4.

2. Terminology

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. Designing a Circuit-Breaker (What makes a good circuit breaker?)

Although circuit breakers have been talked about in the IETF for many years, there has not yet been guidance on the cases where they are need for or the design of circuit breaker mechanisms. This document seeks to offer advise on these topics.

The basic design of a circuit breaker involves communication between the sender and receiver of a network flow. It is assumed that a sender can control the rate of the flow, but the effect of congestion can only be measured at the corresponding receiver (after loss/ marking is experienced across the end-to-end path). The receiver therefore needs to be responsible for either measuring the level of congestion (and returning this measure to the sender to inform a trigger) or for detecting excessive congestion (returning the trigger to the sender). Whether the trigger is generated at the receiver or based on measurements returned to the sender, the result of the trigger (the circuit-breaker action) needs to be applied at the sender.

The set of components needed to implement a circuit breaker are:

- There MUST be a control path from the receiver to the sender. Ideally the CB should trigger if this control path fails. That is, the feedback indicating a congested period is designed so that the sender triggers the CB action when it fails to receive reports from the receiver that indicate an absence of congestion, rather than relying on the successful transmission of a "congested"

signal back to the sender. (The feedback signal could itself be lost under congestion collapse).

- A CB MUST define a measurement period over which the receiver measures the level of congestion. This method does not have to detect individual packet loss, but MUST have a way to know that packets have been lost/marked from the traffic flow. If ECN is enabled, a receiver MAY also count the number of Explicit Congestion Notification (ECN)[RFC3168] marks per measurement interval, but even if ECN is used, the loss MUST still be measured, since this better reflects the impact of excessive congestion. The type of CB will determine how long this measurement period needs to be. The minimum time must be significantly longer than the time that current CC algorithms need to reduce their rate following detection of congestion (i.e. many path RTTs).

- A CB MUST define a threshold to determine whether the measured congestion is considered excessive.

- A CB MUST define a period over which the trigger uses collected measurements.

- A CB MUST be robust to multiple congestion events. This usually will define a number of measured excessive congestion events per triggering period. For example, a CB may combine the results of several measurement periods to determine if the CB is triggered. (e.g. triggered when excessive congestion is detected in 3 measurements within the triggering interval).

- A triggered CB MUST react decisively by reducing traffic at the source (e.g. tunnel egress). A CB SHOULD be constructed so that it does not trigger under light or intermittent congestion, hence the response when triggered needs to be much more severe than that of a CC algorithm. By default, a CB SHOULD disable the flow, it could alternatively significantly reduce the rate of the flow it controls.

- Triggering a CB SHOULD result in a response that continues for a period of time. This by default SHOULD be at least the triggering interval. Manual operator intervention MAY be required to restore the flow. If an automated response is needed to restore the flow, then this MUST NOT be immediate.

- When a CB is triggered, it SHOULD be regarded as an abnormal network event. As such, this event SHOULD be logged. The measurements that lead to triggering of the CB SHOULD also be logged.

3.1. Basic Function

This section provides one example of a suitable method to measure congestion:

1. A sender or a tunnel ingress records the number of packets/bytes sent in each measurement interval. The measurement interval could be every few seconds.

2. The receiver or tunnel egress also records the number/bytes received (at) in each measurement interval.

3. The receiver periodically returns the measured values. (This could be using Operations and Management (OAM), or an in-band signalling datagram).

4. Using the ingress and egress measurements, the loss rate for each measurement interval can be deduced from calculating the difference between these two counter values. Note that accurate measurement intervals are not typically important, since isolated loss events need to be disregard. An appropriate threshold for determining excessive congestion needs to be set (e.g. more than 10% loss, but other methods could also be based on the rate of transmission as well as the loss rate).

5. The transport circuit breaker is triggered when the threshold is exceeded in multiple measurement intervals (e.g. 3 successive measurements). This design is to be robust to single or spurious events resulting in a trigger.

6. The design may also trigger loss when it does not receive receiver measurements for 3 successive measurement periods - this may indicate a loss of control packets.

