Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 3, 2019

Y. Gu
S. Zhuang
Z. Li
Huawei

July 02, 2018

Network Monitoring Protocol (NMP)
draft-gu-network-mornitoring-protol-00

Abstract

To enable automated network OAM (Operations, administration and management), the availability of network protocol running status information is a fundamental step. In this document, a network monitoring protocol (NMP) is proposed to provision the information related to running status of IGP (Interior Gateway Protocol) and other control protocols. It can facilitate the network troubleshooting of control protocols in a network domain. Typical network issues are illustrated as the usecases of NMP for ISIS to showcase the necessity of NMP. Then the operations and the message formats of NMP for ISIS are defined. In this document ISIS is used as the illustration protocol, and the case of OSPF and other control protocols will be included in the future version.

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.

Table of Contents

1. Introduction

1.1. Motivation

The requirement for better network OAM approaches has been greatly driven by the network evolvement. Network OAM provides visibility to the network health conditions, and is beneficial for faster network

troubleshooting and self-healing, network OpEx (operating expenditure) reduction, and network optimization. Network OAM statistics show that a relatively large part of the network issues are caused by the disfunction of various routing protocols and MPLS signalings.

The general troubleshooting logic nowadays 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:

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.

Alternatively, the idea of collecting information from devices and exporting to the centralized controller/server for further analysis is also used to gain more insight on the management plane information for OAM purposes. For example, SNMP (Simple Network Management Protocol) [RFC1157], NETCONF (Network Configuration Protocol) [RFC6241], gNMI/gRPC [I-D.openconfig-rtgwg-gnmi-spec], etc. are used for the purpose. However, the approaches are mainly used for data SET/GET of the management plane which are insufficient for the troubleshooting of control plane issues.

BGP monitoring protocol (BMP) [RFC7854] has been proposed to monitor BGP routes and peer status which provides the control plane information and thus more insight for troubleshooting. This document extends BMP to collect information of other control protocols for monitoring to facilitate the trouble shooting of control plane issues which call as Network Monitoring Protocols (NMP).

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.

ISIS [RFC1195], as one of the most commonly adopted network layer protocols, builds the fundamental network connectivity of an autonomous system (AS). The disfunction of ISIS, e.g., ISIS 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 ISIS, and the availability of information, related to ISIS running status, is the fundamental requirement. In this document, typical network issues are illustrated as the use cases of NMP for ISIS to showcase the necessity of NMP. Then the operations and the message formats of NMP for ISIS are defined. In this document ISIS 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

NMP: Network Monitoring Protocol

IMP: Network Monitoring Protocol for IGP

3. Use Cases

We have identified several typical network issues due to ISIS 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. ISIS Adjacency Issues

ISIS 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 Hello statistics for further analysis.

Using NMP, the ISIS 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 Hello PDUs collected from both devices.

3.2. Forwarding Path Disconnection

Mismatched MTU values for devices along a certain path can lead to packet forwarding failure while the control plane is working properly. The failure may not be detected by Ping, but the forwarding plane appears disconnected for certain size of data packets. It can be quite common since vendors have different understanding and configuration of MTU. There are methods proposed to discover the path MTU. For example, router's link MTU is conveyed in the MPLS LDP/RSVP-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. ISIS LSP Synchronization Failure

It happens that two ISIS 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. Extensions of NMP for ISIS

4.1. Message Types

The variety of ISIS 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 ISIS 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.

- Peer 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 ISIS 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 ISIS PDU or the ISIS statistics.

- Statistic Report Message: A message used to report the statistics of the ongoing ISIS 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.

- ISIS 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 Hello PDUs 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.

4.2. Message Format

4.2.1. 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.

