TSVWG Working Group
Internet-Draft
Intended status: Best Current Practice
Expires: May 21, 2016

G. Fairhurst
University of Aberdeen

November 18, 2015

Network Transport Circuit Breakers
draft-ietf-tsvwg-circuit-breaker-09

Abstract

This document explains what is meant by the term "network transport Circuit Breaker" (CB). It describes the need for circuit breakers for network tunnels and applications when using non-congestion controlled traffic, and explains where circuit breakers are, and are not, needed. It also 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 May 21, 2016.

Copyright Notice

Copyright (c) 2015 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 (http://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 2
1.1. Types of Circuit Breaker 5
2. Terminology 6
3. Design of a Circuit-Breaker (What makes a good circuit breaker?) 6
3.1. Functional Components 6
4. Requirements for a Network Transport Circuit Breaker 9
5. Other network topologies 12
5.1. Use with a multicast control/routing protocol 12
5.2. Use with control protocols supporting pre-provisioned capacity 14
5.3. Unidirectional Circuit Breakers over Controlled Paths . . 14
6. Examples of Circuit Breakers 15
6.1. A Fast-Trip Circuit Breaker 15
6.1.1. A Fast-Trip Circuit Breaker for RTP 15
6.2. A Slow-trip Circuit Breaker 16
6.3. A Managed Circuit Breaker 16
6.3.1. A Managed Circuit Breaker for SAToP Pseudo-Wires . . 17
6.3.2. A Managed Circuit Breaker for Pseudowires (PWs) . . . 17
7. Examples where circuit breakers may not be needed. 18
7.1. CBs over pre-provisioned Capacity 18
7.2. CBs with tunnels carrying Congestion-Controlled Traffic . 19
7.3. CBs with Uni-directional Traffic and no Control Path . . 19
8. Security Considerations 20
9. IANA Considerations 21
10. Acknowledgments 21
11. Revision Notes 21
12. References 23
12.1. Normative References 23
12.2. Informative References 23
Author's Address 24

1. Introduction

[RFC2309] discusses the dangers of congestion-unresponsive flows and also states that "all UDP-based streaming applications should incorporate effective congestion avoidance mechanisms". All applications ought to use a full-featured transport (TCP, SCTP, DCCP), and if not, an application (e.g., those using UDP and its UDP-Lite variant) needs to provide appropriate congestion avoidance. Guidance for applications that do not use congestion-controlled transports is provided in [ID-ietf-tsvwg-RFC5405.bis]. Such mechanisms can be designed to react on much shorter timescales than a

Circuit Breaker, that only observes a traffic envelope. Congestion-control mechanisms can also interact with an application to more effectively control its sending rate. However, not all traffic is known to respond to the onset of congestion.

A network transport Circuit Breaker (CB) is an automatic mechanism that is used to continuously monitor a flow or aggregate set of flows to detect when the flow(s) experience persistent excessive congestion. When this is detected the Circuit Breaker terminates (or significantly reduces the rate of) the flow(s). This is a safety measure to prevent starvation of network resources denying other flows from access to the Internet, such measures are essential for an Internet that is heterogeneous and for traffic that is hard to predict in advance.

The term "Circuit Breaker" originates in electricity supply, and has nothing to do with network circuits or virtual circuits. In electricity supply, a Circuit Breaker is intended as a protection mechanism of last resort. Under normal circumstances, a Circuit Breaker ought not to be triggered; it is designed to protect the supply network and attached equipment 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.

In networking, the Circuit Breaker principle can be used as a protection mechanism of last resort to avoid persistent excessive congestion impacting other flows that share network capacity. Persistent excessive congestion was a feature of the early Internet of the 1980s. This resulted in excess traffic starving another connection from access to the Internet. It was countered by the requirement to use congestion control (CC) by the Transmission Control Protocol (TCP) [Jacobsen88]. These mechanisms operate in Internet hosts to cause TCP connections to "back off" during congestion. The introduction of a congest control 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 non-congestion-controlled traffic to be considered to avoid persistent excessive congestion impacting other flows. This is expected to also help reduce the potential for "Congestion Collapse" [RFC2914].

In contrast, Circuit Breakers are recommended for non-congestion-controlled Internet flows and for traffic aggregates, e.g., traffic sent using a network tunnel. They operate on timescales much longer than the packet RTT, and trigger under situations of abnormal excessive congestion. People have been implementing what this draft characterizes as Circuit Breakers on an ad hoc basis to protect Internet traffic, this draft therefore provides guidance on how to deploy and use these mechanisms. Later sections provide examples of cases where Circuit Breakers may or may not be desirable.

A Circuit Breaker needs to measure (meter) the traffic to determine if the network is experiencing congestion and needs to be designed to trigger robustly when there is persistent excessive congestion.

A Circuit Breaker trigger will often utilize a series of successive sample measurements metered at an ingress point and an egress point (either of which could be a transport endpoint). 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 stabilize after any adjustment. This is to ensure that a Circuit Breaker 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 by disabling the flow or by significantly reducing the level of traffic. This reaction provides the required protection to prevent persistent excessive congestion being experienced by other flows that share the congested part of the network path.

