Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 18, 2019
July 17, 2018

Network Monitoring Protocol (NMP)
draft-gu-network-monitoring-protocol-00

Abstract

To evolve towards automated network OAM (Operations, administration and management), the monitoring of control plane protocols is a fundamental necessity. In this document, a network monitoring protocol (NMP) is proposed to provision the running status information of control plane protocols, e.g., IGP (Interior Gateway Protocol) and other protocols. By collecting the protocol monitoring data and reporting it to the NMP monitoring server in real-time, NMP can facilitate network troubleshooting. In this document, NMP for IGP troubleshooting are illustrated to showcase the necessity of NMP. IS-IS is used as the demonstration protocol, and the case of OSPF (Open Shortest Path First) and other control protocols will be elaborated in the future versions. The operations of NMP are described, and the NMP message types and message formats are defined in the document.

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 3, 2019.

Copyright Notice

Copyright (c) 2018 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction ........................................... 3
   1.1. Motivation ...................................... 3
   1.2. Overview ....................................... 4
2. Terminology ............................................. 5
3. Use Cases ............................................... 5
   3.1. IS-IS Adjacency Issues ........................... 5
   3.2. Forwarding Path Disconnection ..................... 6
   3.3. IS-IS LSP Synchronization Failure .................. 6
4. NMP Message Format ....................................... 7
   4.1. Protocol Selection Options ......................... 7
   4.2. Message Types .................................... 7
   4.3. Message Format ................................... 8
      4.3.1. Common Header ................................ 8
      4.3.2. Per Adjacency Header ......................... 9
      4.3.3. Initiation Message ............................ 10
      4.3.4. Adjacency Status Change Notification ........... 11
      4.3.5. Statistic Report Message ....................... 12
      4.3.6. IS-IS PDU Monitoring Message ................. 14
      4.3.7. Termination Message ........................... 14
5. IANA .................................................... 15
6. Contributors ............................................ 15
7. Acknowledgments ......................................... 15
8. References .............................................. 15
Authors’ Addresses ......................................... 17
1. Introduction

1.1. Motivation

The requirement for better network OAM approaches has been greatly driven by the network evolvement. The concept of network Telemetry has been proposed to meet the current and future OAM demands w.r.t., massive and real-time data storage, collection, process, exportion, and analysis, and an architectural framework of existing Telemetry approaches is introduced in [I-D.song-ntf]. Network Telemetry provides visibility to the network health conditions, and is beneficial for faster network troubleshooting, network OpEx (operating expenditure) reduction, and network optimization. Telemetry can be applied to the data plane, control plane and management plane. There have been various methods proposed for each plane:

- **Management plane**: For example, SNMP (Simple Network Management Protocol) [RFC1157], NETCONF (Network Configuration Protocol) [RFC6241] and gNMI (gRPC Network Management Interface) [I-D.openconfig-rtgw-gnmi-spec] are three typical widely adopted management plane Telemetry approaches. Various YANG modules are defined for network operational state retrieval and configuration management. Subscription to specific YANG datastore can be realized in combination with gRPC/NETCONF.

- **Data plane**: For example, In-situ OAM (iOAM) [I-D.brockners-inband-oam-requirements] embeds an instruction header to the user data packets, and collects the requested data and adds it to the user packet at each network node along the forwarding path. Applications such as path verification, SLA (service-level agreement) assurance can be enabled with iOAM.

- **Control Plane**: BGP monitoring protocol (BMP) [RFC7854] is proposed to monitor BGP sessions and intended to provide a convenient interface for obtaining BGP route views. Data collected using BMP can be further analyzed with big data platforms for network health condition visualization, diagnose and prediction applications.

