



SRv6 的分段标识符结构的改善

唐永林, 刘锋

(UT 斯达康通讯有限公司, 浙江 杭州 310052)

摘要: 基于 IPv6 转发的分段路由 (segment routing, SR) SRv6 是 IETF 标准化组织近几年热烈研讨的新技术。在《SRv6 网络编程》草案中, 对 128 bit 的分段标识符 (segment identifier, SID) 进行了结构定义和节点操作描述。对草案定义的 SID 结构进行了改善, 在 SID 中划出了一段命名为网络 (network), 以标识不同的虚拟子网络。同时指出在分段路由头 (segment routing header, SRH) 中的 SID 列表中不携带网络, 只携带定位、功能和参数, 可以有效地降低多层 SRv6 段标识符列表需求对芯片处理能力的要求, 降低网络开销, 提高网络带宽的有效利用率, 同时为 SRv6 的网络切片运用打下坚实基础, 促使 SRv6 的大规模运用。

关键词: 分段路由; 基于 IPv6 的分段路由; 分段标识符; 网络切片

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2020121

Improvement of SRv6 SID structure

TANG Yonglin, LIU Feng

UTStarcom Telecom Co. Ltd., Hangzhou 310052, China

Abstract: IPv6-based segment routing (SRv6) is a new technology addressed by IETF in recent years. In draft-ietf-spring-srv6-network-programming, the 128-bit segment identifier (SID) was structured, and how to be processed in one node was described. The SID structure defined by the draft was improved. A field named “network” was defined in the SID to identify different virtual sub-networks. It was also pointed out that the SID list in the segment routing header (SRH) did not carry the “Network”, and only carried the locator, function and argument, which could effectively reduce the requirements of the multi-layer SRv6 segment identifier list requirements on the chip processing capability and reduce the overhead in SRv6 networks. To improve the effective utilization of network bandwidth, and at the same time lay a solid foundation for the use of SRv6 network slicing, and promote the large-scale use of SRv6.

Key words: segment routing, SRv6, segment identifier, network slicing

1 引言

分段路由 (segment routing, SR) 技术是 IETF

标准化组织研讨的一种新技术、新标准。目前多数标准还处于草案阶段。SR 是一种源路由技术, 路径信息显示在报文头里面, 每个节点只需要根

据报文头的指示决定下一步操作。根据不同的底层转发技术, SR 又分为 SR MPLS 和 SRv6 两种。SR MPLS 基于 MPLS 转发技术, 分段标识符(segment identifier, SID)是 32 bit 的 MPLS 标签。目前该技术得到了多数电信设备厂商的支持, 相关标准也进展较快。中国移动宣布的下一代承载网采用的 SR 是对 IETF 的 SR MPLS 作了简化和增强的。SR 与 EVPN、FlexE 等技术因将成为下一代智能融合城域网关键技术之一而被广泛关注^[1]。

基于 IPv6 的分段路由 SRv6 是一种通过 SRH(segment routing header) 扩展头部携带 SID 列表的源路由技术, 每一个 SID 用一个 128 bit 的 IPv6 地址来表示。报文转发依赖 SID 地址编码的信息。《IETF IPv6 Segment Routing Header (SRH)》^[2]详细描述了 SRH 的定义, 其中 SID 列表由 128 bit SID 组成。

《IETF 草案 SRv6 网络编程》^[3]对 SID 编码结构进行了详细定义, 如图 1 所示。

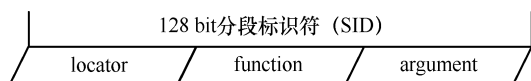


图1 分段标识符(SID)结构示意图

其中, 定位(locator)代表节点标识, 功能(function)代表 locator 匹配的节点收到该报文之后的处理方式。参数 argument 附属在 function 之后, 代表某功能更具体的信息。例如端点功能 END.DT4 需要节点查找路由表, 则其参数代表虚拟转发表 VRF 的标识; 例如端点功能 END.X 需要节点从特定的接口作为下一跳, 其参数代表具体的接口。草案对节点的操作进行了详细描述, 在此不再赘述。

全部 128 bit 的 SID 出现在 SRH 的列表中, 对芯片的处理带来了巨大的压力。目前一些号称支持 SRv6 的芯片单次循环只能压入 4 层 SID, 超过 4 层则需要将报文重新走一遍处理流程, 虽然可以得到超过 4 层 SID 的处理能力, 但代价是降

低芯片交换容量。

这样的规格显然是不能满足实际运用的。比如一个包含 6 个节点的网络里, 从任意节点到任意另外一个节点的可能路径中会出现超过 4 个节点的情况。比如节点 N1 到达节点 N6, 可能经过 N2/N3/N4/N5 到达 N6。如果在某种情况下, 需要采用逐跳严格路由, 当客户报文进入 SRv6 网络的第一个节点时, 芯片可能要压入 5 层 SID, 导致芯片处理上需要两次流程, 大大降低了交换容量和增加了处理时延。

