Network Working Group R. Bush
Internet-Draft Arrcus & IIJ
Intended status: Informational K. Patel
Expires: March 5, 2019 Arrcus
September 1, 2018

# BGP-SPF Topology Discovery Requirements
## draft-ymbk-lsvr-discovery-req-00

Abstract

For wide scale routing protocols to build their topology and reachability databases they need link neighbor discovery, link encapsulation data, and layer two liveness. BGP-LS and its enhancements provide an API to present much of these data to BGP protocols, but do not actually collect these data. This document explores the needs and criteria for the data needed.

Status of This Memo



Copyright Notice

Table of Contents



## 1. Introduction

In a massive scale datacenter or similar environment BGP([RFC4271]) and BGP-like protocols, e.g. BGP-SPF (see [I-D.ietf-lsvr-bgp-spf]), provide massive scale-out without centralization using a tried and tested scalable distributed control plane transport, offering a scalable routing solution. But BGP4 and BGP-SPF need topology discovery, link state liveness, and link addressing data from the network to build and maintain the routing topology.

BGP-LS [RFC7752] and its extensions provide an API which BGP4 and BGP-SPF can use to get the and distribute topology data. But BGP-LS itself does not gather the data, it merely presents it. So the topology data must be gathered.

What topology data do BGP-like protocols actually need? What level of freshness is needed? What are the requirements for scale, extensibility, security, etc?

## 2. Architectural Considerations

Massive Data Centers (MDCs) have on the order of 10,000 racks, often with two Top Of Rack (TOR) devices per rack. To provide this level of scaling reliably, stably, and securely imposes architectural constraints on any discovery protocol.

- o Simple - If it isn't simple, it will not scale. Simplicity requires restraint in design. 'Union Protocols' which are the sum of everyone's desires are complex disasters waiting to happen. Often they do not wait. Prefer 'Intersection Protocols' which include only those things which everyone absolutely needs.

- o Securable - Security properties should be analysed. Again, simplicity is key; complex protocols increase in complexity over time, and security vulnerabilities increase exponentially with complexity. As [RFC5218] 2.2.3 says "The more successful a protocol becomes, the more attractive a target it will be."

- o Extensible - As [RFC5218] Section 2.2.1 said, successful protocols are extensible beyond the original expectation. MDC and similar needs are expanding and we are still learning about the space. Simplicity and extensibility should go a long way to adaptability; complex protocols are hard to extend, especially when they are poorly understood.

- o Implementable - It must be reasonably easy to implement and deploy. Some implications are:

  * Packet formats should be easy to generate and easily parsable. Type/length/Value (TLV) formats are preferred.

  * The protocols should be free to use and deploy; i.e. not be constrained by Intellectual Property Right (IPR) claims.

  * Again, simpler protocols are simpler to implement, deploy, measure, monitor, etc.

  * Performance Problems arise if the protocol was not designed to scale.

- o Protocol Control - It is mandatory that the IETF have full control over the protocol definition. This should not preclude cooperation with other Standards Development Organisations (SDOs); but the final control must rest with the IETF.

## 3. Requirements

The target for the discovery protocol(s) is a massive datacenter scale deployment using BGP or similar routing, e.g. BGP4 or [I-D.ietf-lsvr-bgp-spf]; but should be generally usable by other routing protocols in other environments.

The IETF is very good at finding corner cases which expand needs and complicate protocols. This effort should resist this tendency.

It would be easiest for the BGP-like protocols to consume the data if they are presented via the BGP-LS [RFC7752] API as used in [I-D.ietf-lsvr-bgp-spf] Section 4.

BGP-like protocols will need at least the following information about the topology:

Node Identity:  Each node in the topology must have an identity/identifier which must be unique in the topology.

A node must have one or more links to other nodes or it is, ab definito, not in the topology.

Link Identity:  A link is between two nodes.  Each end of a link is a node/device interface.

Each link in the topology must be uniquely identified and the identities of the nodes on the link must be identified.

L2 Liveness:  Because adjacencies and topology changes must be quickly detected, Layer-2 stability of each link should be monitored and reported.

Encapsulations:  The encapsulation(s) (IPv4, IPv6, ...) on each link must be known.  One or more of the common AFI/SAFIs must be supported on each link, IPv4, IPv6, MPLS, etc.

It is assumed that the set of encapsulations is the same across the entire topology.

Addresses:  The available addresses on the node interfaces for each encapsulation must be known.  More than one address for an encapsulation must be supported.

As BGP-like protocols will be peering between the nodes, there may be a preferred encapsulation and address on an link, or a loopback interface may be used.

## 4.  Security Considerations

While this document has no security considerations per se, it does make a plea for securability in protocol design.

Mis-wires, malicious devices being plugged into ports, and monkey in the middle attacks should be considered.

## 5.  IANA Considerations

This document has no IANA considerations.

## 6. Acknowledgments

The authors thank Victor Kuarsingh and Gunter Van De Velde for reviews.

## 7. References

### 7.1. Normative References

[I-D.ietf-lsvr-bgp-spf]
Patel, K., Lindem, A., Zandi, S., and W. Henderickx, "Shortest Path Routing Extensions for BGP Protocol", draft-ietf-lsvr-bgp-spf-02 (work in progress), August 2018.

[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, <http://www.rfc-editor.org/info/rfc4271>.

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <http://www.rfc-editor.org/info/rfc7752>.

### 7.2. Informative References

[RFC5218] Thaler, D. and B. Aboba, "What Makes for a Successful Protocol?", RFC 5218, DOI 10.17487/RFC5218, July 2008, <http://www.rfc-editor.org/info/rfc5218>.

Authors' Addresses

Randy Bush
Arrcus & IIJ
5147 Crystal Springs
Bainbridge Island, WA 98110
United States of America

Email: randy@psg.com

Keyur Patel
Arrcus
2077 Gateway Place, Suite #250
San Jose, CA  95119
United States of America

Email: keyur@arrcus.com