Workgroup: SPRING
Internet-Draft:
draft-filsfils-spring-net-pgm-extension-srv6-
usid-14
Published: 12 December 2022
Intended Status: Standards Track
Expires: 15 June 2023
Authors: C. Filsfils, Cisco Systems, Inc.
P. Camarillo, Ed., Cisco Systems, Inc.
D. Cai, Alibaba
D. Voyer, Bell Canada
I. Meilik, Broadcom
K. Patel, Arrcus, Inc.
W. Henderickx, Nokia
P. Jonnalagadda, Barefoot Networks
D. Melman, Marvell
Y. Liu, China Mobile
J. Guichard, Futurewei

Network Programming extension: SRv6 uSID instruction

Abstract

The SRv6 "micro segment" (SRv6 uSID or uSID for short) instruction is a straightforward extension of the SRv6 Network Programming model:

* The SRv6 Control Plane is leveraged without any change
* The SRH dataplane encapsulation is leveraged without any change
* Any SID in the SID list can carry micro segments
* Based on the Compressed SRv6 Segment List Encoding in SRH [I-D.ietf-spring-srv6-srh-compression] framework

This enables:

* ultra-scale (e.g. multi-domain 5G deployments)
* minimum MTU overhead
* installed-base reuse

Requirements 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 BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

SRv6 Network Programming [RFC8986] defines a mechanism to build a network program with topological and service segments. It leverages the SRH [RFC8754] to encode a network program together with optional metadata shared among the different SIDs.

This draft extends SRv6 Network Programming with a new type of SRv6 SID behaviors: SRv6 uN, uA, uDT, uDX.

This extension fully leverages the SRv6 network programming solution:

- *The SRv6 Control Plane is leveraged without any change
- *The SRH dataplane encapsulation is leveraged without any change
- *Any SID in the SID list can carry micro segments
- *Based on the Compressed SRv6 Segment List Encoding in SRH [I-D.ietf-spring-srv6-srh-compression] framework

This enables:

- *ultra-scale (e.g. multi-domain 5G deployments)
- *minimum MTU overhead
- *installed-base reuse

2. Terminology

The SRv6 Network Programming, SRH and Compressed SRv6 Segment List Encoding in SRH terminology is leveraged and extended with the following terms:

Term	Definition
uSID block	A block of uSID's. It can be any IPv6 prefix available to the provider.
uSID	A Compressed-SID. In this document a 16-bit ID. A different uSID length may be used.
Active uSID	First uSID after the uSID block.
Next uSID	Next uSID after the Active uSID.
Last uSID	From left to right, the last uSID before the first End-of-Container uSID.
End-of-Container	Reserved uSID used to mark the end of a uSID container. The value 0000 is selected as End-of-Container. All of the empty uSID container positions must be filled with the End-of-Container ID. Hence, the End-of-Container can be present more than once in a uSID container.
uSID container	A CSID container. A 128bit SRv6 SID of format <uSID-Block><Active-uSID><Next-uSID>...<Last-uSID><End-of-Container>...<End-of-Container>. A uSID container can be encoded in the Destination Address of an IPv6 header or at any position in the Segment List of an SRH.

Table 1

3. uSID Allocation within a uSID Block

3.1. GIB, LIB, global uSID and local uSID

GIB: The set of IDs available for global uSID allocation.

LIB: The set of IDs available for local uSID allocation.

3.1.1. Global uSID

A uSID from the GIB.

A Global uSID typically identifies a shortest-path to a node in the SR domain. An IP route (e.g., /64) is advertised by the parent node to each of its global uSID's, under the associated uSID block. The parent node executes a variant of the END behavior.

A node can have multiple global uSID's under the same uSID blocks (e.g. one per IGP flex-algorithm). Multiple nodes may share the same global uSID (anycast).

3.1.2. Local uSID

A uSID from the LIB.

A local uSID may identify a cross-connect to a direct neighbor over a specific interface or a VPN context.

No IP route is advertised by a parent node for its local uSID'.

If N1 and N2 are two different physical nodes of the uSID domain and I is a local uSID value, then N1 and N2 may bind two different behaviors to I.

3.1.3. Reference Illustration

For illustration simplicity, we will use:

* uSID block length: 48 bits

* uSID block: 2001:db8:0::/48

* uSID length: 16 bits

* uSID: 2001:db8:0:XYZW::/64

* GIB: nibble X from hexa(0) to hexa(D)

* LIB: nibble X hexa(E) or hexa(F)

Leveraging our reference illustration,

* A uSID 2001:db8:0:XYZW::/64 is said to be allocated from its block (2001:db8:0::/48).

