Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 12, 2022

Y. Wang
S. Zhuang
Y. Gu
Huawei
R. Pang
China Unicom

July 11, 2021

BGP Extension for Advertising In-situ Flow Information Telemetry (IFIT) Capabilities

draft-wang-idr-bgp-ifit-capabilities-03

Abstract

This document defines extensions to BGP to advertise the In-situ Flow Information Telemetry (IFIT) capabilities. Within an IFIT domain, IFIT-capability advertisement from the tail node to the head node assists the head node to determine whether a particular IFIT Option type can be encapsulated in data packets. Such advertisement would be useful for mitigating the leakage threat and facilitating the deployment of IFIT measurements on a per-service and on-demand basis.

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 RFC 2119 [RFC2119].

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 January 12, 2022.

Table of Contents

1. Introduction

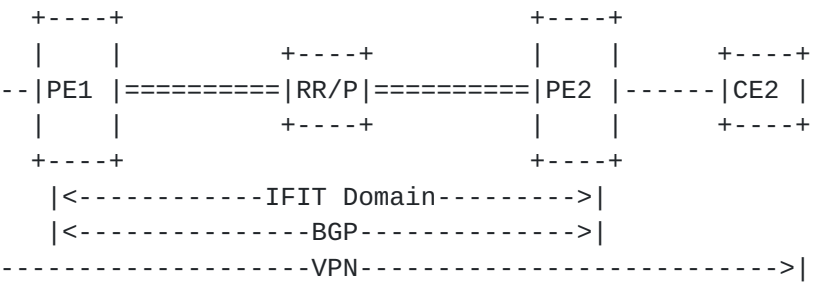
In-situ Flow Information Telemetry (IFIT) denotes a family of flow-oriented on-path telemetry techniques, including In-situ OAM (IOAM) [I-D.ietf-ippm-ioam-data] and Alternate Marking [RFC8321]. It can provide flow information on the entire forwarding path on a per-packet basis in real time.

IFIT is a solution focusing on network domains according to [RFC8799] that introduces the concept of specific domain solutions. A network domain consists of a set of network devices or entities within a single administration. As mentioned in [RFC8799], for a number of reasons, such as policies, options supported, style of network

management and security requirements, it is suggested to limit applications including the emerging IFIT techniques to a controlled domain.

Hence, the family of emerging on-path flow telemetry techniques MUST be typically deployed in such controlled domains. The IFIT solution MAY be selectively or partially implemented in different vendors' devices as an emerging feature for various use cases of application-aware network operations. In addition, for some use cases, the IFIT are deployed on a per-service and on-demand basis.

The figure shows an implementation example of IFIT for VPN scenario.



As the figure shows, a traffic flow is sent out from the customer edge CE1 to another CE2. In order to enable IFIT application for this flow, the IFIT header must be encapsulated in the packet at the ingress node PE1 referred to as the IFIT encapsulating node. Then, transmit nodes in the IFIT domain must be able to support the IFIT capabilities to inspect IFIT command and update the IFIT data fields in the packet. Finally, the IFIT data fields MUST be exported and removed at egress node PE2 that is referred to as the IFIT decapsulating node to avoild IFIT data leakage outside the controlled domain. Hence, a head node needs to know if the IFIT decapsulating node are able to support the IFIT capabilities.

This document defines extensions to Border Gateway Protocol (BGP) to advertise the supported IFIT capabilities of the egress node to the ingress node in an IFIT domain when the egress node distributes a route. Then the ingress node can learn the IFIT node capabilities associated to the routing information distributed between BGP peers and determine whether a particular IFIT Option type can be encapsulated in traffic packets which are forwarded along the path. Such advertisement would be useful for avoiding IFIT data leaking from the IFIT domain and measuring performance metrics on a per-service basis through steering packets of flow into a path where IFIT application are supported.

2. Definitions and Acronyms

- IFIT: In-situ Flow Information Telemetry
- OAM: Operation Administration and Maintenance
- NLRI: Network Layer Reachable Information, the NLRI advertised in the BGP UPDATE as defined in [RFC4271] and [RFC4760].

