Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 15, 2021
Y. Wang
S. Zhuang
Y. Gu
Huawei
July 14, 2020

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

draft-wang-idr-bgp-ifit-capabilities-00

Abstract

This document defines extensions to BGP to advertise the In-situ Flow Information Telemetry (IFIT) capabilities. Within an IFIT measurement domain, the capability is meant to be advertised from the IFIT tail node to the head node to assist the head node to determine whether a particular IFIT Option type can be enabled. This facilitates 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

At present, a family of on-path flow telemetry techniques referred in [I-D.song-opsawg-ifit-framework] are emerging, including In-situ OAM (IOAM) [I-D.ietf-ippm-ioam-data], Postcard-Based Telemetry (PBT) [I-D.song-ippm-postcard-based-telemetry], IOAM Direct Export (DEX) [I-D.ioamteam-ippm-ioam-direct-export], Enhanced Alternate Marking (EAM) [I-D.zhou-ippm-enhanced-alternate-marking].

In-situ Flow Information Telemetry (IFIT) determines network performance by measuring the packet loss and latency of service packets transmitted in an IP network. This feature is easy to deploy and provides an accurate assessment of network performance.

The IFIT model describes how service flows are measured to obtain packet loss and latency. Specifically, IFIT measures the packet loss and latency of service flows on the ingress and egress of the transit network, and summarizes desired performance indicators. The IFIT model is composed of three objects: target flow, transit network, and measurement system. The transit network only bears target flows. The target flows are not generated or terminated on the transit network. The transit network can be a Layer 2 (L2), Layer 3 (L3), or L2+L3 hybrid network. Each node on the transit network must be reachable at the network layer. The measurement system consists of the ingress (configured with IFIT and IFIT parameters) and multiple IFIT-capable devices.

IFIT is a solution focusing on network domains. The "network domain" consists of a set of network devices or entities within a single administration. One network domain MAY consists of multiple IFIT domain. The family of emerging on-path flow telemetry techniques 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 usecases, the IFIT Features are deployed on a per-service and on-demand basis. Within the IFIT domain, one or more IFIT-options are added into packet at the IFIT-enabled head node that is referred to as the IFIT encapsulating node. Then IFIT data fields MAY be updated by IFIT transit nodes that the packet traverses. Finally, the data fields are removed at a device that is referred to as the IFIT decapsulating node. Hence, a head node needs to know if the IFIT decapsulating node is able to support the IFIT capabilities.

This document defines extensions to BGP to advertise the IFIT capabilities to a head node, then the head node can learn the IFIT capabilities and determine whether a particular IFIT Option type can be supported by the sending side as the decapsulating node. This facilitates the deployment of IFIT measurements on a per-service and on-demand basis.

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:

```
 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 1. 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: EAM flag. When set, this indicates that the router is capable of processing Enhanced Alternative Marking packets [I-D.zhou-ippm-enhanced-alternate-marking].

- o Reserved: Must be set to zero upon transmission and ignored upon receipt.

4. Option 1: Extending BGP Extended Community for IFIT Capability

4.1. IPv4-Address-Specific IFIT Extended Community

For IPv4 networks, this section defines a new type of BGP extended community [RFC4360] called IPv4-Address-Specific IFIT Extended Community. The IPv4-Address-Specific IFIT Extended 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 TBA.

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

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Type (0x01)  | Sub-Type (TBA)|     IFIT Capabilities         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Originating IPv4 Address                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2. IPv4-Address-Specific IFIT Extended Community

- o IFIT Capabilities: as defined in previous setion.

- o Originating IPv4 Address field: A IPv4 address of the IFIT decapsulation node.

4.2. IPv6-Address-Specific IFIT Extended Community

For IPv6 networks, this section defines a new type of BGP extended community[RFC4360] called IPv6-Address-Specific IFIT Extended Community. The IPv6-Address-Specific IFIT Extended 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 TBA.

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

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Type (0x00)  | Sub-Type (TBA)|    IFIT Capabilities          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Originating IPv6 Address                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Originating IPv6 Address (cont.)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Originating IPv6 Address (cont.)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Originating IPv6 Address (cont.)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3. IPv6-Address-Specific IFIT Extended Community

- o IFIT Capabilities: as defined in previous setion.

- o Originating IPv6 Address field: A IPv6 address of the IFIT decapsulation node.

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

5. Option 2: Extendng BGP Next-Hop Capability for IFIT Capability

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

```
 0                   1
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Capability Code (2 octets)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Capability Length (2 octets)  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Capability Value (variable)   |
~                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4. 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 TBD1.

- 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 Capability Value: a variable-length field. It is interpreted according to the value of the Capability Code. For the IFIT Next-Hop Capability, Capability Value contains IFIT Capabilities and Originate IP Address, as shown in the following figure.

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| IFIT Capabilities (2 octets)  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Originating IP Address        |
| (4 or 16 octets)              |
~                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5. BGP Capability Value for IFIT

- o IFIT Capabilities: as defined in previous setion.
- o Originating IP Address: An IPv4 or IPv6 Address of the IFIT decapsulation node.

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.

A BGP speaker receiving an IFIT Next-Hop Capability containing the IFIT options that it can supports behaves as defined in the document defining these iFIT options.

6. IANA Considerations

TBD

7. Security Considerations

TBD

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

TBD

10. References

10.1. Normative References

[I-D.ietf-bess-srv6-services]
Dawra, G., Filsfils, C., Raszuk, R., Decraene, B., Zhuang, S., and J. Rabadan, "SRv6 BGP based Overlay services", draft-ietf-bess-srv6-services-03 (work in progress), July 2020.

[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-05 (work in progress), June 2019.

[I-D.ietf-ippm-ioam-data]
Brockners, F., Bhandari, S., and T. Mizrahi, "Data Fields for In-situ OAM", draft-ietf-ippm-ioam-data-10 (work in progress), July 2020.

[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.

[I-D.song-ippm-postcard-based-telemetry]
Song, H., Zhou, T., Li, Z., Shin, J., and K. Lee, "Postcard-based On-Path Flow Data Telemetry", draft-song-ippm-postcard-based-telemetry-07 (work in progress), April 2020.

[I-D.wang-idr-bgp-ls-ifit-node-capability]
Wang, Y., Zhou, T., and R. Pang, "Extensions to BGP-LS for Advertising In-situ Flow Information Telemetry (IFIT) Node Capability", draft-wang-idr-bgp-ls-ifit-node-capability-03 (work in progress), March 2020.

[I-D.zhou-ippm-enhanced-alternate-marking]
Zhou, T., Fioccola, G., Lee, S., Cociglio, M., and W. Li, "Enhanced Alternate Marking Method", draft-zhou-ippm-enhanced-alternate-marking-05 (work in progress), July 2020.

[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.song-opsawg-ifit-framework]
Song, H., Qin, F., Chen, H., Jin, J., and J. Shin, "In-situ Flow Information Telemetry", draft-song-opsawg-ifit-framework-12 (work in progress), April 2020.

[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>.

[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>.

[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>.

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