目前为止在《SRv6 网络编程》草案里规定的 function 只有不到 40 种, 即便是为将来扩展做预留, 用 8 bit 便有 256 种编码可以表达足够多的功能。目前标准规定了 16 bit。根据实际运用情况参数 10 bit 有 1 000 标识, 无论是 VRF 还是接口都足够运用了, 但草案预留了 16 bit 给 argument 用, 这样的容量将达到 64 000。剩下给定位标识 locator 还有 $128-32=96$ 。定位标识的多少实际上代表了某个 SR 网络域内节点或虚拟节点的多少。考虑一定的运用扩展, 1 000 个网元规模的 SR 网络域, 每个网元有 16 个 locator, 则 16 000 个 locator, 分派 16 bit 给 locator 也代表有 64 000 个定位标识。如此看来总共 48 bit 的编码足够满足 SRv6 的全部运用以及将来的扩展。剩下的 80 bit 的信息实际上是有大量的冗余信息携带在 SRH 里面, 从网络带宽和芯片处理能力两方面来看都是不必要的负担。

为此本文提出了一种更加精细化的编码方式, 将大大提升芯片的处理能力, 降低报文开销, 为 SRv6 网络大规模运用奠定了基础。

2 分段标识符 SID 的编码

如上文描述, IETF SRv6 网络编程定义的 128 bit SID, 具有大量的冗余信息携带在 SRH 里, 增加了芯片的处理负担、网络带宽负担, 从而限制了 SRv6 的技术运用。为此本文对分段标识符(SID)进行了新的定义, 如图 2 所示。

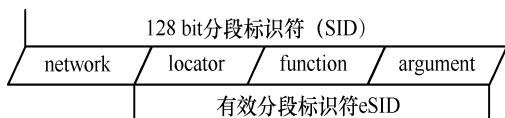


图2 分段标识符 (SID) 定义示意图

新编码实际上可以认为是对 SRv6 网络编程定义的 locator 的细化。网络识别符 network, 可以标识虚拟子网 (virtual network), 或域 (domain)、拓扑 (topology)、协议实例 (instance) 等信息。现在新的 locator 是某个 network 内部的 locator, 指示 network 内部的节点信息。

为了进一步地描述, 在此把不包含网络 (network) 标识的部分称为有效 SID, 即 eSID, 其包括 locator/function/argument。

在一个 SRv6 的大网络内, 采用本文所描述的 SID 结构方法, 技术上便可以将 SRv6 技术改造成天生支持多子网。划分子网可以根据业务的不同 QoS 要求来划分, 也可以根据端口或子端口信息来划分。在 5G 承载网中, 为了适应不同垂直行业的运用, 网络切片是一个很好的解决方案。这里描述的 network 可以对应不同的网络切片, 可以让 SRv6 更加适合切片网络的运用。

因为在同一个网络或者虚拟子网络内, 所有的 SID 的 network 都是一样的。当客户报文进入 SRv6 网络的某虚拟子网络时, 只要在添加新 IPv6 头中的目的 IP 地址里携带网络识别符 network, 而 SRH 中的分段列表只需要将 eSID 放进即可。在 SRH 里面去掉这些冗余信息网络识别符 network 将大大减轻 SRv6 开销。

包含这样 SRH 的 SRv6 报文如图 3 所示。

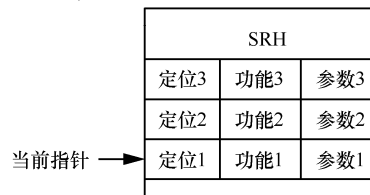
图 3 中, IPv6 头里的目的 IP 地址包含了 network/eSID, 而 SRH 里面的分段列表只携带 eSID。即如果 eSID 是 64 bit, 则原先 128 bit 的空间可以放两个 eSID。

3 eSID 长度通告方法

SRv6 的节点获取 eSID 的长度方法有两种:



(a) IPv6目的地址结构示意图



(b) SRH结构示意图

图3 SRv6 报文示意图

协议通告和静态配置。

eSID 的长度可以通过协议通告。这时需要将 ISIS extension for SRv6 或 OSPF extension for SRv6 进行扩展, 以支持 eSID 长度的全网通告。比如采用 ISIS 来扩展通告 eSID, 可以创建一个新的 TLV。而公告 function 的 TLV 也将扩展携带 network 信息, 这样某节点收到如此 TLV 就可以判断出 function 归属于 network, 并获取到 network 字段长度, 以按照本文描述的方法进行处理和子网内转发。详细的 TLV 定义以及节点处理方法就不再在本文中展开。

eSID 的长度也能以支持静态配置的方式通告全网节点。一般情况下采用集中的 SDN 控制器通过 Netconf 协议告知全网节点。

4 节点处理方法

由于 eSID 的长度是可变的, 因此节点在添加、删除和刷新 SRH 的时候和草案定义的固定长度 eSID 的处理方法有所区别。主要区别包括: 首先节点要获取 eSID 和 network 的长度信息。其次 IPv6 头的目的地址中 network 和 eSID 要分别处理。下面按节点角色来详细描述。

