SPRING WG
Internet-Draft
Intended status: Standards Track
Expires: May 3, 2018

S. Hegde
Juniper Networks, Inc.

October 30, 2017

Traffic Accounting for MPLS Segment Routing Paths
draft-hegde-spring-traffic-accounting-for-sr-paths-01

Abstract

Traffic statistics form an important part of operations and maintenance data that are used to create demand matrices and for capacity planning in networks. Segment Routing (SR) is a source routing paradigm that uses stack of labels to represent a path. The SR path specific state is not stored in any other node in the network except the head-end node of the SR path. Traffic statistics specific to each SR path are an important component of the data which helps the controllers to lay out the SR paths in a way that optimizes the use of network resources. SR paths are inherently ECMP aware.

As SR paths do not have state in the core of the network, it is not possible to collect the SR path traffic statistics accurately on each interface. This document describes an MPLS forwarding plane mechanism to identify the SR path to which a packet belongs and so facilitate accounting of traffic for MPLS SR paths.

The mechanisms described in this document may also be applied to other MPLS paths (i.e., Label Switched Paths) and can be used to track traffic statistics in multipoint-to-point environments such as those where LDP is in use.

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 May 3, 2018.

Table of Contents

1. Introduction

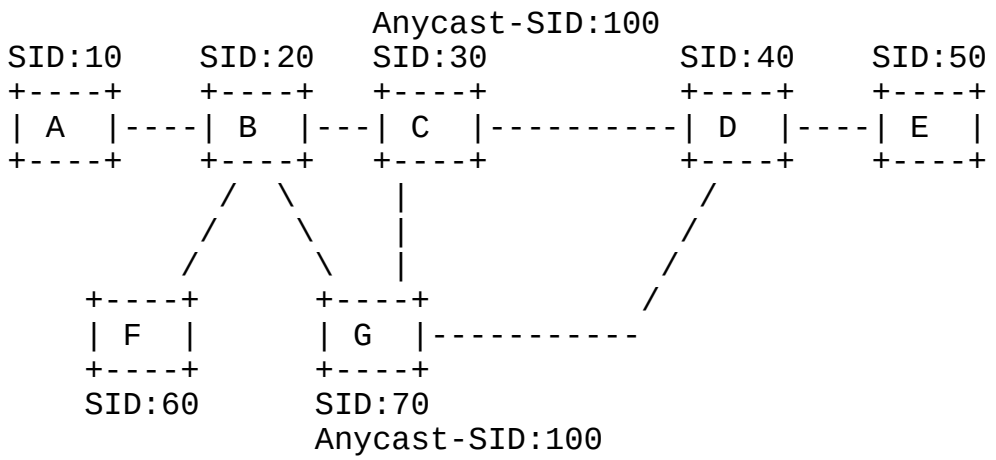
Figure 1 describes an SR enabled network with Node-SIDs and Anycast-SIDs assigned. The SR-Paths with label stacks are as shown in the diagram. The SR-Paths are created (possibly by a central controller) so as to maximize the network resource utilization such as bandwidth. Based on the traffic carried by the SR-Paths, they need to be re-routed occasionally to balance the bandwidth utilization. SR-Paths are inherently ECMP aware.

For example, SR-Path3 in the diagram is balanced across equal cost paths B->C->D and B->G->D. When there is congestion on the link between B and C, the SR path causing the congestion needs to be identified and re-routed. SR paths do not have separate control or forwarding state in any node other than the head-end. Traffic measurement at the head-end node is insufficient to determine the contribution of each SR path to the congestion on the link because of ECMP or Weighted ECMP balancing.

Per-SID traffic measurement on every interface gives some informtion about the traffic carried, but is not sufficient to correctly measure traffic carried by each SR path on the link. If it were possible to identify to which SR path each packet belonged, that information could be used by an external entity to re-route the SR paths to maximize resource utilization.

As SR paths do not have state in the core of the network, it is not possible to collect the SR path traffic statistics accurately on each interface. This document describes an MPLS forwarding plane mechanism to identify the SR path to which a packet belongs and so facilitate accounting of traffic for MPLS SR paths.

The mechanisms described in this document may also be applied to other MPLS paths (i.e., Label Switched Paths) and can be used to track traffic statistics in multipoint-to-point environments such as those where LDP is in use.



SRGB: 1000-2000 on all routers
SR-Path1: A-> 1020,1030
SR-Path2: A-> 1020,1100,1040
SR-Path3: F-> 1020,1040
SR-Path4: A-> 1020,1040,1060

Figure 1: Sample Network

2. Motivation

The motivation of this document is to provide a solution to enable traffic measurement statistics per SR-Path on any node and any link in the network. The objectives listed below help to achieve the requirements in a variety of deployments.

