Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 14, 2016

S. Vinapamula
Juniper Networks
S. Sivakumar
Cisco Systems
M. Boucadair
France Telecom
T. Reddy
Cisco

October 12, 2015

Application-Initiated Flow High Availability Awareness through PCP
draft-vinapamula-flow-ha-13

Abstract

This document specifies a mechanism for a host to signal via Port Control Protocol (PCP) which connections should be protected against network failures. These connections will be elected to be subject to high availability mechanisms enabled at the network side.

This approach assumes that applications/users have more visibility about sensitive connections rather than any heuristic that can be enabled at the network side to guess which connections should be check-pointed.

Requirements Language

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 RFC 2119 [RFC2119].

Status of This Memo

This Internet-Draft will expire on April 14, 2016.

Copyright Notice

Table of Contents

1. Introduction 2
   1.1. Note 3
2. Issues with the Existing Implementations 3
3. CHECKPOINT-REQUIRED PCP Option 4
   3.1. Format 4
   3.2. Operation 5
4. Use cases 7
5. Security Considerations 8
6. IANA Considerations 9
7. References 9
   7.1. Normative references 9
   7.2. Informative References 9
Appendix A. Appendix 11
Acknowledgments 11
Authors' Addresses 11

1. Introduction

The risk of Internet service disruption is critical in service providers and enterprise networking environments. Such a risk is often mitigated with the introduction of active/backup systems. Such designs not only contribute to minimize the risk of service disruption, but also facilitate maintenance operations (e.g., hitless H/W or S/W upgrades).

In addition, the nature of some connections leads to the establishment and the maintenance of connection-specific states by some of the network functions invoked when the connection is established. During active/backup failover in case of a network failure, the said states need to be check-pointed by the backup system. Additional issues are further discussed in Section 2.

Heuristics based on the protocol, mapping lifetime, etc., are used in the network to elect which connections need to be check-pointed (e.g., by means of high availability techniques). This document advocates for an application-initiated approach that would allow applications/users to signal to the network which of their connections are critical.

This document specifies how PCP [RFC6887] can be extended to signal which connection should be check-pointed for high availability (Section 3). A set of use cases are provided for illustration purposes in Section 4. This document does not make any assumption on the PCP-controlled device that will process the PCP-formatted signaling information from PCP clients. These devices are likely to be flow-aware.

The proposed approach is aligned with the networking trends advocating for open network APIs to interact with applications/ services (e.g., [RFC7149]). Policy-decision making process at the network side will be enriched with information signaled by application using PCP for instance.

1.1. Note

The CHECKPOINT-REQUIRED PCP Option (Section 3) is defined in the Specification Required" range (see Section 6). In order to be assigned a code point in that range, a permanent publication is required as per Section 4.1 of [RFC5226]. Publication of an RFC is an ideal means of achieving this requirement and also to ease interoperability.

Note, this work was presented to the Port Control Protocol (pcp) WG but there was no consensus to define this option in the "Standards Action" range despite positive feedback was received from the working group. Technical comments that were received during pcp meetings and those received on the mailing list were addressed.

2. Issues with the Existing Implementations

Regardless of the selected technology or design like HA-based designs, reliably securing connections is expensive in terms of memory, CPU and other resources. Also check-pointing may not be required for all connections as all connections may not be critical. But, this leaves a challenge to identify what connections to check-point.

Typically, long-lived connections are identified and, only the states of such connections are check-pointed.

Typically, this is addressed by identifying long lived connections and check-pointing state of only those connections that lived long enough, to the backup for service continuity.

However, check-pointing long lived connections raises the following issues:

1. It is hard for a network to identify/guess which connection is (business) critical. This characterization is often customer-specific: a flow can be sensitive for a User#1 while it is not for another User#2. Furthermore, this characterization can vary over time: a flow can be sensitive during hour X, while it is not be during other times.

2. Heuristics are not deterministic.

3. A potentially long-lived connection may experience disruption upon failure of the active system, but before it is check-pointed.

4. A connection may not be long lived but critical Voice over IP (VoIP) conversations.

5. Likewise, not all long-lived connections are deemed critical: for example, connections that pertain to free Internet services are usually considered not critical compared to the equivalent connections for paid services. Only the latter need to be check-pointed.

3. CHECKPOINT-REQUIRED PCP Option

3.1. Format

This proposal is based on the assumption that an application or user is the best judge to decide which of its connections are critical.

An application or user may explicitly identify the connections that need to be check-pointed by means of a PCP client, using the CHECKPOINT_REQUIRED option as described in Figure 1.