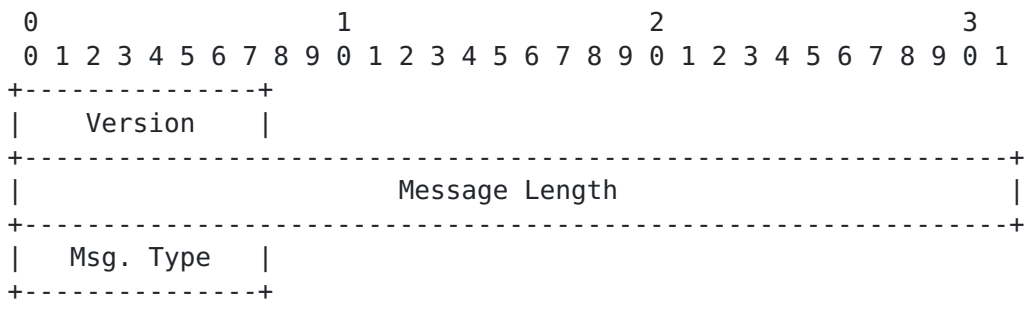
  * Type = 0: Initiation

  * Type = 1: Peer Status Change Notification

  * Type = 2: Statistic Report

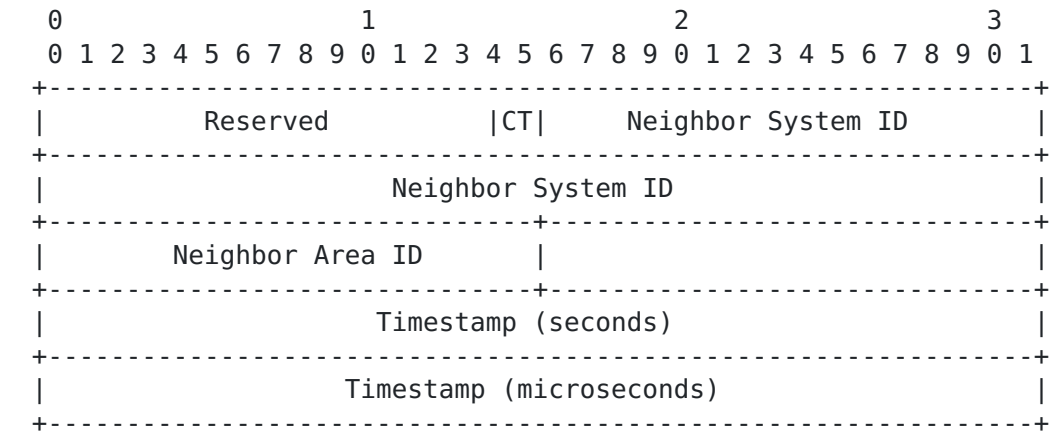
  * Type = 3: ISIS PDU Monitoring

  * Type = 4: Termination Message



4.2.2. Per Peer Header

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



- Peer 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 Peer Header is ignored. This configuration is used when the statistic is not per-peer based, e.g., when reporting the number of adjacencies.

- Neighbor System ID (6 bytes): identifies the system ID of the remote router.

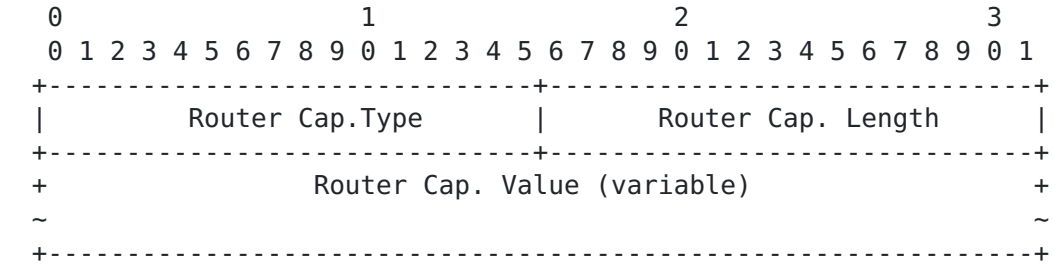
- Neighbor Area ID (2 bytes): identifies the area ID of the remote router.