The general idea of most Telemetry approaches is to collect various information from devices and export to the centralized server for further analysis, and thus providing more network insight. It should not be surprising that any future and even current Telemetry applications may require the fusion of data acquired from more than one single approach/one single plane. For example, for network troubleshooting purposes, it requires the collection of comprehensive information from devices, such system ID/router ID, interface status, PDUs (protocol data units), device/protocol statistics and so on.
Information such as system ID/router ID can be reported by management plane Telemetry approaches, while the protocol related data (especially PDUs) are more fit to be monitored using the control plane Telemetry. With rich information collected in real time at the centralized server, network issues can be localized faster and more accurately, and the root cause analysis can be also provided.

The conventional troubleshooting logic is to log in a faulty router, physically or through Telnet, and by using CLI to display related information/logs for fault source localization and further analysis. There are several concerns with the conventional troubleshooting methods:

1. It requires rich OAM experience for the OAM operator to know what information to check on the device, and the operation is complex;

2. In a multi-vendor network, it requires the understanding and familiarity of vendor specific operations and configurations;

3. Locating the fault source device could be non-trivial work, and is often realized through network-wide device-by-device check, which is both time-consuming and labor-consuming; and finally,

4. The acquisition of troubleshooting data can be difficult under some cases, e.g., when auto recovery is used.

This document proposes the Network Monitoring Protocols (NMP) to monitor the running status of control protocols, e.g., PDUs, protocol statistics and peer status, which have not been systematically covered by any other Telemetry approach, to facilitate network troubleshooting.

1.2. Overview

Like BMP, an NMP session is established between each monitored router (NMP client) and the NMP monitoring station (NMP server) through TCP connection. Information are collected directly from each monitored router and reported to the NMP server. The NMP message can be both periodic and event-triggered, depending on the message type.

IS-IS [RFC1195], as one of the most commonly adopted network layer protocols, builds the fundamental network connectivity of an autonomous system (AS). The disfunction of IS-IS, e.g., IS-IS neighbor down, route flapping, MTU mismatch, and so on, could lead to network-wide instability and service interruption. Thus, it is critical to keep track of the health condition of IS-IS, and the availability of information, related to IS-IS running status, is the fundamental requirement. In this document, typical network issues
are illustrated as the use cases of NMP for IS-IS to showcase the necessity of NMP. Then the operations and the message formats of NMP for IS-IS are defined. In this document IS-IS is used as the illustration protocol, and the case of OSPF and other control protocols will be included in the future version.

2. Terminology

IGP: Interior Gateway Protocol

IS-IS: Intermediate System to Intermediate System

NMP: Network Monitoring Protocol

IMP: Network Monitoring Protocol for IGP

BMP: BGP monitoring protocol

IIH: IS-IS Hello Packet

LSP: Link State Packet

CSNP: Complete Sequence Number Packet

NSNP: Partial Sequence Number Packet

3. Use Cases

We have identified several typical network issues due to IS-IS disfunction that are currently difficult to detect or localize. The usage of NMP is not limited to the solve the following listed issues.

3.1. IS-IS Adjacency Issues

IS-IS adjacency issues are identified as top network issues and may take hours to localize. The adjacency issues can be classified into two situations:

1. An existing established adjacency goes down;

2. An adjacency fails to be established.

In Case 1, the adjacency down can be caused by factors such as circuit down, hold timer expiration, device memory low, user configuration change, and so on. Case 2 can be caused by mismatch link MTU, mismatch authentication, mismatch area ID, system ID conflict, and so on. Typically, such adjacency failure events are logged/recorded in the device, but currently there is no real-time
report/alarm of such issue. The conventional troubleshooting process for adjacency issue is to find the faulty devices and then log in to check the logs or the IIH statistics for further analysis.

Using NMP, the IS-IS adjacency status: up, down and initial, is reported to the NMP server in real time, together with the possible recorded reasons. Then the NMP server can solve such issue in about minutes. For example, for an adjacency set up failure due to different authentications, the NMP server can recognize the difference by comparing the IIHs collected from both devices.