4. Examples of Circuit Breakers

This section provides examples of different types of circuit breaker. There are multiple types of circuit breaker that may be defined for use in different deployment cases:

4.1. A fast-trip Circuit Breaker

A fast-trip circuit breaker is the most responsive It has a response time that is only slightly larger than that of the traffic it controls. It is suited to traffic with well-understood characteristics. It is not be suited to arbitrary network traffic, since it may prematurely trigger (e.g. when multiple congestion-controlled flows lead to short-term overload).

4.1.1. A fast-trip RTP Circuit Breaker

A set of fast-trip CB methods have been specified for use together by a Real-time Transport Protocol (RTP) flow using the RTP/AVP Profile :[RTP-CB] . It is expected that, in the absence of severe congestion, all RTP applications running on best-effort IP networks will be able to run without triggering these circuit breakers.

The RTP congestion control specification is therefore implemented as a fail-safe.

The sender monitors reception of RTCP Reception Report (RR or XRR) packets that convey reception quality feedback information. This is used to measure (congestion) loss, possibly in combination with ECN [RFC6679].

The CB action (shutdown of the flow) is triggered when any of the following trigger conditions are true:

1. An RTP CB triggers on reported lack of progress.

2. An RTP CB triggers when no receiver reports messages are received.

3. An RTP CB uses a TFRC-style check and set a hard upper limit to the long-term RTP throughput (over many RTTs).

4. An RTP CB includes the notion of Media Usability. This circuit breaker is triggered when the quality of the transported media falls below some required minimum acceptable quality.

4.2. A Slow-trip Circuit Breaker

It is expected that most circuit breakers will be slower at responding to loss.

One example where a circuit breaker is needed is where flows or traffic-aggregates use a tunnel or encapsulation and the flows within the tunnel do not all support TCP-style congestion control (e.g. TCP, SCTP, TFRC), see [RFC5405] section 3.1.3. The usual case where this is needed is when tunnels are deployed in the general Internet (rather than "controlled environments" within an ISP or Enterprise), especially when the tunnel may need to cross a customer access router.

4.3. A Managed Circuit Breaker

This type of circuit breaker is implemented in the signalling protocol or management plane that relates to the traffic aggregate being controlled. This type of circuit breaker is typically applicable when the deployment is within a "controlled environment".

4.3.1. A Managed Circuit Breaker for SAToP Pseudo-Wires

[RFC4553], SAToP Pseudo-Wires (PWE3), section 8 describes an example of a managed circuit breaker for isochronous flows.

If such flows were to run over a pre-provisioned (e.g. MPLS) infrastructure, then it may be expected that the Pseudo-Wire (PW) would not experience congestion, because a flow is not expected to either increase (or decrease) their rate. If instead Pseudo-Wire traffic is multiplexed with other traffic over the general Internet, it could experience congestion. [RFC4553] states: "If SAToP PWs run over a PSN providing best-effort service, they SHOULD monitor packet loss in order to detect "severe congestion". The currently recommended measurement period is 1 second, and the trigger operates when there are more than three measured Severely Errored Seconds (SES) within a period.

If such a condition is detected, a SAToP PW should shut down bidirectionally for some period of time..." The concept was that when the packet loss ratio (congestion) level increased above a threshold, the PW was by default disabled. This use case considered fixed-rate transmission, where the PW had no reasonable way to shed load.

The trigger needs to be set at the rate the PW was likely have a serious problem, possibly making the service non-compliant. At this point triggering the CB would remove the traffic prevent undue impact congestion-responsive traffic (e.g., TCP). Part of the rationale, was that high loss ratios typically indicated that something was "broken" and should have already resulted in operator intervention, and should trigger this intervention. An operator-based response provides opportunity for other action to restore the service quality, e.g. by shedding other loads or assigning additional capacity, or to consciously avoid reacting to the trigger while engineering a solution to the problem. This may require the trigger to be sent to a third location (e.g. a network operations centre, NOC) responsible for operation of the tunnel ingress, rather than the tunnel ingress itself.

