PCE Working Group Q. Wu
Internet-Draft D. Dhody
Intended status: Standards Track Huawei
Expires: January 16, 2014 July 15, 2013

Path Computation Element (PCE) Discovery using Domain Name System(DNS) draft-wu-pce-dns-pce-discovery-01

Abstract

Discovery of the Path Computation Element (PCE) within an IGP area or domain is possible using OSPF [RFC5088] and IS-IS [RFC5089]. However, in some deployment scenarios PCEs may not wish, or be able, to participate within the IGP process,therefore it would be beneficial for the Path Computation Client (PCC) (or other PCEs) to discover PCEs via an alternative mechanism to those proposed in [RFC5088] and [RFC5089].

This document specifies the requirements, use cases, procedures and extensions to support discovery via DNS for PCE.

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 January 16, 2014.

Copyright Notice

Table of Contents

1. Introduction

The Path Computation Element Communication Protocol (PCEP) is a transaction-based protocol carried over TCP[RFC4655]. In order to be able to direct path computation requests to the Path Computation Element (PCE), a Path Computation Client (PCC) (or other PCEs) needs to know the location and capability of a PCE.

In a network where an IGP is used and where the PCE participates in the IGP, discovery mechanisms exist for PCC (or PCE) to learn the identity and capability of each PCE. [RFC5088] defines a PCE Discovery (PCED) TLV carried in an OSPF Router LSA. Similarly, [RFC5089] defines the PCED sub-TLV for use in PCE Discovery using IS-IS. Scope of the advertisement is limited to IGP area/level or Autonomous System (AS).

However in certain scenarios not all PCEs will participate in the IGP instance, section 3 (Motivation) outlines a number of use cases. In these cases, current PCE Discovery mechanisms are therefore not appropriate and another PCE discovery function would be required.

1.1. Requirements

As described in [RFC4674], the PCE Discovery information should at least be composed of:

- The PCE location: an IPv4 and/or IPv6 address that is used to reach the PCE. It is RECOMMENDED to use an address that is always reachable if there is any connectivity to the PCE;
- The PCE path computation scope (i.e., intra-layer, inter-area, inter-AS, or inter-layer);
- The set of one or more PCE-Domain(s) into which the PCE has visibility and for which the PCE can compute paths;
- The set of zero, one, or more neighbor PCE-Domain(s) toward which the PCE can compute paths;

that allows PCCs to select appropriate PCEs:

This document specifies an extension to DNS for the above PCE information discovery, which is complements the existing discovery mechanism.

2. Conventions used in this document

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 [RFC2119].

3. Motivation

This section discusses in more detail the motivation and use cases for an alternative DNS based PCE discovery mechanism.

3.1. Load Sharing of Path Computation Requests

Multiple PCE servers can be present in a single network domain for redundancy. However load balance decision is made by PCC,it doesn't enable real load balance across the PCE servers if PCC still tries PCE one by one and PCE doesn't indicate the load status to the PCC.

Inherent DNS based load balancing may be used for inbound load balancing and implemented at the application level in both servers and clients. Multiple host IP addresses are configured in DNS for a single host server name. Also DNS is query-response based mechanism and capable of automatically detecting and reacting to errors. These allow you to provide load balancing across two separate Systems and facilitate PCE system failover and recovery.

Comparing with advertisement based PCE discovery [RFC5088][RFC5089],it can mitigate flooding issue (see section 3.2 of [RFC5088])and avoid unwanted traffic and reduce a large amount of unnecessary advertisement, especially when PCE information needs frequent changes.

3.2. Network Address Translation Gateway

PCEP uses TCP as the transport [RFC5440]. To secure TCP connection that underly PCEP sessions, TLS can be used besides using TCP-MD5. When NAT gateway is in place, a TCP or TCP/TLS connection can be opened by ICE for the purpose of connectivity checks. However the TCP connection cannot be established in cases where one of the agents is behind a NAT with connection-dependent filtering properties [RFC5382]. Therefore IGP discovery is limited within an IGP domain and cannot be used in this case.