The entry to be backed up is indicated by the content of a MAP or PEER message.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Option Code=TBA|  Reserved     |        Option Length          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

         Option Name: CHECKPOINT_REQUIRED
         Number: <TBA>
         Purpose:  Indicate if an entry needs to be check-pointed.
         Valid for Opcodes: MAP, PEER
         Length: 0.
         May appear in: request, response.
         Maximum occurrences: 1.
```

Figure 1: CHECKPOINT_REQUIRED PCP Option

The description of the fields is as follows:

- Option Code: To be assigned by IANA (see Section 6).

- Reserved: This field is initialized as specified in Section 7.3 of [RFC6887].

- Option Length: 0. This means no data is included in the option.

An application or user can take advantage of this PCP option to explicitly indicate which of the connections need to be check-pointed and should not be disrupted. The processing of this option by the PCP server will then yield the check-pointing of the corresponding states by the relevant devices or functions dynamically controlled by the PCP server.

Communication between application/user and PCP client is implementation-specific.

3.2. Operation

Support of the CHECKPOINT_REQUIRED option by PCP servers and PCP clients is optional. This option (Code TBA; see Figure 1) may be included in a PCP MAP/PEER request to indicate a connection is to be protected against network failures.

There is a risk that every PCP client may wish to check-point every connection, which can potentially load the system. Administration SHOULD restrict the number of connections that can be elected to be backed up and the rate of check-pointing on per network attachment point (e.g., CPE, host). To that aim, the PCP server should unambiguously identify the network attachment point a PCP client

belongs to. For example, the PCP server may rely on the PCP identity [RFC7652], the assigned prefix to a CPE/host, the subscriber-mask [I-D.vinapamula-softwire-dslite-prefix-binding], or other identification means.

The PCP client includes a CHECKPOINT_REQUIRED option in a MAP or PEER request to signal that the corresponding mapping is to be protected.

If the PCP client does not receive a CHECKPOINT_REQUIRED option in response to a PCP request that enclosed the CHECKPOINT_REQUIRED option, this means that either the PCP server does not support the option, or the PCP server is configured to ignore the option or the PCP server cannot satisfy the request expressed in this option (e.g., because of a lack of resources).

If the CHECKPOINT_REQUIRED option is not included in the PCP client request, the PCP server MUST NOT include the CHECKPOINT_REQUIRED option in the associated response.

When the PCP server receives a CHECKPOINT_REQUIRED option, the PCP server checks if it can honor this request depending on whether resources are available for check-pointing. If there are no resources available for check-pointing, but there are resources available to honor the MAP/PEER request, a response is sent back to the PCP client without including the CHECKPOINT_REQUIRED option (i.e., the request is processed as any MAP/PEER request that does not convey a CHECKPOINT_REQUIRED option). If check-pointing resources are still available and the quota for this PCP client is not reached, the PCP server tags the corresponding entry as eligible to HA mechanism and sends back the CHECKPOINT_REQUIRED option in the positive answer to the PCP client.

To update the check-pointing behavior of a mapping maintained by the PCP server, the PCP client generates a PCP MAP/PEER renewal request that includes a CHECKPOINT_REQUIRED option to indicate this mapping has to be check-pointed or without including a CHECKPOINT_REQUIRED option to indicate this mapping does not need be check-pointed anymore. Upon receipt of the PCP request, the PCP server proceeds with the same operations to validate a MAP/PEER request updating an existing mapping. If validation checks are passed, the PCP server updates the check-point flag associated with that mapping accordingly (i.e., it is set if a CHECKPOINT_REQUIRED option was included in the update request or it is cleared if no CHECKPOINT_REQUIRED option was included) , and the PCP server returns the response to the PCP client accordingly.

What information to check-point and how to check-point is out of scope of this document, and is left for implementations. Also,

interest to indicate check-pointing by users/applications in a PCP request, may be automatic, semi-automatic, or human intervened. This behavior is also left for application implementations. In case of manges CPEs, a service provider may influence what connections to be check-pointed.

It is RECOMMENDED to check-point state on backup for honored requests before a response is sent to the PCP client.

4. Use cases

Below are provided some examples for illustration purposes:

Example 1: Consider a streaming service such as live TV broadcasting, or any other media streaming, that supports check-pointing signalling functionality. Suppose, this application is installed in three hosts A, B and C. For A it is critical and doesn't want interruption while for B it is not. While for C, only some programs are of interest. At the time of installing this application's software, corresponding preferences can be provisioned. When the application starts streaming:

* All the flows associated with the streaming application are critical for A. Limiting the number of flows to be backed up will ensure that host doesn't exceed the user's limit.

* In case of B, none of these flows are critical for check-pointing. CHECKPOINT_REQUIRED option is not included in the PCP requests.

* In case of C, the user is invited to interact with the application by the means of a configuration option that is provided to dynamically select which streaming to check-point, based on the user's interest.