Section 4 defines requirements for building a Circuit Breaker.

The operational conditions that cause a Circuit Breaker to trigger should be regarded as abnormal. Examples of situations that could trigger a Circuit Breaker include:

- anomalous traffic that exceeds the provisioned capacity (or whose traffic characteristics exceed the threshold configured for the Circuit Breaker);

- traffic generated by an application at a time when the provisioned network capacity is being utilised for other purposes;

- routing changes that cause additional traffic to start using the path monitored by the Circuit Breaker;

- misconfiguration of a service/network device where the capacity available is insufficient to support the current traffic aggregate;

- misconfiguration of an admission controller or traffic policer that allows more traffic than expected across the path monitored by the Circuit Breaker.

In many cases the reason for triggering a Circuit Breaker will not be evident to the source of the traffic (user, application, endpoint, etc). In contrast, an application that uses congestion control will generate elastic traffic that may be expected to regulate the load it introduces under congestion. This will therefore often be a preferred solution for applications that can respond to congestion signals or that can use a congestion-controlled transport.

A Circuit Breaker can be used to limit traffic from applications that are unable, or choose not, to use congestion control, or where the congestion control properties of their traffic cannot be relied upon (e.g., traffic carried over a network tunnel). In such circumstances, it is all but impossible for the Circuit Breaker to signal back to the impacted applications, and it may further be the case that applications may have some difficulty determining that a Circuit Breaker has in fact been tripped, and where in the network this happened. Application developers are advised to avoid these circumstances, where possible, by deploying appropriate congestion control mechanisms.

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 provide protection for network traffic of a single non-responsive flow or related group of non-responsive flows.

- Slow-Trip Circuit Breakers: This Circuit Breaker utilizes a longer timescale and is designed to protect network traffic from congestion by non-responsive traffic aggregates.

- Managed Circuit Breakers: Utilize the operations and management functions that might 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 advice on these two topics.

Circuit Breakers are RECOMMENDED for IETF protocols and tunnels that carry non-congestion-controlled Internet flows and for traffic aggregates. This includes traffic sent using a network tunnel. Designers of other protocols and tunnel encapsulations also ought to consider the use of these techniques as a last resort to protect traffic that shares the network path being used.

This document defines the requirements for design of a Circuit Breaker and provides examples of how a Circuit Breaker can be constructed. The specifications of individual protocols and tunnel encapsulations need to detail the protocol mechanisms needed to implement a Circuit Breaker.

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 transport Circuit Breaker involves communication between an ingress point (a sender) and an egress point (a receiver) of a network flow or set of flows. A simple picture of Circuit Breaker operation is provided in figure 1. This shows a set of routers (each labelled R) connecting a set of endpoints.

In this example, a Circuit Breaker is used to control traffic passing through a subset of these routers, acting between the ingress and a egress point network devices. The path between the ingress and egress could be provided by a tunnel or other network-layer technique. One expected use would be at the ingress and egress of a service, where all traffic being considered terminates beyond the egress point, and hence the ingress and egress carry the same set of flows.