3.2. Forwarding Path Disconnection

The PING test can be used to test the reachability of a destination address. However, there are cases of disconnection that cannot be detected by PING. The PING result may return a connected path, but the forwarding of certain-sized packets always fails. This could be caused by factors, such as mismatched MTU values for devices along the path. It can be quite common since vendors have different understanding and configurations of MTU. There are methods proposed to discover the path MTU. For example, router’s link MTU is conveyed in the MPLS LDP/RSPV-TE path set up signaling, and the path MTU is decided at the ingress or egress node [RFC3988] [RFC3209]. For IPv4 packets, by setting the DF flag bit of the outgoing packet, any device along the path with smaller MTU will drop the packet, and send back an ICMP Fragmentation Needed message containing its MTU, allowing the source to reduce the MTU. The process is repeated until the MTU is small enough to traverse the entire path without fragmentation [RFC1191]. Apparently, such method is too time-consuming.

Using NMP, each device can report its link MTU to the monitoring station directly. The mismatch can be recognized at the NMP server in seconds.

3.3. IS-IS LSP Synchronization Failure

It happens that two IS-IS neighbors fail to learn the LSPs sent from each other in the following two cases: in Case 1, the LSP fails to be received, and in Case 2, the LSP is received but the LSP information shown in the receiver’s LSDB is not the same as the one sent from the transmitter (e.g., one or more prefixes missing, the LSP sequence number modified). Case 1 can be caused by link failure, similar to the adjacency down issue. In Case 2, the received LSP can be processed incorrectly due to hardware/software bugs. In fact, the LSDB synchronization issue is usually hard to localize once happens.
Using NMP, the NMP server can detect the failure by comparing the sent/received LSP statistics from the two neighbors. In the case that the received LSPs are improperly processed within the device, the NMP monitoring station can recognize the LSP synchronization failure by comparing the LSPs sent out from the two neighbors.

4. NMP Message Format

4.1. Protocol Selection Options

Regarding the NMP/IMP monitoring data exportation, BMP has been a good option. First of all, BMP serves similar purposes of NMP that reports routes, route statistics and peer status. In addition, BMP has already been implemented in major vendor devices and utilized by operator. Thus, we propose the following two options for the NMP data exportation.

- Option 1: Extending BMP with new message types to carry NMP/IMP data: Reusing the BMP framework saves certain implementation cost for both vendors and operators. Besides, the monitoring data exportation of different routing protocols (e.g., BGP, ISIS, OSPF) can be unified.

- Option 2: Defining NMP to carry NMP/IMP data: This option defines a brand new framework to carry protocol monitoring data, similar to BMP. Defining a new framework provides advantages such as more flexible and customized features for IGP and other protocols, since the monitoring data and troubleshooting of different protocols vary from one another.

In this document, we take Option 2 as the illustration example to define the NMP message types and message formats. The decision of the protocol selection may be further clarified in futures versions.

4.2. Message Types

The variety of IS-IS troubleshooting use cases requires a systematic information report of NMP, so that the NMP server or any third party analyzer could efficiently utilize the reported messages to localize and recover various network issues. We define NMP messages for IS-IS uses the following types:

- Initiation Message: A message used for the monitored device to inform the NMP monitoring station of its capabilities, vendor, software version and so on. For example, the link MTU can be included within the message. The initiation message is sent once the TCP connection between the monitoring station and monitored
router is set up. During the monitoring session, any change of the initiation message could trigger an Initiation Message update.

- Adjacency Status Change Notification Message: A message used to inform the monitoring station of the adjacency status change of the monitored device, i.e., from up to down, from down/initiation to up, with possible alarms/logs recorded in the device. This message notifies the NMP server of the ongoing IS-IS adjacency change event and possible reasons. If no reason is provided or the provided reason is not specific enough, the NMP server can further analyze the IS-IS PDU or the IS-IS statistics.

