IETF Internet Draft PCE Working Group
Proposed Status: Informational
Expires: November 2005

Jerry Ash (AT&T)
Editor
J.L. Le Roux (France Telecom)
Editor

May 2005

draft-ash-pce-comm-protocol-gen-reqs-01.txt

PCE Communication Protocol Generic Requirements

Status of this Memo

Copyright Notice

Abstract

Constraint-based path computation is a fundamental building block for traffic engineering systems such as multiprotocol label switching (MPLS) and generalized multiprotocol label switching (GMPLS) networks. Path computation in large, multi-domain or multi-layer networks is highly complex and may require special computational components and cooperation between the different network domains.

There are multiple components in the Path Computation Element (PCE)-

based path computation model, including PCE discovery and the PCE communication protocol. The PCE model is described in the "PCE Architecture" document and facilitates path computation requests from Path Computation Clients (PCCs) to PCEs. This document specifies generic requirements for a communication protocol between PCCs and PCEs, and between PCEs where cooperation between PCEs is desirable. Subsequent documents will specify application-specific requirements for the PCE communication protocol.

Table of Contents

1. Contributors

This document is the result of the PCE Working Group PCE communication protocol requirements design team joint effort. The following are the design team member authors that contributed to the present document:

Jerry Ash (AT&T)
Alia Atlas (Avici)
Arthi Ayyangar (Juniper)
Nabil Bitar (Verizon)
Igor Bryskin (Independent Consultant)
Dean Cheng (Cisco)
Durga Gangisetti (MCI)
Kenji Kumaki (KDDI)
Jean-Louis Le Roux (France Telecom)
Eiji Oki (NTT)
Raymond Zhang (BT Infonet)

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 RFC 2119 [RFC2119].

3. Introduction

The path computation element (PCE) capability [PCE-ARCH] supports requests for path computation issued by a path computation client (PCC), which may be co-located or remote from a PCE. When the PCC is remote from the PCE, a request/response communications protocol is required to carry the path computation request and return the response. In order for the PCC and PCE to communicate, the PCC must discover the location of the PCE, as described in [PCE-DISC-REQ]. The PCE operates on a network graph in order to compute paths based on the path computation request issued by the PCC, which will normally include the source, destination, and a set of constraints. The PCE response includes the computed paths or the reason for a failed computation.

This document lists a set of generic requirements for the PCE communication protocol, where the PCE communications protocol solution MUST satisfy these requirements. Application-specific requirements are beyond the scope of this document, and will be addressed in separate documents.

4. Terminology

Domain: any collection of network elements within a common sphere of

address management or path computational responsibility. Examples of domains include IGP areas, Autonomous Systems (ASs), multiple ASs

within a service provider network, or multiple ASs across multiple service provider networks.

GMPLS: generalized multiprotocol label switching

LSP: MPLS Label Switched Path.

MPLS: multiprotocol label switching

PCC: Path Computation Client: any client application requesting a Path computation to be performed by the PCE.

PCE: Path Computation Element: an entity (component, application or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints (see further description in [PCE-ARCH]).

TED: Traffic Engineering Database, which contains the topology and resource information of the network or network segment used by a PCE.

TE LSP: Traffic Engineering MPLS Label Switched Path.

See [PCE-ARCH] for further definitions of terms.

5. Overview of PCE Communication Protocol

In the PCE model, path computation requests are issued by a PCC to a PCE that may be co-located or situated at a remote site. If the PCC and PCE are not co-located a request/response communications protocol is required to carry the request and return the response. If the PCC and PCE are co-located a communications protocol is not required, but implementations may choose to utilize a protocol for exchanges between the components.

In order that a PCC and PCE can communicate, the PCC must know the location of the PCE. This can be configured or discovered. The PCE discovery mechanism is out of scope of this document, but requirements are documented in [PCE-DISC-REQ].

The PCE operates on a network graph built from the TED in order to compute paths. The mechanism by which the TED is populated is out of scope for the PCE Communications Protocol.

A path computation request issued by the PCC will include a specification of the path(s) needed. The information supplied will include at a minimum the source and destination for the path(s), but may also include a set of further requirements (known as constraints) as described in Section 6.

