

Network Working Group  
Internet-Draft  
Intended status: Informational  
Expires: August 2, 2019

F. Templin, Ed.  
G. Saccone  
Boeing Research & Technology  
G. Dawra  
LinkedIn  
A. Lindem  
V. Moreno  
Cisco Systems, Inc.  
January 29, 2019

Scalable De-Aggregation for Overlays Using the Border Gateway Protocol  
(BGP)  
draft-templin-rtgwg-scalable-bgp-01.txt

## Abstract

The Border Gateway Protocol (BGP) has well-known limitations in terms of the numbers of routes that can be carried and stability of the routing system. This is especially true when mobile nodes frequently change their network attachment points, which in the past has resulted in excessive announcements and withdrawals of de-aggregated prefixes. This document discusses a means of accommodating scalable de-aggregation of IPv6 prefixes for overlay networks using BGP.

## 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 August 2, 2019.

## Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

Internet-Draft

Scalable De-Aggregation for BGP

January 2019

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

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">2.</a>	Overview and Analysis . . . . .	<a href="#">2</a>
<a href="#">3.</a>	Opportunities and Limitations . . . . .	<a href="#">4</a>
<a href="#">4.</a>	Use Cases . . . . .	<a href="#">4</a>
<a href="#">5.</a>	Implementation Status . . . . .	<a href="#">4</a>
<a href="#">6.</a>	IANA Considerations . . . . .	<a href="#">5</a>
<a href="#">7.</a>	Security Considerations . . . . .	<a href="#">5</a>
<a href="#">8.</a>	Acknowledgements . . . . .	<a href="#">5</a>
<a href="#">9.</a>	References . . . . .	<a href="#">5</a>
<a href="#">9.1.</a>	Normative References . . . . .	<a href="#">5</a>
<a href="#">9.2.</a>	Informative References . . . . .	<a href="#">5</a>
<a href="#">Appendix A.</a>	Change Log . . . . .	<a href="#">6</a>
	Authors' Addresses . . . . .	<a href="#">6</a>

## [1.](#) Introduction

The Border Gateway Protocol (BGP) [[RFC4271](#)] has well-known limitations in terms of the numbers of routes that can be carried and the stability of the routing system. This is especially true for routing systems that include mobile nodes that frequently change their network attachment points, which in the past have resulted in excessive announcements and withdrawals of de-aggregated prefixes. This document discusses a means of accommodating scalable de-aggregation of IPv6 prefixes [[RFC8200](#)] for overlay networks using BGP.

## [2.](#) Overview and Analysis

As discussed in [[I-D.ietf-rtgwg-atn-bgp](#)] and [[I-D.templin-intarea-6706bis](#)], the method for accommodating de-aggregation is to institute an overlay network instance of BGP that

is separate and independent from the global Internet BGP routing system. The overlay is presented to the global Internet as a small number of aggregated IPv6 prefixes (also known as Mobility Service Prefixes (MSPs)) that never change. In this way, the Internet BGP routing system sees only stable aggregated MSPs (e.g., 2001:db8::/32)

and is completely unaware of any de-aggregation or mobility-related churn that may be occurring within the overlay.

The overlay is operated by an Overlay Service Provider (OSP), and consists of a core Autonomous System (AS) with core AS Border Routers (c-ASBRs) that connect to stub ASes with stub ASBRs (s-ASBRs) in a hub-and-spokes fashion. Mobile nodes associate with nearby (i.e., regional) stub ASes for extended timeframes, and change to new stub ASes only after movements of significant topological or geographical distance. Mobility-related changes between stub ASes are therefore normally infrequent.

The s-ASBRs use eBGP to announce de-aggregated Mobile Network Prefixes (MNPs) of mobile nodes (e.g., 2001:db8:1:2::/64, etc.) to their neighboring c-ASBRs, but do not announce fine-grained mobility events such as a mobile node moving to a new network attachment point. Instead, mobile nodes coordinate with stub ASes using mobility protocols such as MIPv6, LISP, AERO, etc. and stub ASes accommodate these localized mobility events without disturbing the c-ASBRs.

The c-ASBRs originate "default" to their neighboring s-ASBRs but do not announce any MNP routes. In this way, MNP announcements and withdrawals are unidirectional from s-ASBRs to c-ASBRs only, thereby suppressing BGP updates on the reverse path. The c-ASBRs in turn use iBGP to maintain a consistent view of the full topology. BGP Route Reflectors (RRs) [[RFC4456](#)] can also be used to support increased c-ASBR scaling.

Each c-ASBR should be able to carry at least as many routes as a typical core router in the global public Internet BGP routing system. Since the number of active routes in the Internet is rapidly approaching 1 million (1M), viable c-ASBRs must be capable of carrying at least 1M MNP routes (this has been proven even for BGP running on lightweight virtual machines). The method for increasing scaling therefore is to divide the MSP into longer sub-MSPs, and to

assign a different set of c-ASBRs for each sub-MSP.