* More specifically, a uSID is allocated from the GIB or LIB of block 2001:db8:0::/48 depending on the value of the "X" nibble: 0-D for GIB, and E-F for LIB.

* With the above allocation scheme, the uSID Block 2001:db8:0::/48 supports up to 57k global uSID's (e.g. routers) while each router would support up to 8k local uSID's.

Another illustration could assume a 32-bit uSID length and a LIB restricted to the uSIDs with the first byte set to FF. In this context, the network as a whole would support 2^32-2^24 global uSID's (e.g. routers) while each router would support up to 2^24 local uSID's.

4. SRv6 behaviors associated with a uSID

The SRv6 SRH encapsulation and its network programming model are extended with the following functions:

4.1. uSID behaviors related to the IGP

4.1.1. uN

The uN is a short notation for the End behavior with NEXT-CSID, PSP and USD flavors as defined in [I-D.ietf-spring-srv6-srh-compression].

As a reminder the pseudo-code of the End behavior with NEXT-CSID flavor, when applied to a 48b uSID block and a 16b uSID length is as follows:

```
2001:db8:0:0N00::/64 bound to the pseudocode shift-and-lookup:
    1.   Copy DA[64..127] into DA[48..111]                  ;; Ref1
    2.   Set DA[112..127] to 0x0000
    3.   Forward the packet to the new DA

2001:db8:0:0N00::/80 bound to the End behavior with PSP & USD flavors
```

Ref 1: DA[X..Y] refers to the bits from position X to Y (included) in the IPv6 Destination Address of the received packet. The bit 0 is the MSB, while the bit 127 is the LSB.

4.1.1.1. Control-plane representation

In ISIS [I-D.ietf-lsr-isis-srv6-extensions], a uN is advertised with the following information:

*Value = 2001:db8:0:0N00::

*Behavior = uN

*Structure =

-LBL = 48

-LNL = 16

-FL = 0

-AL = 64

*Algorithm = 0 (or other)

4.1.2. uA

The uA local behavior is a short notation for the End.X behavior with NEXT-CSID, PSP and USD flavors [I-D.ietf-spring-srv6-srh-compression].

An instance of the uA SRv6 uSID behavior is associated with a set, J, of one or more Layer-3 adjacencies.

As a reminder the pseudo-code of the End.X behavior with NEXT-CSID flavor, when applied to a 48b uSID block and a 16b uSID length is as follows:

```
2001:db8:0:FNAJ::/64 bound to the pseudocode shift-and-xconnect:
    1.   Copy DA[64..127] into DA[48..111]                  ;; Ref1
    2.   Set DA[112..127] to 0x0000
    3.   Forward to layer-3 adjacency J

2001:db8:0:FNAJ::/80 bound to the End.X behavior w PSP & USD flavors
```

Ref 1: DA[X..Y] refers to the bits from position X to Y (included) in the IPv6 Destination Address of the received packet. The bit 0 is the MSB, while the bit 127 is the LSB.

4.1.2.1. Control-plane representation

In ISIS [I-D.ietf-lsr-isis-srv6-extensions], a uA is advertised with the following information:

*Value = 2001:db8:0:0N00:FNAJ::

*Behavior = uA

*Structure =

-LBL = 48

-LNL = 16

-FL = 16

-AL = 48

*Algorithm = 0 (or other)

Note: From a formal viewpoint, a uA SID of node N is defined by the local FIB entry B:uA/64 of N (i.e. this definition is independent from any uN SID of node N). In order to signal in ISIS a container SID with the same routable semantics as End.X, the ISIS advertisement of a uA SID is done as uN+uA. uN provides the global route to the node like the End behavior. uA provides the cross-connect function like the "X" of the End.X.

4.2. uSID Behaviors related to BGP

4.2.1. uDT

A local uDT behavior of Node D 2001:db8:0:FNVT:: is defined by the following single FIB entry and pseudo-code:

2001:db8:0:FNVT::/80 bound to the same pseudocode as End.DT4/End.DT6/End.DT2*

4.2.1.1. Control-plane representation

In BGP [I-D.ietf-bess-srv6-services], a uDT is advertised with the following information:

*Value = 2001:db8:0:0N00:FNVT::

*Behavior = uDT

*Structure =

-LBL = 48

-LNL = 16

-FL = 16

-AL = 0

-TL = 16

-TO = 64

*Algorithm = 0 (or other)

Note: the advertised SID value includes the uN SRv6 uSID of the parent.