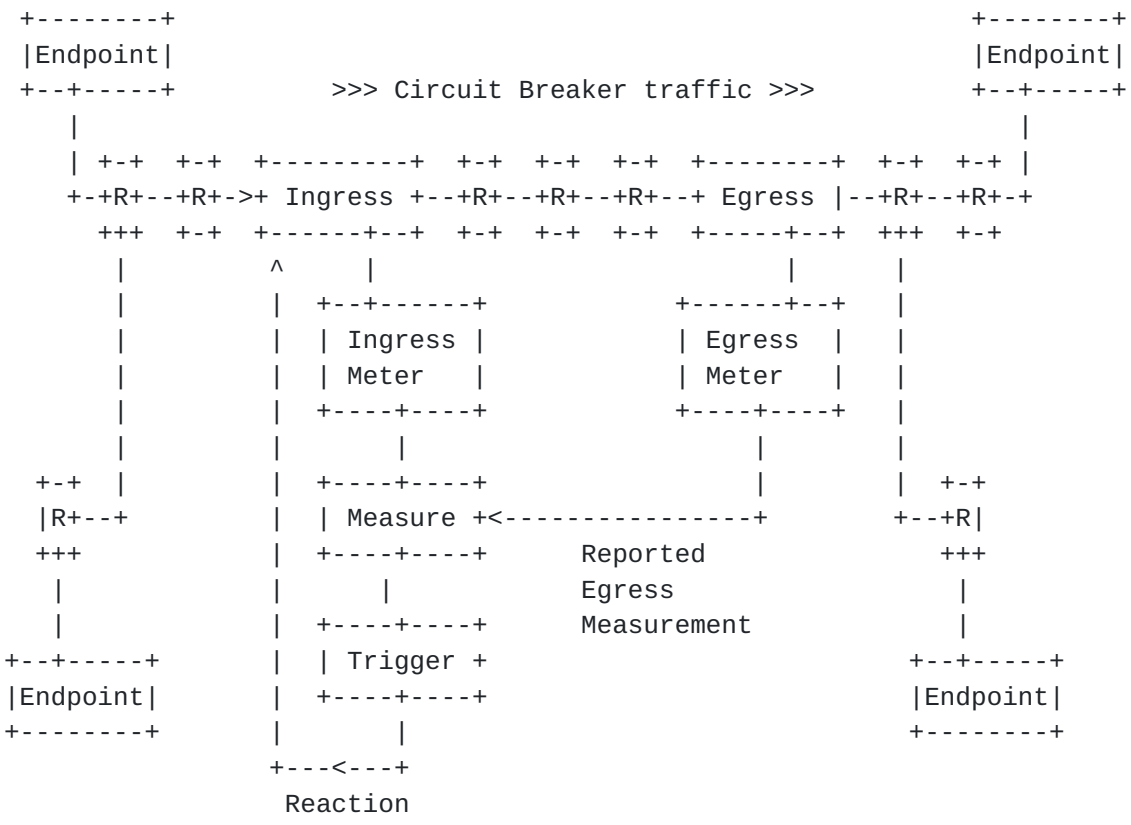

Figure 1: A CB controlling the part of the end-to-end path between an ingress point and an egress point. (Note: In some cases, the trigger and measure functions could alternatively be located at other locations (e.g., at a network operations centre.)

In the context of a Circuit Breaker, the ingress and egress functions could be implemented in different places. For example, they could be located in network devices at a tunnel ingress and at the tunnel egress. In some cases, they could be located at one or both network endpoints (see figure 2), e.g., implemented as components within a transport protocol.

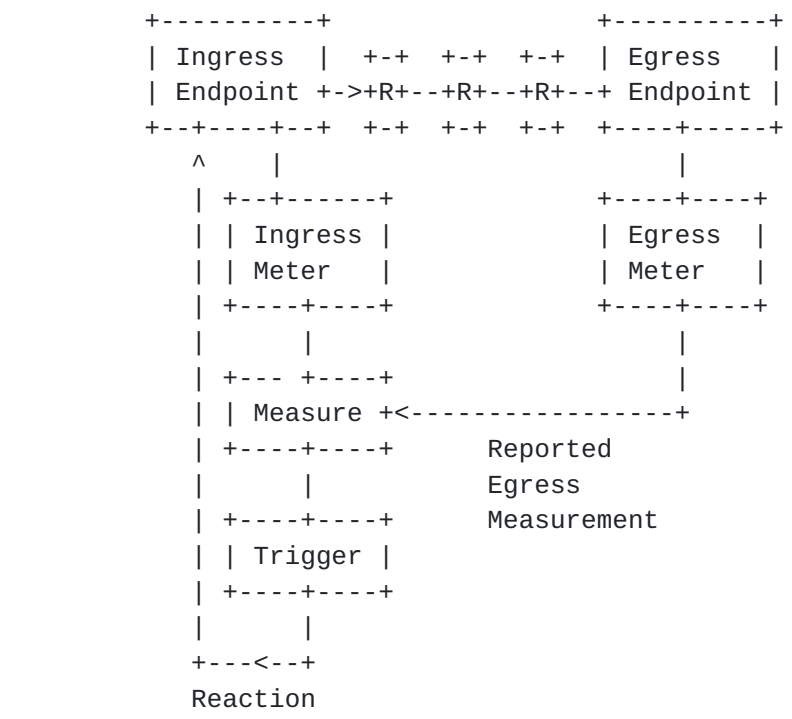

Figure 2: An endpoint CB implemented at the sender (ingress) and receiver (egress).

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 for a flow or set of flows. For example, the measurement interval could be many seconds (or every few tens of seconds or a series of successive shorter measurements that are combined by the Circuit Breaker Measurement function).

2. An egress meter (at the receiver or tunnel egress) records the number/bytes received in each measurement interval. This measures the supported load for the flow or set of flows, and could utilize other signals to detect the effect of congestion (e.g., loss/marking experienced over the path). The measurements at the egress could be synchronised (including an offset for the time of flight of the data, or referencing the measurements to a particular packet) to ensure any counters refer to the same span of packets.

3. The measured values (measurements) at the ingress and egress are communicated to the Circuit Breaker Measurement function. This could 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 be sending another in-band signalling datagram to the trigger function. This could also be implemented purely as a control plane function, e.g., 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 ingress and egress counter values.) Note that methods do not require high accuracy for the period of the measurement interval (or therefore the measured value), since isolated and/or infrequent loss events need to be disregarded.

5. A trigger function determines whether the measurements indicate persistent excessive congestion. This function defines an appropriate trigger interval and threshold for determining that there is persistent excessive congestion between the ingress and egress. This preferably considers a rate or ratio, rather than an absolute value (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 Circuit Breaker is triggered when the threshold is exceeded in multiple measurement intervals (e.g., 3 measurements within the triggering interval [RFC4553]). Designs need to be robust so that single or spurious events do not trigger a reaction.

6. A reaction that is applied at the Ingress when the Circuit Breaker is triggered. This seeks to automatically remove the traffic causing persistent excessive congestion.

7. A method for control communication control between the components that provides appropriate security and is robust when ingress and egress measurements are not available.

