Routing Area Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 3, 2018

J. Dong
S. Bryant
Huawei Technologies

October 30, 2017

Protocol Considerations for Enhanced VPN
draft-dong-rtgwg-enhanced-vpn-protocol-00

Abstract

This document describes the candidate protocol mechanisms which may be used to meet the requirements of enhanced virtual private networks (VPN+). The gaps and limitations of existing mechanisms are analyzed, then a proposed mechanism is briefly described. The proposed mechanism can be used to achieve network slicing to meet the stringent requirement of emerging 5G services, but it can also be useful in other general network scenarios.

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

Virtual networks, often referred to as virtual private networks (VPNs) have been widely deployed to provide different groups of users with logically isolated access to a common network. The common network that is used to provide the VPNs is often referred to as the underlay, and the VPN is often called an overlay.

As described in [I-D.bryant-rtgwg-enhanced-vpn], the enhanced virtual private networks (VPN+) refers to a virtual network which has dedicated network resources allocated from the underlay network, so that can achieve greater isolation and guaranteed performance than traditional VPNs. VPN+ aims to provide a set of enhancements to existing VPN services, among which greater isolation is one of the key requirements of many emerging services, such as financial and vertical industrial services. Apparently such level of isolation cannot be met with pure overlay networks, as it requires tighter coordination and integration between the overlay and the underlay network, also it may rely on necessary enhancements to both the data plane and control plane of networks.

This document describes the candidate protocol mechanisms to meet the requirements of enhanced virtual private networks (VPN+), analyses

the limitations of some existing mechanisms, and proposes the protocol mechanisms needed to provide VPN+ service.

2. 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 [RFC2119].

3. Role of the Underlay and Overlay

Basically the VPN based multi-tenant networks consist of two layers: the overlay and the underlay, each layer plays a different role. The underlay is responsible for establishing the network connectivity based on the physical network infrastructure, also the management of physical network resources. The overlay is used to setup various customized virtual network topologies, and the logical network separation between different tenants.

The overlay and the underlay can be loosely coupled, in which case the overlay only requires the underlay to provide the connectivity between specific service nodes, without the control of the underlay path and resources, which means network resources in underlay can be shared by all the overlay networks. This can be desirable when scalability is the primary goal, as it minimizes the amount of state in the network.

However, as many emerging services require guaranteed performance and greater isolation from each other in the network, this loose-couple mode can not meet such requirements. The overlay and underlay need to be further integrated with each other. Normally this means more states need to be maintained in the network, which may cause scalability problem with some mechanisms. To overcome the scalability problem, VPN+ requires an efficient mechanism for network resource allocation and the mapping between the overlay and underlay networks.

4. Considerations about Network Isolation

The requirement of enhanced VPN is described in [I-D.bryant-rtgwg-enhanced-vpn], in which isolation is identified as one of the most important requirement. When a network is used to carry different types of services of multiple tenants, it is required to provide some levels of isolation between different services or different tenants. Based on different dimensions of isolation, network isolation is categorized and described in following sections.

4.1. Data Plane Isolation

Isolation in data plane is the fundamental requirement of services which are deployed as virtual private networks in a shared network infrastructure. Depends on the level of data plane isolation, the requirement is classified as soft isolation and hard isolation.

Soft isolation means that traffic of one application or tenant cannot be received or inspected by any other application or tenant. Usually soft isolation does not have strict resource or performance requirement, the underlying network resource can be shared by multiple applications or tenants, which is useful to achieve better economy with statistical multiplexing. With soft isolation, when service in one of the virtual networks experience some event such as traffic burst or congestion, this may result in negative impacts to other virtual networks in terms of bandwidth, latency, jitter, etc.

On the other hand, hard isolation means that any event happened to the traffic in one virtual network will not interfere any other application or tenant on the same network, which means the characteristics of service can be more predictable. To achieve this, at least some of the network resource need to be dedicated rather than shared, which may reduce the economy due to statistical multiplexing. Hard isolation is required by services that previously have their own dedicated private networks and expect to have the same network characteristics in a shared network.

