Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 25 October 2025

G. Fioccola
Huawei
R. Pang
China Unicom
S. Wang
China Telecom
B. Decraene
Orange
S. Zhuang
H. Wang
Huawei
23 April 2025

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

draft-ietf-idr-bgp-ifit-capabilities-07

Abstract

In-situ Flow Information Telemetry (IFIT) refers to network OAM data plane on-path telemetry techniques, in particular In-situ OAM (IOAM) and Alternate Marking. This document defines a new Characteristic to advertise the In-situ Flow Information Telemetry (IFIT) capabilities. Within an IFIT domain, the IFIT capabilities 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 helps mitigating the leakage threat and facilitating the deployment of IFIT measurements on a per-service and on-demand basis.

Status of This Memo

Copyright Notice

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) [RFC9197] and Alternate Marking [RFC9341]. 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 describes 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, IFIT methods are deployed on a per-service and on-demand basis.

[I-D.ietf-idr-entropy-label] defines the BGP Next Hop Dependent Characteristics attribute (NHC). This document introduces a new NHC Characteristic 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, such as EVPNv4, EVPNv6, L2EVPN(EVPN VPWS and EVPN VPLS) routes, etc. 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 is also 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.

The IFIT NHC Characteristic, defined in this document, allows a distributed solution, while [I-D.ietf-idr-sr-policy-ifit] allows to centrally distribute Segment Routing (SR) Policies and can be considered as a centralized control solution. Therefore, this document enables the IFIT application in networks where no controller is introduced and it helps network operators to deploy IFIT in their networks.

Since BGP can be used to advertise a candidate path of a SR Policy ([I-D.ietf-idr-sr-policy-safi]), in a SR network it may be convenient to advertise IFIT capabilities in BGP as well, as specified in this document. While, in other scenarios, ICMPv6 can also be an alternative solution ([RFC9359]).

1.1. 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], RFC 8174 [RFC8174].

1.2. Definitions and Acronyms

* IFIT: In-situ Flow Information Telemetry. This term refers to the on-path telemetry techniques also known as In-situ OAM (IOAM) [RFC9197] and Alternate Marking [RFC9341].

* OAM: Operation Administration and Maintenance

* NLRI: Network Layer Reachable Information, the NLRI advertised in the BGP UPDATE as defined in [RFC4271] and [RFC4760].

2. IFIT Domain

IFIT deployment modes can include monitoring at node-level, tunnel-level, and service-level. The requirement of this document is to provide IFIT deployment at service-level, since different services may have different IFIT requirements. With the service-level solution, different IFIT methods can be deployed for different VPN services.

The figure shows an implementation example of IFIT application in a VPN scenario.

```
                  +----+                          +----+
      +----+      |    |          +----+          |    |      +----+
      |CE1 |------|PE1 |==========|RR/P|==========|PE2 |------|CE2 |
      +----+      |    |          +----+          |    |      +----+
                  +----+                          +----+
                   |<------------IFIT Domain--------->|
                   |<---------------BGP-------------->|
      |<----------------------------VPN--------------------------->|
```

Figure 1. Example of IFIT application in a VPN scenario

Figure 1

As Figure 1 shows, a traffic flow is sent out from the customer edge node CE1 to another customer edge node CE2. In order to enable IFIT application for this flow, the IFIT header must be encapsulated in the packet at the ingress provider edge node PE1, referred to as the

IFIT encapsulating node. Then, transit nodes in the IFIT domain may be able to support the IFIT capabilities in order to inspect IFIT extensions and, if needed, to update the IFIT data fields in the packet. Finally, the IFIT data fields must be exported and removed at egress provider edge node PE2 that is referred to as the IFIT decapsulating node. This is essential to avoid IFIT data leakage outside the controlled domain.

Since the IFIT decapsulating node MUST be able to handle and remove the IFIT header, the IFIT encapsulating node MUST know if the IFIT decapsulating node supports the IFIT application and, more specifically, which capabilities can be enabled.

