Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: 2 November 2025

J. Haas
Juniper Networks, Inc.

1 May 2025

Distribution of Source Address Validation State in BGP
draft-haas-savnet-bgp-sav-distribution-01

Abstract

Source Address Validation (SAV) is a technique that can be used to mitigate source address spoofing. draft-huang-savnet-sav-table codifies the concept of source address validation on a per-incoming interface basis, the validation mode, and the traffic handling policy as a "SAV Table".

This document defines a mechanism for distributing logical SAV Tables in BGP. This mechanism is addresses inter-domain SAV use cases. These SAV Tables may be used to implement appropriate device-specific validation for source addresses.

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 2 November 2025.

Table of Contents

1. Introduction

Introductory text

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Terminology

The following terminology is defined in [I-D.huang-savnet-sav-table]:

SAV rule: The entry specifying the valid incoming interfaces of specific source addresses or source prefixes.

Traffic handling policy: The policy taken on the packets validated by SAV.

Validation mode: The mode that describes the typical applications of SAV in a specific kind of scenarios. Different modes take effect in different scales and treat the default prefix differently.

The following terminology is defined in this document:

BGP SAV-dist Entry (SAV-D Entry): A tuple consisting of a source address prefix, validation mode, and traffic handling policy.

BGP SAV-dist Target (SAV-D Target): An attribute declaring membership of a BGP SAV-dist Entry in one or more BGP SAV-dist Tables.

BGP SAV-dist Table (SAV-D Table): A logical table defined by a set of BGP SAV-dist Targets and the BGP SAV-dist Entries matching those targets.

3. A model for SAV distribution

Source address validation can be modeled on a per-interface basis as a match operation on a list of prefixes. Upon a match, or failure to match, a traffic handling policy is applied to the traffic. An instance of a prefix, its validation mode, and its traffic handling policy is a SAV rule.

In the inter-domain SAV use case, this set of prefixes consists of the set of upstream networks that may receive the advertised BGP routes, either directly or transitively, where the forwarding element is a BGP next hop. Determining this set of upstream prefixes is the subject of a number of savnet proposals.

A given SAV rule may be applied to a number of forwarding elements on the same device, or across multiple devices.

It can be observed that a given SAV rule that might be applied to a given interface by a provisioning mechanism follows several common membership criteria:

* This rule is for this specific interface on a device.

* This rule is for a set of interfaces on this device that may share a common property. For example, "all customer interfaces".

* This rule is for a set of interfaces across the network (Autonomous System) that may share a common property. For example, all interfaces for a given service provider.

* This rule is for a prefix from an upstream BGP Autonomous System.

A "SAV Table" may be modeled as the interface-specific list of memberships for SAV rules that share those membership criteria.

These properties have some similarity with Layer 3 VPNs. A SAV entry may be modeled as a BGP route that is a member of some number of SAV-D Tables via one or more SAV-D Targets.

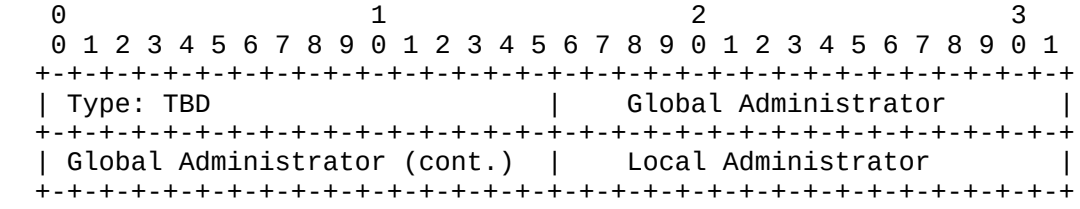
4. SAV-D Routes

A SAV-D Route implements an instance of a SAV rule. As noted above, SAV rules consist of a prefix, a validation mode, and a traffic handling policy.