The response from the PCE may be positive in which case it will include the paths that have been computed. If the computation fails

or cannot be performed, a negative response is required with an indication of the type of and reason(s) for the failure. A negative response may also include further details of the reason(s) for the failure, and potentially advice about which constraints might be relaxed to be more likely to achieve a positive result. That is, the PCE SHOULD provide sufficient information for the PCC to know whether it has to relax constraints or query another PCE.

A request/response protocol is also required for a PCE to communicate path computation requests to another PCE and for the PCE to return the path computation response. As described in [PCE-ARCH], there is no reason to assume that two different protocols are needed, and this document assumes that a single protocol will satisfy all requirements for PCC-PCE and PCE-PCE communications.

[PCE-ARCH] describes four models of PCE: composite, external, multiple PCE path computation and multiple PCE path computation with inter-PCE communication. In all cases except the composite PCE model, a communication protocol is required. The requirements defined in this document therefore are applicable to all models described in the [PCE-ARCH] except the composite PCE model.

6. PCE Communication Protocol Generic Requirements

The designers of a PCE communication protocol MUST take the requirements set out in this document and discuss them widely within the IETF and particularly within the Applications Area to determine whether a suitable protocol already exists. The results of this investigation MUST be published on the PCE mailing list.

6.1 Basic Protocol Requirements

6.1.1 Client-Server Communication

PCC-PCE and PCE-PCE communication is by nature client-server based. The communication protocol MUST allow for a PCC or a PCE to send a path request message to a PCE, and for a PCE to reply with a path response message to the requesting PCC or PCE, once the path has been computed. In addition to this request-response model, there may be cases where there is unsolicited communication from the PCE to PCC (see Requirement 6.1.6).

The protocol MUST be capable of returning any explicit path that would be acceptable for use for MPLS and GMPLS LSPs once converted to an Explicit Route Object for use in RSVP-TE signaling. Note that the resultant path(s) may be made up of a set of strict or loose hops, or any combination of strict and loose hops. Moreover, a hop may have the form of a non-explicit abstract node. See RFC 3209 for the definition of strict hop, loose hop, and abstract node.

It MUST be possible to send multiple path computation requests,

correlated or not, within the same path request message. There are various motivations for doing so (optimality, path diversity, etc.).

It MUST be possible to limit by configuration the number of requests that can be carried within a single message. The transport protocol MUST allow sending unlimited size messages, but MUST be able to limit message size, to avoid a big message from unduly delaying a small message. Maximum message size MAY be negotiated at session initialization. If the number of correlated requests exceeds the maximum message size, then separate messages MAY be sent with an indication that they are correlated.

The path request message MUST include, at least, a source and a destination, and MAY include a set of one or more path constraints, such as the requested bandwidth or resources (hops, affinities, etc.) to include/exclude (e.g., a PCC requests the PCE to exclude points of failure in the computation of the new path if an LSP setup fails).

The path request message MUST support the ability to prefer/customize various path computation objective functions, policies and optimization criteria. For example, a PCC may be aware of and would like to choose from among various objective functions that a PCE may offer, and the PCE communication protocol SHOULD allow this to be specified per path computation request. This capability to prefer certain objective functions depends on the fact that the PCE advertises this to a PCC or that the PCC requests one of a set of objective functions defined as a minimal subset that MUST be supported by any PCE.

The requester MUST be allowed to select from the advertised list or minimal subset of standard objective functions and functional options. The requester SHOULD also be able to select a vendor-specific or experimental objective function or functional option. Furthermore, the requester MUST be allowed to customize the objective function/options in use. That is, individual objective functions will often have parameters to be set in the request from PCC to PCE. Specification of objective functions and objective function parameters is required in the protocol extensibility specified in Section 6.1.9.

If a PCC selects an objective function that the PCE does not support, the PCE response MUST be negative.

Note that a PCC MAY send a request that is based on the set of TE parameters carried by the MPLS/GMPLS LSP setup signaling protocol, and as long as those parameters are satisfied, the PCC MAY not care about which objective function is used. Also, the PCE MAY execute objective functions not advertised to the PCC, for example, policy

based routing path computation for load balancing instructed by the management plane.

A PCC or PCE MUST be able to cancel a pending request.