3. IFIT Capabilities

This document defines the IFIT Capabilities as a 32-bit bitmap. The following format is used:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-----------------------------------------------------+
|P|I|D|E|M|                      Reserved                       |
+-+-+-+-+-+-----------------------------------------------------+
```

Figure 2. 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 [RFC9197].

* I-Flag: IOAM Incremental Trace Option Type flag. When set, this indicates that the router is capable of IOAM Incremental Tracing [RFC9197].

* D-Flag: IOAM DEX Option Type flag. When set, this indicates that the router is capable of IOAM DEX [RFC9326].

* E-Flag: IOAM E2E Option Type flag. When set, this indicates that the router is capable of IOAM E2E processing [RFC9197].

* M-Flag: Alternate Marking flag. When set, this indicates that the router is capable of processing Alternative Marking packets Alternate Marking [RFC9341].

* Reserved: Reserved for future use. They MUST be set to zero on transmission and ignored upon receipt.

3.1. IFIT Capabilities Advertisement

The NHC Attribute is defined in [I-D.ietf-idr-entropy-label]. It is an optional, transitive BGP attribute with type code 39. The NHC has as its data a network layer address, representing the next hop of the route the NHC accompanies. The NHC signals potentially useful optimizations, so it is desirable to make it transitive; the next hop data is to ensure correctness if it traverses BGP speakers that do not understand the NHC.

The Attribute Data field of the NHC attribute is encoded as a header portion that identifies the originator of the attribute, followed by one or more Characteristic TLVs.

It is modified or deleted when the next-hop is changed, to reflect the characteristic of the new next-hop.

The IFIT Characteristic described above can be encoded as an NHC Characteristic in the NHC attribute. It can be included in a BGP UPDATE message and indicates that the BGP Next-Hop supports the IFIT capabilities for the NLRI advertised in this BGP UPDATE.

The Network Address of Next Hop, as part of the NHC, is the IPv4 or IPv6 Address of the IFIT decapsulating node.

The IFIT NHC Characteristic is defined below and is a triple (Characteristic Code, Characteristic Length, Characteristic Value) aka a TLV:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Characteristic Code = 4    |   Characteristic Length = 4   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   IFIT Characteristic Value                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3. IFIT NHC Characteristic

* Characteristic Code: a two-octet unsigned binary integer that indicates the type of Characteristic advertised and unambiguously identifies an individual characteristic. This document defines a new NHC Characteristic Code called IFIT Characteristic. The Characteristic Code is 4 (as allocated by [I-D.ietf-idr-entropy-label]).

* Characteristic Length: a two-octet unsigned binary integer that indicates the length, in octets, of the Characteristic Value field. The length MUST be four octets.

* IFIT Characteristic Value: IFIT Capabilities as defined in Section 3.

3.2. Error handling

The IFIT NHC Characteristic TLV is considered malformed and must be disregarded if its length is other than four.

3.3. Operation

A BGP speaker that sends an UPDATE with the NHC Attribute MAY include the IFIT Characteristic if IFIT is configured and enabled. The inclusion of the IFIT Characteristic 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 in the characteristic value. This is applied for all routes indicated in the same NRLI.

The IFIT Characteristic MUST reflect the capabilities of the router indicated in the BGP Next-Hop. If a BGP speaker sets the BGP Next-Hop to an address of a different router, it MUST NOT advertise the IFIT Characteristic not supported by this router. Therefore the IFIT Characteristic MUST be re-advertised according to the new BGP Next-Hop.

In case of large networks, the IFIT domain may span across multiple Autonomous Systems (ASes) and hence the IFIT Characteristic needs to be able to cross AS boundaries if configured to do so. In this case, it is also possible to pass this information between BGP clusters to keep the IFIT methods consistent. BGP Link-State (BGP-LS) may allow to bring the information back to a centralized controller as well.

4. IANA Considerations

The IFIT NHC Characteristic Code has been allocated by [I-D.ietf-idr-entropy-label] from the proposed "BGP Next Hop Dependent Characteristic Codes" within the Border Gateway Protocol (BGP) Parameters group. IANA is requested to update the reference to this document.