4.2.2. uDX

A local uDX behavior of Node D 2001:db8:0:FNXJ:: is defined by the following single FIB entry and pseudo-code:

2001:db8:0:FNXJ::/80 bound to the same pseudocode as End.DX4/End.DX6/End.DX2

4.2.2.1. Control-plane representation

In BGP [I-D.ietf-bess-srv6-services], a uDX is advertised with the following information:

*Value = 2001:db8:0:0N00:FNXJ::

*Behavior = uDX

*Structure =

-LBL = 48

-LNL = 16

-FL = 16

-AL = 0

-TL = 16

-TO = 64

*Algorithm = 0 (or other)

Note: the advertised SID value includes the uN SRv6 uSID of the parent.

5. FIB entry at originating node for performant support of global-local sequence

Any originating parent node may install the sequence of <Global, Local> uSID to perform more efficient processing given the LPM lookup.

For example, a parent node N that has the following FIB entries:

*2001:db8:0:0N00::/64 bound to the pseudocode shift-and-lookup

*2001:db8:0:0N00:0000::/80 bound to the End behavior with PSP&USD flavors

*2001:db8:0:FNAJ::/64 bound to the pseudocode shift-and-xconnect

*2001:db8:0:FNAJ:0000:/80 bound to the End.X behavior with PSP&USD flavors

may install the following additional FIB entries:

*2001:db8:0:0N00:FNAJ::/80 bound to the pseudocode shift-and-xconnect (with 32b shifting)

*2001:db8:0:0N00:FNAJ:0000::/96 bound to the End.X behavior with PSP&USD flavors

6. Routing

If Node 1 is configured with a uN SID 2001:db8:0:0100::/64 then the operator must ensure that Node 1 advertises 2001:db8:0:0100::/64 in the routing protocol.

7. Benefits

*Leverages SRv6 Network Programming with NO change

-SRv6 uSID is a flavor of the SRv6 network programming model

*Leverages SRv6 dataplane (SRH) with NO change

-Any SID in DA or SRH can be an SRv6 uSID container

*Leverages SRv6 Control-Plane with NO change

*Ultra-Scale

-6 uSID's per uSID container

-18 source routing waypoints in only 40bytes of overhead

oH.Encaps.Red with an SRH of 40 bytes (8 fixed + 2 * 16 bytes)

o6 uSID's in DA and 12 in SRH

*Lowest MTU overhead

-In apple to apple comparison, the SRv6 solution outperforms any alternative (VxLAN with SR-MPLS, CRH).

*Scalable number of globally unique nodes in the domain

-16-bit uSID: 65k uSIDs per domain block

-32-bit uSID: 4.3M uSIDs per domain block

*Proven Hardware-friendliness

-Leverages mature hardware capabilities (Inline DA edit, DA longest match)

-Avoids any extra lookup in indexed mapping table

- Demonstrated by the number of linerate interoperable hardware implementations at the first Interop report in February 2020, less than 9 months after the first public version of this document.

- Public operator report of leverage of installed base

- A micro-program which requires less than 6 uSID's only requires legacy IPinIP encapsulation behavior

* Scalable Control-Plane

  - No indexed mapping table is required

  - Summarization at area/domain boundary provides massive scaling advantage

  - No routing extension is required: a simple prefix advertisement suffices

* Seamless Deployment

  - A uSID may be used as a SID: i.e. the container holds a single uSID

  - The inner structure of an SR Policy can stay opaque to the source: i.e. a container with uSID's is just seen as a SID by the policy headend

* Security

  - Leverages SRv6's native SR domain security

* Large-Scale DC

  - SID's may be used to address applications on hosts (scale in 2^128)

  - Hardware friendliness of uSID's may be used to specify billions of waypoints in cost/power-optimized DC fabric

8. Running code

8.1. NANOG78 interoperability testing

The hardware and software platforms listed have participated in a joint interoperability testing of the uN instruction defined in this document.

Hardware implementations (in alphabetical order):

* Arrcus ArcOS (based on Broadcom Jericho2)

* Barefoot Tofino P4-programmable Ethernet switch ASIC

* Cisco 8000 Series Routers (based on Cisco Silicon One Q100)

* Cisco ASR9000 platform (with 3rd gen Tomahawk and 4th gen Lightspeed line-cards)

* Cisco NCS5500 platform (based on Broadcom Jericho/Jericho+)

* Marvell Prestera Packet Processor

Software open-source implementations (in alphabetical order):

* FD.io VPP

* Linux Kernel

Further details are available in the [NANOG78].

