Link State Routing Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 24, 2020

Y. Wang
T. Zhou
M. Liu
Huawei
R. Pang
China Unicom

March 23, 2020

In-situ Flow Information Telemetry (IFIT) Node Capability Advertisement

draft-wang-lsr-ifit-node-capability-advertisement-00

Abstract

For advertising In-situ Flow Information Telemetry (IFIT) node capabilities within the entire routing domain, this document extends a new optional TLV to the OSPF RI Opaque LSA, a new optional sub-TLV to the IS-IS Router CAPABILITY TLV, and a new Node Attribute TLV that is encoded in the BGP-LS attribute with Node NLRIs to carry IFIT node capabilities information. Such advertisement allows entities (e.g. a centralized controller) to determine whether a particular IFIT functionality can be supported in a given network.

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

IFIT provides a complete framework architecture and a reflection-loop working solution for on-path flow telemetry [I-D.song-opsawg-ifit-framework]. At present, there are a family of emerging on-path flow telemetry techniques, 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], etc. IFIT is a solution focusing on network domains. The "network domain" consists of a set of network devices or entities within a single Atonomous System (AS). The part of the network which employs IFIT is referred to as the IFIT domain. One network domain may consist of multiple IFIT domains. An IFIT domain may cross multiple network domains.

The family of emerging on-path flow telemetry techniques may be selectively or partially enabled in different vendors' devices as an emerging feature for various use cases of application-aware network operations. So that in order to dynamically enable IFIT functionality in a network domain, it is necessary to advertise the information of IFIT option types supported in each device.

BGP-LS defines a way to advertise topology and associated attributes and capabilities of the nodes in that topology to a centralized controller [RFC7752]. Typically, BGP-LS is configured on a small number of nodes that do not necessarily act as head-ends. In order for BGP-LS to signal IFIT node capabilities for all the devices in the network, IFIT node capabilities SHOULD be advertised by every IGP router in the network. Therefore, this document defines extensions to OSPF, IS-IS, and BGP-LS to advertise the IFIT node capabilities. Entities (e.g. centralized controllers) that can use this information to determine whether a particular IFIT functionality can be enabled in a given IFIT domain. An application to this information advertisement is described in detail in Section 4.

2. Terminology

OSPF: Open Shortest Path First

IS-IS: Intermediate System to Intermediate System

RI: Router Information

LSA: Link State Advertisement

BGP-LS: Advertisement of Link-State and TE Information using Border Gateway Protocol

3. IFIT Node Capability Advertisement

3.1. IFIT Node Capability Information

Each IFIT node is configured with a node-id which uniquely identifies a node within the associated IFIT domain. To accommodate the different use cases or requirements of in-situ flow information telemetry, IFIT data fields updated by network nodes fall into different categories which are referred as different IFIT option types, including IOAM Trace Option-Types [I-D.ietf-ippm-ioam-data], IOAM Edge-to-Edge (E2E) Option-Type [I-D.ietf-ippm-ioam-data], IOAM DEX Option-Type [I-D.ioamteam-ippm-ioam-direct-export] and Enhanced Alternate Marking (EAM) Option-Type [I-D.zhou-ippm-enhanced-alternate-marking]. And a subset or all the IFIT-Option-Types and their corresponding IFIT-Data-Fields can be

associated to an IFIT-Namespce. The Namespace identifiers allow devices which are IFIT capable to determine whether IFIT-Option-Types need to be processed. So that IFIT-Option-Types and Namespace-ID SHOULD be carried in IFIT node capability information for advertisement.

In this document, the IFIT-Node-Capability information consists of one or more pairs of a 2-octet Namespace-ID and 16-bit Option-Type enabled Flag.

```
 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
+-----------------------------------------------------------------+
|          Namespace-ID_1       |  Option-Type enabled Flag_1   |
+-----------------------------------------------------------------+
|          Namespace-ID_2       |  Option-Type enabled Flag_2   |
+-----------------------------------------------------------------+
|              ...              |              ...              |
+-----------------------------------------------------------------+
```

Fig. 4 IFIT-Node-Capability Format

Where:

Namespace-ID: A 16-bit identifier, which MUST be present and populated in all IFIT-Option-Types. The definition is the same as described in [I-D.ietf-ippm-ioam-data].