3.3. Multiple-Provider Domains

Backward recursive path computation (BRPC) [RFC5441] MAY be used by cooperating PCEs to compute inter-domain path. In which case these cooperating PCEs should known to other PCEs. In case of inter-AS where the PCE do not participate in a common IGP, the existing IGP discovery mechanism cannot be used to discover inter-AS PCE.

Also in the case of multiple ASes within different service provider networks, the H-PCE [RFC6805] architecture does not require disclosure of internals of a child domain to the parent PCE. It may

be necessary for a third party to manage the parent PCEs according to commercial and policy agreements from each of the participating service providers.

[RFC6805] specifies that a child PCE must be configured with the address of its parent PCE in order for it to interact with its parent PCE. However handling changes in parent PCE identities and coping with failure events would be an issue for a configured system.

There is no scope for parent PCEs to advertise their presence, however there is potential for directory systems (such as DNS [RFC4848] as used in the ALTO discovery function [I-D.ietf-alto-server-discovery]).

3.4. Multiple PCE Servers

In some cases, each network domain may have multiple PCE server, only one main PCE sever is responsible for Establish topology database by participating in OSPF/ISIS routing protocol, the other PCE server gains knowledge of Topology information either from TED maintained by the main PCE server or some management system(e.g.,NMS/OSS). In such cases, it is desirable to use DNS based mechanism to discover PCE.

3.5. End to End Path Computation

To compute end to end paths across domains, PCE has the following limitations:

- Within a single area, the PCE can not offers enhanced computational power for end to end path computation,e.g., coordination of computation across the whole area.

- A single router participating in IGP area lacks visibility of complete topology with its own TED.

Per domain path computation mechanism[RFC5152]can be used to compute end to end path, however it may lead to sub-optimal paths or result in no end to end path to be found when the PCE only has visibility into the IGP area it serves. This issue can be resolved when one powerful PCE is responsible for multiple areas,i.e., PCE sits in one area it serves and also can get access to topology information provided by PCE server in other IGP area using BGP. In such case, it will be desirable to use DNS based mechanism to discover those PCE that has visibility to multiple areas.

4. Discovering a Path Computation Element

The Dynamic Delegation Discovery System (DDDS) [RFC3401] is used to implement lazy binding of strings to data, in order to support dynamically configured delegation systems. The DDDS functions by mapping some unique string to data stored within a DDDS database by iteratively applying string transformation rules until a terminal condition is reached. When DDDS uses DNS as a distributed database of rules, these rules are encoded using the Naming Authority Pointer (NAPTR) Resource Record (RR). One of these rules is the First Well Known Rule, which says where the process starts.

In current specifications, the First Well Known Rule in a DDDS application [RFC3403] is assumed to be fixed, i.e., the domain in the tree where the lookups are to be routed to, is known. This document proposes the input to the First Well Known Rule to be dynamic, based on the search path the resolver discovers or is configured to use.

The search path of the resolver can either be pre-configured, or discovered using DHCP.

When the PCC or other PCEs needs to discover PCEs in the domain into which the PCEP speaker has visibility (e.g.,local domain), the input to the First Well Known Rule MUST be the domain the PCC knows, which is assumed to be pre- configured in the PCC or discovered using DHCP.

When the PCC needs to discover PCE in the other domain (e.g., AS, Parent PCE in the parent domain)into which the PCC has no visibility, it SHOULD know the domain name of that domain and use DHCP to discover IP address of the PCE in that domain that provides path computation service along with some PCE location information useful to a PCC for PCE selection, and contact it directly. In some instances, the discovery may result in a per protocol/application list of domain names that are then used as starting points for the subsequent NAPTR lookups. If neither the IP address or PCE location information can be discovered with the above procedure, the PCC MAY request a domain search list, as described in [RFC3397] and [RFC3646], and use it as input to the DDDS application.

When the PCC does not find valid domain names using the procedures above, it MUST stop the attempt to discover any PCE.