8.2. L3VPN interoperability testing with control-plane

In December 2020 the following routing platforms have participated in a successful interoperability testing including the uDT instruction and its BGP control-plane signalling.

* Arrcus ArcOS

* Cisco ASR9000 with IOS-XR

* Cisco NCS5500 with IOS-XR

* Cisco XRv9k with IOS-XR

* FD.io VPP with GoBGP

Further details are available in [L3VPN-INTEROP].

8.3. Dataplane traffic engineering interoperability testing

In November 2020, the following hardware and software platforms have participated in a joint interoperability testing of the uN instruction defined in this document. This interoperability testing was hosted by China Mobile.

* Hardware implementation in Cisco ASR 9000 running IOS XR

* Software implementation in Cisco IOS XRv9000 virtual appliance

*Hardware implementation in Huawei NE40E running VRP

*Hardware implementation in Huawei NE5000E running VRP

Further details are available in [I-D.ietf-spring-srv6-srh-compression] Section 11.

9. Security

The security rules defined in Section 7 of [RFC8986], protect intra-domain deployments that includes SRv6 uSID.

10. IANA Considerations

This document requests IANA to allocate the following codepoints within the "SRv6 Endpoint Behaviors" sub-registry under the top-level "Segment Routing Parameters" registry.

Value	Hex	Endpoint behavior	Reference
42	0x002A	End with NEXT-ONLY-CSID	[This.ID]
43	0x002B	End with NEXT-CSID	[This.ID]
44	0x002C	End with NEXT-CSID & PSP	[This.ID]
45	0x002D	End with NEXT-CSID & USP	[This.ID]
46	0x002E	End with NEXT-CSID, PSP & USP	[This.ID]
47	0x002F	End with NEXT-CSID & USD	[This.ID]
48	0x0030	End with NEXT-CSID, PSP & USD	[This.ID]
49	0x0031	End with NEXT-CSID, USP & USD	[This.ID]
50	0x0032	End with NEXT-CSID, PSP, USP & USD	[This.ID]
51	0x0033	End.X with NEXT-ONLY-CSID	[This.ID]
52	0x0034	End.X with NEXT-CSID	[This.ID]
53	0x0035	End.X with NEXT-CSID & PSP	[This.ID]
54	0x0036	End.X with NEXT-CSID & USP	[This.ID]
55	0x0037	End.X with NEXT-CSID, PSP & USP	[This.ID]
56	0x0038	End.X with NEXT-CSID & USD	[This.ID]
57	0x0039	End.X with NEXT-CSID, PSP & USD	[This.ID]
58	0x003A	End.X with NEXT-CSID, USP & USD	[This.ID]
59	0x003B	End.X with NEXT-CSID, PSP, USP & USD	[This.ID]
60	0x003C	End.DX6 with NEXT-CSID	[This.ID]
61	0x003D	End.DX4 with NEXT-CSID	[This.ID]
62	0x003E	End.DT6 with NEXT-CSID	[This.ID]
63	0x003F	End.DT4 with NEXT-CSID	[This.ID]
64	0x0040	End.DT46 with NEXT-CSID	[This.ID]
65	0x0041	End.DX2 with NEXT-CSID	[This.ID]
66	0x0042	End.DX2V with NEXT-CSID	[This.ID]
67	0x0043	End.DT2U with NEXT-CSID	[This.ID]
68	0x0044	End.DT2M with NEXT-CSID	[This.ID]

Table 2: IETF - SRv6 Endpoint Behaviors

11. Acknowledgements

The authors would like to acknowledge Francois Clad, Peter Psenak, Ketan Talaulikar, Jakub Horn, Swadesh Agrawal, Zafar Ali, Darren Dukes, Kiran Sasidharan, Junaid Israr, Lakshmanan Srikanth, Asif Islam, Saleem Hafeez, Michael MacKenzie, Sushek Shekar, YuanChao Su, Alexander Preusche, Alberto Donzelli, Miya Kohno, David Smith, Ianik Semco, Bertrand Duvivier, Frederic Trate, Kris Michielsen, Eyal Dagan, Eli Stein, Ofer Iny, Elad Naor, Guy Caspari, Mel Tsai, Anand Sridharan, Aviad Behar, Joseph Chin.

