Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: November 10, 2022

G. Fioccola
Huawei
R. Pang
China Unicom
S. Zhuang
H. Wang
Huawei

May 9, 2022

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

draft-wang-idr-bgp-ifit-capabilities-05

Abstract

This document defines extensions to BGP [RFC4271] 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

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.

This document introduces 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, 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 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.

This document defines an IFIT Next-Hop Capability Attribute according to [I-D.ietf-idr-next-hop-capability]. It allows a distributed solution that does not require the participation of centralized control element, while [I-D.ietf-idr-sr-policy-ifit] allows to centrally distribute 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.

1.1. Definitions and Acronyms

- o IFIT: In-situ Flow Information Telemetry
- o OAM: Operation Administration and Maintenance
- o 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

As the figure 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 formed of a 16-bit bitmap. The following format is used:

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

Figure 2. IFIT Capabilities

- o 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].

- o 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].

- o 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].

- o 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].

- o M-Flag: Alternate Marking flag. When set, this indicates that the router is capable of processing Alternative Marking packets [RFC8321].

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

4. BGP Next-Hop IFIT Capability Advertisement

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

The IFIT Capabilities described above can be encoded as a BGP Next-Hop IFIT Capability Attribute. It can be included in a BGP UPDATE message and indicates that the BGP Next-Hop supports the IFIT capability for the NLRI advertised in this BGP UPDATE.

The IFIT Next-Hop Capability is defined below and is a triple (Capability Code, Capability Length, Capability 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
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |     Capability Code (TBA1)    |        Capability Length      |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |       IFIT Capabilities       |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |            ORIG. IP Address(4/16 octets)                      |
     |            ...                                                |
     |                                                               |
     |                                                               |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3. BGP Next-Hop Capability

- o 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 TBA1.

- o 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.

- o IFIT Capabilities: as defined in previous section.

- o 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.

5. Hop-by-Hop and Head-to-Tail Mechanisms

When all devices are upgraded to support IFIT, the hop-by-hop mechanism can be suitable. In the current stage, where new and old devices are deployed together, we must first ensure that the tail node can properly decapsulate the IFIT header, so we need an advertisement mechanism from the head node to the tail node.

Further, different services on the egress node may have different IFIT requirements, so the capability advertisement from the head node to the tail node is always required.

However, hop-by-hop and head-to-tail mechanisms can eventually be used together without conflict.

6. IANA Considerations

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

Value	Description	Reference
TBA1	IFIT Capabilities	This document

7. Security Considerations

This document defines extensions to BGP Next-Hop Capability to advertise the IFIT capabilities. It does not introduce any new security risks to BGP, as also mentioned in [I-D.ietf-idr-next-hop-capability].

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 [I-D.ietf-ippm-ioam-data] and Alternate Marking method [I-D.ietf-6man-ipv6-alt-mark] 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:

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

9. Acknowledgements

The authors would like to thank 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

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-14 (work in progress), April 2022.

[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-07 (work in progress), December 2021.

[I-D.ietf-idr-sr-policy-ifit]
Qin, F., Yuan, H., Zhou, T., Fioccola, G., and Y. Wang, "BGP SR Policy Extensions to Enable IFIT", draft-ietf-idr-sr-policy-ifit-03 (work in progress), January 2022.

[I-D.ietf-ippm-ioam-data]
Brockners, F., Bhandari, S., and T. Mizrahi, "Data Fields for In-situ OAM", draft-ietf-ippm-ioam-data-17 (work in progress), December 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>.

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

Giuseppe Fioccola
Huawei
Munich
Germany

Email: giuseppe.fioccola@huawei.com

Ran Pang
China Unicom
Beijing
China

Email: pangran@chinaunicom.cn

Shunwan Zhuang
Huawei
Beijing
China

Email: zhuangshunwan@huawei.com

Hiabo Wang
Huawei
Beijing
China

Email: rainsword.wang@huawei.com