3. IFIT Capabilities

This document defines the IFIT Capabilities formed of a 16-bit bitmap. The following format is used:

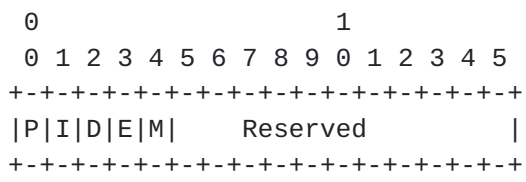


Figure 1. IFIT Capabilities

- P-Flag: IOAM Pre-allocated Trace Option Type flag. When set, this indicates that the router is capable of IOAM Pre-allocated Trace [I-D.ietf-ippm-ioam-data].
- I-Flag: IOAM Incremental Trace Option Type flag. When set, this indicates that the router is capable of IOAM Incremental Tracing [I-D.ietf-ippm-ioam-data].
- D-Flag: IOAM DEX Option Type flag. When set, this indicates that the router is capable of IOAM DEX [I-D.ioamteam-ippm-ioam-direct-export].
- E-Flag: IOAM E2E Option Type flag. When set, this indicates that the router is capable of IOAM E2E processing [I-D.ietf-ippm-ioam-data].
- M-Flag: Alternate Marking flag. When set, this indicates that the router is capable of processing Alternative Marking packets [RFC8321].
- Reserved: Reserved for future use. They MUST be set to zero upon transmission and ignored upon receipt.

4. Option 1: Extension to BGP Extended Community for IFIT-Capability Advertisement

4.1. IPv4-Address-Specific IFIT Tail Community

For IPv4 networks [RFC4360], this section defines a new type of BGP extended community called IPv4-Address-Specific IFIT Extended Community. The IPv4-Address-Specific IFIT Tail Community can be used by the IFIT decapsulation node to notify the IFIT Capabilities to its parterner (as the IFIT encapsulation node). It is a transitive extended community with type 0x01 and sub-type TBA1.

The format of this extended community is shown in Figure 2.

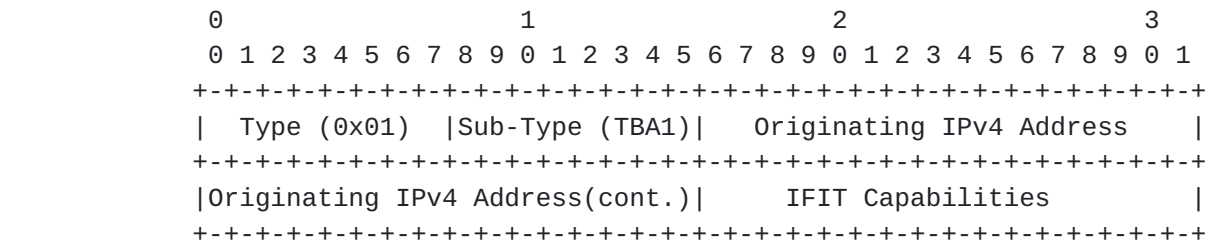


Figure 2. IPv4-Address-Specific IFIT Tail Community

- Originating IPv4 Address field: A 4 octets field. A IPv4 address of the IFIT decapsulation node. It is an IPv4 unicast address assigned by one of the Internet registries

- IFIT Capabilities: A 2 octets field. as defined in previous setion.

4.2. IPv6-Address-Specific IFIT Tail Community

For IPv6 networks [RFC5701], this section defines a new type of BGP extended community called IPv6-Address-Specific IFIT Extended Community. The IPv6-Address-Specific IFIT Tail Community can be used by the IFIT decapsulation node to notify the IFIT Capabilities to its parterner (as the IFIT encapsulation node). It is a transitive IPv6 address specific extended community with type 0x00 and sub-type TBA2.

The format of this extended community is shown in Figure 3.