12. Contributors

Jisu Bhattacharyaa Cisco Systems, Inc. United States of America

Email: jisu@cisco.com

Kamran Raza Cisco Systems, Inc. Canada

Email: skraza@cisco.com

John Bettink Cisco Systems, Inc. United States of America

Email: jbettink@cisco.com

Tomonobu Niwa KDDI Japan

Email: to-niwa@kddi.com

Luay Jalil Verizon United States of America

Email: luay.jalil@one.verizon.com

Zhichun Jiang Tencent China

Email: zcjiang@tencent.com

Ahmed Shawky Saudi Telecom Company Saudi Arabia

Email: ashawky@stc.com.sa

Nic Leymann Deutsche Telekom Germany

Email: N.Leymann@telekom.de

Dirk Steinberg Lapishills Consulting Limited Cyprus

Email: dirk@lapishills.com

Shawn Zandi LinkedIn United States of America

Email: szandi@linkedin.com

Gaurav Dawra LinkedIn United States of America

Email: gdawra@linkedin.com

Jim Uttaro AT&T United States of America

Email: ju1738@att.com

Ning So Reliance United States of America

Email: Ning.So@ril.com

Michael Fiumano Sprint United States of America

Email: michael.f.fiumano@sprint.com

Mazen Khaddam Cox United States of America

Email: Mazen.Khaddam@cox.com

Jichun Ma China Unicom China

Email: majc16@chinaunicom.cn

Satoru Matsushima Softbank Japan

Email: satoru.matsushima@g.softbank.co.jp

Francis Ferguson CenturyLink United States of America

Email: Francis.Ferguson@centurylink.com

Takuya Miyasaka KDDI Japan

Email: ta-miyasaka@kddi.com

Kentaro Ebisawa Toyota Motor Corporation Japan

Email: ebisawa@toyota-tokyo.tech

Yukito Ueno NTT Communications Corporation Japan

Email: yukito.ueno@ntt.com

13. References

13.1. Normative References

[I-D.ietf-spring-srv6-srh-compression]

Cheng, W., Filsfils, C., Li, Z., Decraene, B., Cai, D., Voyer, D., Clad, F., Zadok, S., Guichard, J., Liu, A., Raszuk, R., and C. Li, "Compressed SRv6 Segment List Encoding in SRH", Work in Progress, Internet-Draft, draft-ietf-spring-srv6-srh-compression-02, 11 July 2022, <https://www.ietf.org/archive/id/draft-ietf-spring-srv6-srh-compression-02.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>.

[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>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

[RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <https://www.rfc-editor.org/info/rfc8986>.

13.2. Informative References

[I-D.ietf-bess-srv6-services]
Dawra, G., Talaulikar, K., Raszuk, R., Decraene, B., Zhuang, S., and J. Rabadan, "BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)", Work in Progress, Internet-Draft, draft-ietf-bess-srv6-services-15, 22 March 2022, <https://www.ietf.org/archive/id/draft-ietf-bess-srv6-services-15.txt>.

[I-D.ietf-lsr-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extensions to Support Segment Routing over IPv6 Dataplane", Work in Progress, Internet-Draft, draft-ietf-lsr-isis-srv6-extensions-19, 14 November 2022, <https://www.ietf.org/archive/id/draft-ietf-lsr-isis-srv6-extensions-19.txt>.

[L3VPN-INTEROP] Cisco Systems, Inc. and Arrcus, "SRv6 uSID L3VPN Interopability Testing", L3VPN Interop , December 2020, <https://www.segment-routing.net/demos/2020-12-22-SRv6-uSID-L3VPN-interoperability/>.

[NANOG78]
Filsfils, C., "SRv6 Technology and Deployment Use-cases", NANOG78 , February 2020, <https://storage.googleapis.com/site-media-prod/meetings/NANOG78/2097/20200212_Mcdougall_Srv6_Technology_And_v1.pdf>.

Authors' Addresses

Clarence Filsfils
Cisco Systems, Inc.
Belgium

Email: cf@cisco.com

Pablo Camarillo Garvia (editor)
Cisco Systems, Inc.
Spain

Email: pcamaril@cisco.com

Dennis Cai
Alibaba
China

Email: d.cai@alibaba-inc.com

Daniel Voyer
Bell Canada
Canada

Email: daniel.voyer@bell.ca

Israel Meilik
Broadcom
Israel

Email: israel.meilik@broadcom.com

Keyur Patel
Arrcus, Inc.
United States of America

Email: keyur@arrcus.com

Wim Henderickx
Nokia
Belgium

Email: wim.henderickx@nokia.com

Prem Jonnalagadda
Barefoot Networks
United States of America

Email: prem@barefootnetworks.com

David Melman
Marvell
Israel

Email: davidme@marvell.com

Yisong Liu
China Mobile
China

Email: liuyisong@chinamobile.com

James Guichard
Futurewei
United States of America

Email: james.n.guichard@futurewei.com