A prefix for a SAV-D Route may have a large number of memberships across a deployment. Since baseline BGP is limited by its PDU size to 4096 bytes, the prefix alone cannot be the "NLRI key". In order to permit a large number of memberships for a given prefix, this document defines a new Route Distinguisher. The prefix is encoded in a BGP NLRI with a new SAFI, defined below. The validation mode and traffic handling policies are encoded in new BGP Extended Communities [RFC4360].

4.1. SAV-D RD-Originator

Type Field: TBD (2 octets)
Value Field: 6 octets



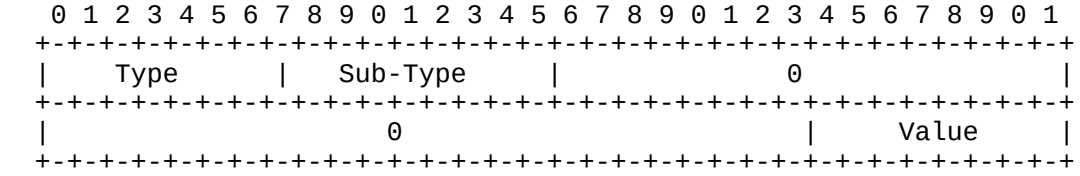
The Global Administrator field contains the BGP Identifier of the BGP speaker originating these SAV-D routes.

The Local Administrator contains a locally chosen value that permits unique combinations of SAV-D Routes to be distributed across the deployment.

4.2. SAV-D NLRI

This document defines two new BGP NLRI types, AFI=1/2, SAFI=TBD, which can be used to distribute SAV Rules in BGP. The NLRI format for this address family consists of a fixed-length 8 octet SAV-D RD-Originator followed by a variable-length IPv4 or IPv6 prefix whose length depends on AFI.

4.3. SAV-D Match Extended Community



Type: TBD
Sub-Type: TBD

The first five octets of this Extended Community's Value field MUST be zero.

The Value of this extended community can be:

1: Allow. Traffic matching this source address has passed the local validation criteria and will continue to be processed in the forwarding pipeline.

2: Block. Traffic matching this source address has failed

validation and will be blocked.

3: Action. Traffic matching this source address is subject to further evaluation based on the SAV-D Route Traffic Handling procedures.

4.4. Encoding SAV-D Route Traffic Handling Actions

When traffic has the "Action" match criteria, the traffic is subject to additional processing when supported by the local forwarding element.

The following extended communities are re-used from Section 7 of [RFC8955]:

* traffic-rate-bytes - rate-limiting bps.

* traffic-rate-packets - rate-limit pps.

* rt-redirect AS-2octet - traffic redirection by VRF.

* rt-redirect AS-4octet - traffic redirection by VRF.

* rt-redirect IPv4 - traffic redirection by VRF.

* traffic-marking - traffic marking DSCP.

* traffic-action, S-bit only - sampling. All other bits MUST be zero.

The following extended communities are re-used from Section 7 of [RFC8956]:

* rt-redirect IPv6 - traffic redirection by VRF.

The following extended communities are re-used from [I-D.ietf-idr-flowspec-redirect-ip]:

* Flow-spec Redirect to IPv4

* Flow-spec Redirect to IPv6

4.4.1. Error handling for SAV-D Traffic Handling Actions

A BGP speaker might not be able to process some or any of the traffic handling actions. Since SAV tables are intended to be provisioned on a per-device per-interface basis, BGP speakers generating SAV-D routes should have an understanding of what features can be implemented by the receiving device and avoid encoding actions that cannot be implemented by that device.

The following error handling behaviors are specified for SAV-D routes containing actions:

* A SAV-D route MUST NOT contain more than one redirection action. If a BGP speaker receives more than one redirection action it MAY locally permit traffic to be allowed or blocked based on provisioning.

* A BGP speaker unable to implement redirection MAY locally permit traffic to be allowed or blocked based on provisioning.

* A BGP speaker unable to implement rate-limiting MAY locally permit traffic to be allowed or blocked based on provisoning.

* A BGP speaker unable to implement traffic-marking MAY locally permit traffic to be allowed or blocked based on provisoning.