Option-Type enabled Flag: A 16-bit field, which is defined as:

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-------------------------------+
|p|i|d|e|m|       Reserved      |
+-------------------------------+
```

Where:

p-Flag: IOAM Pre-allocated Trace Option Type-enabled flag. If bit p is set (1), the router is capable of IOAM Pre-allocated Trace [I-D.ietf-ippm-ioam-data].

i-Flag: IOAM Incremental Trace Option Type-enabled flag. If bit i is set (1), the router is capable of IOAM Incremental Tracing [I-D.ietf-ippm-ioam-data].

d-Flag: IOAM DEX Option Type-enabled flag. If bit d is set (1), the router is capable of IOAM DEX [I-D.ioamteam-ippm-ioam-direct-export].

e-Flag: IOAM E2E Option Type-enabled flag. If bit e is set (1), the router is capable of IOAM E2E processing [I-D.ietf-ippm-ioam-data].

m-Flag: Enhanced Alternative Marking enabled flag. If bit m is set (1), then the router is capable of processing Enhanced Alternative Marking packets [I-D.zhou-ippm-enhanced-alternate-marking].

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

An IFIT node SHALL be capable of more than one IFIT option types. So in this case, Option-Type enabled Flag can has more than one bit being set.

3.2. OSPF Extension IFIT Node Capability TLV

Given that OSPF uses the options field in LSAs and hello packets to advertise optional router capabilities [RFC7770], this document defnes a new IFIT Node Capability TLV within the body of the OSPF RI Opaque LSA [RFC7770] to carry the IFIT node capabilities of the router originating the RI LSA. The IFIT Node Capability TLV is composed of three fields, a 2-octet Type field, a 2-octet Length field, and 4-octet Value field. The Type field indicates the type of items in the Value field. The Length field indicates the length of the Value field in octets. The Value field carries the IFIT Node Capability information, which is a multiple of 4 octets field.

The IFIT node capability TLV has the following format:

```
 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            |             Length            |
+---------------------------------------------------------------+
|                      IFIT-Node-Capability                     |
~                                                               ~
+---------------------------------------------------------------+
```

Fig. 1 OSPF IFIT Node Capability TLV Format

Type: To be assigned by IANA

Length: A 2-octet field that indicates the length of the value field.

IFIT-Node-Capability: A multiple of 4 octets field, which is as defined in Section 3.1.

3.3. IS-IS Extension IFIT Node Capability Sub-TLV

The IS-IS Extensions for Advertising Router Information TLV named IS-IS Router CAPABILITY TLV [RFC7981], which allows a router to announce its capabilities within an IS-IS level or the entire routing domain, has been chosen for IFIT node capabilities advertisement. IS-IS Router CAPABILITY TLV is formed of multiple sub-TLVs [RFC5305].

According to the format of IS-IS Router CAPABILITY TLV [RFC7981], the IFIT Node Capability sub-TLV is composed of three fields, a one-octet Type field, a one-octet Length field, and zero or more octets of Value. The Type field indicates the type of items in the Value field. The Length field indicates the length of the Value field in octets. The Value field indicates the IFIT Node Capability, which is a multiple of 4 octets field.

The IS-IS IFIT Node-capability Sub-TLV has the following format:

```
 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     |    Length     |
+---------------+---------------+-------------------------------+
|                      IFIT-Node-Capability                     |
~                                                               ~
+---------------------------------------------------------------+
```

Fig. 2 IS-IS IFIT Node Capability Sub-TLV Format

Type: To be assigned by IANA

Length: A 8-bit field that indicates the length of the value portion in octets.

IFIT-Node-Capability: A multiple of 4 octets field, which is as defined in Section 3.1.

3.4. BGP-LS Extension IFIT Node Capability TLV

This document describes extensions enabling BGP-LS speakers to announce the IFIT node capabilities of routers in a network to a BGP-LS consumer (e.g. a centralized controller). The centralized controller can leverage this information in enabling IFIT applications in network domains based on IFIT node capabilities and OAM use cases.

IFIT Node-Capability TLV is defined as a new Node Attribute TLV that is encoded in the BGP-LS attribute with Node NLRIs [RFC7752]. The

IFIT Node Capability TLV is defined as a TLV triplet, i.e. a 2-octet Type field, a 2-octet Length field, and 4-octet Value field. The Type field indicates the type of items in the Value field. The Length field indicates the length of the Value field in octets. The Value field indicates the IFIT Node Capability, which is a multiple of 4 octets field.

The BGP-LS IFIT Node Capability TLV has the following format:

```
 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            |             Length            |
+---------------------------------------------------------------+
|                      IFIT-Node-Capability                     |
~                                                               ~
+---------------------------------------------------------------+
```

Fig. 3 BGP-LS IFIT Node Capability TLV Format

Type: To be assigned by IANA

Length: A 2-octet field that indicates the length of the value.

IFIT-Node-Capability: A multiple of 4 octets field, which is as defined in Section 3.1.

4. Application

Within an IFIT domain, one or more IFIT-Option-Types are added into packets at the IFIT-capable 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, IFIT-Option-Types are removed at the IFIT-capable end node that is referred to as the IFIT decapsulating node. The role of an IFIT-encapsulating, IFIT-transit or IFIT-decapsulating node is always performed within a specific Namespce.

As any packet with IFIT-specific header and metadata MUST not leak out from the IFIT domain, the IFIT decapsulating node MUST be able to capture packets with IFIT-specific header and metadata and recover their format before forwarding them out of the IFIT domain. So that entities (e.g., centralized controllers) can use IFIT node capabilities information to avoid the leak of IFIT-specific header and metadata.

Besides, in order to adapt to different network conditions and different application requirements, a centralized controller needs to switch between different underlying techniques. As different IFIT option types have different encapsulation format in packets and have different processing procedure when packets travers to encapsulating, transit, and decapsulating nodes. For example, for IOAM Trace Option-Types, IOAM tracing data is expected to be collected at every IOAM transit node that a packet traverses to ensure visibility into the entire path a packet takes within an IOAM-domain. If not all nodes within a domain are IOAM Trace Option-Type capable, IOAM-Data-Fields will only be changed on those nodes which are IOAM Trace Option-Type capable and IOAM tracing information will only be collected by those IOAM-capable nodes. For IOAM DEX Option-Type, the requred IOAM data is expected to be exported at every transit node that process a packet with the DEX option.

Therefore, this advertisement allows entities (e.g., centralized controllers) to determine whether a specific IFIT functionality can be supported by all devices in a network domain, then enable the IFIT-Data-Fields encapsulation at the head node.

5. IANA Considerations

This document makes the following registrations for a TLV type of the new IFIT Node Capability TLV within the body of the OSPF RI Opaque LSA, a Sub-TLV type of the new Sub-TLV proposed from the "Sub-TLVs for TLV 242 (IS-IS Router CAPABILITY TLV)" registry, and a BGP-LS Node Attribute TLV code point for the IFIT Node Capability TLV.

Type	Description
TBD	OSPF Extension IFIT Node Capability TLV

Type	Description
TBD	IS-IS Extension IFIT Node Capability Sub-TLV

Code Point	Description
TBD	BGP-LS Extension IFIT Node Capability TLV

6. Security Considerations

This document introduces a new TLV within the existing OSPF RI Opaque LSA, a new sub-TLV for the existing IS-IS Router capability TLV, and a new Node Attribute TLV for the existing Node NLRIs. It does not introduce any new security risks to OSPF, IS-IS and BGP-LS.

7. Acknowledgements

TBD.

8. References

8.1. Normative References

[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>.

[RFC5305] "IS-IS Extensions for Traffic Engineering", <https://www.rfc-editor.org/info/rfc5305>.

[RFC7752] "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", <https://datatracker.ietf.org/doc/rfc7752/>.

[RFC7770] "Extensions to OSPF for Advertising Optional Router Capabilities", <https://www.rfc-editor.org/info/rfc7770>.

[RFC7981] "IS-IS Extensions for Advertising Router Information", <https://www.rfc-editor.org/info/rfc7981>.

8.2. Informative References

[I-D.ietf-ippm-ioam-data] "Data Fields for In-situ OAM".

[I-D.ioamteam-ippm-ioam-direct-export] "In-situ OAM Direct Exporting", <https://datatracker.ietf.org/doc/draft-ioamteam-ippm-ioam-direct-export/>.

[I-D.song-ippm-postcard-based-telemetry] "Postcard-based On-Path Flow Data Telemetry", <https://datatracker.ietf.org/doc/draft-song-ippm-postcard-based-telemetry/>.

[I-D.song-opsawg-ifit-framework]
"In-situ Flow Information Telemetry Framework",
<https://datatracker.ietf.org/doc/draft-song-opsawg-ifit-framework/>.

[I-D.zhou-ippm-enhanced-alternate-marking]
"Enhanced Alternate Marking Method",
<https://datatracker.ietf.org/doc/draft-zhou-ippm-enhanced-alternate-marking/>.

Authors' Addresses

Yali Wang
Huawei
156 Beiqing Rd., Haidian District
Beijing
China

Email: wangyali11@huawei.com

Tianran Zhou
Huawei
156 Beiqing Rd., Haidian District
Beijing
China

Email: zhoutianran@huawei.com

Min Liu
Huawei
156 Beiqing Rd., Haidian District
Beijing
China

Email: lucy.liumin@huawei.com

Ran Pang
China Unicom
9 Shouti South Rd., Haidian District
Beijing
China

Email: pangran@chinaunicom.cn