1. The control plane MUST be free of any per SR path state.

2. The forwarding plane MUST be free of any per SR path state.

3. The number of counters created to measure traffic SHOULD be optimized.

4. The additional information carried in each packet SHOULD be minimized.

5. The mechanism SHOULD be applicable to all MPLS environments.

3. Terminology

Source-SID: The (globally unique) Node-SID of the head-end node which places traffic on the SR path. This is a 20 bit number excluding 0-15 and may be encoded in an MPLS label field.

SR-Path-Identifier: An SR-Path-Identifier is an identifier for each SR path in the network. It is unique within the scope of the node that allocated the identifier. If the identifier is allocated by the head-end node (the source) the combination of Source-SID and SR-Path Identifier uniquely identifies an SR path within a network. If the identifier is allocated by a central controller then the SR-Path Identifier is network unique. The SR-Path Identifier is a 19 bit number excluding the values 0-15 and may be encoded in an MPLS label field. See Section 4.

SR-Path-Indicator: The SR-Path-Indicator is an MPLS Special Purpose Label [RFC7274]. This label indicates the presence of an SR-Path Identifier and an Source Node-SID encoded in MPLS label stack entries and situated immediately below this label stack entry in the label stack.

SR-Path-Stats Labels: The SR-Path-Indicator, SR-Path-Identifier, and Source-SID together are termed as the SR-Path-Stats Labels.

4. SR-Path Identifier

4.1. Centrally Managed SR Paths

In controller-based deployments, a controller creates an SR policy, associates a segment list and a Binding SID to the policy, and sends it to the head-end of the SR path as described in [I-D.filsfils-spring-segment-routing-policy]. The controller may also allocate a network-unique SR-Path-Identifier and send it to the head-end along with the policy. When the head-end node receives this policy, if it has not been supplied with an SR-Path-Identifier, it creates a locally-unique identifier for each the SR path network and associates it with SR-TE Policy and advertizes it back to the controller using mechanisms described in [I-D.ietf-idr-te-lsp-distribution].

The SR-Path-Identifier is used for the purpose of traffic accounting as described in Section 5.

4.2. Locally Managed SR Paths

Deployments which do not use a central controller for managing the network configure locally manage SR-Paths on the head-end router. Every SR path in the network is identified using a Source-SID and a source-unique SR-Path-Identifier. The head-end node generates the SR-Path-Identifier for each SR path and associates it with the SR path. An Operator MAY also configure 19-bit globally unique Identifiers on each SR-Path and use it for accounting traffic as described in Section 5

5. Use of the SR-Path-Identifier and Source-SID

The SR-Path-Identifier is a 19 bit number created by the head-end node as described in Section 4. The SR-Path-Identifier and Source-SID are inserted in the packet below a Special Purpose Label called the SR-Path-Indicator. The three values are each carried in a label stack entry as shown in Figure 2.

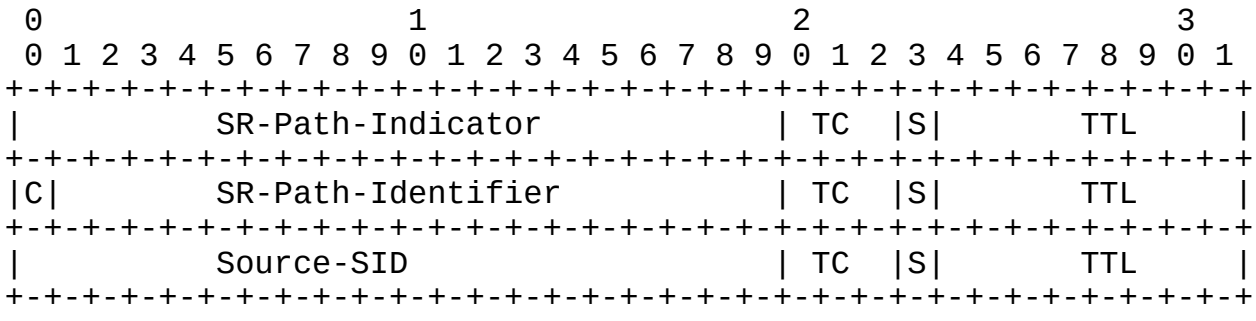


Figure 2: The SR-Path-Stats Labels Encoded in Label Stack Entries

The SR-Path-Indicator label value is TBD-1 to be assigned by IANA.