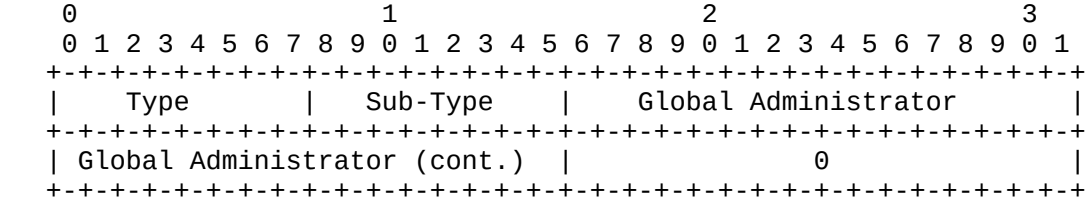
* A BGP speaker unable to implement sampling SHOULD permit traffic to be allowed.

5. SAV-D Targets

SAV-D Routes are members of one or more SAV-D Tables. This membership is signaled in SAV-D Routes using one or more SAV-D Target Extended Communities.

Some SAV-D Targets are node-specific. In order to clearly encode such node-specific SAV-D Extended Communities and appropriately scope them, routes containing SAV-D node-specific Extended Communities MUST contain a single instance of a Node Target Extended Community, [I-D.ietf-idr-node-target-ext-comm].

5.1. SAV-D Interface Member Extended Community

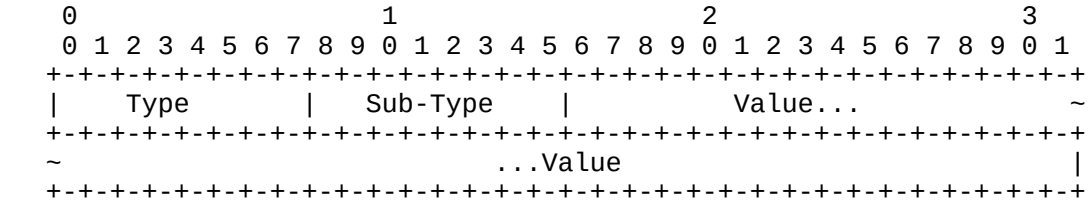


Type: TBD - non-transitive
Sub-Type: TBD

This is a SAV-D node-specific Extended Community type and requires a Node Target Extended Community for context.

The Global Administrator field is set to the ifIndex [RFC1213] of the interface that the SAV Rule is a member of. The Local Administrator field MUST be set to zero.

5.2. SAV-D Device-Specific Group Member Extended Community

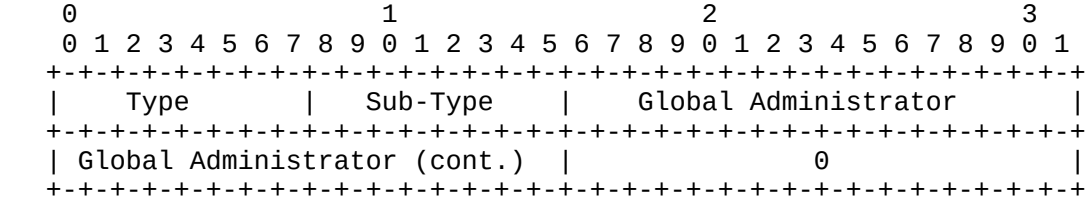


Type: TBD - non-transitive
Sub-Type: TBD

This is a SAV-D node-specific Extended Community type and requires a Node Target Extended Community for context.

The 6-octet value is has local significance to the node.

5.3. SAV-D Neighbor-AS Member Extended Community

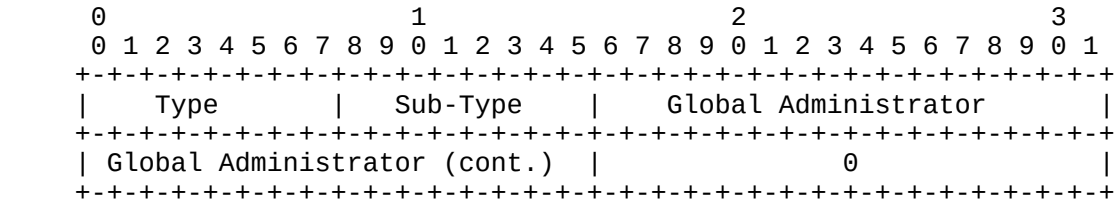


Type: TBD - non-transitive
Sub-Type: TBD

The SAV rules are associated with a given neighbor AS of the service provider.