Example 2: Consider a streaming service offered by a provider. Suppose, three levels of subscriptions are offered by that provider: e.g., gold, silver, bronze. To guarantee a certain level of quality of service for each subscription, policies are configured such that:

* All flows associated with a gold subscription should be check-pointed.

* Only some flows associated with a silver subscription are check-pointed.

- None of the flows associated with a bronze subscription are check-pointed.

When a user invokes the streaming service, he/she may fall into one of those buckets, and according to the configured policy, his/her associated streaming flows are automatically check-pointed. Login credentials can be used as a trigger to determine the subscription level (and therefore the associated check-pointing behavior).

Example 3: Consider a VoIP application that is able to request its flows to be check-pointed. No matter what is configured by the user, some calls such as emergency calls should be check-pointed. The application has to identify such calls.

Example 4: In the context of an enterprise network, applications are customized by the administrator. Instructions whether a CHECKPOINT_REQUIRED option is to be included is determined by the administrator. Only the subset of applications identified by the administrator will make use of this option in conformance with the enterprise network management policies. Any mis-behavior can be considered as an abuse.

In order to avoid that every application includes a CHECKPOINT_REQUIRED option in its PCP requests, the following items are assumed:

- Applications may be delivered with some default settings for check-pointing, and these settings should be programmable by end user.

- Exposing and enforcing these settings is application specific.

- End user may customize these settings on need basis based on his preferences.

5. Security Considerations

PCP-related security considerations are discussed in [RFC6887].

CHECKPOINT_REQUIRED option can be used by an attacker to identify critical flows, which is sensitive from a privacy standpoint. Also, an attacker can cause critical flows to not be check-pointed by stripping the CHECKPOINT_REQUIRED option or by consuming the quota by adding the option to other flows.

These two issues can be mitigated if the network on which the PCP messages are to be sent is fully trusted. Means to defend against

attackers who can intercept packets between the PCP server and the PCP client should be enabled. In some deployments, access control lists (ACLs) can be installed on the PCP client, PCP server, and the network between them, so those ACLs allow only communications between trusted PCP elements. If the networking environment between the PCP client and the PCP server is not secure, PCP authentication [RFC7652] MUST be enabled.

A network device can always override the end-user signalling, i.e., what is signaled by the PCP client, if the instructions are conflicting with the network policies.

6. IANA Considerations

The following PCP Option Code is to be allocated in the "Specification Required" range (64-95,192-223) (the registry is maintained in http://www.iana.org/ assignments/pcp-parameters):

CHECKPOINT_REQUIRED set to TBA (see Section 3.1)

7. References

7.1. Normative references

[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>.

[RFC6887] Wing, D., Ed., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", RFC 6887, DOI 10.17487/RFC6887, April 2013, <http://www.rfc-editor.org/info/rfc6887>.

[RFC7652] Cullen, M., Hartman, S., Zhang, D., and T. Reddy, "Port Control Protocol (PCP) Authentication Mechanism", RFC 7652, DOI 10.17487/RFC7652, September 2015, <http://www.rfc-editor.org/info/rfc7652>.

7.2. Informative References

[I-D.vinapamula-softwire-dslite-prefix-binding]
Vinapamula, S. and M. Boucadair, "Recommendations for Prefix Binding in the Softwire DS-Lite Context", draft-vinapamula-softwire-dslite-prefix-binding-11 (work in progress), August 2015.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, DOI 10.17487/RFC5226, May 2008, <http://www.rfc-editor.org/info/rfc5226>.

[RFC7149] Boucadair, M. and C. Jacquenet, "Software-Defined Networking: A Perspective from within a Service Provider Environment", RFC 7149, DOI 10.17487/RFC7149, March 2014, <http://www.rfc-editor.org/info/rfc7149>.

Appendix A. Appendix

It was tempting to include additional fields in the option but this would lead to a more complex design that is not justified, e.g.,:

- Define a dedicated field to indicate a priority level. This priority is intended to be used by the PCP server as a hint when processing a request with a CHECKPOINT_REQUIRED option. Nevertheless, an applications may systematically choose to set the priority level to the highest value so that it increases its chance to be serviced!

- Return a more granular failure error code to the requesting PCP client. Nevertheless this would require extra processing at both the PCP client and server sides for handling the various error codes without any guarantee for the PCP client to have its mappings check-pointed.

Acknowledgments

Thanks to Reinaldo Penno, Stuart Cheshire, Dave Thaler, Prashanth Patil, and Christian Jacquenet for their comments.

Authors' Addresses

Suresh Vinapamula
Juniper Networks
1194 North Mathilda Avenue
Sunnyvale, CA 94089
USA

Phone: +1 408 936 5441
EMail: sureshk@juniper.net

Senthil Sivakumar
Cisco Systems
7100-8 Kit Creek Road
Research Triangle Park, NC 27760
USA

Phone: +1 919 392 5158
EMail: ssenthil@cisco.com

Mohamed Boucadair
France Telecom
Rennes 35000
France

EMail: mohamed.boucadair@orange.com

Tirumaleswar Reddy
Cisco Systems, Inc.
Cessna Business Park, Varthur Hobli
Sarjapur Marathalli Outer Ring Road
Bangalore, Karnataka 560103
India

EMail: tireddy@cisco.com