- Statistic Report Message: A message used to report the statistics of the ongoing IS-IS process at the monitored device. For example, abnormal LSP count of the monitored device can be a sign of route flapping. This message can be sent periodically or event triggered. If sent periodically, the frequency can be configured by the operator depending on the monitoring requirement. If it’s event triggered, it could be triggered by a counter/timer exceeding the threshold.

- IS-IS PDU Monitoring Message: A message used to update the NMP server of any PDU sent from and received at the monitored device. For example, the IIHs collected from two neighbors can be used for analyzing the adjacency set up failure issue. The LSPs collected from two neighbors can be analyzed for the LSP synchronization issue.

- Termination Message: A message for the monitored router to inform the monitoring station of why it is closing the NMP session. This message is sent when the monitoring session is to be closed.

4.3. Message Format

4.3.1. Common Header

The common header is encapsulated in all NMP messages. It includes the Version, Message Length and Message Type fields.

- Version (1 byte): Indicates the NMP version and is set to ‘1’ for all messages.

- Message Length (4 bytes): Length of the message in bytes (including headers, data, and encapsulated messages, if any).

- Message Type (1 byte): This indicates the type of the NMP message, which are listed as follows.
* Type = 0: Initiation
* Type = 1: Adjacency Status Change Notification
* Type = 2: Statistic Report
* Type = 3: IS-IS PDU Monitoring
* Type = 4: Termination Message

4.3.2. Per Adjacency Header

Except the Initiation and Termination Message, all the rest messages are per adjacency based. Thus, a per adjacency header is defined as follows.

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+---------------+
|    Version    |
+---------------+
|   Message Length   |
+---------------+
|   Msg. Type     |
+---------------+
```

- Adjacency Flag (2 bytes): The Circuit Type (2 bits) flag specifies if the router is an L1(01), L2(10), or L1/L2(11). If both bits are zeroes (00), the Per Adjacency Header SHALL be ignored. This configuration is used when the statistic is not per-adjacency based, e.g., when reporting the number of adjacencies.
4.3.3. Initiation Message

The Initiation Message indicates the monitored router’s capabilities, vendor, software version and so on. It consists of the Common Header and the Router Capability TLV. The Common Header can be followed by multiple Router Capability TLVs.

The Router Capability TLV is defined as follows.

