Workgroup: MPLS Working Group
Internet-Draft:
draft-andersson-mpls-mna-fwk-01
Published: 27 April 2022
Intended Status: Informational
Expires: 29 October 2022

Authors: L. Andersson S. Bryant
Bronze Dragon Consulting University of Surrey 5GIC
M. Bocci T. Li
Nokia Juniper Networks

MPLS Network Actions Framework

Abstract

This document specifies an architectural framework for the MPLS Network Actions (MNA) technologies. MNA technologies are used to indicate actions for Label Switched Paths (LSPs) and/or packets and to transfer data needed for these actions.

The document describes a common set of protocol actions and information elements supporting additional operational models and capabilities of MPLS networks. Some of these actions are defined in existing MPLS specifications, while others require extensions to existing specifications to meet the requirements found in "Requirements for MPLS Label Stack Indicators and Ancillary Data".

This document is the result of work started in MPLS Open Desgign Team, with participation by the MPLS, PALS and DETNET working groups.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction
  1.1. Requirement Language
  1.2. Terminology
    1.2.1. Normative Definitions
    1.2.2. Abbreviations
2. Structure
  2.1. Scopes
  2.2. Partial Processing
  2.3. Signaling
  2.4. Positioning
  2.5. State
3. Encoding
  3.1. The MNA Label
    3.1.1. Existing Base SPL
    3.1.2. New Base SPL
    3.1.3. New Extended SPL
    3.1.4. User-Defined Label
  3.2. TC and TTL
    3.2.1. TC and TTL retained
    3.2.2. TC and TTL Repurposed
  3.3. Length of the NAS
    3.3.1. Last/Continuation Bits
    3.3.2. Length Field
  3.4. Encoding of Scopes
  3.5. Encoding a Network Action
    3.5.1. Bit Catalogs
    3.5.2. Operation Codes
  3.6. Encoding of Post-Stack Data
    3.6.1. First Nibble Considerations
4. Definition of a Network Action
5. Management Considerations
6. Security Considerations

7. IANA Considerations
8. Acknowledgements
9. Editorial attic
9.1. Process Note on E2E
9.2. Concepts used in this Framework
9.3. LSE
9.4. MPLS Forwarding model
9.4.1. Orginal Model
10. References
10.1. Normative References
10.2. Informative References
Authors' Addresses

1. Introduction

This document specifies an architectural framework for the MPLS Network Actions (MNA) technologies. MNA technologies are used to indicate actions for LSPs and/or packets and to transfer data needed for these actions.

The document describes a common set of protocol actions and information elements supporting additional operational models and capabilities of MPLS networks. Some of these actions are defined in existing MPLS specifications, while others require extensions to existing specifications to meet the requirements found in [I-D.bocci-mpls-miad-adi-requirements]. [Ed.: In a future draft, the language in the requirements draft will be changed to align with the terminology found here.]

Forwarding actions are instructions to MPLS routers to apply additional actions when forwarding a packet. These might include load-balancing a packet given its entropy, whether or not to perform fast reroute on a failure, and whether or not a packet has metadata relevant to the forwarding decisions along the path.

This document generalizes the concept of "forwarding actions" into "network actions" to include any action that an MPLS router is requested to take on the packet. That includes any forwarding action, but may include other operations (such as security functions, OAM procedures, etc.) that are not directly related to forwarding of the packet.

This document is the result of work started in MPLS Open Desgign Team, with participation by the MPLS, PALS and DETNET working groups.

1.1. Requirement Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and

"OPTIONAL" in this document are to be interpreted as described in BCP14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

1.2. Terminology

1.2.1. Normative Definitions

*Ancillary Data (AD): Data relating to the MPLS packet that may be used to affect the forwarding or other processing of that packet, either at an Label Edge Router (LER) [RFC4221] or Label Switching Router (LSR). This data may be encoded within a network action sub-stack (see below) (in-stack data), and/or after the bottom of the label stack (post-stack data).

*Network Action: An operation to be performed on a packet. A network action may affect router state, packet forwarding, or it may affect the packet in some other way. A network action is said to be present if there is an indicator in the packet that invokes the action.

*Network Action Indication (NAI): An indication in the packet that a certain network action is to be perfomed. There may be associated ancillary data in the packet.

*Network Action Sub-Stack (NAS): A set of related, contiguous Label Stack Entries (LSEs). The first LSE is the Network Action Sub-stack Indicator. The TC and TTL values in the sub-stack may be redefined. The label field in the second and following LSE may be redefined. Solutions MUST NOT redefine the S bit. See Section 3.1 through Section 3.5.