4.1 SRv6 网络的边缘节点 ingress PE

在划分逻辑子网的 SRv6 网络内, 端口或逻辑端口归属于某个逻辑子网, network 识别符和该逻辑子网一一对应。业务报文封装某 network, 其报

文只会在归属于该 network 的节点端口内转发。

对于业务报文进入 SRv6 网络的 ingress PE 节点, 是处理客户报文的起始节点。该节点首先要通过控制或管理平面获取该业务报文是进入哪个子网 (network)、哪条路径 (eSID 清单)。客户业务归属于哪个子网是需要事先规划的, 在本子网内的路径可以通过 PCE (path computation element) 计算并通过 PCEP (path computation element communication protocol) 下发给网元节点。ingress PE 只需要将 network 和下一跳 eSID 一起封装进入 IPv6 目的地址头。并将 eSID 列表封装进入 SRH。SRH 里其他字段信息的填写处理一般遵循草案标准所规定的方法。其编码方式按第 3 节映射表处理即可。

4.2 SRv6 网络内节点 P 和 egress PE

携带 SRv6 的报文, 达到 SRv6 的网络内节点, 则需要按如图 4 所示的流程处理。

图 4 中, 节点收到报文后, 首先要增加识别

network 字段处理, network 字段的信息和端口所属的 network 不一致则丢弃该报文。然后比较 locator, 是本地还是远端, 如果是远端, 则和草案描述的方法一样, 查 IP 地址表然后转发; 如果是本地 locator, 则要检查 function 字段, 如果是如下这些功能: End/End.x/End.T/End.B6/ End.B6.Insert/End.B6.Encaps/End.B6.Encaps.Red/End.BM, 则进入刷新目的 IP 地址和更新 SRH 的流程, 并且按 programming 定义的功能类型查表转发。

需要注意 SRH 里封装的是 eSID, 因此刷新 IPv6 的 eSID 字段, 节点需要明确获知 eSID 的长度信息, 这样才能正确提取和处理 SRH 里的 eSID。