4. Requirements for a Network Transport Circuit Breaker

The requirements for implementing a Circuit Breaker are:

- A Circuit Breaker REQUIRED to define a measurement function to measure the level of congestion or loss. This does not have to detect individual packet loss, but MUST specify 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 excessive congestion. In this context, loss represents a reliable indication of congestion, as opposed to the finer-grain marking of incipient congestion that can be provided via ECN.

- A Circuit Breaker is REQUIRED to define the period over each measurement is made by the Circuit Breaker measurement function. The measurement period MUST be longer than the time that current congestion control mechanisms need to reduce their rate following detection of congestion. This is important because end-to-end congestion control mechanisms require at least one RTT to notify and adjust the traffic to experienced congestion, and congestion bottlenecks can share traffic with a diverse range of RTTs. A sufficiently long period is needed to avoid additionally penalizing flows with a long path RTT. The type of Circuit Breaker will determine how long this measurement period needs to be, but it needs to be significantly longer than the RTT experienced by the Circuit Breaker itself.

- If necessary, the measurement period MAY combine successive individual meter samples from the ingress and egress to ensure observation over a sufficiently long interval. (Note when meter samples need to be combined, the combination needs to reflect the sum of the individual sample counts divided by the total time/ volume over which the samples were measured. Individual samples over different intervals can not be directly combined to generate an average value.)

- A Circuit Breaker is REQUIRED to define the triggering interval. This is the period over which the trigger uses the collected measurements.

- A Circuit Breaker is REQUIRED to define a threshold to determine whether the measurements indicate that congestion is excessive. This SHOULD be constructed so that it does not trigger under light or intermittent congestion and MUST be robust to multiple congestion events per triggering period. For example, a Circuit Breaker is expected to monitor over several measurement periods to determine whether the Circuit Breaker is to be triggered. (e.g., triggered when persistent excessive congestion is detected in at least 3 of the measurement periods within the triggering interval).

- Once triggered, the Circuit Breaker MUST react decisively by disabling or significantly reducing traffic at the source (e.g., ingress). The reaction needs to be much more severe than that of a congestion control mechanism (such as TCP's congestion control [RFC5681] or TCP-Friendly Rate Control, TFRC [RFC5348]), because

the Circuit Breaker reacts to more persistent excessive congestion and operates over longer timescales (i.e., the overload condition will have persisted for a longer time before the Circuit Breaker is triggered).

- The default response to a trigger SHOULD cause the ingress to disable all of the traffic flows managed by the Circuit Breaker.

- A reaction that instead results in a reduction SHOULD reduce the traffic by at least an order of magnitude. A response that achieves the reduction by terminating flows, rather than uniformally dropping packets across multiple flows, will often be more desirable to users of the service. A Circuit Breaker that reduces the rate of a flow, MUST continue to monitor the level of congestion and MUST further react to reduce the rate if the Circuit Breaker is again triggered.

- The reaction to a triggered Circuit Breaker MUST continue for a period that is at least the triggering interval. Operator intervention will usually be required to restore a flow. If an automated response is needed to reset the trigger, then this needs to not be immediate. The design of an automated reset mechanism needs to be sufficiently conservative that it does not adversely interact with other mechanisms (including other Circuit Breaker mechanisms that control traffic over a common path). It SHOULD NOT perform an automated reset when there is evidence of continued congestion.

- When a Circuit Breaker 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 Circuit Breaker SHOULD also be logged.

- A Circuit Breaker needs a communication path for control between the ingress and the egress meters and other components. The source and integrity of control information (measurements and triggers) MUST be protected from off-path attacks (Section 8). When there is a risk of on-path attack, a cryptographic authentication mechanism for all control/measurement messages is RECOMMENDED (Section 8).

- Control communication can be in-band or out-of-band. In-band communication is RECOMMENDED when either design would be possible. If this traffic is sent over a shared path, it is RECOMMENDED that this control traffic is prioritized to reduce the probability of loss under congestion.

in-Band: An in-band control method SHOULD assume that loss of control messages is an indication of potential congestion on the path, and repeated loss (e.g., failure to receive measurement reports) ought to cause the Circuit Breaker to be triggered. (Because the feedback signal could itself be lost under congestion, this needs to confirm the absence of congestion, rather than relying on the successful transmission of a "congested" signal back to the sender.) This design has the advantage that it provides fate-sharing of the traffic flow(s) and the control communications.

Out-of-Band: An out-of-band control method SHOULD NOT trigger Circuit Breaker reaction when there is loss of control messages (e.g., a loss of measurement reports). This avoids failure amplification/propagation when the measurement and data paths fail independently. A failure of an out-of-band communication path SHOULD be regarded as abnormal network event and be handled as appropriate for the network, e.g., this event SHOULD be logged, and additional network operator action might be appropriate, depending on the network and the traffic involved.

- The Circuit Breaker MUST be designed to be robust to loss of control messages that can also be experienced during congestion/ overload. This does not imply that it is desirable to provide reliable delivery (e.g., over TCP), since this can incur additional delay in responding to congestion. Appropriate mechanisms could duplicate control messages over time to provide increased robustness to loss, or/and to regard a lack of control traffic as an indication that excessive congestion may be being experienced [ID-ietf-tsvwg-RFC5405.bis].