4.2. Control Plane Isolation

There are many aspects in control plane, from router's perspective, isolation in control plane can be achieved in different levels: isolation of routing tables and isolation of routing protocols.

Isolation of routing tables is the preliminary requirement of multi-tenancy, and can be achieved with many existing VPN mechanisms. It usually can be done using a common control plane protocl such as BGP, and the scalability has been proved by the wide deployment in the field networks.

Isolation of routing protocols can provide further customization and flexibility, as different tenants or applications can choose their preferred protocols and provision it independently with customized parameters. The cost of routing protocol isolation is that it requires further complexity and more resource overhead, in some cases the scalability of control protocol isolation can be challenging.

With the introduction of SDN, isolation of control plane can be achieved by using separate controllers for different tenants or

applications. For example, each tenant can have his own controller for network information allocation, path computation and service provisioning. Note in this case, each tenant's controller can only see information specific to this tenant, and has no access to information of any other tenants. The tenant's controller may have limited access to information and states of the network infrastructure.

5. Analysis of Existing Mechanisms

This section analyses several existing mechanisms which are considered as the candidate protocol mechanisms for VPN+, and illustrates the gaps in meeting the requirements as described in section 4.

5.1. Overlay Virtual Networks

In this document, the conventional VPNs (L3VPN, L2VPN, EVPN etc.) and the overlay technologies developed in [NVO3] working group are classified as overlay virtual networks. These mechanisms aim at providing multi-tenant overlay connectivity using a unified control plane. The underlay provides no resource of performance commitment to the tenants of the overlays, thus the tenant only gets the best effort provided to any traffic carried with the same traffic class in the network.

According to operator's policy, the overlay connections of different tenants can either share the same underlay tunnel, or use separate dedicated tunnels to provide some degree of data plane isolation. If there is a requirement to provide guaranteed network bandwidth from a particular tenant, the approach is to establish a set of dedicated RSVP-TE [RFC3205] LSPs to carry the traffic of this tenant. However, such tunnels only provide bandwidth reservation for the tenant but no other guarantees. The mechanisms under development in [DETNET] working group may be needed for guaranteed low latency and packet loss.

However, as the number of tenants requiring guaranteed performance rises, so does the number of RSVP-TE LSPs, which ultimately leads to scalability problems in the network. There are ongoing efforts to improve the scalability of RSVP-TE LSPs both in control plane [I-D.ietf-teas-rsvp-te-scaling-rec] and in data plane [I-D.sitaraman-mpls-rsvp-shared-labels].

5.2. Multiple-Topology Routing and Segment Routing

Multi-topology Routing (MTR) [RFC4915][RFC5120] has been designed to provide multiple customized network topologies for different services. When native IP forwarding is used as the data plane, there is limitation in mapping the incoming packets of one interface to different MTR topologies. The major use cases of multi-topology routing is to provide different topologies for different address families, e.g. IPv4 and IPv6 with different topologies, or use particular topology for non-forwarding purpose, e.g. RPF check of multicast.

Segment Routing (SR) [I-D.ietf-spring-segment-routing] leverages the source routing paradigm and allows for flexible designation of forwarding paths by encoding the paths as sequences of "segments" in the data packet. In the IGP extensions for segment routing, multi-topology has been taken into consideration, which may be used to solve the forwarding plane issues of multi-topology, although it is not specified in what scenarios segment routing and multi-topology routing need to be used together.

Although MTR can create multiple customized topologies in the network, it was not designed for resource reservation and isolation between different topologies. When some nodes or links belongs to multiple topologies, network resources on these nodes and links are shared by all those topologies. Thus it is not possible to provide tenants with isolation and guaranteed performance based on multi-topology routing.

