TSVWG Working Group
Internet-Draft
Intended status: Standards Track
Expires: November 6, 2014

G. Fairhurst
University of Aberdeen

May 5, 2014

# Network Transport Circuit Breakers
## draft-fairhurst-tsvwg-circuit-breaker-01

## Abstract

This note explains what is meant by the term "network transport circuit breaker" (CB). It describes the needs for circuit breakers when using network tunnels, and other non-congestion controlled applications. It defines requirements for building a circuit breaker and the expected outcomes of using a circuit breaker within the Internet.

## 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 November 6, 2014.

Table of Contents



## 1. Introduction

A network 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 persistent 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 Transmission Control Protocol (TCP) [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, including many applications of the User Datagram Protocol (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 non-congestion-controlled Internet flows and 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 measure (meter) the traffic to determine if the network

is experiencing congestion and must be designed to trigger robustly when there is persistent congestion. This means the trigger needs to operate on a timescale much longer than the path round trip time (e.g. seconds to possibly many tens 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 any adjustment. A CB trigger will often be based on a series of successive sample measurements 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). Once triggered, a control function needs to remove traffic from the network, either disabling the flow or significantly reducing the level of traffic. This reaction provides the required protection to prevent persistent congestion being experienced by other flows that share the congested part of the network path.

## 1.1. Types of Circuit-Breaker

There are various forms of network transport circuit breaker. These 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. Design of 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 circuit breakers are needed or upon the design of circuit breaker mechanisms. This document seeks to offer advise on these two topics.

Section 3.1 describes the functional components of a circuit breaker and section 3.2 defines requirements for implementing a circuit breaker.

### 3.1. Functional Components

The basic design of a circuit breaker involves communication between an ingress point (a sender) and an egress point (a receiver) of a network flow. A simple picture of CB operation is provided in figure





1. This shows a set of routers (each labelled R) connecting a set of endpoints. A CB is used to control traffic passing through a subset of these routers, acting between an ingress and a egress point. In some cases the ingress and egress may be in one or both endpoints, in other cases they will be in the network, for example one expected use would be at the ingress and egress of a tunnel service.

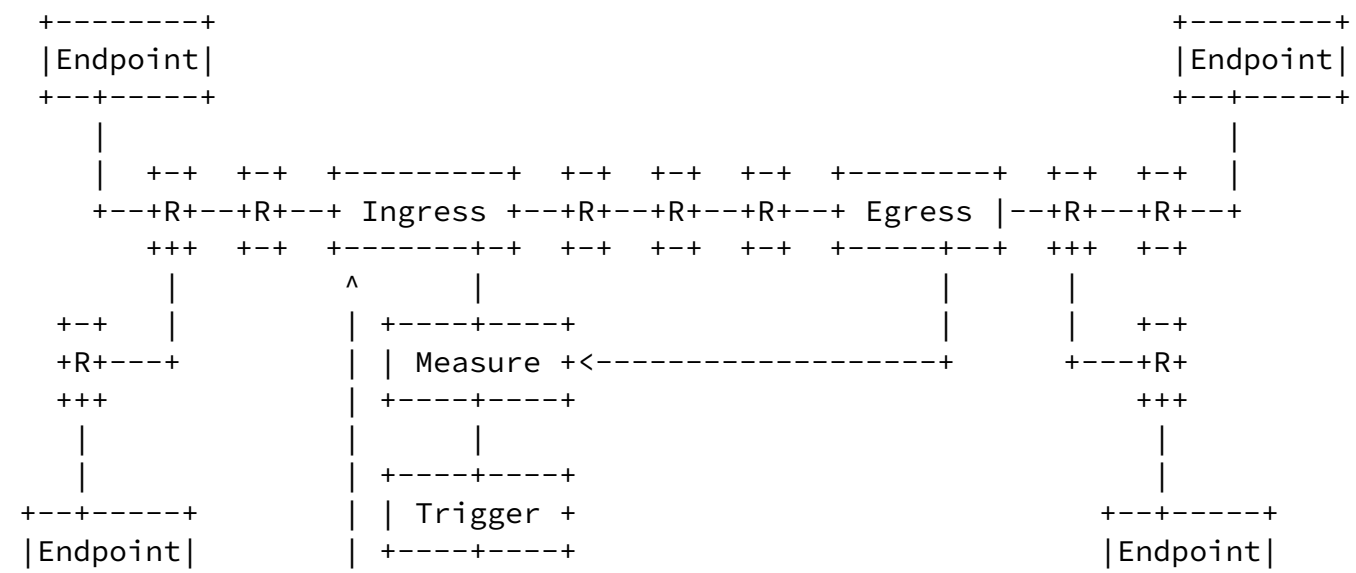


Figure 1: A CB controlling the part of the end-to-end path between an ingress point and an egress point.

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

1. An Ingress meter (at the sender or tunnel ingress) records the number of packets/bytes sent in each measurement interval. This measures the offered network load. The measurement interval could be every few seconds.