The SR-Path-Indicator label indicates that the MPLS label stack entries that follow carry an identifier of SR path. These label stack entries MUST NOT be used for forwarding, and if they are encountered at the top of the label stack (for example, at the egress node) they MUST be stripped.

The SR-Path-Identifier label stack entry is inserted immediately below the SR-Path-Indicator. The label field contains two elements:

- The C-flag indicates whether the SR-Path-Identifier is allocated by a central controller or not. If the C-flag is set (one) then this indicates that the SR-Path-Identifier was allocated by a central controller and has global scope, and that a Source-SID is not included. If the C-flag is clear (zero) then the SR-Path-Identifier is scoped by the Source-SID that is included after the SR-Path-Identifier.

- The SR-Path-Identifier identifies the SR path as described in Section 4.

The Source-SID is inserted immediately below the SR-Path-Identifier and is present only if indicated by the setting of the C-flag in the SR-Path-Identifier label stack entry. If present the Source-SID

gives scope to the SR-Path-Identifier. The Source-SID is described in Section 4.

An intermediate node in the network can look into the packet and account the traffic based on the SR-Path-Identifier and Source-SID.

Because it is necessary that the SR-Path-Stats labels are removed when they are found at the top of the label stack, the node imposing the label stack (the ingress) must know which nodes are capable of stripping the labels. This ability is advertised in IGP advertisements defined in TBD and TBD.

6. Inserting the SR-Path-Identifier in Packets

The SR-Path-Identifier and Source-SID are used as a key to account the SR path traffic. The forwarding plane entities should look up the SR-Path-Identifier and Source-SID (if present) values to account the traffic against the right path counters.

The SR-Path-Stats Labels are normally placed at the bottom of the label stack.

Forwarding hardware may have limitations and not support accessing the label stack beyond certain depth. In such cases, the hardware will not be able to find the SR-Path-Stats Labels at the bottom of the label stack if the stack is too deep. To support traffic accounting in such cases it is necessary to insert the SR-Path-Stats Labels within the Readable Label Stack Depth Capability (RLDC) of the nodes in the SR path. The extensions defined in [I-D.ietf-ospf-segment-routing-msd] and [I-D.ietf-isis-segment-routing-msd] describe how the MSD supported by each node is advertised. The head-end node SHOULD insert the SR-Path-Stats Labels at a depth in the label stack such that the nodes in the SR path can access the SR-Path-Identifier for accounting. The SR-Path-Stats Labels may be present multiple times in the label stack of a packet.

In general, if all the nodes in the network support RLDC which is more than the label-stack depth being pushed at the head-end node then the SR-Path-Stats Labels SHOULD be pushed at the bottom of the label-stack. If there are service labels to be inserted, they MUST be pushed at the bottom of the stack. If entropy labels [RFC6790] are to be inserted they SHOULD be pushed next. The SR-Path-Stats Labels SHOULD be pushed next.

It is possible to partially deploy this feature when not all the nodes in the network support the extensions defined in this document. In such scenarios, the special labels MUST NOT get exposed on the top

of the label stack at a node that does not support the extensions defined in this document. This may require multiple blocks of SR-Path-Stats Labels to be inserted in the packet header.

If the egress has not indicated that it is capable of removing the SR-Path-Stats Labels, then they MUST NOT be placed at the bottom of the label stack. In this case the SR-Path-Stats Labels SHOULD be placed at a point in the label stack such that they will be found at the top of stack by the latest node in the SR path that is capable of removing them. In this way, traffic accounting can be performed along as much of the SR path as possible.

7. Traffic-Accounting for Sub SR-Paths in the Network

SR paths may require large label stacks. Some hardware platforms do not support creating such large label stacks (i.e., imposing a large number of labels at once). To overcome this limitation sub-paths are created within the network, and Binding-SIDs are allocated to these sub-paths. When the label representing a Binding-SID is processed it is swapped for a stack of labels. When a head-end node builds the label stack for an SR path, it may use these Binding-SIDs to reduce the depth of the label stack it has to impose and effectively constructs the end-to-end SR path from a series of sub-paths

The sub-paths are not accounted separately. Accounting is performed on the end-to-end SR paths. However, edge routers MAY create Binding-SIDs for BGP-SR-TE Policies as described in [I-D.ietf-idr-segment-routing-te-policy]. Traffic accounting for the traffic carried on the SR paths indicated by these Binding-SIDs can be done separately by allocating separate SR-Path-Identifiers for these sub-paths.