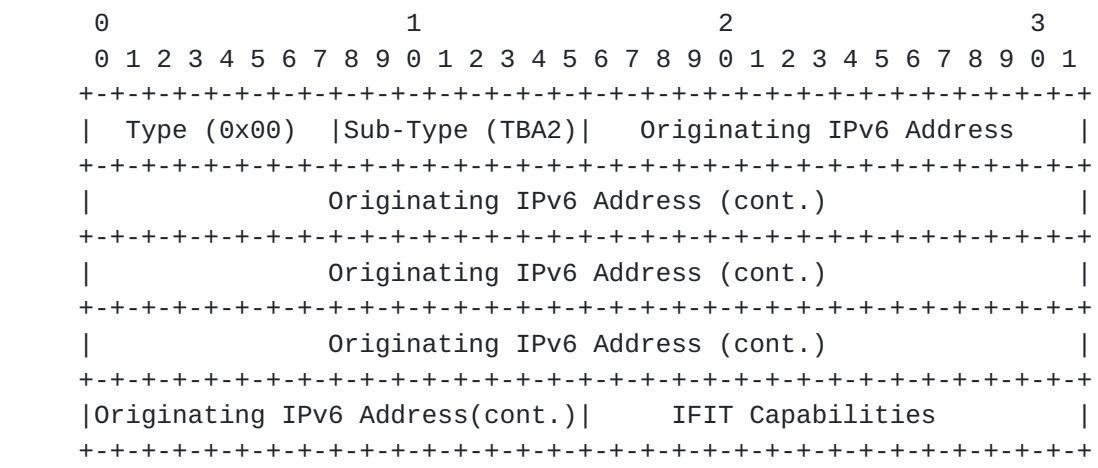


Figure 3. IPv6-Address-Specific IFIT Tail Community

- Originating IPv6 Address field: A IPv6 address of the IFIT decapsulation node. It is an IPv6 unicast address assigned by one of the Internet registries.

- IFIT Capabilities: as defined in previous setion.

In this option, the Originating IP Address (inclue IPv4 and IPv6) in the extended community attribute is used as the IFIT decapsulation node.

5. Option 2: Extension to BGP Next-Hop Capability for IFIT-Capability Advertisement

The BGP Next-Hop Capability Attribute [I-D.ietf-idr-next-hop-capability] is a non-transitive BGP attribute, which is modified or deleted when the next-hop is changed, to reflect the capabilities of the new next-hop. The attribute consists of a set of Next-Hop Capabilities.

A IFIT Next-Hop Capability is a triple (Capability Code, Capability Length, Capability Value) aka a TLV:

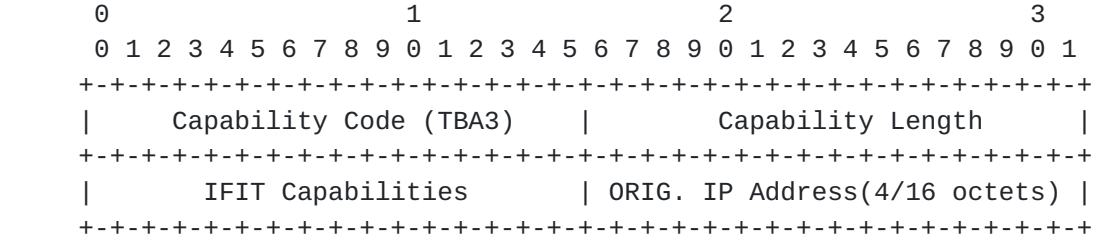


Figure 4. BGP Next-Hop Capability

- Capability Code: a two-octets unsigned binary integer which indicates the type of "Next-Hop Capability" advertised and unambiguously identifies an individual capability. This document defines a new Next-Hop Capability, which is called IFIT Next-Hop Capability. The Capability Code is TBA3.

- Capability Length: a two-octets unsigned binary integer which indicates the length, in octets, of the Capability Value field. A length of 0 indicates that no Capability Value field is present.

- IFIT Capabilities: as defined in previous setion.

- ORIG. IP Address: An IPv4 or IPv6 Address of the IFIT decapsulation node. It is an IPv4 or IPv6 unicast address assigned by one of the Internet registries.