*Network Action Sub-Stack Indicator (NSI): An LSE that contains a special label that indicates the start of a Network Action Sub-stack.

*Scope: The set of nodes that should perform a given action.

1.2.2. Abbreviations

Abbreviation	Meaning	Reference
AD	Ancillary Data	[I-D.bocci-mpls-miad-adi-requirements]
bSPL	Base Special Purpose Label	[RFC9017]
ECMP	Equal Cost Multipath	
eSPL		[RFC9017]

Abbreviation	Meaning	Reference
	Extended Special Purpose Label	
HBH	Hop by hop	In the MNA context, this document.
I2E	Ingress to Egress	In the MNA context, this document.
ISD	In stack data	[I-D.bocci-mpls-miad-adi-requirements]
LSE	Label Stack Entry	[RFC3032]
MNA	MPLS Network Actions	This documnent
NAI	Network Action Indicator	[I-D.bocci-mpls-miad-adi-requirements]
NAS	Network Action Sub-Stack	This document
PSD	Post stack data	[I-D.bocci-mpls-miad-adi-requirements] and Section 3.6
SPL	Special Purpose Label	[RFC9017]

Table 1: Abbreviations

2. Structure

An MNA solution is envisioned as a set of network action sub-stacks that indicate the network actions being invoked, plus possible post-stack data. A solution must specify where in the label stack the network actions sub-stacks occur, if and how frequently they should be replicated, and how network action sub-stack and post-stack data are encoded.

A network action sub-stack contains:

*Label: A special label is used to indicate the start of a network action sub-stack.

*Indicators: A set of indicators that describes the set of network actions.

*In-Stack Data: A set of zero or more LSEs that carry ancillary data for the present network actions.

Each network action present in the network action sub-stack may have zero or more LSEs of in-stack data. The ordering of the in-stack data LSEs corresponds to the ordering of the network action indicators. The encoding of the in-stack data, if any, for a network action must be specified in the document that defines the network action.

Certain network actions may also specify that data is carried after the label stack. This is called post-stack data. The encoding of the post-stack data, if any, for a network action must be specified in the document that defines the network action. If multiple network actions are present and have post-stack data, the ordering of their post-stack data corresponds to the ordering of the network action indicators.

A solution must specify the order that network actions are to be applied to the packet.

2.1. Scopes

A network action may need to be processed by every node along the path, or some subset of the nodes along its path. Some of the scopes that an action may have are:

*Hop-by-hop (HBH): Every node along the path will perform the action.

*Ingress-to-Egress (I2E): Only the last node on the path will perform the action.

*Select: Only specific nodes along the path will perform the action.

If a solution supports the select scope, it must describe how it specifies the set of nodes to perform the actions.

2.2. Partial Processing

Legacy devices that do not recognize the MNA label will discard the packet as described in [RFC3031].

Devices that do recognize the MNA label may not implement all of the present network actions. A solution must specify how unrecognized present network actions should be handled.

One alternative is that an implementation should stop processing network actions when it encounters an unrecognized network action. Subsequent present network actions would not be applied. The result is dependent on the solution's order of operations.

Another alternative is that an implementation should drop any packet that contains any unrecognized present network actions.

A third alternative is that an implementation should perform all recognized present network actions, but ignore all unrecognized present network actions.

Other alternatives may also be possible and should be specified by the solution.

2.3. Signaling

A node that wishes to make use of MNA and apply network actions to a packet must understand the nodes that the packet will transit and whether or not the nodes support MNA and the network actions that are to be invoked. These capabilities are presumed to be signaled by protocols that are out-of-scope for this document and are presumed to have per-network action granularity. If a solution requires alternate signaling, it must specify so explicitly.

A node that pushes a NAS onto the label stack is responsible for determining that all nodes that should process the NAS will have the NAS within its Readable Label Depth (RLD). A node should use signaling (e.g., [RFC9088]) to determine this.

2.4. Positioning

A network action sub-stack should never occur at the top of the MPLS label stack. A node that is responsible for popping a forwarding label immediately above a network action sub-stack must also pop any network action sub-stacks that immediately follow.

2.5. State

A network action can affect state in the network. This implies that a packet may affect how subsequent packets are handled.

3. Encoding

Several possibilities to carry NAI's have been discussed in MNA drafts and in the MPLS Open DT. In this section, we enumerate the possibilities and some considerations for the various alternatives.