The 4-octet Global Administrator field is set to the neighbor AS that the SAV table is attached to. The Local Administrator field MUST be set to zero.

5.4. SAV-D Origin-AS Member Extended Community



Type: TBD - non-transitive
Sub-Type: TBD

The SAV rules are associated with a given originating AS. This membership enables such use cases as including the AS numbers in a customer cone.

BGP speakers implementing [RFC8210] and ensuring that all Origin AS state used for SAV rules is present in their RPKI ROA cache MAY obtain the prefixes for their SAV Rules from that mechanism rather than BGP.

The 4-octet Global Administrator field is set to the neighbor AS that the SAV table is attached to. The Local Administrator field MUST be set to zero.

6. SAV-D Operational Model

The method by which a set of SAV Rules is determined for a given interface remains a somewhat contentious discussion. For purposes of this document, it is presumed that a "SAV Controller" exists to calculate the appropriate SAV rules for the deployment and that such a controller implements the extensions defined in this document to provision SAV within the network.

Each forwarding element implementing SAV will be provisioned with the per-interface SAV-D memberships for the SAV Rules implemented by that device. This state constitutes a logical SAV Table.

6.1. SAV-D Route Distribution and Scaling

The SAV Controller originates BGP SAV-D Routes with appropriate membership Extended Communities to each forwarding element. When the device has matching membership for such SAV-D routes in a SAV Table, it installs that route in that logical SAV Table using the standard BGP Decision Process to choose the best SAV Route when multiple candidate routes exist.

[RFC4684] SHOULD be deployed between BGP speakers distributing SAV-D Routes.

SAV-D BGP speakers SHOULD originate [RFC4684] routes for SAV-D Neighor-AS and Origin-AS Member Extended Communities. Such routes will attract SAV Rules that may be used in-common across multiple SAV Tables in a deployment.

6.1.1. Leveraging RPKI-RTR

In deployments where it is known that the SAV Controller would be emitting SAV Routes with Origin-AS Member Extended Communities that are identical with the RPKI ROA state carried in RPKI-RTR ([RFC8210]), implementations MAY exclude originating [RFC4684] routes for these Origin-AS Member Extended Communities and instead derive the local SAV Rules for their SAV Tables from the received RPKI-RTR state.

6.1.2. Node-specific Extended Community scaling considerations

Some SAV Routes have membership that is only only node-specific, i.e. contains only node-specific SAV-D Extended Communities. Distribution of these routes to all iBGP speakers within an AS, either via full-mesh iBGP or route reflection, will result in SAV-D BGP speakers receiving SAV Routes they do not care about.

SAV-D BGP speakers SHOULD originate [RFC4684] routes for the Node Target Extended Communities. At the same time, such speakers SHOULD NOT originate [RFC4684] routes for SAV-D node-specific Extended Communities. The Node Target Extended Community is sufficient to attract the appropriate SAV-D Routes without needing to expose internal membership details.

As an optimization, implementations MAY distribute such node-specific SAV-D Routes using "targeted" BGP sessions. In this case, a dedicated BGP session between a SAV Controller and a SAV-D speaker is used to distribute node-specific BGP routes. Such routes SHOULD be marked with the [RFC1997] NO_ADVERTISE community to prevent further redistribution.

6.2. Completeness of SAV information and enabling enforcement

Even presuming pervasive and instantaneous deployment of SAV in a network, many documents discuss the problems associated with incomplete SAV information. Such issues can include incorrectly passing traffic that should have been blocked or blocking traffic that should have been passed.

