

RTGWG
Internet-Draft
Intended status: Informational
Expires: May 3, 2018

J. Tantsura
Individual
D. Afanasiev
YANDEX
K. Patel
Arccus
P. Lapukhov
Facebook
P. Przygienda
Juniper
R. White
LinkedIn
Y. Qu
Huawei
J. Uttaro
ATT
October 30, 2017

Requirements for the DataCenter Routing
draft-dt-rtgwg-dcrouting-requirements-00

Abstract

This document describes list of functional requirments towards a Routing Protocol for Data Center Networks.

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 May 3, 2018.

Internet-Draft

October 2017

Copyright Notice

Copyright (c) 2017 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 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	2
2.	Requirements Language	3
3.	Recommended Reading	3
4.	Goals and Requirements	3
5.	For further study	5
6.	Network Topologies	5
7.	IANA Considerations	5
8.	Security Considerations	5
9.	Acknowledgements	5
10.	Change Log	5
11.	References	5
11.1.	Normative References	5
11.2.	Informative References	5
	Authors' Addresses	6

[1.](#) Introduction

It is common to host and build a network of more than tens of thousands of end points inside a data center. Data Center Networks of such size have unique set of requirements with emphasis on scale, convergence, network stability and operational simplicity. Existing or new set of protocols and routing infrastructure needs to be augmented to support higher scale, faster convergence with increased optional simplicity in order to address the requirements of these networks.

This document describes list of functional requirements that are required towards addressing scale, convergence and operational maintainance of such scaled networks. The requirements described in this document can be used to augment existing solutions or used to design a new set of solutions.

Internet-Draft

October 2017

[2.](#) Requirements Language

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 [\[RFC2119\]](#) only when they appear in all upper case. They may also appear in lower or mixed case as English words, without any normative meaning.

[3.](#) Recommended Reading

This document assumes knowledge of existing data center networks and data center network topologies [\[CLOS\]](#), [\[RFC7938\]](#). This document also assumes knowledge of data center routing protocols like BGP [\[RFC4271\]](#), OSPF [\[RFC2328\]](#) and BFD [\[RFC5880\]](#) as well as data center layer 2 link layer protocols like LLDP [\[RFC4957\]](#).

[4.](#) Goals and Requirements

Following are general requirements for the Data Center Network Fabric and its Routing Protocols:

- o The Fabric provides basic connectivity, with possibility to carry one or more overlays.
The Fabric provides no domain separation, if needed, to be handled by an overlay.
The Fabric MAY provide interconnect facility for other fabrics.
The Fabric MUST support non equidistant end-points.
- o The Fabric MUST support Spine and Leaf [\[CLOS\]](#) + isomorphic topologies within its network.
The Fabric MAY support non Spine and Leaf topologies
- o The KPI's below are single-dimensional and expected to be changed, as the document progresses, baseded on more feedback, we ask community to communicate their views. Combination of # of routes vs # of paths vs desired convergence time will be discussed in a

later version.

The Fabric SHOULD support 250k routes @ 5k fabric nodes with convergence time below 250ms.

The Fabric SHOULD support 500k routes @ 7.5k fabric nodes with convergence time below 500ms.

The Fabric SHOULD support 1M routes @ 10k fabric nodes with convergence time below 1s.

- o The Fabric routing protocol MUST support load balancing using ECMP, wECMP and UCMP.

The Fabric routing protocol MAY support any custom or adaptive load balancing algorithms.

The Fabric routing protocol MUST support and provide facility for topology-specific algorithms that enable correct operations in that specific topology.

- o The Fabric routing protocol SHOULD support route scale and convergence times of a Fabric mentioned above.
The Fabric routing protocol SHOULD support ECMP as wide as 256 paths.
- o The Fabric routing protocol MUST support various address families that covers IP as well as MPLS forwarding.
The Fabric routing protocol MUST support extensions to carry 3rd party data and Opaque data.
- o The Fabric routing protocol MUST support Traffic Engineering paths that are host and/or router based paths.
The Fabric routing protocol MUST provide facility to address constituents in an ECMP bundle.
- o The Fabric routing protocol MUST support inband as well as out of band management.
- o The Fabric routing protocol MUST support Zero Touch Provisioning (ZTP).
The Fabric routing protocol MUST support Neighbor Discovery to facilitate ZTP.