如果 function 不是上述功能, 则进入 egress PE 节点流程。egress PE 点的处理流程一般遵循草案规定。

5 结束语

SRv6 网络编程定义了 128 bit SID 的结构, 并

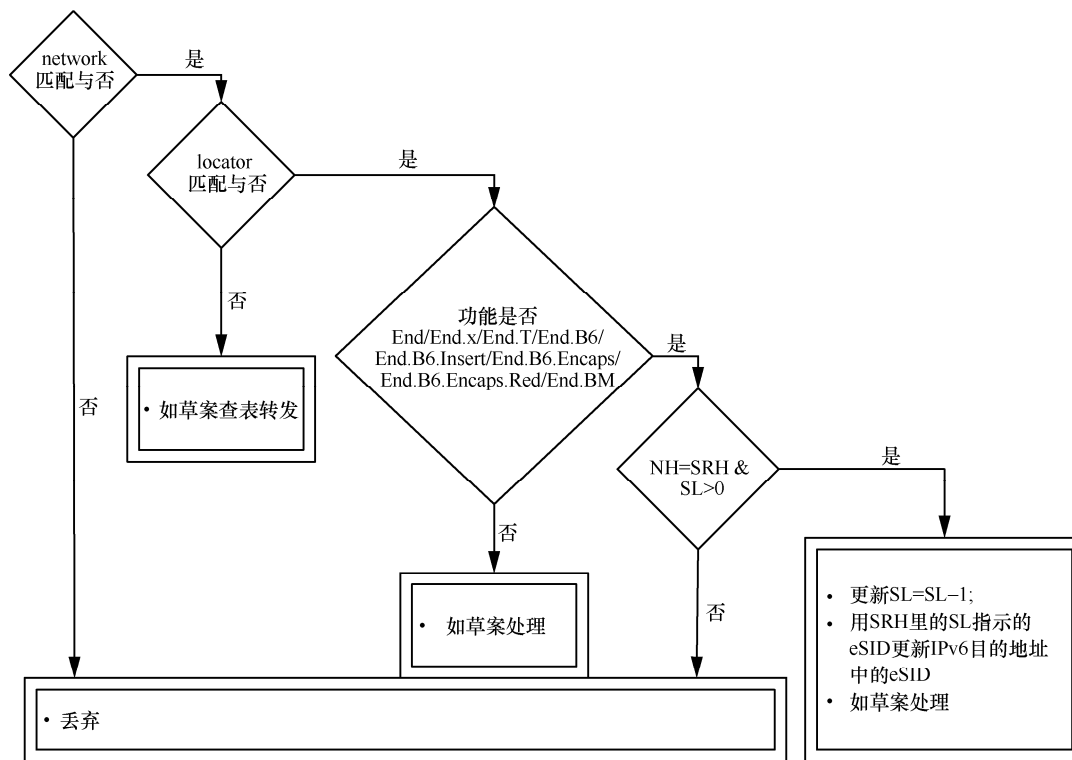


图 4 报文处理流程示意图



将完整的 128 bit 信息放进 SRH 里。对于复杂拓扑的大网络，多层 128 bit 的 SID 给芯片处理和网络带宽都带来没有必要的负担。针对这个问题，本文对 SID 的结构进行了改善，增加一个 network 信息段，即 SID 被重新划分成 network+eSID。

eSID 的长度可以灵活配置，可以适应不同大小的网络。eSID 的长度可以通过配置获得，全网统一。SRH 里只携带有效 eSID。eSID 可能 bit 数目只有标准的一半甚至四分之一，因此这样的封装方法和传输方式将大大减小芯片处理压力和网络带宽压力。

本文重构的 SID，方便地实现了业务报文在 SRv6 网络中的某一个 network 内不传输。但在有些场合下可能存在跨 network 子网络的传输可能性，为了更灵活地进行跨 network 子网络传输可能需要另外定义跨网络的 function。

参考文献：

- [1] 曹畅, 张帅, 唐雄燕. 下一代智能融合城域网方案[J]. 电信科学, 2019, 35(10): 51-59.
CAO C, ZHANG S, TANG X Y. Next generation intelligent converged metro network[J]. Telecommunications Science, 2019, 35(10): 51-59.
- [2] FILSFILS C, DUKES E D, PREVIDI E S, et al. draft-ietf-6man-segment-routing-header-21[EB]. 2019.
- [3] FILSFILS C, CAMARILLO P, LEDDY E J, et al. draft-ietf-spring-srv6-network-programming[EB]. 2019.

[作者简介]



唐永林（1970—），男，博士，UT 斯达康通讯有限公司高级（副教授）工程师、研发总监，主要从事 PTN、SPN、SRv6 和时钟同步等相关技术研究。参与了多个与 PTN、SPN 和时钟同步相关的设备和技术的行业标准制定。

刘锋（1982—），男，UT 斯达康通讯有限公司主任级系统工程师，主要研究方向为 PTN/IP-RAN、5G 承载网、SDN 解决方案以及如 SR MPLS、SRv6、网络切片、FlexE 等新技术。

收录声明

本刊已被中国核心期刊遴选数据库(万方)、中国学术期刊数据库(知网)全文收录,被本刊录用的文稿,将一律由编辑部统一纳入中国核心期刊遴选数据库(万方)、中国学术期刊数据库(知网)。若作者不同意被收录,将在来稿时向本刊说明,本刊将做适当处理。若无特别声明,视同将录用文稿的电子网络版、汇编或合订本的使用权授予本期刊社,本刊支付的稿费中已包含上述费用。

The articles of *Telecommunications Science* have been embodied by WANFANG DATA and CNKI. The accepted articles will be unified into WANFANG DATA and CNKI. If the author does not agree to his or her article to be included, the author must illustrate when the article is submitted. The editorial department will properly handle. If the authors do not give a special announcement, the editorial department believes that the authors have granted the right of their articles to *Telecommunications Science*. The expenses have been included in the fee paid by editorial department.

道德声明

本刊发表的文稿是作者独立取得的原创性研究成果,无一稿多投;文稿内容不涉及国家机密;未曾以任何形式用任何文种在国内外公开发表过;文稿内容不侵犯他人著作权和其他权利。若发生一稿多投、侵权、泄密等问题,文稿作者将承担全部责任。

The authors of *Telecommunications Science* guarantee that their submitted articles are original and contain nothing confidential. The said article is only submitted to *Telecommunications Science*. The said article has not been published before and has not been submitted elsewhere for print or electronic publication consideration. The said article is in no way whatever a violation or an infringement of any existing copyright or license from the third party. Otherwise, the authors of the said article shall take the blame for the violation or infringement of the related copyright and the leakage of secrets.

电信科学

Telecommunications Science



邮发代号：2-397 国外代号：M841 定价：68.00元

ISSN 1000-0801