```
+-------------------------------+-------------------------------+
|         Router Cap. Type      |       Router Cap. Length      |
+-------------------------------+-------------------------------+
| Router Cap. Value (variable) |
+---------------------------------------------------------------+
```

- **Router Capability Type**: provides the type of the router capability information. Currently defined types are:
  - *Type = 0*: sysDescr. The corresponding Router Capability Value field should contain an ASCII string whose value MUST be set to be equal to the value of the sysDescr MIB-II [RFC1213] object.
  - *Type = 1*: sysName. The corresponding Router Capability Value field should contain an ASCII string whose value MUST be set to be equal to the value of the sysName MIB-II [RFC1213] object.
  - *Type = 2*: Local System ID. The corresponding Router Capability Value field SHALL indicate the router’s System ID
  - *Type = 3*: Link MTU. The corresponding Router Capability Value field SHALL indicate the router’s link MTU.
  - *Type = 4*: String. The corresponding Router Capability Value field contains a free-form UTF-8 string whose length is given by the Information Length field.
4.3.4. Adjacency Status Change Notification

The Adjacency Status Change Notification Message indicates an IS-IS adjacency status change: from up to down or from initiation/down to up. It consists of the Common Header, Per Adjacency Header and the Reason TLV. The Notification is triggered whenever the status changes. The Reason TLV is optional, and is defined as follows. More Reason types can be defined if necessary.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>+-------------------------------+-------------------------------+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reserved</td>
<td>S</td>
<td>Reason Type</td>
</tr>
<tr>
<td>+-------------------------------+-------------------------------+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reason Value (variable)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| +---------------------------------------------------------------+

- **Reason Flags (1 byte):** The S flag (1 bit) indicates if the Adjacency status is from up to down (set to 0) or from down/initial to up (set to 1). The rest bits of the Flag field are reserved. When the S flag is set to 1, the Reason Type SHALL be set to all zeroes (i.e., Type 0), the Reason Length fields SHALL be set to all zeroes, and the Reason Value field SHALL be set empty.

- **Reason Type (1 byte):** indicates the possible reason that caused the adjacency status change. Currently defined types are:
  - Type = 0: Adjacency Up. This type indicates the establishment of an adjacency. For this reason type, the S flag MUST be set to 1, indicating it’s a adjacency-up event. There’s no further reason to be provided. The reason Length field SHALL be set to all zeroes, and the Reason Value field SHALL be set empty.
  - Type = 1: Circuit Down. For this data type, the S flag MUST be set to 0, indicating it’s a adjacency-down event. The length field is set to all zeroes, and the value field is set empty.
  - Type = 2: Memory Low. For this data type, the S flag MUST be set to 0, indicating it’s a adjacency-down event. The length field is set to all zeroes, and the value field is set empty.
  - Type = 3: Hold timer expired. For this data type, the S flag MUST be set to 0, indicating it’s a adjacency-down event. The length field is set to all zeroes, and the value field is set empty.
* Type = 4: String. For this data type, the S flag MUST be set to 0, indicating it’s a adjacency-down event. The corresponding Reason Value field indicates the reason specified by the monitored router in a free-form UTF-8 string whose length is given by the Reason Length field.

- Reason Length (2 bytes): indicates the length of the Reason Value field.
- Reason Value (variable): includes the possible reason why the Adjacency is down.

4.3.5. Statistic Report Message

The Statistic Report Message reports the statistics of the parameters that are of interest to the operator. The message consists of the NMP Common Header, the Per Adjacency Header and the Statistic TLV. The message include both per-adjacency based statistics and non per-adjacency based statistics. For example, the received/sent LSP counts are per-adjacency based statistics, and the local LSP change times count and the number of established adjacencies are non per-adjacency based statistics. For the non per-adjacency based statistics, the CT Flag (2 bits) in the Per Adjacency Header MUST be set to 00. Upon receiving any message with CT flag set to 00, the Per Adjacency Header SHALL be ignored (the total length of the Per Adjacency Header is 18 bytes as defined in Section 3.2.2, and the message reading/analysis SHALL resume from the Statistic TLV part.

The Statistic TLV is defined as follows.