All types of network actions are represented in the MPLS label stack by a set of LSEs termed a network action sub-stack (NAS). An NAS consists of a special label, followed by LSEs that specify which network actions are to be performed on the packet, and the in-stack ancillary data for each indicated network action.

[I-D.bocci-mpls-miad-adi-requirements] requires that a solution not add unnecessary LSEs to the sub-stack (Section 3.1, requirement 5). Accordingly, solutions should also make efficient use of the bits within the sub-stack, as inefficient use of the bits will result in the addition of unnecessary LSEs.

3.1. The MNA Label

The first LSE in a network action sub-stack contains a special label that indicates a network action sub-stack. A solution has several choices for this special label.

3.1.1. Existing Base SPL

A solution may reuse an existing Base SPL (bSPL). If it elects to do so, it must explain how the usage is backwards compatible, including in the case where there is ISD.

3.1.2. New Base SPL

A solution may select a new bSPL.

3.1.3. New Extended SPL

A solution may select a new eSPL. If it elects to do so, it must address the requirement for the minimal number of LSEs.

3.1.4. User-Defined Label

A solution may allow the network operator to define the label that indicates the network action sub-stack. This creates management overhead for the network operator to coordinate the use of this label across all nodes on the path using management or signaling protocols. If a solution elects to use a user-defined label, the solution should justify this overhead.

3.2. TC and TTL

In the first LSE of the network action sub-stack, only the 20 bits of Label Value and the Bottom of Stack bit are significant, the TC field (3 bits) and the TTL (8 bits) are not used. This leaves 11 bits that could be used for other purposes.

3.2.1. TC and TTL retained

If the solution elects to retain the TC and TTL field, then the first LSE of the network action sub-stack would appear as:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Label                   | TC  |S|      TTL      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
              Label:  Label value, 20 bits
              TC:     Traffic Class, 3 bits
              S:      Bottom of Stack, 1 bit
              TTL:    Time To Live
```

Further LSEs would be needed to encode NAIs. If a solution elects to retain these fields, it must address the requirement for the minimal number of LSEs.

3.2.2. TC and TTL Repurposed

If the solution elects to reuse the TC and TTL field, then the first LSE of the network action sub-stack would appear as:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Label                  |x x x|S|x x x x x x x x|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
              Label:  Label value, 20 bits
              x:      Bit available for solution definition
              S:      Bottom of Stack, 1 bit
```

The solution may use more LSEs to contain NAIs.

3.3. Length of the NAS

A solution must have a mechanism to indicate the length of the NAS. This must be easily processed even by implementations that do not understand the full contents of the NAS. Two options are described below, other solutions may be possible.

3.3.1. Last/Continuation Bits

A solution may use a bit per LSE to indicate whether the NAS continues into the next LSE or not. The bit may indicate continuation by being set or by being clear. The overhead of this approach is one bit per LSE and has the advantage that it can effectively encode an arbitrarily sized NAS. This approach is efficient if the NAS is small.

3.3.2. Length Field

A solution may opt to have a fixed size length field at a fixed location within the NAS. The fixed size of the length field may not be large enough to support all possible NAS contents. This approach may be more efficient if the NAS is longer, but not longer than can be described by the length field.

Advice from hardware designers advocates a length field as this minimizes branching in the logic.

3.4. Encoding of Scopes

A solution may choose to explicitly encode the scope of the actions contained in a network action sub-stack. A solution may also choose to have the scope encoded implicitly, based on the actions present in the network action sub-stack. This choice may have performance implications as an implementation might have to parse the network actions that are present in a network action sub-stack only to discover that there are no actions for it to perform.

Solutions need to consider the order of scoped NAIs and their associated AD within individual sub-stacks and the order of per-scope sub-stacks in order that network actions and the AD can be most readily found and not need to processed by nodes that are not required to handle those actions.

3.5. Encoding a Network Action

Two options for encoding NAIs are described below, other solutions may be possible. Any solution should allow encoding of an arbitrary number of NAIs.

3.5.1. Bit Catalogs

A solution may opt to encode the set of network actions as a list of bits, sometimes known as a catalog. The solution must provide a mechanism to determine how many LSEs are devoted to the catalog. A set bit in the catalog would indicate that the corresponding network action is present.

Catalogs are efficient if the number of present network actions is relatively high and if the size of the necessary catalog is small. For example, if the first 16 actions are all present, a catalog can encode this in 16 bits. However, if the number of possible actions is large, then a catalog can become inefficient. Selecting only one action that is the 256th action would require a catalog of 256 bits, which would require more than one LSE.