- The volume of control traffic ought to be considered when provisioning a network that uses a Circuit Breaker.

5. Other network topologies

A Circuit Breaker can be deployed in networks with topologies different to that presented in figure 2. This section describes examples of such usage, and possible places where functions may be implemented.

5.1. Use with a multicast control/routing protocol

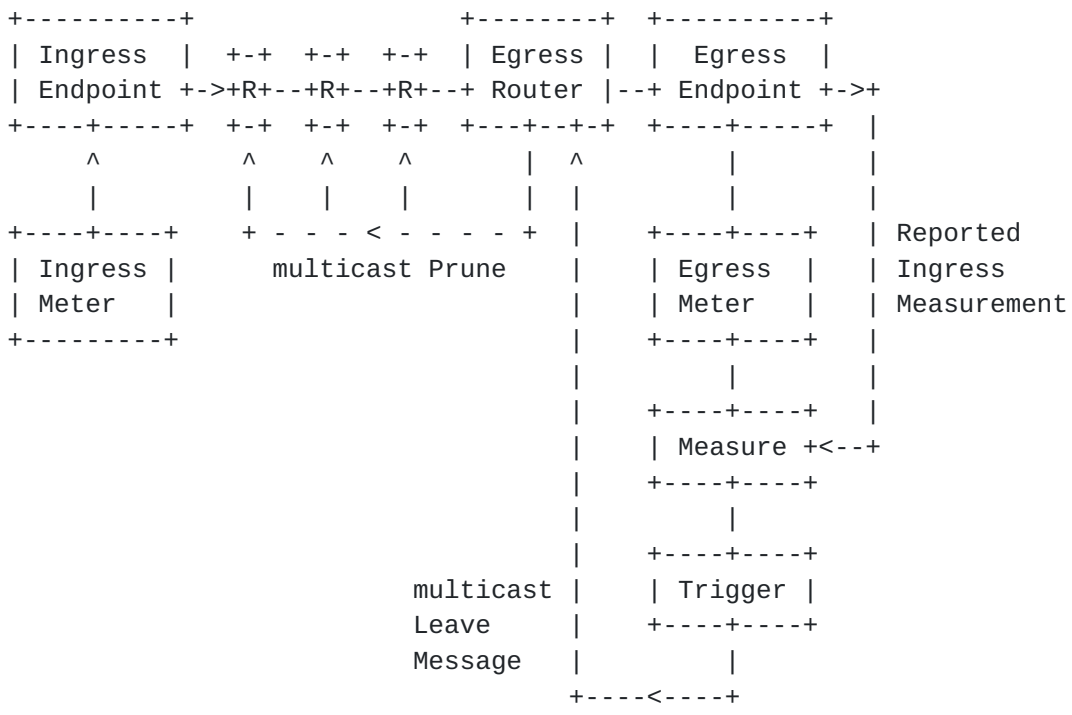

Figure 3: An example of a multicast CB controlling the end-to-end path between an ingress endpoint and an egress endpoint.

Figure 3 shows one example of how a multicast Circuit Breaker could be implemented at a pair of multicast endpoints (e.g., to implement a Fast-Trip Circuit Breaker, Section 6.1). The ingress endpoint (the sender that sources the multicast traffic) meters the ingress load, generating an ingress measurement (e.g., recording timestamped packet counts), and sends this measurement to the multicast group together with the traffic it has measured.

Routers along a multicast path forward the multicast traffic (including the ingress measurement) to all active endpoint receivers. Each last hop (egress) router forwards the traffic to one or more egress endpoint(s).

In this figure, each endpoint includes a meter that performs a local egress load measurement. An endpoint also extracts the received ingress measurement from the traffic, and compares the ingress and egress measurements to determine if the Circuit Breaker ought to be triggered. This measurement has to be robust to loss (see previous section). If the Circuit Breaker is triggered, it generates a multicast leave message for the egress (e.g., an IGMP or MLD message sent to the last hop multicast router), which causes the upstream multicast router to cease forwarding traffic to the egress endpoint.

Any multicast router that has no active receivers for a particular multicast group will prune traffic for that group, sending a prune message to its upstream router. This starts the process of releasing the capacity used by the traffic and is a standard multicast routing function (e.g., using the PIM-SM routing protocol). Each egress operates autonomously, and the Circuit Breaker reaction is executed by the multicast control plane (e.g., by the PIM multicast routing protocol), requiring no explicit signalling by the Circuit Breaker along the communication path used for the control messages. Note: In this example, there is no direct communication back to the Ingress, and hence a triggered Circuit Breaker only controls traffic downstream of the first hop router. It does not stop traffic flowing from the sender to the first hop router; this is however the common practice for multicast deployment.

The method could also be used with a multicast tunnel or subnetwork (e.g., Section 6.2, Section 6.3), where a meter at the ingress generates additional control messages to carry the measurement data towards the point where the egress metering is implemented.