- Timestamp (4 bytes): records the time when the message is sent/ received, expressed in seconds and microseconds since midnight (zero hour), January 1, 1970 (UTC).

4.2.3. 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.

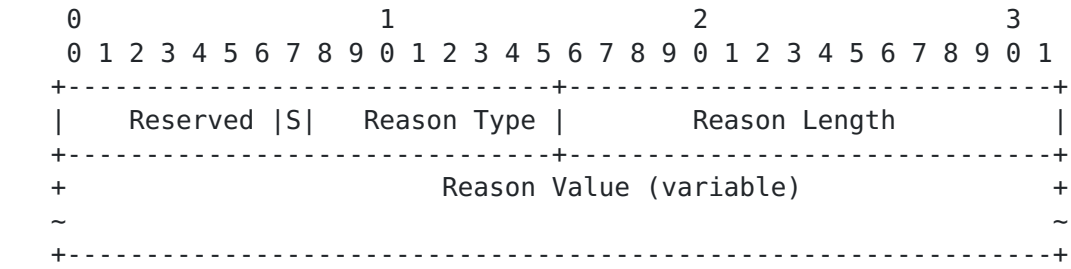
The Router Capability TLV is defined as follows.



- Router 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 should indicate the router's System ID
  - Type = 3: Link MTU. The corresponding Router Capability Value field should 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.

4.2.4. Peer Status Change Notification

The Peer Status Change Notification Message indicates an ISIS adjacency status change: from up to down or from initiation/down to up. It consists of the Common Header, Per Peer 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.



- Reason Flags (1 byte): The S flag (1 bit) indicates if the Peer 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 should be set to all zeroes (i.e., Type 0), the Reason Length fields should be set to all zeroes, and the Reason Value field should be set empty.
- Reason Type (1 byte): indicates the possible reason that caused the peer 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 peer-up event. There's no further reason to be provided. The reason Length field should be set to all zeroes, and the Reason Value field should be set empty.

  * Type = 1: Circuit Down. For this data type, the S flag MUST be set to 0, indicating it's a peer-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 peer-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 peer-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 peer-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.

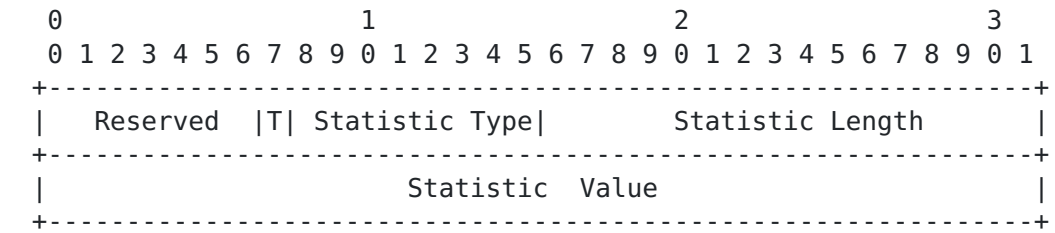
- Reason Length (2 bytes): indicates the length of the Reason Value field.

- Reason Value (variable): includes the possible reason why the Adjacency is down.

4.2.5. 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-peer based statistics and non per-peer based statistics. For example, the received/sent LSP counts are per-peer based statistics, and the local LSP change times count and the number of established adjacencies are non per-peer based statistics. For the non per-peer based statistics, the CT Flag (2 bits) in the Per Peer Header MUST be set to 00. Upon receiving any message with CT flag set to 00, the Per Peer Header should be ignored (the total length of the Per Peer Header is 18 bytes as defined in Section 3.2.2, and the message reading/analysis should resume from the Statistic TLV part.

The Statistic TLV is defined as follows.