The dynamic rule described above SHOULD NOT be used for discovering services other than Path computation services described in this document, unless stated otherwise by a future specification.

The procedures defined here result in an IP address, PCE domain, neighboring PCE domain and PCE Computation Scope where the PCC can

contact the PCE that hosts the service the PCC is looking for.

4.1. Determining the PCE Service and transport protocol

The PCC should know the service identifier for the Path Computation Discovery service. The service identifier for the Path Computation Discovery service is defined as "PCED", The PCE supporting "PCED" service MUST support only TCP as transport, as described in [RFC5440].

The services relevant for the task of transport protocol selection are those with NAPTR service fields with values "ID+M2X", where ID is the service identifier defined in the previous section, and X is a letter that corresponds to a transport protocol supported by the domain. This specification only defines M2T for TCP. This document also establishes an IANA registry for mappings of NAPTR service name to transport protocol.

These NAPTR [RFC3403] records provide a mapping from a domain to the SRV [RFC2782] record for contacting a PCE with the specific transport protocol in the NAPTR services field. The resource record MUST contain an empty regular expression and a replacement value, which indicates the domain name where the SRV record for that particular transport protocol can be found. As per [RFC3403], the client discards any records whose services fields are not applicable.

The PCC MUST discard any service fields that identify a resolution service whose value is not "M2T", for values of T that indicate TCP transport protocols supported by the client. The NAPTR processing as described in RFC 3403 will result in the discovery of the most preferred transport protocol of the PCE that is supported by the client, as well as an SRV record for the PCE.

4.2. Determining the IP Address of the PCE

As an example, consider a client that wishes to find "PCED" service in the example.com domain. The client performs a NAPTR query for that domain, and the following NAPTR records are returned:

```
Order Pref Flags  Service     Regexp       Replacement
 IN NAPTR  50   50   "s"  "PCED"    ""
 _PCED._tcp.example.com
 IN NAPTR  90   50   "s"  "PCED+M2T"    ""
 _PCED._tcp.example.com
```

This indicates that the domain does have a PCE providing Path Computation services over TCP, in that order of preference. Since the client only supports TCP, TCP will be used, targeted to a host

determined by an SRV lookup of _PCED._tcp.example.com. That lookup would return:

```
   ;;  Priority  Weight    Port        Target
IN  SRV    0        1      XXXX   server1.example.com
IN  SRV    0        2      XXXX   server2.example.com
```

where XXXX represents the port number at which the service is reachable.

Note that the regexp field in the NAPTR example above is empty. The regexp field MUST NOT be used when discovering path computation services, as its usage can be complex and error prone. Also, the discovery of the path computation service does not require the flexibility provided by this field over a static target present in the TARGET field.

If the client is already configured with the information about which transport protocol is used for a path computation service in a particular domain, it can directly perform an SRV query for that specific transport using the service identifier of the path computation Service. For example, if the client knows that it should be using TCP for path computation service, it can perform a SRV query directly for_PCED._tcp.example.com.

Once the server providing the desired service and the transport protocol has been determined, the next step is to determine the IP address.

According to the specification of SRV RRs in [RFC2782], the TARGET field is a fully qualified domain name (FQDN) that MUST have one or more address records; the FQDN must not be an alias, i.e., there MUST NOT be a CNAME or DNAME RR at this name. Unless the SRV DNS query already has reported a sufficient number of these address records in the Additional Data section of the DNS response (as recommended by [RFC2782]), the PCC needs to perform A and/or AAAA record lookup(s) of the domain name, as appropriate. The result will be a list of IP addresses, each of which can be contacted using the transport protocol determined previously.

4.3. Determining path computation scope,the PCE domains and Neighbor PCE domains

DNS servers MAY use DNS TXT record and add new RRsets to the additional information section that are relevant to the answer and have the same authenticity as the data (the IP Address of the PCE)in the answer section. RRsets include path computation scope, the PCE domains and Neighbor PCE domains associated with the PCE. the PCC MAY

inspect those Additional Information section and be capable of handling responses from nameservers that never fill in the Additional Information part of a response.