8. Forwarding Plane Procedures

To support per-path traffic accounting, the forwarding plane in a router MUST look through the label stack of a packet for the first instance of the SR-Path-Indicator. The label value in the next label stack entry is the SR-Path-Identifierand the C-flag indicates whether a Source-SID label stack entry is also present. The label values are used as the key for accounting SR path traffic. If the Source-SID label stack entry is absent, an implementation may find it helpful to use a mock Source-SID value of zero for accounting purposes.

The SR-Path-Identifier may be located at different depth in the packet based on the RLDC of nodes in the network as described in Section 6. Finding the SR-Path-Identifier in the packet may be a costly operation and MUST NOT be done unless if SR path accounting is enabled on the device. Implementations MUST include a device-wide

configuration option to enable and disable SR path accounting, and this option MUST default to "off". Implementations SHOULD include more granular configuration (such as per-interface).

A further configuration option is to limit the type of packets to which the procedures described in this section are applied. Thus, the forwarding plane could be configured to inspect only SR packets, or only MPLS packets established using a specific control plane technique (such as LDP). The top label on the incoming packet can be used to determine the nature of the packet and whether to search for the SR-Path-Identifier. The SR labels are predictable and are mostly assigned from SRGB or SRLB. If the top label belongs to any of these label blocks the procedures described in this section may be applied. If the SR label is allocated dynamically as in case of dynamic Adjacency-SIDs, it may be difficult to identify whether the label belongs to SR. It is RECOMMENDED to use configured Adjacency-SIDs when SR path traffic accounting is enabled.

If the top label of the incoming packet is of the right type for accounting and if other appropriate configuration options are enabled, then packet’s label stack MUST be examined label by label until an SR-Path-Indicator label is found. The label below SR-Path-Indicator label is the SR-Path-Identifier label and the Source-SID label follows according to the setting of the C-flag. The {incoming interface, SR-Path-Identifier, Source SID} together are the key for traffic accounting. If the Source-SID label stack entry is absent, an implementation may find it helpful to use a mock Source-SID value of zero for accounting purposes.

If a counter does not already exist for that three-tuple, a new counter SHOULD be created. If a counter already exists, it MUST be incremented.

There is no requirement to preemptively create counters for every incoming interface and every SID: the counters need only be created, when a packet is received with the new SR-Path-identifier. This will significantly reduce the number of counters that need to be instantiated as not every interface will receive traffic for any particular SR path.

If the SR-Path-Indicator is the top label in a packet, the SR-Path-Stats labels are popped and further processing is based on the remaining labels in the label stack. Implementations MUST make sure the traffic accounting is carried out before the SR-Path-Stats labels are popped.

9. Consideration of Protection Mechanisms

SR paths typically consist of one or more Node-SIDs, Adjacency-SIDs, Anycast-SIDs, and Binding-SIDs. A variety of protection mechanisms may be in place for these SIDs as described in [I-D.ietf-spring-resiliency-use-cases]. When the head-end node inserts the SR-Path-Stats labels in the label stack, the place in the stack is decided based on whether the node where the special label gets exposed is capable of popping those labels.

When link protection is enabled, the traffic reaches the next-hop node before moving to towards the destination. With link-protection enabled, there is no risk of exposing the special labels at a node that does not support the extensions.

When node-protection is enabled, the traffic skips the next-hop node and reaches the next-next-hop towards the destination. In this case there is a possibility of special labels getting exposed at a node (the Merge Point) that does not support the extensions described in this document. In such cases, the node that receives the packet with special label at the top will discard the packet according to the processing rules of Section 3.18 of [RFC3031]. When using extensions described in this document for traffic accounting and with node-protection enabled in the network, it is RECOMMENDED to make sure all the nodes in the network support the extension.

10. Backward Compatibility

The extensions described in this document are backward compatible. Nodes that do not support the extensions defined in this document will not account the traffic (they will not search for the SR-Path-Indicator), but will forward traffic as normal.

While inserting the SR-Path-Stats labels, the head-end router MUST ensure that the labels are not exposed to the nodes that do not support them. If an error is made such that the SR-Path-Stats labels are exposed at the top of the label stack at a node that does not support this document then that node will discard the packets according to [RFC3031]. While the packets will be black-holed, no further harm will be caused to the network, and since this is a configuration or implementation error, this is an acceptable situation.

If an appropriate point in the label stack cannot be found for the insertion of the SR-Path-Stats labels, the head-end node, head-end MUST NOT insert the SR-Path-Stats labels, but SHOULD continue to label and transmit data. Under such circumstances the head-end node

SHOULD also log the event. A head-end or central controller MAY seek an alternate SR path that allows traffic accounting.