5.2. Use with control protocols supporting pre-provisioned capacity

Some network paths are provisioned using a control protocol, e.g., flows provisioned using the Multi-Protocol Label Switching (MPLS) services, path provisioned using the Resource reservation protocol (RSVP), networks utilizing Software Defined Network (SDN) functions, or admission-controlled Differentiated Services.

Figure 1 shows one expected use case, where in this usage a separate device could perform the measurement and trigger functions. The reaction generated by the trigger could take the form of a network control message sent to the ingress and/or other network elements causing these elements to react to the Circuit Breaker. Examples of this type of use are provided in section Section 6.3.

5.3. Unidirectional Circuit Breakers over Controlled Paths

A Circuit Breaker can be used to control uni-directional traffic where the traffic does not itself provide congestion control (e.g., there is no feedback of congestion information at the transport or higher layers), providing that there is a communication path that can be used for control messages to connect the functional components at the Ingress and Egress. This communication path for the control messages can exist in networks for which the traffic flow is purely unidirectional. For example, a multicast stream that sends packets across an Internet path and can use multicast routing to prune flows to shed network load. Some other types of subnetwork also utilize control protocols that can be used to control traffic flows.

6. Examples of Circuit Breakers

There are multiple types of Circuit Breaker that could be defined for use in different deployment cases. This section provides examples of different types of Circuit Breaker:

6.1. A Fast-Trip Circuit Breaker

A fast-trip Circuit Breaker is the most responsive form of Circuit Breaker. It has a response time that is only slightly larger than that of the traffic that it controls. It is suited to traffic with well-understood characteristics (and could include one or more trigger functions specifically tailored the type of traffic for which it is designed). It is not suited to arbitrary network traffic and may be unsuitable for traffic aggregates, since it could prematurely trigger (e.g., when multiple congestion-controlled flows lead to short-term overload).

Although the mechanisms can be implemented in a RTP-aware network devices, these mechanisms are also suitable for implementation in endpoints (e.g., as a part of the transport system), where they can also compliment end-to-end congestion control mechanism. A shorter response time enables these mechanisms to triggers before other forms of Circuit Breaker (e.g., Circuit Breakers operating on traffic aggregates at a point along the network path).

6.1.1. A Fast-Trip Circuit Breaker for RTP

A set of fast-trip Circuit Breaker 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 Circuit Breaker is therefore implemented as a fail-safe that when triggered will terminate RTP traffic.

The sending endpoint monitors reception of in-band RTP Control Protocol (RTCP) reception report blocks, as contained in SR or RR packets, that convey reception quality feedback information. This is used to measure (congestion) loss, possibly in combination with ECN [RFC6679].

The Circuit Breaker reaction (shutdown of the flow) is triggered when any of the following trigger conditions are true:

1. An RTP Circuit Breaker triggers on reported lack of progress.

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

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

4. An RTP Circuit Breaker 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.

6.2. A Slow-trip Circuit Breaker

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

One example where a slow-trip 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 [ID-ietf-tsvwg-RFC5405.bis] section 3.1.3. A use case is where tunnels are deployed in the general Internet (rather than "controlled environments" within an Internet service provider or enterprise network), especially when the tunnel could need to cross a customer access router.

6.3. A Managed Circuit Breaker

A managed 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".

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 excessive congestion.

The control messages can use either in-band or out-of-band communications.

6.3.1. 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., Multi-Protocol Label Switching, MPLS) infrastructure, then it could be expected that the Pseudowire (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 ought to 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 that the PW was likely to experience a serious problem, possibly making the service non-compliant. At this point, triggering the Circuit Breaker would remove the traffic preventing undue impact on congestion-responsive traffic (e.g., TCP). Part of the rationale, was that high loss ratios typically indicated that something was "broken" and ought to have already resulted in operator intervention, and therefore need to 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 could 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.

6.3.2. A Managed Circuit Breaker for Pseudowires (PWs)

Pseudowires (PWs) [RFC3985] have become a common mechanism for tunneling traffic, and may compete for network resources both with other PWs and with non-PW traffic, such as TCP/IP flows.

[ID-ietf-pals-congcons] discusses congestion conditions that can arise when PWs compete with elastic (i.e., congestion responsive) network traffic (e.g, TCP traffic). Elastic PWs carrying IP traffic (see [RFC4488]) do not raise major concerns because all of the traffic involved responds, reducing the transmission rate when network congestion is detected.

In contrast, inelastic PWs (e.g., a fixed bandwidth Time Division Multiplex, TDM) [RFC4553] [RFC5086] [RFC5087]) have the potential to harm congestion responsive traffic or to contribute to excessive congestion because inelastic PWs do not adjust their transmission rate in response to congestion. [ID-ietf-pals-congcons] analyses TDM PWs, with an initial conclusion that a TDM PW operating with a degree of loss that may result in congestion-related problems is also operating with a degree of loss that results in an unacceptable TDM service. For that reason, the draft suggests that a managed Circuit Breaker that shuts down a PW when it persistently fails to deliver acceptable TDM service is a useful means for addressing these congestion concerns.

7. Examples where circuit breakers may not be needed.

A Circuit Breaker is not required for a single congestion-controlled flow using TCP, SCTP, TFRC, etc. In these cases, the congestion control mechanisms are already designed to prevent persistent excessive congestion.