- 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: Hello PDU count. The T flag indicates if it's a sent or received Hello PDU. It is a per-peer based statistic type, and the CT flag in the Per Peer Header MUST NOT be set to 00.
  - Type = 1: Incorrect Hello PDU received count. For this type, the T flag MUST be set to 1. It is a per-peer based statistic type, and the CT flag in the Per Peer 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-peer based statistic type, and the CT flag in the Per Peer 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-peer based statistic type, and the CT flag in the Per Peer 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-peer based statistic type, and the CT flag in the Per Peer 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-peer based statistic type, and the CT flag in the Per Peer 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-peer based statistic type, and the CT flag in the Per Peer Header MUST NOT be set to 00.

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

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

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

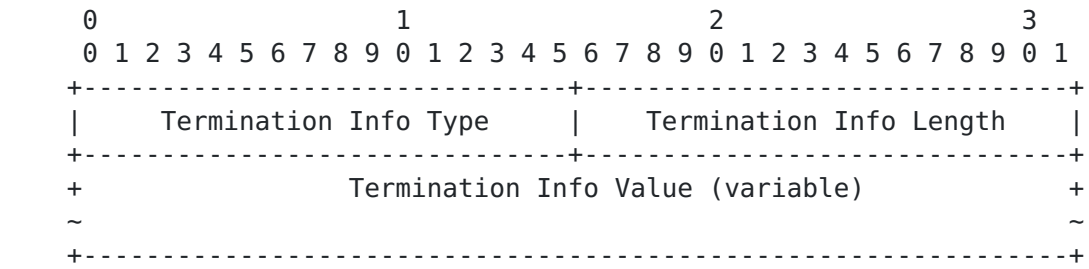
o Statistic Value (4 bytes): specifies the counter value, which is a non-negative integer.

4.2.6. ISIS PDU Monitoring Message

The ISIS 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 Peer Header is the ISIS 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.

4.2.7. 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 should be terminated upon receiving this message. It consists of the Common Header and the Termination Info TLVs, defined as follows.



o 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.

o 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.ietf-netconf-yang-push]
Clemm, A., Voit, E., Prieto, A., Tripathy, A., Nilsen-Nygaard, E., Bierman, A., and B. Lengyel, "YANG Datastore Subscription", draft-ietf-netconf-yang-push-17 (work in progress), July 2018.

[I-D.openconfig-rtgwg-gnmi-spec]
Shakir, R., Shaikh, A., Borman, P., Hines, M., Lebsack, C., and C. Morrow, "gRPC Network Management Interface (gNMI)", draft-openconfig-rtgwg-gnmi-spec-01 (work in progress), March 2018.

[RFC1157] Case, J., Fedor, M., Schoffstall, M., and J. Davin, "Simple Network Management Protocol (SNMP)", RFC 1157, DOI 10.17487/RFC1157, May 1990, <https://www.rfc-editor.org/info/rfc1157>.

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, DOI 10.17487/RFC1191, November 1990, <https://www.rfc-editor.org/info/rfc1191>.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, DOI 10.17487/RFC1195, December 1990, <https://www.rfc-editor.org/info/rfc1195>.

[RFC1213] McCloghrie, K. and M. Rose, "Management Information Base for Network Management of TCP/IP-based internets: MIB-II", STD 17, RFC 1213, DOI 10.17487/RFC1213, March 1991, <https://www.rfc-editor.org/info/rfc1213>.

[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>.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <https://www.rfc-editor.org/info/rfc3209>.

[RFC3988] Black, B. and K. Kompella, "Maximum Transmission Unit Signalling Extensions for the Label Distribution Protocol", RFC 3988, DOI 10.17487/RFC3988, January 2005, <https://www.rfc-editor.org/info/rfc3988>.

[RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, <https://www.rfc-editor.org/info/rfc6241>.

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <https://www.rfc-editor.org/info/rfc7752>.

[RFC7854] Scudder, J., Ed., Fernando, R., and S. Stuart, "BGP Monitoring Protocol (BMP)", RFC 7854, DOI 10.17487/RFC7854, June 2016, <https://www.rfc-editor.org/info/rfc7854>.

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