11. Scalability Considerations

The counter space is a limited resource in hardware. As described in Section 8 counters need only be created, when a packet is received with the an SR-Path-Identifier. Furthermore, counters need only be maintained where collection of statistics is configured.

Head-end nodes MUST NOT insert SR-Path-Stats labels by default. Careful configuration of which SR paths have statistics collection enabled will help to minimize the number of counters that need to be maintained at transit nodes.

Transit nodes that are constrained for the number of counters that they can support MAY implement mechanisms that sacrifice some under-used counters to create new counters.

As previously noted, the label stack is a prescious resource itself. That means that under some circumstances it is desirable to only use two labels in the SR-Path-Stats label sequence rather than three. This can be achieved by using a central controller to allocate SR-Path-Identifier values and set the C-flag to indicate that no Source-SID is used.

Conversely, in a large network with a central controller the SR-Path-Identifier may be a prescious resource. That is, there may be more than 2^19 SR paths that need identifiers to be allocated. In this case, a central controller may use knowledge of label stack depth and network node capabilities to allocate SR-Path-Indicators that include a Source-SID (set to indicate the controller, itself) where that would not cause a problem in the network.

12. Security Considerations

As noted in Section 11 the counter space is a limited resource in hardware. This document introduces dynamic creation of counters based on packet headers of the incoming packets. There is the possibility that a DOS attack is mounted by requesting new counter creation on each packet. Implementations SHOULD monitor the counter space and generate appropriate warnings if the counter space is getting exhausted. Implementations SHOULD control the rate at which the counters get created to mitigate DOS attacks.

13. IANA Considerations

IANA maintains a registry called the "Multiprotocol Label Switching Architecture (MPLS) Label Values" registry. IANA is requested to make a new assignment from this registry as follows:

Value	Description	Reference
TBD-1	SR Path Indicator	[This.I-D]

14. Acknowledgements

Thanks to John Drake, Harish Sitaraman, and Ron Bonica for helpful discussions.

15. Contributors

Adrian Farrel
Juniper Networks

Email: afarrel@juinper.net

16. References

16.1. Normative References

[I-D.ietf-idr-te-lsp-distribution]
Previdi, S., Dong, J., Chen, M., Gredler, H., and j. jefftant@gmail.com, "Distribution of Traffic Engineering (TE) Policies and State using BGP-LS", draft-ietf-idr-te-lsp-distribution-07 (work in progress), July 2017.

[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>.

16.2. Informative References

[I-D.filsfils-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., Raza, K., Liste, J., Clad, F., Lin, S., bogdanov@google.com, b., Horneffer, M., Steinberg, D., Decraene, B., and S. Litkowski, "Segment Routing Policy for Traffic Engineering", draft-filsfils-spring-segment-routing-policy-01 (work in progress), July 2017.

[I-D.ietf-idr-segment-routing-te-policy]
Previdi, S., Filsfils, C., Mattes, P., Rosen, E., and S. Lin, "Advertising Segment Routing Policies in BGP", draft-ietf-idr-segment-routing-te-policy-00 (work in progress), July 2017.

[I-D.ietf-isis-segment-routing-msd]
Tantsura, J., Chunduri, U., Aldrin, S., and L. Ginsberg, "Signaling MSD (Maximum SID Depth) using IS-IS", draft-ietf-isis-segment-routing-msd-04 (work in progress), June 2017.

[I-D.ietf-ospf-segment-routing-msd]
Tantsura, J., Chunduri, U., Aldrin, S., and P. Psenak, "Signaling MSD (Maximum SID Depth) using OSPF", draft-ietf-ospf-segment-routing-msd-05 (work in progress), June 2017.

[I-D.ietf-spring-resiliency-use-cases]
Filsfils, C., Previdi, S., Decraene, B., and R. Shakir, "Resiliency use cases in SPRING networks", draft-ietf-spring-resiliency-use-cases-11 (work in progress), May 2017.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <https://www.rfc-editor.org/info/rfc6790>.

[RFC7274] Kompella, K., Andersson, L., and A. Farrel, "Allocating and Retiring Special-Purpose MPLS Labels", RFC 7274, DOI 10.17487/RFC7274, June 2014, <https://www.rfc-editor.org/info/rfc7274>.

Author's Address

Shraddha Hegde
Juniper Networks, Inc.
Embassy Business Park
Bangalore, KA 560093
India

Email: shraddha@juniper.net