

idr
Internet-Draft
Updates: 9085 (if approved)
Intended status: Standards Track
Expires: 3 September 2026

K. Kompella
Z. Zhang
HPE
2 March 2026

Using BGP to Maintain and Update a Traffic Engineering Database
draft-kompella-idr-bgp-ted-00

Abstract

In some networking situations, most commonly in Data Centers, no IGP protocol is run. The preferred option is to run BGP to exchange reachability information. If it is also desirable to run Traffic Engineering, a Traffic Engineering Database is needed; this information is typically exchanged via IGP TE extensions. This memo proposes elements of BGP procedure to carry this information when there is no IGP.

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 <https://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 3 September 2026.

Copyright Notice

Copyright (c) 2026 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 (<https://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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	2
1.1. Mode of Operation	3
2. Theory of Operation	3
2.1. Generating Topology/TE Information	3
2.2. Propagating the TED	3
2.3. Using the TED	4
3. Security Considerations	4
4. IANA Considerations	4
5. References	4
5.1. Normative References	4
5.2. Informative References	4
Authors' Addresses	5

1. Introduction

BGP is the protocol of choice to carry reachability information in some networks (such as Data Centers (DCs)). In such cases, no IGP is run (it would be nice to write down the rationale for this decision, which seems quite pervasive). If, however, Traffic Engineering (TE) is deemed useful in these same networks (consider [I-D.kompella-rtgwg-mlnwsched]), simple reachability information is insufficient. What is needed is topology information, including node and link attributes, with which to compute TE paths or Directed Acyclic Graphs (DAGs). This information comprises the Traffic Engineering Database (TED).

Fortunately, there is a way to carry the TED via BGP, namely, BGP Link State (BGP-LS [RFC9085]); however, this typically requires an IGP from which BGP-LS sources its information. BGP-LS is primarily used to propagate the TED to "listeners" such as PCE engines [RFC5440].

This memo describes how BGP-LS can be used without a backing IGP.

1.1. Mode of Operation

In DCs, BGP is typically as "external BGP", i.e., each device has a unique ASN, and has an eBGP session with each of its neighbors. This session has the aforementioned reachability information whereby each device can communicate with every other device. This same session can be used to carry BGP-LS using the magic of Multiprotocol BGP [RFC4760]. Every device in the network would thus obtain the node and link attributes of all devices, and thus can build up the TED.

However, there is some chance that the information carried in MP-BGP can cause one or more sessions to go down. As the sessions also carry reachability information, this could cause one or more devices to lose connectivity with others. This in turn could lead to packets being dropped (or looped) and DC workloads being delayed or aborted.

To avoid overloading the eBGP reachability sessions with TED information, one can have a parallel eBGP session for the TED. More efficiently, one can have an iBGP session between each device and a small number of Route Reflectors (RRs) (for redundancy). This avoids doubling the number of BGP sessions and alleviates the burden of propagating TED information from all the devices -- this task is limited to the RRs. Such iBGP sessions can also be used for signaling TE DAGs ([I-D.zhang-idr-mp-te-signaling]), which can benefit even more by the use of RRs.

2. Theory of Operation

2.1. Generating Topology/TE Information

Each device that wants to contribute to the TED must be configured to advertise its own node and link attributes. Typically, such configuration is sent via an IGP, but in this case, this information is sent via BGP over each of its sessions.

The BGP-LS AFI/SAFI is used to carry the device's TE information.

2.2. Propagating the TED

In the case eBGP sessions are used, in addition to generating its own TE information, each device is responsible for propagating the TE information it receives from its neighbors. If iBGP session with a set of RRs is the mode of operation, then this is not necessary.

2.3. Using the TED

Note that "regular" SPF is not the objective here. Rather, some set of devices are configured with TE constraints that define the requirements of the TE tunnels. Having received all the BGP-LS updates, these nodes will (eventually) have the full TED, and can proceed to compute TE DAGs that meet the requirements.

3. Security Considerations

To be added.

4. IANA Considerations

To be added.

5. References

5.1. Normative References

- [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <<https://www.rfc-editor.org/info/rfc4760>>.
- [RFC9085] Previdi, S., Talaulikar, K., Ed., Filsfils, C., Gredler, H., and M. Chen, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing", RFC 9085, DOI 10.17487/RFC9085, August 2021, <<https://www.rfc-editor.org/info/rfc9085>>.

5.2. Informative References

- [I-D.kompella-rtgwg-mlnwsched] Kompella, K., Beeram, V. P., Mahale, A., Bhargava, R., and N. Geyer, "Scheduling Network Resources for Machine Learning Clusters", Work in Progress, Internet-Draft, draft-kompella-rtgwg-mlnwsched-02, 1 March 2026, <<https://datatracker.ietf.org/doc/html/draft-kompella-rtgwg-mlnwsched-02>>.
- [I-D.zzhang-idr-mp-te-signaling] Zhang, Z., Kompella, K., Mahale, A., Bhargava, R., and A. Zhang, "BGP Signaling for Multipath Traffic Engineering Junction States", Work in Progress, Internet-Draft, draft-zzhang-idr-mp-te-signaling-00, 2 March 2026, <<https://datatracker.ietf.org/doc/html/draft-zzhang-idr-mp-te-signaling-00>>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <<https://www.rfc-editor.org/info/rfc5440>>.

Authors' Addresses

Kireeti Kompella
HPE
Email: kireeti.ietf@gmail.com

Zhaohui Zhang
HPE
Email: zhaohui.zhang@hpe.com