3.5.2. Operation Codes

A solution may opt to encode the set of present network actions as a list of operation codes (opcodes). Each opcode is a fixed number of bits. The size of the opcode bounds the number of network actions that the solution can support.

Opcodes are efficient if there are only one or two active network actions. For example, if an opcode is 8 bits, then two active network actions could be encoded in in 16 bits. However, if there are 16 actions required, then opcodes would consume 128 bits. Opcodes are efficient at encoding a large number of possible actions. If only the 256th action is to be selected, that still requires 8 bits.

3.6. Encoding of Post-Stack Data

If there are multiple instances of post-stack data, they should occur in the same order as their relevant network action sub-stacks and then in the same order as their relevant network functions occur within the network action sub-stacks.

3.6.1. First Nibble Considerations

The first nibble after the label stack has been used to convey information in certain cases.

For example, in [RFC4928] this nibble is investigated to find out if it has the value "4" or "6", if it is not, it is assumed that the packet payload is not IPv4 or IPv6 and Equal Cost Multipath (ECMP) is not performed.

It should be noted that this is an inexact method, for example an Ethernet Pseudowire without a control word might have "4" or "6" in the first nibble and thus will be ECMP'ed.

Nevertheless, the method is implemented and deployed, it is used today and will be for the foreseeable future.

The use of the first nibble for BIER is specified in [RFC8296]. Bier sets the first nibble to 5. The same is true for BIER payload, as for any use of the first nibble, it is not possible from the first nibble itself being set to 5, conclude that the payload is BIER. However, it achieves the design goal of [RFC8296], to exclude that the payload is IPv4, IPv6 or a pseudowire.

There are possibly more examples, they will be added if we find that they further highlight the issue with using the first nibble.

[Ed. Outstanding comments from Adrian:

Shouldn't we include RFC4385 for 0b0000 for the PW control word and 0b0001 for the PW ACH?

This section is all very well, but it doesn't give any direction to the solution developer for what they should do with the first nibble in the post stack data.

Is it also relevant to note that there may be other post-stack information that comes before the payload (such as the PW control word, and that the solution must consider the location of the post-stack data in relaiton to that (e.g., immeidately after the LSE with the S bit set) etc.]

4. Definition of a Network Action

Network actions should be defined in a document and must contain:

- *Name: The name of the network action.

- *Network Action Indicator: The bit position or opcode that indicates that the network action is active.

- *Scope: The document should specify which nodes should perform the network action. The action may apply to each transit node (HBH), only the egress node that pops the final label off of the label stack, or specific nodes along the label switched path.

- *State: The document should specify if the network action can modify state in the network, and if so, the state that may be modified and its side-effects.

- *Required/Optional: The document should specify whether a node is required to perform the network action.

- *In-Stack Data: The number of LSEs of in-stack data. If this is of a variable length, then the solution must specify how an implementation can determine this length without implementing the network action.

- *Post-Stack Data: The encoding of post-stack data, if any. If this is of a variable length, then the solution must specify how an implementation can determine this length without implementing the network action.

A solution should create an IANA registry for network actions.

5. Management Considerations

Network operators will need to be cognizant of which network actions are supported by which nodes and will need to ensure that this is signalled appropriately. Some solutions may require network-wide configuration to synchronize the use of the labels that indicate the start of an NAS. Solution documents must make clear what management considerations apply to the solutions they are describing. Solutions documents must describe mechanisms for performing network diagnostics in the presence of MNAs.

6. Security Considerations

The forwarding plane is insecure. If an adversary can affect the forwarding plane, then they can inject data, remove data, corrupt data, or modify data. MNA additionally allows an adversary to make packets perform arbitrary network actions.

Link-level security mechanisms can help mitigate some on-link attacks, but does nothing to preclude hostile nodes.

End-to-end encryption of an LSP can help provide security, but would make it impossible to process post-stack data.

7. IANA Considerations

This document does not make any allocations of code points from IANA registries.

As long as the "does not make any allocations ..." from IANA is true, this pragraph shoukd be removed by the RFC-Editor. If it turns out that we will need to do IANA allocation, a proper IANA section will be added.

8. Acknowledgements

The authors would like to thank Adrian Farrel for his contributions and to John Drake for his comments.

9. Editorial attic

This section contains old material that will be discarded before publication, assuming we don't find it useful between now and then.

9.1. Process Note on E2E

There has been some discussion on the of the E2E abbreviation. 1. In a mail to the MPLS Working group mailing list Joel Halpern pointed out that the abbreviation E2E has been used in several different meanings. Joel suggested to use another abbreviation.