```
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
+---------------------------------------------------------------+
|   Reserved  |T| Statistic Type|        Statistic Length       |
+---------------------------------------------------------------+
|                       Statistic  Value                        |
+---------------------------------------------------------------+
```

- Statistic Flags (1 byte): provides information for the reported statistics.
  * T flag (1 bit): indicates if the statistic is for the received-from direction (set to 1) or sent-to direction the neighbor (set to 0)

- Statistic Type (1 byte): specifies the statistic type of the counter. Currently defined types are:
* Type = 0: IIH count. The T flag indicates if it’s a sent or received Hello PDU. It is a per-adjacency based statistic type, and the CT flag in the Per Adjacency Header MUST NOT be set to 00.

* Type = 1: Incorrect IIH received count. For this type, the T flag MUST be set to 1. It is a per-adjacency based statistic type, and the CT flag in the Per Adjacency Header MUST NOT be set to 00.

* Type = 2: LSP count. The T flag indicates if it’s a sent or received LSP. It is a per-adjacency based statistic type, and the CT flag in the Per Adjacency Header MUST NOT be set to 00.

* Type = 3: Incorrect LSP received count. For this type, the T flag MUST be set to 1. It is a per-adjacency based statistic type, and the CT flag in the Per Adjacency Header MUST NOT be set to 00.

* Type = 4: Retransmitted LSP count. For this type, the T flag MUST be set to 0. It is a per-adjacency based statistic type, and the CT flag in the Per Adjacency Header MUST NOT be set to 00.

* Type = 5: CSNP count. The T flag indicates if it’s a sent or received CSNP. It is a per-adjacency based statistic type, and the CT flag in the Per Adjacency Header MUST NOT be set to 00.

* Type = 6: PSNP count. The T flag indicates if it’s a sent or received PSNP. It is a per-adjacency based statistic type, and the CT flag in the Per Adjacency Header MUST NOT be set to 00.

* Type = 7: Number of established adjacencies. It’s a non per-adjacency based statistic type, and thus for the monitoring station to recognize this type, the CT flag in the Per Adjacency Header MUST be set to 00.

* Type = 8: LSP change time count. It’s a non per-adjacency based statistic type, and thus for the monitoring station to recognize this type, the CT flag in the Per Adjacency Header MUST be set to 00.

- Statistic Length (2 bytes): indicates the length of the Statistic Value field.

- Statistic Value (4 bytes): specifies the counter value, which is a non-negative integer.
4.3.6. IS-IS PDU Monitoring Message

The IS-IS PDU Monitoring Message is used to update the monitoring station of any PDU sent from and received at the monitored device per neighbor. Following the Common Header and the Per Adjacency Header is the IS-IS PDU. To tell whether it’s a sent or received PDU, the monitoring station can analyze the source and destination addresses in the reported PDUs.

4.3.7. Termination Message

The Termination Message is sent when the NMP session is to be closed, and is used to indicate the termination reason to the monitoring station. The TCP session between the monitored router and the monitoring station SHALL be terminated upon receiving this message. It consists of the Common Header and the Termination Info TLVs, defined as follows.

```
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
+-------------------------------+-------------------------------+
|     Termination Info Type     |    Termination Info Length    |
+-------------------------------+-------------------------------+
+                 Termination Info Value (variable)             +
~                                                               ~
+---------------------------------------------------------------+
```

- **Termination Info Type (2 bytes):** Provides the termination reason type. Currently defined types are:

  * **Type = 0:** Unknown. This reason type specifies that the NMP session is closed for an unknown or unspecified reason. For this data type, the length field is filled with all zeroes, and the value field is set empty.

  * **Type = 1:** Memory Low. This reason indicates that the monitored router lacks resources for the NMP session. For this data type, the length field is filled with all zeroes, and the value field is set empty.

  * **Type = 2:** Administratively Closed. This reason specifies that the session is closed due to administrative reasons. The corresponding Termination Info Value field may include more details about the reason expressed in a free-form UTF-8 string whose length is given by the Termination Info Length field.

  * **Type = 3:** String. The corresponding Termination Info Value field may include details about the reason expressed in a free-
form UTF-8 string whose length is given by the Termination Info Length field.

Termination Info Length (2 bytes): indicates the length of the Termination Info Reason Value field.

- Termination Info Value (variable): includes more detailed reason for the session termination.

5. IANA
   TBD

6. Contributors
   TBD

7. Acknowledgments
   TBD

8. References

[I-D.brockners-inband-oam-requirements]
   Brockners, F., Bhandari, S., Dara, S., Pignataro, C., Gredler, H., Leddy, J., Youell, S., Mozes, D., Mizrahi, T., <> , P., and r. remy@barefootnetworks.com, "Requirements for In-situ OAM", draft-brockners-inband-oam-requirements-03 (work in progress), March 2017.

[I-D.ietf-netconf-yang-push]

[I-D.openconfig-rtgwg-gnmi-spec]

[I-D.song-ntf]


Authors’ Addresses

Yunan Gu
Huawei
156 Beiqing Road
Beijing, 100095
P.R. China

Email: guyunan@huawei.com

Shunwan Zhuang
Huawei
156 Beiqing Road
Beijing, 100095
P.R. China

Email: zhuangshunwan@huawei.com

Zhenbin Li
Huawei
156 Beiqing Road
Beijing, 100095
P.R. China

Email: lizhenbin@huawei.com