The path response message MUST allow returning various elements including, at least, the computed path. It MUST be possible to return multiple paths within the same path response message, corresponding either to the same request (e.g. load balancing) or to distinct requests of the same path request message or distinct path request messages.

6.1.2 PCC-PCE and PCE-PCE Communication

A single protocol MUST be defined for PCC-PCE and PCE-PCE communication. A PCE requesting a path from another PCE can be considered as a PCC.

6.1.3 Reliable Message Exchange

The PCE communication protocol MUST run on top of a reliable transport protocol. In particular, it MUST allow for the detection and recovery of lost messages to occur quickly and not impede the operation of the communication protocol. Here the PCE communication protocol includes a number of application-specific capabilities, all of which run on top of a common, reliable transport protocol layer.

In some particular cases (e.g. link failure), a large number of PCCs may simultaneously send a request to a PCE, leading potentially to a saturation of request buffers on PCEs. The PCE communication protocol MUST properly handle such overload situations without a significant decrease in performance, such as through throttling of such requests.

The PCE communication-protocol transport MUST provide:

- acknowledged message delivery with retransmission, as discussed in Section 6.1.1
- in order message delivery. For the set of requests between a given PCC and a PCE, the ordering is already there relying on the reliable transport layer. For requests between a set of PCCs and a given PCE, the ordering of responses SHOULD be based on the PCE's own handling policy, as well as the priority of the requests.
- message corruption detection
- flow control and back-pressure, as specified above with the throttling of requests.

These requirements SHOULD be satisfied by an existing reliable transport protocol, and functionality SHOULD only be added where the transport protocol does not provide it (e.g., rapid partner failure detection). With regard to the rapid partner failure detection, the PCC MUST be informed of any failed PCE (or PCE connection) when it

happens.

6.1.4 Secure Message Exchange

The PCC-PCE and PCE-PCE communication MUST be secure. In particular, it MUST support mechanisms to prevent spoofing (e.g., authentication), snooping (e.g., encryption) and DOS attacks.

6.1.5 Request Prioritization

The communication protocol MUST support the notion of request priority, allowing a PCC to specify the degree of urgency of a particular request. This is used to serve some requests before others, and would require global prioritization. That is, a request from one PCC can have a higher priority than a request from another PCC to the same PCE. However, there is no intention or need for a PCE to preempt (i.e., discard) a given request from one PCC if it receives a higher-priority request from another PCC; the PCE just delays the lower-priority request.

If, for example, the PCE is processing a low priority request that will take extended computation time (e.g., for full re-optimization of 1000 protected LSPs through a complex algorithm), it is RECOMMENDED that the low priority request to set up a new LSP be suspended/interrupted until the high priority request can be completed. The PCE must consider, however, in addition to the priority of the path computations, the PCE policy based on its system resources, configurations, etc. That is, the handling of priority on the PCE is not entirely in the purview of the PCE communication protocol design.

The PCE communication protocol design MUST consider whether request if starvation can occur for particular priorities, whether that is acceptable, and how that is handled.

6.1.6 Unsolicited Notifications

The PCE communication protocol SHOULD support unsolicited notifications from PCE to PCC or from PCE to PCE. That is, the normal mode is for the PCC to make path computation requests to the PCE. This requirement includes cases of PCEs computing paths without being asked by a PCC, and the PCE sending those unsolicited paths to PCCs. This could also include PCE overload notifications.

6.1.7 Asynchronous Communication

The PCC-PCE protocol MUST allow for asynchronous communication. A client MUST NOT have to wait for a response to make another request. Also it MUST be possible to have the order of some responses differ from the order of their corresponding requests. This may occur, for instance, when path request messages have distinct priorities (see

Requirement 6.1.5).

6.1.8 Communication Overhead Minimization

The request and response messages SHOULD be designed so that the communication overhead is minimized. Particular attention SHOULD be given to the message size. Other considerations in overhead minimization include the following:

- the number of messages exchanged to arrive at a computation answer
- the amount of background messages to keep the session up
- the processing cost at the PCE (or PCC) associated with requests/responses.

