Abstract

In order to populate network performance information like link latency, latency variation, packet loss and bandwidth into Traffic Engineering Database (TED) and ALTO server, this document describes extensions to BGP protocol, that can be used to distribute network performance information (such as link delay, delay variation, packet loss, residual bandwidth, available bandwidth and utilized bandwidth).

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 July 8, 2015.
1. Introduction

As specified in [RFC4655], a Path Computation Element (PCE) is an entity that is capable of computing a network path or route based on a network graph, and of applying computational constraints during the computation. In order to compute an end to end path, the PCE needs to have a unified view of the overall topology[I-D.ietf-pce-pcep-service-aware]. [I.D-ietf-idr-ls-distribution] describes a mechanism by which links state and traffic engineering information can be collected from networks and shared with external components using the BGP routing protocol. This mechanism can be used by both PCE and ALTO server to gather information about the topologies and capabilities of the network.
With the growth of network virtualization technology, the Network performance or QoS requirements such as latency, limited bandwidth, packet loss, and jitter, for real traffic are all critical factors that must be taken into account in the end to end path computation and selection ([I-D.ietf-pce-pcep-service-aware]) which enable optimizing resource usage and degrading gracefully during period of heavy load.

In order to populate network performance information like link latency, latency variation, packet loss and bandwidth into TED and ALTO server, this document describes extensions to BGP protocol, that can be used to distribute network performance information (such as link delay, delay variation, packet loss, residual bandwidth, available bandwidth, and utilized bandwidth). The network performance information can be distributed in the same way as link state information distribution, i.e., either directly or via a peer BGP speaker (see figure 1 of [I.D-ietf-idr-1s-distribution]).

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. Use Cases

3.1. MPLS-TE with H-PCE

For inter-AS path computation the Hierarchical PCE (H-PCE) [RFC6805] may be used to compute the optimal sequence of domains. Within the H-PCE architecture, the child PCE communicates domain connectivity information to the parent PCE, and the parent PCE will use this information to compute a multi-domain path based on the optimal TE links between domains [I.D-ietf-pce-hierarchy-extensions] for the end-to-end path.

The following figure demonstrates how a parent PCE may obtain TE performance information beyond that contained in the LINK_STATE attributes [I.D-ietf-idr-1s-distribution] using the mechanism described in this document.
3.2. ALTO Server Network API

The ALTO Server can aggregate information from multiple systems to provide an abstract and unified view that can be more useful to applications.

The following figure shows how an ALTO Server can get TE performance information from the underlying network beyond that contained in the LINK_STATE attributes [I.D-ietf-idr-ls-distribution] using the mechanism described in this document.

Figure 2: ALTO Server using network performance information
4. Carrying TE Performance information in BGP

This document proposes new BGP TE performance TLVs that can be announced as attribute in the BGP-LS attribute (defined in [I-D.ietf-idr-ls-distribution]) to distribute network performance information. The extensions in this document build on the ones provided in BGP-LS [I.D-ietf-idr-ls-distribution] and BGP-4 [RFC4271].

BGP-LS attribute defined in [I.D-ietf-idr-ls-distribution] has nested TLVs which allow the BGP-LS attribute to be readily extended. This document proposes seven additional TLVs as its attributes:

<table>
<thead>
<tr>
<th>Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD1</td>
<td>Unidirectional Link Delay</td>
</tr>
<tr>
<td>TBD2</td>
<td>Min/Max Unidirectional Link Delay</td>
</tr>
<tr>
<td>TBD3</td>
<td>Unidirectional Delay Variation</td>
</tr>
<tr>
<td>TBD4</td>
<td>Unidirectional Packet Loss</td>
</tr>
<tr>
<td>TBD5</td>
<td>Unidirectional Residual Bandwidth</td>
</tr>
<tr>
<td>TBD6</td>
<td>Unidirectional Available Bandwidth</td>
</tr>
<tr>
<td>TBD7</td>
<td>Unidirectional Utilized Bandwidth</td>
</tr>
</tbody>
</table>

As can be seen in the list above, the TLVs described in this document carry different types of network performance information. Some of these TLVs include a bit called the Anomalous (or "A") bit at the left-most bit after length field of each TLV defined in figure 4 of [[I.D-ietf-idr-ls-distribution]]. The other bits in the first octets after length field of each TLV is reserved for future use. When the A bit is clear (or when the TLV does not include an A bit), the TLV describes steady state link performance. This information could conceivably be used to construct a steady state performance topology for initial tunnel path computation, or to verify alternative failover paths.