A BGP speaker S that sends an UPDATE with the BGP Next-Hop Capability Attribute MAY include the IFIT Next-Hop Capability. The inclusion of the IFIT Next-Hop Capability with the NLRI advertised in the BGP UPDATE indicates that the BGP Next-Hop can act as the IFIT decapsulating node and it can process the specific IFIT encapsulation format indicated per the capability value. This is applied for all routes indicated in the same NRLI.

6. IANA Considerations

The IANA is requested to make the assignments for IPv4-Address-Specific IFIT Tail Community and IPv6-Address-Specific IFIT Tail Community:

Value	Description	Reference
TBA1	IPv4-Address-Specific IFIT Tail Community	This document
TBA2	IPv6-Address-Specific IFIT Tail Community	This document

The IANA is requested to make the assignments for IFIT Next-Hop Capability:

Value	Description	Reference
TBA3	IFIT Capabilities	This document

7. Security Considerations

This document defines extensions to BGP Extended Community and BGP Next-Hop Capability to advertise the IFIT capabilities. It does not introduce any new security risks to BGP.

Solutions to ensure the IFIT data propagated in a controlled domain and avoild malicious attacks, both of iOAM method [I-D.ietf-ippm-ioam-data]and Alternate Marking method [I-D.ietf-6man-ipv6-alt-mark] have respectively discussed that the implementation of both methods must be within a controlled domain.

8. Contributors

The following people made significant contributions to this document:

Weidong Li
Huawei
Email: poly.li@huawei.com

Haibo Wang
Huawei
Email: rainsword.wang@huawei.com

Tianran Zhou
Huawei
Email: zhoutianran@huawei.com

9. Acknowledgements

The authors would like to thank Giuseppe Fioccola for their reviews and suggestions

10. References

10.1. Normative References

[I-D.ietf-6man-ipv6-alt-mark]
Fioccola, G., Zhou, T., Cociglio, M., Qin, F., and R. Pang, "IPv6 Application of the Alternate Marking Method", draft-ietf-6man-ipv6-alt-mark-04 (work in progress), March 2021.

[I-D.ietf-idr-next-hop-capability]
Decraene, B., Kompella, K., and W. Henderickx, "BGP Next-Hop dependent capabilities", draft-ietf-idr-next-hop-capability-06 (work in progress), October 2020.

[I-D.ietf-ippm-ioam-data]
Brockners, F., Bhandari, S., and T. Mizrahi, "Data Fields for In-situ OAM", draft-ietf-ippm-ioam-data-12 (work in progress), February 2021.

[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>.

[RFC5701] Rekhter, Y., "IPv6 Address Specific BGP Extended Community Attribute", RFC 5701, DOI 10.17487/RFC5701, November 2009, <https://www.rfc-editor.org/info/rfc5701>.

10.2. Informative References

[I-D.ioamteam-ippm-ioam-direct-export]
Song, H., Gafni, B., Zhou, T., Li, Z., Brockners, F., Bhandari, S., Sivakolundu, R., and T. Mizrahi, "In-situ OAM Direct Exporting", draft-ioamteam-ippm-ioam-direct-export-00 (work in progress), October 2019.

[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>.

[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>.

[RFC8321] Fioccola, G., Ed., Capello, A., Cociglio, M., Castaldelli, L., Chen, M., Zheng, L., Mirsky, G., and T. Mizrahi, "Alternate-Marking Method for Passive and Hybrid Performance Monitoring", RFC 8321, DOI 10.17487/RFC8321, January 2018, <https://www.rfc-editor.org/info/rfc8321>.

[RFC8799] Carpenter, B. and B. Liu, "Limited Domains and Internet Protocols", RFC 8799, DOI 10.17487/RFC8799, July 2020, <https://www.rfc-editor.org/info/rfc8799>.

Authors' Addresses

Yali Wang
Huawei
Beijing
China

Email: wangyali11@huawei.com

Shunwan Zhuang
Huawei
Beijing
China

Email: zhuangshunwan@huawei.com

Yunan Gu
Huawei
Beijing
China

Email: guyunan@huawei.com

Ran Pang
China Unicom
Beijing
China

Email: pangran@chinaunicom.cn