1. Some variants has been proposed, for example.

   *Ingress to Egress (I2E); alernative abbreviation (i2e)

   *Egress

*LSP Ingress to LSP Egress (LI2LE)

*Egress (because the Ingress has already done its thing)

*Ultimate Hop

*Destination

*Start-to-End

*Last-LSR

*Head to Tail

In a few days (counting from the publication date of this document) the working group chairs will take an initiative to poll the working groups for consensus on this.

9.2. Concepts used in this Framework

Concept	Meaning	Reference	Note
E2E concept	E2E in MNA context is defined in...	this document	-
concept	free text	this document	-

Table 2: Concepts

Not complete, help appreciated. [Ed. This section is planned for removal as it seems unhelpful so far.]

9.3. LSE

An individual LSE has the following format [RFC3032]:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Label                  |  TC |S|       TTL     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

            Label:  Label Value, 20 bits
            TC:     Traffic Class, 3 bits
            S:      Bottom of Stack, 1 bit
            TTL:    Time to Live, 8 bits
```

Figure 1: A Label Stack Entry (LSE)

9.4. MPLS Forwarding model

This is section here to basically to have a place holder where to discuss the development of the MPLS forwrding model. It might be removed. [Ed. So far, it adds no value. Wave bye-bye.]

9.4.1. Orginal Model

```
+-----------------------------------------------------------------+
|                                                                 |
|  +---------------------+                                        |
|  | +------------+      |                                        |
|  | | MPLS Label |  LSE |                                        |
|  | +---|--------+      |                                        |
|  +-----|---------------+                                        |
|        |                                                        |
|        |  +----------------------+                              |
|        |  |                 FIB |                              |
|        |  |                      |                              |
|        |  |     +------------+    |     +----------------------+ |
|        +-------->|FIB Entry   |-----+-->|Forwarding Code       | |
|           |     +------------+   | |   +----------------------+ |
|           +----------------------| |                            |
|                                  | |   +----------------------+ |
|                                    +-->|Forwarding Parameters | |
|                                        +----------------------+ |
|                                                                 |
|                                                                 |
| LSE = Label Stack Entry (what many people call a label)         |
| FIB = Forwarding Information (date)Base                         |
+-----------------------------------------------------------------+
```

Figure 2: MPLS Original Forwarding Model

10. References

10.1. Normative References

[I-D.bocci-mpls-miad-adi-requirements] Bocci, M. and S. Bryant, "Requirements for MPLS Network Action Indicators and MPLS Ancillary Data", Work in Progress, Internet-Draft, draft-bocci-mpls-miad-adi-requirements-04, 11 April 2022, <https://www.ietf.org/archive/id/draft-bocci-mpls-miad-adi-requirements-04.txt>.

[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>.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <https://www.rfc-editor.org/info/rfc3031>.

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, DOI 10.17487/RFC3032, January 2001, <https://www.rfc-editor.org/info/rfc3032>.

[RFC4221] Nadeau, T., Srinivasan, C., and A. Farrel, "Multiprotocol Label Switching (MPLS) Management Overview", RFC 4221, DOI 10.17487/RFC4221, November 2005, <https://www.rfc-editor.org/info/rfc4221>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC9017] Andersson, L., Kompella, K., and A. Farrel, "Special-Purpose Label Terminology", RFC 9017, DOI 10.17487/RFC9017, April 2021, <https://www.rfc-editor.org/info/rfc9017>.

[RFC9088] Xu, X., Kini, S., Psenak, P., Filsfils, C., Litkowski, S., and M. Bocci, "Signaling Entropy Label Capability and Entropy Readable Label Depth Using IS-IS", RFC 9088, DOI 10.17487/RFC9088, August 2021, <https://www.rfc-editor.org/info/rfc9088>.

10.2. Informative References

[RFC4928] Swallow, G., Bryant, S., and L. Andersson, "Avoiding Equal Cost Multipath Treatment in MPLS Networks", BCP 128, RFC 4928, DOI 10.17487/RFC4928, June 2007, <https://www.rfc-editor.org/info/rfc4928>.

[RFC8296] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks", RFC 8296, DOI 10.17487/RFC8296, January 2018, <https://www.rfc-editor.org/info/rfc8296>.

Authors' Addresses

Loa Andersson
Bronze Dragon Consulting

Email: loa@pi.nu

Stewart Bryant
University of Surrey 5GIC

Email: sb@stewartbryant.com

Matthew Bocci
Nokia

Email: matthew.bocci@nokia.com

Tony Li
Juniper Networks

Email: tony.li@tony.li