For example, the MSP 2001:db8::/32 could be sub-divided into sub-MSPs such as 2001:db8:0010::/44, 2001:db8:0020::/44, 2001:db8:0030::/44, etc. with each sub-MSP assigned to a different set of c-ASBRs. Each s-ASBR peers with at least one member of each c-ASBR set and uses route filters such that BGP updates are only sent to the c-ASBR(s) that aggregate the specific sub-MSP. Then, assuming 1 thousand (1K) or more sub-MSPs (each with its own set of c-ASBRs) the entire BGP overlay routing system should be able to service 1 billion (1B) MNPs or more.

### [3.](#) Opportunities and Limitations

Since a lightweight virtual machine (e.g., a linux image running quagga in the cloud) can service up to 1M MNPs using BGP, it is likely that dedicated high-performance IPv6 router hardware could support even more. With such dedicated high-performance hardware, the number of MNPs could be increased further.

The deployed numbers of s-ASBRs even for very large overlays should not exceed a c-ASBR's capacity for BGP peering sessions. For example, c-ASBRs should be capable of servicing 1K or more BGP peering sessions, with the upper bound limited by keepalive and update control messaging overhead. Conversely, s-ASBRs should be capable of supporting even more sessions since they only receive keepalives and only send updates for mobile nodes within their local stub ASes.

Mobile nodes should refrain from moving rapidly between stub ASes for no good reason, since the objective is only to reduce routing stretch due to movement of significant distances. OSPs could employ disincentives such as surcharge penalties for gratuitous mobility, but intentional abuse would also yield little reward since only the bad actor (i.e., and not others) would be subject to MNP instability.

Packets sent between mobile nodes that associate with different stub ASes would initially need to be forwarded through the core AS, which presents a forwarding bottleneck. For this reason, a route optimization function is needed to reduce congestion in the core. Since c-ASBRs should be commercial off-the-shelf (COTS) dedicated high-performance IPv6 routers, however, they should not be required

to participate directly in any out-of-band route optimization signaling. Instead, route optimization should be coordinated by stub AS network elements and/or the mobile nodes themselves.

#### [4.](#) Use Cases

Use cases include Unmanned Air Systems (UAS) in controlled and uncontrolled airspaces, Intelligent Transportation Systems (ITS) in urban air/ground mobility environments, aviation networks, enterprise mobile device users, and cellular network users. Any other use cases in which an OSP services large numbers of mobile nodes are also in scope.

#### [5.](#) Implementation Status

The arrangement of stub and core ASes described in this document has been implemented using standards-compliant linux operating systems and BGP routing protocol implementations (i.e., quagga). No new code

was included, and all requirements were satisfied through standard configuration options.

#### [6.](#) IANA Considerations

This document does not introduce any IANA considerations.

#### [7.](#) Security Considerations

Security considerations are discussed in the references.

#### [8.](#) Acknowledgements

This work is aligned with the FAA as per the SE2025 contract number DTFAWA-15-D-00030.

This work is aligned with the NASA Safe Autonomous Systems Operation (SASO) program under NASA contract number NNA16BD84C.

This work is aligned with the Boeing Information Technology (BIT) MobileNet program.

## [9.](#) References

### [9.1.](#) Normative References

- [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>>.
- [RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", [RFC 4456](#), DOI 10.17487/RFC4456, April 2006, <<https://www.rfc-editor.org/info/rfc4456>>.
- [RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, [RFC 8200](#), DOI 10.17487/RFC8200, July 2017, <<https://www.rfc-editor.org/info/rfc8200>>.

### [9.2.](#) Informative References

- [I-D.ietf-rtgwg-atn-bgp]  
Templin, F., Saccone, G., Dawra, G., Lindem, A., and V. Moreno, "A Simple BGP-based Mobile Routing System for the Aeronautical Telecommunications Network", [draft-ietf-rtgwg-atn-bgp-01](#) (work in progress), January 2019.

Templin, et al. Expires August 2, 2019 [Page 5]

---

Internet-Draft Scalable De-Aggregation for BGP January 2019

- [I-D.templin-intarea-6706bis]  
Templin, F., "Asymmetric Extended Route Optimization (AERO)", [draft-templin-intarea-6706bis-03](#) (work in progress), December 2018.

## [Appendix A.](#) Change Log

<< RFC Editor - remove prior to publication >>

Changes from -00 to -01:

- o added Route Reflectors
- o introduced term "Overlay Service Provider (OSP)"

- o removed estimate of number of routes for high-performance routers
- o revised text on route optimization
- o added use case and implementation sections

Status as of 01/23/2018:

- o -00 draft published

#### Authors' Addresses

Fred L. Templin (editor)  
Boeing Research & Technology  
P.O. Box 3707  
Seattle, WA 98124  
USA

Email: fltemplin@acm.org

Greg Saccone  
Boeing Research & Technology  
P.O. Box 3707  
Seattle, WA 98124  
USA

Email: gregory.t.saccone@boeing.com

Templin, et al.

Expires August 2, 2019

[Page 6]

---

Internet-Draft

Scalable De-Aggregation for BGP

January 2019

Gaurav Dawra  
LinkedIn  
USA

Email: gdawra.ietf@gmail.com

Acee Lindem

Cisco Systems, Inc.  
USA

Email: [acee@cisco.com](mailto:acee@cisco.com)

Victor Moreno  
Cisco Systems, Inc.  
USA

Email: [vimoreno@cisco.com](mailto:vimoreno@cisco.com)