It is well accepted that segment routing can provide traffic engineering (TE) with better scalability than RSVP-TE, however all that current SR can do is to create a set of non-shortest paths in the network, with network resource planning executed in the controller. Different service traffic can be mapped to different paths to achieve some service differentiation. Currently SR does not provide any mechanism for resource reservation and isolation in the network data plane, thus all network resources are shared by the services carried by the same set of links and nodes.

6. Proposed Mechanism

This section describes the proposed mechanisms to meet the isolation requirements of VPN+.

The overall solution is segment routing based, with necessary extensions to create multiple isolated logical networks using a common network infrastructure. Each logical network can have its own customized topology and guaranteed network resources. Hard isolation

can be achieved between different logical networks, so that services in different logical networks will not interfere with each other.

Segment Routing uses different types of Segment Identifiers (SIDs) to build the SR path, in which the adjacency SID and the node SID are the basic building blocks. Taking advantage of the SR architecture, with some proper extensions, each logical network can be constructed with a dedicated set of SIDs, each of which represents a subset of resource reserved on a specific link or node. Based on the SIDs with resource reservation, isolation between different logical networks can be achieved. Compared to the resource reservation using RSVP-TE, which is per end-to-end path based reservation, the SR based resource reservation is per-hop per-virtual- network based, which could significantly reduce the amount of states introduced to the network, thus can avoid the scalability problem of RSVP-TE.

6.1. Per-link/Node Resource Partitioning

To achieve resource reservation with SR, resources of the links and nodes needs to be partitioned into isolated pieces, so that different pieces of the node and link resources can be allocated to different logical networks independently. Normally a network controller is responsible for the collection of network resource information, and the computation of the subset of network resources needed on each node or link for a particular virtual network based on tenant's service requirement. The controller may also be used to trigger the allocation of network resources on the network equipments using some appropriate protocol.

Some enhancement in the data plane may be needed to meet the requirement of hard resource isolation. For example, FlexE [FLEXE] can provide time slot based link channelization, which could be used as one mechanism for link resource partitioning and reservation. Also there are efforts for resource partitioning inside the routing nodes. The mechanisms of link and node resource partitioning can be implementation specific, which are outside the scope of this document. While the capability and information about link and node resource partitioning needs to be advertised using some control protocol, so that different partitions of link and node resources can be used to set up different isolated networks.

When a link is partitioned into several pieces, information about each piece of the link needs to be advertised, so that there is no ambiguity about which particular piece of the link resource is allocated for which particular logical network. Using segment routing paradigm, each link partition needs to be assigned with a dedicated adj-SID. In order to ensure that a particular link partition would only be used for the path computation and data

forwarding of a particular logical network, each link partition needs to be associated with the unique identifier of the logical network.

[I-D.ietf-isis-l2bundles] specifies a mechanism to advertise the link attributes of the member links of a link bundle. Such mechanism can be reused for the link partitioning case described in this document, while some further extensions are needed to advertise the mapping between the link attributes and the Logical Network Identifiers (LN-ID).

Similarly, for node resource partitioning, different node-SIDs can be assigned for each partition of node resource. Different Node-SIDs are also needed for loose path forwarding of SR, service of different logical networks uses different node-SIDs of the same node to identify the logical networks it belongs to, so that the node could steer different service inside their own logical networks using the dedicated resource reserved for the logical network. In this way, network isolation can also be achieved with SR loose path forwarding.

6.2. Construction of Isolated Logical Networks

With the mechanism described in section 6.1, each link or node partition can be identified with a (SID, LN-ID) tuple, which associate the SID and the resource it represents with a particular logical network. Such information needs to be distributed both in the network and to the network controller using protocol extensions of IGP and BGP-LS, so that every node in the network and the controller obtain the same link-state information of different logical networks, so as to create the logical network topology using the subset of the adj-SIDs and node-SIDs which associate with the same LN-ID. From network resource perspective, each logical network is constructed with the cluster of reserved network resources the SIDs point to, and different logical networks are isolated from each other. The SR path computation of each logical network SHOULD be constrained to only use the SIDs belong to the logical network.