Value	Description	Reference
4	IFIT	This document

Table 1

5. Security Considerations

This document defines a new NHC Characteristic to advertise the IFIT capabilities. It does not introduce any new security considerations beyond the one described in [I-D.ietf-idr-entropy-label].

IFIT methods are applied within a controlled domain and solutions MUST be taken to ensure that the IFIT data are properly propagated to avoid malicious attacks. Both IOAM method [RFC9197] and Alternate Marking [RFC9341] [RFC9343] respectively discusses that the implementation of both methods MUST be within a controlled domain.

The NHC Characteristic Attribute being a transitive attribute in order to facilitate early deployments it may leak outside of the domain if both the NLRI carrying this characteristic is advertised outside of the domain and the ASBR does not support [I-D.ietf-idr-entropy-label]. In general, it is not an issue for IFIT because the only information about the capabilities would be leaked. However if any characteristic leakage must be avoided, one must ensure that all the border routers must support the NHC Characteristic feature.

6. Contributors

The following people made significant contributions to this document:

Yali Wang
Huawei
Email: wangyali11@huawei.com

Yunan Gu
Huawei
Email: guyunan@huawei.com

Tianran Zhou
Huawei
Email: zhoutianran@huawei.com

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

7. Acknowledgements

The authors would like to thank John Scudder, Ketan Talaulikar, Haoyu Song, Jie Dong, Robin Li, Jeffrey Haas, Robert Raszuk, Zongpeng Du, Yisong Liu, Yongqing Zhu, Aijun Wang, Fan Yang for their reviews and suggestions.

8. References

8.1. Normative References

[I-D.ietf-idr-entropy-label]
Decraene, B., Scudder, J., Kompella, K., Satya, M. R., Wen, B., Wang, K., and S. Krier, "BGP Next Hop Dependent Characteristics Attribute", Work in Progress, Internet-Draft, draft-ietf-idr-entropy-label-17, 30 March 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-entropy-label-17>.

[I-D.ietf-idr-sr-policy-ifit]
Qin, F., Yuan, H., Yang, S., Zhou, T., and G. Fioccola, "BGP SR Policy Extensions to Enable IFIT", Work in Progress, Internet-Draft, draft-ietf-idr-sr-policy-ifit-10, 18 April 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-policy-ifit-10>.

[I-D.ietf-idr-sr-policy-safi]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-sr-policy-safi-13, 6 February 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-policy-safi-13>.

[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>.

[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>.

[RFC9197] Brockners, F., Ed., Bhandari, S., Ed., and T. Mizrahi, Ed., "Data Fields for In Situ Operations, Administration, and Maintenance (IOAM)", RFC 9197, DOI 10.17487/RFC9197, May 2022, <https://www.rfc-editor.org/info/rfc9197>.

[RFC9326] Song, H., Gafni, B., Brockners, F., Bhandari, S., and T. Mizrahi, "In Situ Operations, Administration, and Maintenance (IOAM) Direct Exporting", RFC 9326, DOI 10.17487/RFC9326, November 2022, <https://www.rfc-editor.org/info/rfc9326>.

[RFC9341] Fioccola, G., Ed., Cociglio, M., Mirsky, G., Mizrahi, T., and T. Zhou, "Alternate-Marking Method", RFC 9341, DOI 10.17487/RFC9341, December 2022, <https://www.rfc-editor.org/info/rfc9341>.

[RFC9343] Fioccola, G., Zhou, T., Cociglio, M., Qin, F., and R. Pang, "IPv6 Application of the Alternate-Marking Method", RFC 9343, DOI 10.17487/RFC9343, December 2022, <https://www.rfc-editor.org/info/rfc9343>.

8.2. Informative 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>.

[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>.

[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>.

[RFC9359] Min, X., Mirsky, G., and L. Bo, "Echo Request/Reply for Enabled In Situ OAM (IOAM) Capabilities", RFC 9359, DOI 10.17487/RFC9359, April 2023, <https://www.rfc-editor.org/info/rfc9359>.

Authors’ Addresses

Giuseppe Fioccola
Huawei
Viale Martesana, 12
20055 Vimodrone (Milan)
Italy
Email: giuseppe.fioccola@huawei.com

Ran Pang
China Unicom
9 Shouti South Rd.
Beijing
100089
China
Email: pangran@chinaunicom.cn

Subin Wang
China Telecom
Guangzhou
China
Email: wangsb6@chinatelecom.cn

Bruno Decraene
Orange
Email: bruno.decraene@orange.com

Shunwan Zhuang
Huawei
Huawei Building, No.156 Beiqing Road

Beijing
100095
China
Email: zhuangshunwan@huawei.com

Hiabo Wang
Huawei
Huawei Building, No.156 Beiqing Road
Beijing
100095
China
Email: rainsword.wang@huawei.com