7.1. CBs over pre-provisioned Capacity

One common question is whether a Circuit Breaker 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 ought not to result in persistent excessive congestion. A Circuit Breaker will hence only be triggered when there is non-compliant traffic. It could be argued that this event ought never to happen - but it could also be argued that the Circuit Breaker equally ought never to be triggered. If a Circuit Breaker were to be implemented, it will provide an appropriate response if persistent excessive congestion occurs in an operational network.

Implementing a Circuit Breaker will not reduce the performance of the flows, but in the event that persistent excessive congestion occurs, it protects network traffic that shares network capacity with these flows. A Circuit Breaker also could be used to protect other sharing network traffic from a failure that causes the Circuit Breaker traffic to be routed over a non-pre-provisioned path.

7.2. CBs with tunnels carrying Congestion-Controlled 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 [ID-ietf-tsvwg-RFC5405.bis]. A question therefore arises when people deploy a tunnel that is thought to only carry an aggregate of TCP (or some other congestion control) traffic: Is there advantage in this case in using a Circuit Breaker?

For sure, traffic in a such a tunnel will respond to congestion. However, the answer to the question is not always obvious, because the overall traffic formed by an aggregate of flows that implement a congestion control mechanism does not necessarily prevent persistent excessive congestion. For instance, most congestion control mechanisms require long-lived flows to react to reduce the rate of a flow, an aggregate of many short flows could 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 congestion-controlled traffic (e.g., Inspecting packet headers could not be possible). The important thing to note is that if the aggregate of the traffic does not result in persistent excessive congestion (impacting other flows), then the Circuit Breaker will not trigger. This is the expected case in this context - so implementing an appropriately configured Circuit Breaker will not reduce performance of the tunnel, but in the event that persistent excessive congestion occurs this protects other network traffic that shares capacity with the tunnel traffic.

7.3. CBs with Uni-directional Traffic and no Control Path

A one-way forwarding path could have no associated communication path for sending control messages, and therefore cannot be controlled using an automated process. This service could be provided using a path that has dedicated capacity and does not share this capacity with other elastic Internet flows (i.e., flows that vary their rate and respond to congestion indications).

A way to mitigate the impact on other flows when capacity could be shared is to manage the traffic envelope by using ingress policing.S.

Supporting this type of traffic in the general Internet requires operator monitoring to detect and respond to persistent excessive congestion.

8. Security Considerations

All Circuit Breaker mechanisms rely upon coordination between the ingress and egress meters and communication with the trigger function. This is usually achieved by passing network control information (or protocol messages) across the network. Timely operation of a Circuit Breaker depends on the choice of measurement period. If the receiver has an interval that is overly long, then the responsiveness of the Circuit Breaker decreases. This impacts the ability of the Circuit Breaker to detect and react to congestion.

A Circuit Breaker could potentially be exploited by an attacker to mount a Denial of Service (DoS) attack against the traffic being measured. Mechanisms therefore need to be implemented to prevent attacks on the network control information that would result in DoS. The source and integrity of control information (measurements and triggers) MUST be protected from off-path attacks. Without protection, it could be trivial for an attacker to inject fake or modified control/measurement messages (e.g., indicating high packet loss rates) causing a Circuit Breaker to trigger and to therefore mount a DoS attack that disrupts a flow.

Simple protection can be provided by using a randomized source port, or equivalent field in the packet header (such as the RTP SSRC value and the RTP sequence number) expected not to be known to an off-path attacker. Stronger protection can be achieved using a secure authentication protocol. This attack is relatively easy for an on-path attacker when the messages are neither encrypted nor authenticated. When there is a risk of on-path attack, a cryptographic authentication mechanism for all control/measurement messages is RECOMMENDED to mitigate this concern. There is a design trade-off between the cost of introducing cryptographic security for control messages and the desire to protect control communication. For some deployment scenarios the value of additional protection from DoS attack will therefore lead to a requirement to authenticate all control messages.

Transmission of network control information consumes network capacity. This control traffic needs to be considered in the design of a Circuit Breaker and could potentially add to network congestion. If this traffic is sent over a shared path, it is RECOMMENDED that this control traffic is prioritized to reduce the probability of loss under congestion. Control traffic also needs to be considered when provisioning a network that uses a Circuit Breaker.

The Circuit Breaker MUST be designed to be robust to packet loss that can also be experienced during congestion/overload. Loss of control

messages could be a side-effect of a congested network, but also could arise from other causes Section 4.

The security implications depend on the design of the mechanisms, the type of traffic being controlled and the intended deployment scenario. Each design of a Circuit Breaker MUST therefore evaluate whether the particular Circuit Breaker mechanism has new security implications.

9. IANA Considerations

This document makes no request from IANA.

10. Acknowledgments

There are many people who have discussed and described the issues that have motivated this draft. Contributions and comments included: Lars Eggert, Colin Perkins, David Black, Matt Mathis, Andrew McGregor, Bob Briscoe and Eliot Lear. This work was part-funded by the European Community under its Seventh Framework Programme through the Reducing Internet Transport Latency (RITE) project (ICT-317700).