Forwarding resources utilized by an implementation to enforce SAV may take time to program. Such programming MUST be in-place with full state and correct prior to enforcement. Failure to do so may result in incorrect enforcement. Implementations SHOULD NOT enable enforcement of SAV until the state is fully programmed.

One possible mechanism for controlling the activation of enforcement is careful control of a SAV-D "default" route for the node. In circumstances where SAV Rules are primarily of the form of an allow-list, enforcement of block activity can be modeled as a "default" block action. As long as the default is signaled last and its processing is deferred until other installation of SAV Table state is complete, inappropriate dropping of traffic may be minimized.

Note that BGP does not permit careful sequencing of NLRI distribution within the protocol. Routes advertised "last" are only done so between a pair of BGP speakers on a session. Implementations might otherwise re-order advertised NLRI while propagating BGP routes according to their own desires.

7. IANA Considerations

TBD

8. Security Considerations

A huge list to come. In particular, incorrect or incomplete installation of SAV in a network while having attracted traffic to that network by normal BGP means is harmful to the Internet.

9. References

9.1. Normative References

[RFC1213] McCloghrie, K. and M. Rose, "Management Information Base for Network Management of TCP/IP-based internets: MIB-II", STD 17, RFC 1213, DOI 10.17487/RFC1213, March 1991, <https://www.rfc-editor.org/info/rfc1213>.

[RFC1997] Chandra, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, DOI 10.17487/RFC1997, August 1996, <https://www.rfc-editor.org/info/rfc1997>.

[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>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <https://www.rfc-editor.org/info/rfc4360>.

[RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, DOI 10.17487/RFC4684, November 2006, <https://www.rfc-editor.org/info/rfc4684>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8210] Bush, R. and R. Austein, "The Resource Public Key Infrastructure (RPKI) to Router Protocol, Version 1", RFC 8210, DOI 10.17487/RFC8210, September 2017, <https://www.rfc-editor.org/info/rfc8210>.

[RFC8955] Loibl, C., Hares, S., Raszuk, R., McPherson, D., and M. Bacher, "Dissemination of Flow Specification Rules", RFC 8955, DOI 10.17487/RFC8955, December 2020, <https://www.rfc-editor.org/info/rfc8955>.

[RFC8956] Loibl, C., Ed., Raszuk, R., Ed., and S. Hares, Ed., "Dissemination of Flow Specification Rules for IPv6", RFC 8956, DOI 10.17487/RFC8956, December 2020, <https://www.rfc-editor.org/info/rfc8956>.

[I-D.huang-savnet-sav-table] Huang, M., Cheng, W., Li, D., Geng, N., Liu, Chen, L., and C. Lin, "General Source Address Validation Capabilities", Work in Progress, Internet-Draft, draft-huang-savnet-sav-table-08, 10 December 2024, <https://datatracker.ietf.org/doc/html/draft-huang-savnet-sav-table-08>.

[I-D.ietf-idr-flowspec-redirect-ip]
Uttaro, J., Haas, J., akarch@cisco.com, Ray, S., Mohapatra, P., Henderickx, W., Simpson, A., and M. Texier, "BGP Flow-Spec Redirect-to-IP Action", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-redirect-ip-03, 8 September 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-flowspec-redirect-ip-03>.

[I-D.ietf-idr-node-target-ext-comm]
Dong, J., Zhuang, S., Van de Velde, G., and J. Tantsura, "BGP Extended Community for Identifying the Target Nodes", Work in Progress, Internet-Draft, draft-ietf-idr-node-target-ext-comm-02, 4 March 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-node-target-ext-comm-02>.

9.2. Informative References

[RFC8704] Sriram, K., Montgomery, D., and J. Haas, "Enhanced Feasible-Path Unicast Reverse Path Forwarding", BCP 84, RFC 8704, DOI 10.17487/RFC8704, February 2020, <https://www.rfc-editor.org/info/rfc8704>.