When service traffic is carried in a particular logical network, the data packet is encapsulated with a sequence of Adj-SIDs and Node-SIDs dedicated to this logical network, so that in forwarding plane the packet will be steered through the link and node resources allocated for those SIDs. Service of different logical networks always use different SIDs when traversing the same physical links or nodes. This ensures that service always use the network resources allocated for its logical network.

6.3. Mapping Service to Logical Network

Mapping of service to logical networks can be quite flexible. According to the different isolation requirements, one tenant who requires hard isolation can be mapped to a dedicated logical network, so that the network resource of the logical network are dedicated to this tenant. If this tenant have multiple services, the resources can be shared by all the services of this tenant, but the service performance will never be impacted by service of other tenants. Some service may require stringent performance in terms of bounded latency and packet loss, then mechanisms of [DETNET] may be applied to this service.

For tenants who only expect soft isolation and resource sharing or competing is allowed between these tenants, these tenants can be mapped to the same logical network for better economy and scalability. Service traffic of tenants which mapped to the same logical network may compete for some shared resources, but they will never impact another tenant who owns a separate logical network. According to the customized requirements, different group of tenants can be mapped to different logical networks.

7. Security Considerations

The security concerns about segment routing [I-D.ietf-spring-segment-routing] applies here.

8. IANA Considerations

There are no requested IANA actions.

9. References

9.1. Normative References

[I-D.bryant-rtgwg-enhanced-vpn]
Bryant, S. and J. Dong, "Enhanced Virtual Private Networks (VPN+)", draft-bryant-rtgwg-enhanced-vpn-00 (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, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P. Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF", RFC 4915, DOI 10.17487/RFC4915, June 2007, <https://www.rfc-editor.org/info/rfc4915>.

[RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, DOI 10.17487/RFC5120, February 2008, <https://www.rfc-editor.org/info/rfc5120>.

9.2. Informative References

[DETNET] "IETF Detnet Working Group", 2017, <https://datatracker.ietf.org/wg/detnet/>.

[FLEXE] "Flex Ethernet Implementation Agreement", March 2016, <http://www.oiforum.com/wp-content/uploads/OIF-FLEXE-01.0.pdf>.

[I-D.ietf-detnet-architecture] Finn, N., Thubert, P., Varga, B., and J. Farkas, "Deterministic Networking Architecture", draft-ietf-detnet-architecture-02 (work in progress), June 2017.

[I-D.ietf-isis-l2bundles] Ginsberg, L., Bashandy, A., Filsfils, C., Nanduri, M., and E. Aries, "Advertising L2 Bundle Member Link Attributes in IS-IS", draft-ietf-isis-l2bundles-07 (work in progress), May 2017.

[I-D.ietf-spring-segment-routing] Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-12 (work in progress), June 2017.

[I-D.ietf-teas-rsvp-te-scaling-rec] Beeram, V., Minei, I., Shakir, R., Pacella, D., and T. Saad, "Techniques to Improve the Scalability of RSVP Traffic Engineering Deployments", draft-ietf-teas-rsvp-te-scaling-rec-07 (work in progress), September 2017.

[I-D.sitaraman-mpls-rsvp-shared-labels] Sitaraman, H., Beeram, V., Parikh, T., and T. Saad, "Signaling RSVP-TE tunnels on a shared MPLS forwarding plane", draft-sitaraman-mpls-rsvp-shared-labels-02 (work in progress), September 2017.

[NVO3] "IETF NVO3 Working Group", 2017,
<https://datatracker.ietf.org/wg/nvo3/>.

[RFC3205] Moore, K., "On the use of HTTP as a Substrate", BCP 56,
RFC 3205, DOI 10.17487/RFC3205, February 2002,
<https://www.rfc-editor.org/info/rfc3205>.

Authors' Addresses

Jie Dong
Huawei Technologies

Email: jie.dong@huawei.com

Stewart Bryant
Huawei Technologies

Email: stewart.bryant@gmail.com