6.1.9 Extensibility

The PCE communication protocol MUST provide a way for introduction of new path computation constraints, diversity types, objective functions, optimization methods and parameters, etc., without requiring modifications in the protocol. In particular, the PCE communication protocol SHOULD allow supporting future applications not currently in the scope of the PCE working group, such as, for instance, P2MP path computations.

The communication protocol MUST allow supporting various PCE based applications that have been currently identified and MAY be identified in the future, such as:

- intra-area path computation
- inter-area path computation
- inter-AS intra provider and inter-AS inter-provider path computation
- multi-layer and virtual network topology computation

Note that application specific requirements are out of the scope of this document and will be addressed in separate requirements documents.

6.1.10 Scalability

The PCE communication protocol MUST scale well with an increase of any of the following parameters:

- number of PCCs
- number of PCEs
- number of PCCs communicating with a single PCE
- number of PCEs communicated to by a single PCC
- number of PCEs communicated to by another PCE.
- TED size (number of links/nodes, which may drive up path computation time)
- number of domains

- number of path requests
- handling bursts of requests

Bursts of requests may arise, for example, after a network outage when multiple recomputations are requested as a result. It is RECOMMENDED that the protocol handle the congestion in a graceful way so that it does not unduly impact the rest of the network, and so that it does not gate the ability of the PCE to perform computation.

6.2 Deployment Support Requirements

6.2.1 Support for Various Service Provider Environments and Applications

The communication protocol MUST operate in various service provider network environments, where the IP control plane is deployed, such as

- MPLS-TE and GMPLS networks
- centralized and distributed PCE path computation
- single and multiple PCE path computation

Definitions of centralized, distributed, single, and multiple PCE path computation can be found in [PCE-ARCH].

6.2.2 Confidentiality

The communication protocol MUST allow minimizing the amount of topological information exchanged between a PCC and PCE, and between PCEs. This is of particular importance in inter-PCE communication, where the PCEs are located in distinct service-provider domains. For example, the protocol design SHOULD enable policies to be implemented such that domain-specific topology information is excluded on inter-PCE, inter-domain communication.

6.2.3 Policy Support

The communication protocol MUST allow for policies to accept/reject requests, and include the ability for a PCE to reject requests with sufficient detail to allow the PCC to determine the reason for rejection or failure. For example, filtering could be required for intra-AS PCE path computation such that all requests are rejected that come from another AS. However, specific policy details are left to application-specific communication protocol requirements. Furthermore, the communication protocol MUST allow for the notification of a policy violation. Actual policies, configuration of policies, and applicability of policies are out of scope.

6.3 Detection & Recovery Requirements

6.3.1 Aliveness Detection

The PCE communication protocol MUST allow a PCC to check the

liveliness of PCEs it is using for path computation and a PCE to check the liveliness of PCCs it is serving. The PCE communication

protocol MUST provide partner failure detection.

Depending on the design, this requirement MAY be met by the PCE communication protocol design or the transport protocol design.

6.3.2 PCC/PCE Failure Response

Appropriate PCC and PCE procedures MUST be defined to deal with PCE and PCC failures. A PCC MUST be able to clear any pending request to a PCE. That is, the PCC MAY cancel a previously-made path computation request to a PCE.

Similarly, a PCE MUST be able to clear pending requests from a PCC, for instance, when it detects the failure of the requesting PCC or when its buffer of requests is full. It is RECOMMENDED that a PCC select another PCE upon detection of PCE failure or unreachability of a PCE but note that PCE selection procedure are out of the scope of this document.

It is assumed that the underlying reliable communication mechanism ensures reciprocal knowledge of PCE and PCC liveness. Therefore it NOT possible for the PCC/PCE to believe that the PCE/PCC is unreachable, but not vice versa.

6.3.3 Protocol Recovery

Information distributed in asynchronous/unsolicited messages SHOULD be allowed to persist at the recipient in the event of the failure of the sender or of the communications channel. Upon recovery, the communications protocol MUST support resynchronization of information between the sender and the receiver, and this SHOULD be arranged so as to minimize repeat data transfer.

For example, the communication protocol SHOULD allow a stateful PCE to resynchronize and recover states (e.g., LSP status, paths, etc.) after a restart. Recovery would require the PCE communication protocol to support recovery of state information in the PCE. This would be of particular importance when local PCE recovery is not supported or fails.

