

IDR Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 13, 2014

Q. Wu
D. Wang
Huawei
S. Previdi
Cisco
H. Gredler
Juniper
S. Ray
Cisco
October 10, 2013

**BGP attribute for North-Bound Distribution of Traffic Engineering (TE)
performance Metrics
draft-wu-idr-te-pm-bgp-02**

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, and available 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 April 13, 2014.

Copyright Notice

Copyright (c) 2013 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 [3](#)
- [2.](#) Conventions used in this document [4](#)
- [3.](#) Use Cases [5](#)
 - [3.1.](#) MPLS-TE with PCE [5](#)
 - [3.2.](#) ALTO Server Network API [5](#)
- [4.](#) Carrying TE Performance information in BGP [7](#)
- [5.](#) Attribute TLV Details [9](#)
- [6.](#) Security Considerations [10](#)
- [7.](#) IANA Considerations [11](#)
- [8.](#) References [12](#)
 - [8.1.](#) Normative References [12](#)
 - [8.2.](#) Informative References [12](#)
- [Appendix A.](#) Change Log [13](#)
 - [A.1.](#) [draft-wu-idr-te-pm-bgp-02](#) [13](#)
- Authors' Addresses [14](#)

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 needs for inter-connection between various overlay technologies (e.g. Enterprise BGP/MPLS IP VPNs) in the Wide Area Network (WAN) become important. The Network performance or QoS requirements such as latency, limited bandwidth, packet loss, and jitter, are all critical factors that must be taken into account in the end to end path computation ([[I-D.ietf-pce-pcep-service-aware](#)]) and selection which enable establishing segment overlay tunnel between overlay nodes and stitching them together to compute end to end path.

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, and available bandwidth).

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 PCE

In inter-AS path computation, PCE in each AS participant in different IGP. In Hierarchy of PCE, A child PCE must be configured with the address of its parent PCE[RFC6805]. Configuration system is challenged by handling changes in parent PCE identities and coping with failure events, especially when parent PCE and child PCE are not a part of the same routing domain.

The following figure shows how a PCE can get its TE performance information beyond that contained in the LINK_STATE attributes [I.D-ietf-idr-ls-distribution] using the mechanism described in this document.

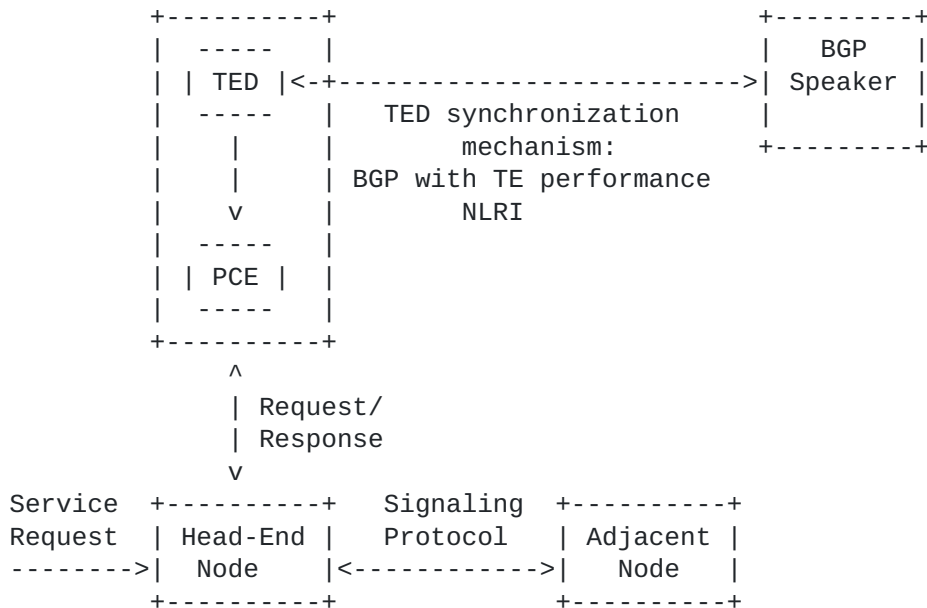


Figure 1: External PCE node using a TED synchronization mechanism

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.

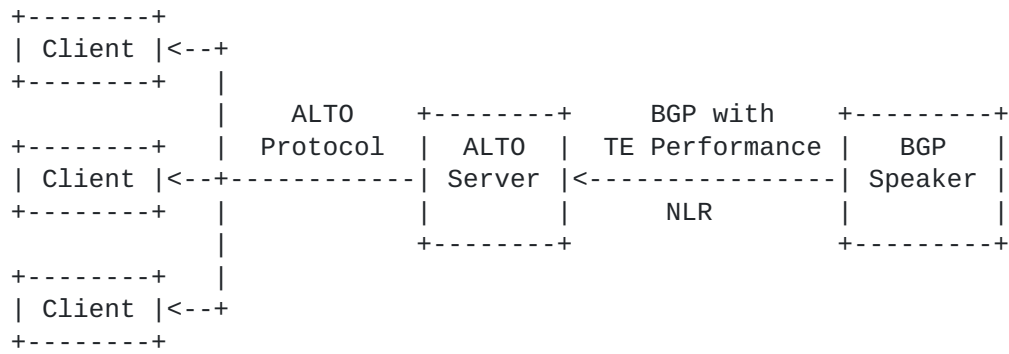


Figure 2: ALTO Server using network performance information

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 six additional TLVs as its attributes:

Type	Value
TBD1	Unidirectional Link Delay
TBD2	Min/Max Unidirectional Link Delay
TBD3	Unidirectional Delay Variation
TBD4	Unidirectional Packet Loss
TBD5	Unidirectional Residual Bandwidth
TBD6	Unidirectional Available Bandwidth

[Editor Note: When this draft(v-01) was presented in the IDR WG session of Berlin meeting, John Scudder suggested to define new attributes(i.e., link utilization attribute, channel throughput attribute) added in the previous version of this draft in the [draft-ietf-isis-te-metric-extensions](#). After Berlin meeting, Hannes Gredler help initiate discussion with authors of IGP drafts(i.e., [draft-ietf-isis-te-metric-extensions](#) and [draft-ietf-ospf-te-metric-extensions](#)) on why two additional attributes should be added into IGP draft. After a few offline discussion with authors of IGP drafts, specially with John Drake, David Ward, Alia Atlas, Stefano Previdi, it was roughly agreed that

- o drop channel throughput attribute since it is node attribute rather than link attribute.
- o and add link utilization attribute into IGP drafts.

However the open issue is whether defining total Link Utilization as Currently Utilized Bandwidth or as Currently Utilized Bandwidth / Maximum Bandwidth. Until this open issue is resolved, the link utilization attribute will be added into the update of this draft as seventh additional TLV.]

As can be seen in the list above, the TLVs described in this document carry different types of network performance information. These TLVs include a bit called the Anomalous (or "A") bit at the left-most bit after length field of each TLV. 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, packetloss 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 calculated as rolling averages where the averaging period MUST be a configurable period of time.

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 some '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:

TLV Code Point	Description	IS-IS TLV/Sub-TLV	Defined in:
xxxx	Unidirectional Link Delay	22/xx	[ISIS-TE]/4.1
xxxx	Min/Max Unidirectional Link Delay	22/xx	[ISIS-TE]/4.2
xxxx	Unidirectional Delay Variation	22/xx	[ISIS-TE]/4.3
xxxx	Unidirectional Link Loss	22/xx	[ISIS-TE]/4.4
xxxx	Unidirectional Residual Bandwidth	22/xx	[ISIS-TE]/4.5
xxxx	Unidirectional Available Bandwidth	22/xx	[ISIS-TE]/4.6

Table 1: Link Attribute TLVs

[Editor Note: The open issue is whether defining total Link Utilization as Currently Utilized Bandwidth or as Currently Utilized Bandwidth / Maximum Bandwidth? We will add link utilization attribute as seventh additional attribute(e.g.,Currently Utilized Bandwidth) when the open issue is resolved.]

6. Security Considerations

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

7. 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.

8. References

8.1. Normative References

- [I-D.ietf-idr-ls-distribution]
Gredler, H., "North-Bound Distribution of Link-State and TE Information using BGP",
ID [draft-ietf-idr-ls-distribution-03](#), May 2013.
- [I-D.ietf-pce-pcep-service-aware]
Dhruv, D., "Extensions to the Path Computation Element Communication Protocol (PCEP) to compute service aware Label Switched Path (LSP)",
ID [draft-ietf-pce-pcep-service-aware-01](#), July 2013.
- [ISIS-TE-METRIC]
Giacalone, S., "ISIS Traffic Engineering (TE) Metric Extensions", ID [draft-ietf-isis-te-metric-extensions-00](#),
June 2013.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", March 1997.
- [RFC4271] Rekhter, Y., "A Border Gateway Protocol 4 (BGP-4)",
[RFC 4271](#), January 2006.
- [RFC5305] Li, T., "IS-IS Extensions for Traffic Engineering",
[RFC 5305](#), October 2008.

8.2. Informative References

- [ALTO] Yang, Y., "ALTO Protocol",
ID <http://tools.ietf.org/html/draft-ietf-alto-protocol-16>,
May 2013.
- [RFC4655] Farrel, A., "A Path Computation Element (PCE)-Based Architecture", [RFC 4655](#), August 2006.

Appendix A. Change Log

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

A.1. draft-wu-idr-te-pm-bgp-02

The following are the major changes compared to previous version 01:

- o Taking out link utilization metric and channel throughput metric from this version and will add link utilization metric back to the update when there was agreement on what measurement unit is used for link utilization.
- o Some additional texts in BGP extension [section 4](#) to explain how to position 'A' bit in the BGP TE performance TLV.
- o Add two editor notes to explain the status of this draft and open issue that need be resolved.
- o Some additional text in the use case sections to clarify how to use these TE performance metrics.

Authors' Addresses

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

Email: sunseawq@huawei.com

Danhua Wang
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: wangdanhua@huawei.com

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

Email: sprevidi@cisco.com

Hannes Gredler
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: hannes@juniper.net

Saikat Ray
Cisco Systems, Inc.
170, West Tasman Drive
San Jose, CA 95134
US

Email: sairay@cisco.com