[I-D.ietf-sidrops-bar-sav]
Sriram, K., Lubashev, I., and D. Montgomery, "Source Address Validation Using BGP UPDATEs, ASPA, and ROA (BAR-SAV)", Work in Progress, Internet-Draft, draft-ietf-sidrops-bar-sav-06, 15 March 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-sidrops-bar-sav-06>.

[I-D.ietf-idr-flowspec-v2]
Hares, S., Eastlake, D. E., Yadlapalli, C., and S. Maduschke, "BGP Flow Specification Version 2", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-v2-04, 28 April 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-flowspec-v2-04>.

[I-D.ietf-idr-flowspec-interfaceset]
Litkowski, S., Simpson, A., Patel, K., Haas, J., and L. Yong, "Applying BGP flowspec rules on a specific interface set", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-interfaceset-05, 18 November 2019, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-flowspec-interfaceset-05>.

Appendix A. Encoding choices, why not Flowspec version 2?

SAV enforcement of SAV rules in many ways resembles an access control list (ACL), which are often implemented via firewalls. BGP Flowspec [RFC8955] is capable of distributing IP source address filtering to be instantiated in firewalls. However, Flowspec's original design for mitigating distributed denial of service (DDoS) attacks generally has meant that rules distributed via Flowspec are intended to be applied to the interfaces of all Flowspec routers in a network.

The [I-D.ietf-idr-flowspec-interfaceset] feature provides scoped application of Flowspec filters. This feature inspires some of the applicability of the SAV-D Device-Specific Group Member Extended Community.

Given the overlapping nature of SAV enforcement behavior, and firewalls that may be programmed via Flowspec, should SAV enforcement be encoded there?

* SAV enforcement is not required to be implemented in a forwarding element's firewall infrastructure. Encoding SAV state in firewall functionality implies that this might be the case. (Namespace confusion.)

* Flowspec provides for ordered filter installation. If a SAV rule is present in flowspec, and implemented in something other than the firewall infrastructure, and filtering is not executed in the expected order, that's a problem. Minimally, it is a matter of violating the "Principle of Least Astonishment". Moreover, firewall applications built on the expectation that rules are executed in a particular order may be fail to execute as expected due to the firewall term dependencies.

* [I-D.ietf-idr-flowspec-v2] proposes a user-managed term order mechanism in addititon to the default term order that Flowspec generates based on filtering components. This could be leveraged to order SAV rules in a more appropriately logical place in the filter relative to an implementations forwarding pipeline; e.g. "last". However, this still presumes to know how a given platform that does not instantiate SAV filtering will perform enforcement.

* The placement of SAV-D in a different AFI/SAFI than Flowspec also permits for the distribution of SAV rules distinctly from a BGP routing topology that may carry Flowspec. Similarly, a separate AFI/SAFI means that implementations do not conflate distribution of firewall state vs. SAV state in terms of ordering and prioritization. SAV state will, in the majority of cases, significantly exceed the number of entries typically carried in Flowspec for firewall by several orders of magnitude.

* Flowspec does have a form of encoding that carries a Route Distinguisher for VPN Flowspec purposes in a VPN-specific AFI/SAFI. However, that format currently carries the expectation that it is accompanied by a BGP Route Target for installing the routes in a VRF context.

* Finally, a significant amount of the fragility that Flowspec is subject to is a result of the complexity needed to encode the various components in a term in the NLRI. SAV-D state has the much simpler need to encode only a RD, and a IPv4/IPv6 prefix. This also has the benefit of being not only "fit for purpose", but can pack better in BGP Updates since Flowspec v1/v2 NLRI have more overhead.

It is a valid observation that the Group Member Extended Communities defined for SAV-D may have overlapping usefulness for general-purpose BGP Flowspec applications. Future versions of this document may repurpose the Extended Communities for such general use.

Acknowledgements

Discussions with Nan Geng and Susan Hares at IETF-119 were helpful in formulating the requirements for this draft.

Contributors

TBD.

Author’s Address

Jeffrey Haas
Juniper Networks, Inc.
1133 Innovation Way
Sunnyvale, CA 94089
United States of America
Email: jhaas@juniper.net