When network performance downgrades and exceeds configurable maximum thresholds, a TLV with the A bit set is advertised. These TLVs could be used by the receiving BGP peer to determine whether to redirect failing traffic to a backup path, or whether to calculate an entirely new path. If link performance improves later and falls below a configurable value, that TLV can be re-advertised with the Anomalous bit cleared. In this case, a receiving BGP peer can conceivably do...
whatever re-optimization (or failback) it wishes to do (including nothing).

Note that when a TLV does not include the A bit, that TLV cannot be used for failover purposes. The A bit was intentionally omitted from some TLVs to help mitigate oscillations.

Consistent with existing ISIS TE specifications [ISIS-TE-METRIC], the bandwidth advertisements, the delay and delay variation advertisements, packet loss defined in this document MUST be encoded in the same unit as one defined in IS-IS Extended IS Reachability sub-TLVs [ISIS-TE-METRIC]. All values (except residual bandwidth) MUST be obtained by a filter that is reasonably representative of an average or calculated as rolling averages where the averaging period MUST be a configurable period of time. The measurement interval, any filter coefficients, and any advertisement intervals MUST be configurable per sub-TLV in the same way as ones defined in section 5 of [ISIS-TE-METRIC].

5. Attribute TLV Details

Link attribute TLVs defined in section 3.2.2 of [I-D.ietf-idr-ls-distribution] are TLVs that may be encoded in the BGP-LS attribute with a link NLRI. Each ‘Link Attribute’ is a Type/Length/Value (TLV) triplet formatted as defined in Section 3.1 of [I-D.ietf-idr-ls-distribution]. The format and semantics of the ‘value’ fields in ‘Link Attribute’ TLVs correspond to the format and semantics of value fields in IS-IS Extended IS Reachability sub-TLVs, defined in [RFC5305]. Although the encodings for ’Link Attribute’ TLVs were originally defined for IS-IS, the TLVs can carry data sourced either by IS-IS or OSPF.

The following ’Link Attribute’ TLVs are valid in the LINK_STATE attribute:
Table 1: Link Attribute TLVs

6. Manageability Considerations

Manageability Considerations described in section 6.2 of [I-D.ietf-idr-ls-distribution] can be applied to Traffic Engineering (TE) performance Metrics as well.

7. Security Considerations

This document does not introduce security issues beyond those discussed in [I.D-ietf-idr-ls-distribution] and [RFC4271].

8. IANA Considerations

IANA maintains the registry for the TLVs. BGP TE Performance TLV will require one new type code per TLV defined in this document.
9. References

9.1. Normative References

[I-D.ietf-idr-ls-distribution]

[I-D.ietf-pce-pcep-service-aware]

[ISIS-TE-METRIC]


9.2. Informative References


[I.D-ietf-pce-hierarchy-extensions]

Appendix A. Change Log

Note to the RFC-Editor: please remove this section prior to publication as an RFC.

A.1. draft-ietf-idr-te-pm-bgp-00

The following are the major changes compared to previous version draft-wu-idr-te-pm-bgp-03:

- Update PCE case in section 3.1.
- Add some texts in section 1 and section 4 to clarify from where to distribute pm info and measurement interval and method.

A.2. draft-ietf-idr-te-pm-bgp-02

The following are the major changes compared to previous version draft-wu-idr-te-pm-bgp-03:

- Some Editorial changes.

Authors’ Addresses

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

Email: bill.wu@huawei.com

Stefano Previdi
Cisco Systems, Inc.
Via Del Serafico 200
Rome 00191
Italy

Email: sprevidi@cisco.com