5. Examples where circuit breakers may not be needed.

A CB is not required for a single CC-controlled flow using TCP, SCTP, TFRC, etc. In these cases, the CC methods are designed to prevent congestion collapse.

5.1. CBs and uni-directional Traffic

A CB can not be used to control uni-directional UDP traffic. The lack of feedback prevents automated triggering of the CB. Supporting this type of traffic in the general Internet requires operator monitoring to detect and respond to congestion collapse or the use of dedicated capacity - e.g. Using per-provisioned MPLS services, RSVP, or admission-controlled Differentiated Services.

5.2. CBs over pre-provisioned Capacity

One common question is whether a CB is needed when a tunnel is deployed in a private network with pre-provisioned capacity? In this case, compliant traffic that does not exceed the provisioned capacity should not result in congestion. The CB will hence only be triggered when there is non-compliant traffic. It could be argued that this event should never happen - but it may also be argued that the CB equally should never be triggered. If a CB were to be implemented, it would provide an appropriate response should this excessive congestion occur in an operational network.

5.3. CBs with CC Traffic

IP-based traffic is generally assumed to be congestion-controlled, i.e., it is assumed that the transport protocols generating IP-based traffic at the sender already employ mechanisms that are sufficient to address congestion on the path [RFC5405]. A question therefore arises when people deploy a tunnel that is thought to only carry an aggregate of TCP (or some other CC-controlled) traffic: Is there advantage in this case in using a CB? For sure, traffic in a such a tunnel will respond to congestion. However, the answer to the question is not obvious, because the overall traffic formed by an aggregate of flows that implement a CC mechanism does not necessarily prevent congestion collapse. For instance, most CC mechanisms require long-lived flows to react to reduce the rate of a flow, an aggregate of many short flows may result in many terminating before they experience congestion. It is also often impossible for a tunnel service provider to know that the tunnel only contains CC-controlled traffic (e.g. Inspecting packet headers may not be possible). The important thing to note is that if the aggregate of the traffic does not result in persistent congestion (impacting other flows), then the CB will not trigger. This is the expected case in this context - so

implementing a CB will not reduce performance of the tunnel, but offers protection should congestion collapse occur.

6. Security Considerations

This section will describe security considerations.

7. IANA Considerations

This document makes no request from IANA.

8. Acknowledgments

There are many people who have discussed and described the issues that have motivated this draft.

9. Revision Notes

RFC-Editor: Please remove this section prior to publication

Draft 00

This was the first revision. Help and comments are greatly appreciated.

10. References

10.1. Normative References

[Jacobsen88] European Telecommunication Standards, Institute (ETSI), "Congestion Avoidance and Control", SIGCOMM Symposium proceedings on Communications architectures and protocols", August 1998.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC5405] Eggert, L. and G. Fairhurst, "Unicast UDP Usage Guidelines for Application Designers", BCP 145, RFC 5405, November 2008.

[RTP-CB] and , "Multimedia Congestion Control: Circuit Breakers for Unicast RTP Sessions", February 2014.

10.2. Informative References

[RFC1112] Deering, S., "Host extensions for IP multicasting", STD 5, RFC 1112, August 1989.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, September 2001.

[RFC4553] Vainshtein, A. and YJ. Stein, "Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)", RFC 4553, June 2006.

[RFC5681] Allman, M., Paxson, V., and E. Blanton, "TCP Congestion Control", RFC 5681, September 2009.

[RFC6040] Briscoe, B., "Tunnelling of Explicit Congestion Notification", RFC 6040, November 2010.

[RFC6679] Westerlund, M., Johansson, I., Perkins, C., O'Hanlon, P., and K. Carlberg, "Explicit Congestion Notification (ECN) for RTP over UDP", RFC 6679, August 2012.

Author's Address

Godred Fairhurst
University of Aberdeen
School of Engineering
Fraser Noble Building
Aberdeen, Scotland AB24 3UE
UK

Email: gorry@erg.abdn.ac.uk
URI: http://www.erg.abdn.ac.uk