7. Security Considerations

The impact of the use of a PCE-based architecture MUST be considered in the light of the impact that it has on the security of the existing routing and signaling protocols and techniques in use within the network. There is unlikely to be any impact on intra-domain security, but an increase in inter-domain information flows and the facilitation of inter-domain path establishment may increase the vulnerability to security attacks.

Of particular relevance are the implications for confidentiality

inherent in a PCE-based architecture for multi-domain networks. It is not necessarily the case that a multi-domain PCE solution will compromise security, but solutions MUST examine their impacts in this area.

Applicability statements for particular combinations of signaling, routing and path computation techniques are expected to contain detailed security sections.

It should be observed that the use of a non-local PCE (that is, not co-resident with the PCC) does introduce additional security issues. Most notable amongst these are:

- interception of PCE requests or responses
- impersonation of PCE
- falsification of TE information
- denial of service attacks on PCE or PCE communication mechanisms

It is expected that PCE solutions will address these issues in detail using authentication and security techniques.

8. Manageability Considerations

Manageability of the PCE communication protocol MUST address the following considerations:

- need for a MIB module for control and monitoring
- need for built-in diagnostic tools (e.g., partner failure detection, OAM, etc.)
- configuration implications for the protocol

9. IANA Considerations

This document makes no requests for IANA action.

10. Acknowledgements

The authors would like to extend their warmest thanks to (in alphabetical order) Adrian Farrel, Thomas Morin, and JP Vasseur for their review and suggestions.

11. Normative References

[PCE-ARCH] Farrel, A., Vasseur, JP, Ash, J., "Path Computation Element (PCE) Architecture", work in progress.

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

[RFC3667] Bradner, S., "IETF Rights in Contributions", BCP 78, RFC 3667, February 2004.

[RFC3668] Bradner, S., "Intellectual Property Rights in IETF Technology", BCP 79, RFC 3668, February 2004.

12. Informational References

[PCE-DISC-REQ] Le Roux, JL, et. al., "Requirements for Path Computation Element (PCE) Discovery," work in progress.

[RFC3209] Awduche, D., et. al., "RSVP-TE: Extensions to RSVP for LSP Tunnels," RFC 3209, December 2001.

13. Authors' Addresses

Jerry Ash
AT&T
Room MT D5-2A01
200 Laurel Avenue
Middletown, NJ 07748, USA
Phone: +1-(732)-420-4578
Email: gash@att.com

Alia K. Atlas
Avici Systems, Inc.
101 Billerica Avenue
N. Billerica, MA 01862, USA
Phone: +1 978 964 2070
Email: aatlas@avici.com

Arthi Ayyangar
Juniper Networks, Inc.
1194 N.Mathilda Ave
Sunnyvale, CA 94089 USA
Email: arthi@juniper.net

Nabil Bitar
Verizon
40 Sylvan Road
Waltham, MA 02145
Email: nabil.bitar@verizon.com

Igor Bryskin
Independent Consultant
Email: i_bryskin@yahoo.com

Dean Cheng
Cisco Systems Inc.
3700 Cisco Way
San Jose CA 95134 USA

Phone: +1 408 527 0677
Email: dcheng@cisco.com

Durga Gangisetti
MCI
Email: durga.gangisetti@mci.com

Kenji Kumaki
KDDI Corporation
Garden Air Tower
Iidabashi, Chiyoda-ku,
Tokyo 102-8460, JAPAN
Phone: +81-3-6678-3103
Email: ke-kumaki@kddi.com

Jean-Louis Le Roux
France Telecom
2, avenue Pierre-Marzin
22307 Lannion Cedex, FRANCE
Email: jeanlouis.leroux@francetelecom.com

Eiji Oki
NTT
Midori-cho 3-9-11
Musashino-shi, Tokyo 180-8585, JAPAN
Email: oki.eiji@lab.ntt.co.jp

Raymond Zhang
BT INFONET Services Corporation
2160 E. Grand Ave.
El Segundo, CA 90245 USA
Email: Raymond_zhang@bt.infonet.com

14. Intellectual Property Statement

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary

rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

Copyright Statement