5. IANA Considerations

The usage of NAPTR records described here requires well-known values for the service fields for the transport supported by Path Computation Services. The table of mappings from service field values to transport protocols is to be maintained by IANA.

The registration in the RFC MUST include the following information:

Service Field: The service field being registered.

Protocol: The specific transport protocol associated with that service field. This MUST include the name and acronym for the protocol, along with reference to a document that describes the transport protocol.

Name and Contact Information: The name, address, email address, and telephone number for the person performing the registration.

The following values have been placed into the registry:

Service Fields	Protocol
PCED+M2T	TCP

New Service Fields are to be added via Standards Action as defined in [RFC5226].

IANA is also requested to register PCED as service name in the Protocol and Service Names registry.

6. Security Considerations

It is believed that this proposed DNS extension introduces no new security considerations (i.e., A list of known threats to services using DNS) beyond those described in [RFC3833]. For most of those identified threats, the DNS Security Extensions [RFC4033] does provide protection. It is therefore recommended to consider the usage of DNSSEC [RFC4033] and the aspects of DNSSEC Operational Practices [RFC4641] when deploying Path Computation Services.

In deployments where DNSSEC usage is not feasible, measures should be taken to protect against forged DNS responses and cache poisoning as much as possible. Efforts in this direction are documented in [RFC5452].

Where inputs to the procedure described in this document are fed via DHCP, DHCP vulnerabilities can also cause issues. For instance, the inability to authenticate DHCP discovery results may lead to the Path Computation service results also being incorrect, even if the DNS process was secured.

7. Acknowledges

The author would like to thank Claire Bi,Ning Kong and Liang Xia for their review and comments that help improvement to this document.

8. References

8.1. Normative References

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

[RFC2782] Gulbrandsen, A., "A DNS RR for specifying the location of services (DNS SRV)", RFC 2782, February 2000.

[RFC3397] Aboba, B., "Dynamic Host Configuration Protocol (DHCP) Domain Search Option", RFC 3397, November 2002.

[RFC3403] Mealling, M., "Dynamic Delegation Discovery System (DDDS) Part Three: The Domain Name System (DNS) Database", RFC 3403, October 2002.

[RFC3646] Droms, R., "DNS Configuration options for Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3646, December 2003.

[RFC4033] Arends, R., "DNS Security Introduction and Requirements", RFC 4033, March 2005.

[RFC4641] Kolkman, O., "DNSSEC Operational Practices", RFC 4641, September 2006.

[RFC4674] Droms, R., "Requirements for Path Computation Element (PCE) Discovery", RFC 4674, December 2003.

[RFC4848] Daigle, D., "Domain-Based Application Service Location Using URIs and the Dynamic Delegation Discovery Service (DDDS)", RFC 4848, April 2007.

[RFC5226] Narten, T., "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 5226, May 2008.

[RFC5440] Le Roux, JL., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, April 2007.

[RFC6805] King, D. and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, November 2012.

8.2. Informative References

[ALTO] Kiesel, S., "ALTO Server Discovery", ID draft-ietf-alto-server-discovery-08, March 2013.

[RFC3401] Mealling, M., "Dynamic Delegation Discovery System (DDDS) Part One: The Comprehensive DDDS", RFC 3401, October 2002.

[RFC3833] Atkins, D., "Threat Analysis of the Domain Name System (DNS)", RFC 3833, August 2004.

[RFC5088] Le Roux, JL., "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.

[RFC5089] Le Roux, JL., "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.

[RFC5382] Guha, S., "NAT Behavioral Requirements for TCP", RFC 5382, October 2008.

[RFC5452] Hubert, A., "Measures for Making DNS More Resilient against Forged Answers", RFC 5452, January 2009.

Authors' Addresses

Qin Wu
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: sunseawq@huawei.com

Dhruv Dhody
Huawei
Leela Palace
Bangalore, Karnataka 560008
INDIA

Email: dhruv.ietf@gmail.com