- o The Fabric routing protocol MUST be able to leverage BFD [[RFC5880](#)] for neighbor state.
The Fabric routing protocol SHOULD be capable of bootstrapping a BFD session [[RFC5882](#)].
- o The Fabric routing protocol MUST be able to support real time state notifications of routes and its neighbors state to facilitate control plane telemetry.
The Fabric routing protocol MUST be able to support on-demand snapshots of protocol state and real time state notifications of routes and its neighbors state to remote node(s) to facilitate control plane telemetry.
- o The Fabric routing protocol MUST be able handle commission/decommission of a node as well as any node restart with a minimal data plane impact.

[5.](#) For further study

Following topics have been identified to be studied at a later time:

- o gRPC/THRIFT/similar encodings.
- o Ability to function as an overlay.
- o Flowlets signaling.
- o Multicast.
- o State representation NB.
- o Integration with PCE/SDNc.

[6.](#) Network Topologies

[7.](#) IANA Considerations

[8.](#) Security Considerations

This document describes list of functional requirements towards a routing protocol for Data Center Networks. It does not raise any security concerns or issues in addition to ones common to a routing protocol for Data Center Networks.

[9.](#) Acknowledgements

[10.](#) Change Log

Initial Version: October 31 2017

[11.](#) References

[11.1.](#) Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.

[11.2.](#) Informative References

[CLOS] "A Study of Non-Blocking Switching Networks", The Bell System Technical Journal, Vol. 32(2), DOI 10.1002/j.1538-7305.1953.tb01433.x, March 1953.

Tantsura, et al.

Expires May 3, 2018

[Page 5]

Internet-Draft

October 2017

[RFC2328] Moy, J., "OSPF Version 2", STD 54, [RFC 2328](#), DOI 10.17487/RFC2328, April 1998, <<https://www.rfc-editor.org/info/rfc2328>>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", [RFC 4271](#), DOI 10.17487/RFC4271, January 2006, <<https://www.rfc-editor.org/info/rfc4271>>.

[RFC4957] Krishnan, S., Ed., Montavont, N., Njedjou, E., Veerepalli, S., and A. Yegin, Ed., "Link-Layer Event Notifications for Detecting Network Attachments", [RFC 4957](#), DOI 10.17487/RFC4957, August 2007, <<https://www.rfc-editor.org/info/rfc4957>>.

- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", [RFC 5880](#), DOI 10.17487/RFC5880, June 2010, <<https://www.rfc-editor.org/info/rfc5880>>.
- [RFC5882] Katz, D. and D. Ward, "Generic Application of Bidirectional Forwarding Detection (BFD)", [RFC 5882](#), DOI 10.17487/RFC5882, June 2010, <<https://www.rfc-editor.org/info/rfc5882>>.
- [RFC7938] Lapukhov, P., Premji, A., and J. Mitchell, Ed., "Use of BGP for Routing in Large-Scale Data Centers", [RFC 7938](#), DOI 10.17487/RFC7938, August 2016, <<https://www.rfc-editor.org/info/rfc7938>>.

Authors' Addresses

Jeff Tantsura
Individual
USA

Email: jefftant.ietf@gmail.com

Dmitry Afanasiev
YANDEX
RU

Email: dmitry.afanasiev@gmail.com

Keyur Patel
Arccus
San Jose
USA

Email: keyur@arccus.com

Petr Lapukhov
Facebook
USA

Email: petr@fb.com

Tony Przygienda
Juniper
1137 Innovation Way
Sunnyvale, CA
USA

Email: prz@juniper.net

Russ White
LinkedIn
USA

Email: russ@riw.us

Yingzhen Qu
Huawei
Santa Clara
USA

Email: yingzhen.ietf@gmail.com

Jim Uttaro
ATT
USA

Email: juttaro@ieee.org