11. Revision Notes

XXX RFC-Editor: Please remove this section prior to publication XXX

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.

WG Draft 00

Approved as a WG work item on 28th Aug 2014.

WG Draft 01

Incorporates feedback after Dallas IETF TSVWG meeting. This version is thought ready for WGLC comments. Definitions of abbreviations.

WG Draft 02

Minor fixes for typos. Rewritten security considerations section.

WG Draft 03

Updates following WGLC comments (see TSV mailing list). Comments from C Perkins; D Black and off-list feedback.

A clear recommendation of intended scope.

Changes include: Improvement of language on timescales and minimum measurement period; clearer articulation of endpoint and multicast examples - with new diagrams; separation of the controlled network case; updated text on position of trigger function; corrections to RTP-CB text; clarification of loss v ECN metrics; checks against submission checklist 9use of keywords, added meters to diagrams).

WG Draft 04

Added section on PW CB for TDM - a newly adopted draft (D. Black).

WG Draft 05

Added clarifications requested during AD review.

WG Draft 06

Fixed some remaining typos.

Update following detailed review by Bob Briscoe, and comments by D. Black.

WG Draft 07

Additional update following review by Bob Briscoe.

WG Draft 08

Updated text on the response to lack of meter measurements with managed Circuit Breakers. Additional comments from Eliot Lear (APPs area).

WG Draft 09

Updated text on applications from Eliot Lear. Additional feedback from Bob Briscoe. Comments from David Black and Mirja Kuehlewind. Resulted in change of terminology to describe this as reacting to "persistent excessive congestion", and more consistent use of "congestion control mechanisms". The requirements section was

reordered and repetition removed to ease reading. Moved text on value of CC to front of document.

12. References

12.1. Normative References

[ID-ietf-tsvwg-RFC5405.bis]
Eggert, L., Fairhurst, G., and G. Shepherd, "UDP Usage Guidelines (Work-in-Progress)", 2015.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, DOI 10.17487/RFC3168, September 2001, <http://www.rfc-editor.org/info/rfc3168>.

12.2. Informative References

[ID-ietf-pals-congcons]
Stein, YJ., Black, D., and B. Briscoe, "Pseudowire Congestion Considerations (Work-in-Progress)", 2015.

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

[RFC1112] Deering, S., "Host extensions for IP multicasting", STD 5, RFC 1112, DOI 10.17487/RFC1112, August 1989, <http://www.rfc-editor.org/info/rfc1112>.

[RFC2309] Braden, B., Clark, D., Crowcroft, J., Davie, B., Deering, S., Estrin, D., Floyd, S., Jacobson, V., Minshall, G., Partridge, C., Peterson, L., Ramakrishnan, K., Shenker, S., Wroclawski, J., and L. Zhang, "Recommendations on Queue Management and Congestion Avoidance in the Internet", RFC 2309, DOI 10.17487/RFC2309, April 1998, <http://www.rfc-editor.org/info/rfc2309>.

[RFC2914] Floyd, S., "Congestion Control Principles", BCP 41, RFC 2914, DOI 10.17487/RFC2914, September 2000, <http://www.rfc-editor.org/info/rfc2914>.

[RFC3985] Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, DOI 10.17487/RFC3985, March 2005, <http://www.rfc-editor.org/info/rfc3985>.

[RFC4488] Levin, O., "Suppression of Session Initiation Protocol (SIP) REFER Method Implicit Subscription", RFC 4488, DOI 10.17487/RFC4488, May 2006, <http://www.rfc-editor.org/info/rfc4488>.

[RFC4553] Vainshtein, A., Ed. and YJ. Stein, Ed., "Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)", RFC 4553, DOI 10.17487/RFC4553, June 2006, <http://www.rfc-editor.org/info/rfc4553>.

[RFC5086] Vainshtein, A., Ed., Sasson, I., Metz, E., Frost, T., and P. Pate, "Structure-Aware Time Division Multiplexed (TDM) Circuit Emulation Service over Packet Switched Network (CESoPSN)", RFC 5086, DOI 10.17487/RFC5086, December 2007, <http://www.rfc-editor.org/info/rfc5086>.

[RFC5087] Stein, Y(J)., Shashoua, R., Insler, R., and M. Anavi, "Time Division Multiplexing over IP (TDMoIP)", RFC 5087, DOI 10.17487/RFC5087, December 2007, <http://www.rfc-editor.org/info/rfc5087>.

[RFC5348] Floyd, S., Handley, M., Padhye, J., and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification", RFC 5348, DOI 10.17487/RFC5348, September 2008, <http://www.rfc-editor.org/info/rfc5348>.

[RFC5681] Allman, M., Paxson, V., and E. Blanton, "TCP Congestion Control", RFC 5681, DOI 10.17487/RFC5681, September 2009, <http://www.rfc-editor.org/info/rfc5681>.

[RFC6679] Westerlund, M., Johansson, I., Perkins, C., O'Hanlon, P., and K. Carlberg, "Explicit Congestion Notification (ECN) for RTP over UDP", RFC 6679, DOI 10.17487/RFC6679, August 2012, <http://www.rfc-editor.org/info/rfc6679>.

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

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