2. An Egress meter (at the receiver or tunnel egress) records the number/bytes received in each measurement interval. This measures the supported load and may utilise other signals to detect the effect of congestion (e.g. loss/marking experienced over the path).

3. The measured values at the ingress and egress are communicated to the CB Measurement function. This may use several methods including: Sending return measurement packets from a receiver to a trigger function at the sender; An implementation using Operations, Administration and Management (OAM), or another in-band signalling datagram to send to the trigger function; It





could also be implemented purely as a control plane function using a software-defined network controller.

4. The Measurement function combines the Ingress and Egress measurements to assess the present level of network congestion. (For example, the loss rate for each measurement interval could be deduced from calculating the difference between counter values. Note that accurate measurement intervals are not typically important, since isolated loss events need to be disregarded.)

5. A Trigger function determines if the measurements indicate persistent congestion. This defines an appropriate threshold for determining there is persistent congestion between the ingress and egress (e.g. more than 10% loss, but other methods could also

be based on the rate of transmission as well as the loss rate). The transport CB is triggered when the threshold is exceeded in multiple measurement intervals (e.g. 3 successive measurements). This design needs to be robust to single or spurious events triggering a reaction.

6. A Reaction that is applied at the Ingress when the CB is triggered. This seeks to automatically remove the traffic causing persistent congestion.

7. The CB also triggers when it does not receive both sender and receiver measurements, since this also could indicate a loss of control packets (also a symptom of heavy congestion or inability to control the load).

## 3.2. Requirements for implementing a CB

The requirements for implementing a CB are:

- There MUST be a control path from the Ingress meter and the Egress meter to the point of measurement. The CB MUST trigger if this control path fails. That is, the feedback indicating a congested period is designed so that the CB is triggered when it fails to receive measurement reports 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 Explicit Congestion Notification (ECN) is enabled [RFC3168], an egress





meter MAY also count the number of ECN congestion marks/event per measurement interval, but even if ECN is used, loss MUST still be measured, since this better reflects the impact of persistent 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 is REQUIRED to define a threshold to determine whether the measured congestion is considered excessive.

- A CB is REQUIRED to define a period over which the Trigger uses the collected measurements.

- A CB MUST be robust to multiple congestion events. This usually will define a number of measured persistent 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 persistent congestion is detected in 3 measurements within the triggering interval).

- A triggered CB MUST react decisively by disabling (or significantly reducing) traffic at the source (e.g. tunnel ingress). The CB SHOULD be constructed so that it does not trigger under light or intermittent congestion, with a default response to a trigger that disables all traffic that contributed to congestion.

- Some circuit breaker designs use a reaction that reduces, rather that disables, the flows it control. This response MUST be much more severe than that of a CC algorithm, because the CB reacts to more persistent congestion and operates over longer timescales. A CB that reduces the rate of a flow, MUST continue to monitor the level congestion and MUST further reduce the rate if the CB is again triggered.

- The reaction to a triggered CB MUST continue for a period of time of at least the triggering interval. Manual operator intervention will usually be required to restore the flow. If an automated response is needed to reset the trigger, then this MUST NOT be immediate. The design of this release mechanism needs to be sufficiently conservative that it does not adversely interact with other mechanisms (including other CB algorithms that control traffic over a common path.

- 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.

## 4. Examples of Circuit Breakers

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

### 4.1. A Fast-Trip Circuit Breaker

A fast-trip circuit breaker is the most responsive form of CB. 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 Circuit Breaker for RTP

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. A fast-trip RTP CB 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

A slow-trip CB may be implemented in an endpoint or network device. This type of CB is much slower at responding to congestion than a fast-trip CB and is expected to be more common.

One example where a slow-trip CB 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. A use case is where 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

A managed CB 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".

A Circuit Breaker requires more than the ability to determine that a network path is forwarding data, or to measure the rate of a path - which are often normal network operational functions. There is an additional need to determine a metric for congestion on the path and to trigger a reaction when a threshold is crossed that indicates persistent congestion.

#### 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.

XX NOTE: Comments on this section are particularly welcome to establish clearer understanding of the operational conditions under which circuit breakers should or must be deployed.

### 5.1. CBs and uni-directional Traffic

A CB can be used to control uni-directional UDP traffic, providing that there is a control path to connect the functional components at the Ingress and Egress. This control path can exist in networks for which the traffic flow is purely unidirectional (e.g. a multicast stream that sends packets across an Internet path).

A one-way physical link may have no associated control path, and therefore cannot be controlled using an automated process. This could be managed by policing traffic to ensure it does not exceed the

available capacity. Supporting this type of traffic in the general Internet requires operator monitoring to detect and respond to persistent congestion 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. A 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 persistent 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 may not be 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 persistent congestion 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. Contributions and comments are appreciated, including: Lars Eggert, Colin Perkins, David Black, Matt Mathis.

## 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.

Draft 01

Contained clarifications and changes in response to received comments, plus addition of diagram and definitions. Comments are welcome.

## 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