TEAS Working Group
Internet-Draft
Intended status: Informational
Expires: February 16, 2019

J. Dong
S. Bryant
Huawei
Z. Li
China Mobile
T. Miyasaka
KDDI Corporation

August 15, 2018

A Framework for Enhanced Virtual Private Networks (VPN+) draft-dong-teas-enhanced-vpn-01

Abstract

This document specifies a framework for using existing, modified and potential new networking technologies as components to provide an enhanced VPN (VPN+) service. The purpose is to enable virtual private networks (VPNs) to support the needs of new applications, particularly applications that are associated with 5G services. A network enhanced with these properties can form the underpinning of network slicing, but will also be of use in its own right.

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 February 16, 2019.

Copyright Notice

Table of Contents

1. Introduction

Virtual private networks (VPNs) have served the industry well as a means of providing different groups of users with logically isolated access to a common network. The common or base network that is used to provide the VPNs is often referred to as the underlay, and the VPN is often called an overlay.

Driven largely by needs surfacing from 5G, the concept of network slicing has gained traction. Network slicing requires the transport network to support a set of virtual networks, each of which can provide the client with dedicated (private) networking, computing and storage resources drawn from a shared pool. There is a need to create virtual networks with enhanced characteristics. The tenant of such a virtual network can require a degree of isolation and performance that previously could only be satisfied by dedicated networks. Additionally the tenant may ask for some level of control to their virtual network e.g. to customize the service paths in the network slice.

These properties cannot be met with pure overlay networks, as they require tighter coordination and integration between the underlay and the overlay network. This document introduces a new network service called enhanced VPN (VPN+). VPN+ refers to a virtual network which has dedicated network resources allocated from the underlay network. Unlike a traditional VPN, an enhanced VPN can achieve greater isolation and guaranteed performance. These new properties, which have general applicability, may also be of interest as part of a network slicing solution.

This document specifies a framework for using existing, modified and potential new networking technologies as components to provide an enhanced VPN (VPN+) service. Specifically we are concerned with:

- The design of the enhanced data plane
- The necessary protocols in both underlay and the overlay of enhanced VPN
- The mechanisms to achieve integration between overlay and underlay
- The necessary OAM methods to instrument an enhanced VPN to make sure that the required SLA are met, and to take any required action to avoid SLA violation, such as switching to an alternate path

The required network layered structure to achieve this is shown in Section 4.1.

One use for enhanced VPNs is to create network slices with different isolation requirements. Such network slices may be used to provide different tenants of vertical industrial markets with their own virtual network with the explicit characteristics required. These network slices may be "hard" slices providing a high degree of confidence that the VPN+ characteristics will be maintained over the slice life cycle, or they may be "soft" slices in which case some interaction may be experienced.

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. Overview of the Requirements

In this section we provide an overview of the requirements of an enhanced VPN.

3.1. Isolation between Virtual Networks

The requirement is to provide both hard and soft isolation between the tenants/applications using one enhanced VPN and the tenants/ applications using another enhanced VPN. Hard isolation is needed so that applications with exacting requirements can function correctly, despite a flash demand being created on another VPN competing for the underlying resources. An example of hard isolation is a network supporting both emergency services and public broadband multi-media services.

During a major incident the VPNs supporting these services would both be expected to experience high data volumes, and it is important that both make progress in the transmission of their data. In these circumstances the VPNs would require an appropriate degree of isolation to be able to continue to operate acceptably.

We introduce the terms hard and soft isolation to cover cases such as the above. A VPN has soft isolation if the traffic of one VPN cannot be inspected by the traffic of another. Both IP and MPLS VPNs are examples of soft isolated VPNs because the network delivers the traffic only to the required VPN endpoints. However, the traffic from one or more VPNs and regular network traffic may congest the network resulting in delays for other VPNs operating normally. The ability for a VPN to be sheltered from this effect is called hard isolation, and this property is required by some critical applications. Although these isolation requirements are triggered by

the needs of 5G networks, they have general utility. In the remainder of this section we explore how isolation may be achieved in packet networks.

In order to provide the required isolation, resources has to be reserved in the data plane. This may introduce scalability concerns, thus some trade-off needs to be considered to provide the required isolation between network slices while still allows reasonable sharing inside each network slice.

An optical layer can offer a high degree of isolation, at the cost of allocating resources on a long term and end-to-end basis. Such an arrangement means that the full cost of the resources must be borne by the service that is allocated with the resources. On the other hand, where adequate isolation can be achieved at the packet layer, this permits the resources to be shared amongst many services and only dedicated to a service on a temporary basis. This in turn, allows greater statistical multiplexing of network resources and thus amortizes the cost over many services, leading to better economy. However, the degree of isolation required by network slicing cannot be entirely met with existing mechanisms such as TE-LSPs. This is because most implementations enforce the bandwidth in the data-plane only at the PEs, but at the P routers the bandwidth is only reserved in the control plane, thus bursts of data can accidentally occur at a P router with higher than committed data rate.

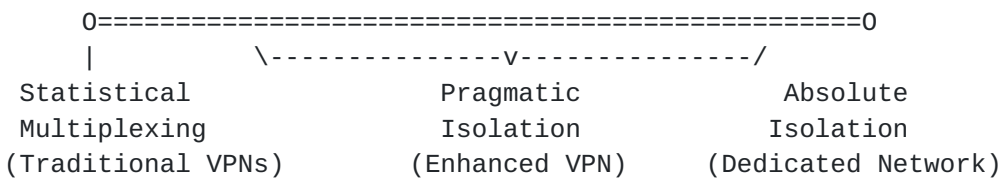
There are several new technologies that provide some assistance with these data plane issues. Firstly there is the IEEE project on Time Sensitive Networking [TSN] which introduces the concept of packet scheduling of delay and loss sensitive packets. Then there is [FLEXE] provides the ability to multiplex multiple channels over one or more Ethernet links in a way that provides hard isolation. Finally there are advanced queueing approaches which allow the construction of virtual sub-interfaces, each of which is provided with dedicated resource in a shared physical interface. These approaches are described in more detail later in this document.

3.1.1. A Pragmatic Approach to Isolation

A key question is whether it is possible to achieve hard isolation in packet networks, which were never designed to support hard isolation. On the contrary, they were designed to provide statistical multiplexing, a significant economic advantage when compared to a dedicated, or a Time Division Multiplexing (TDM) network. However there is no need to provide any harder isolation than is required by the application. Pseudowires [RFC3985] emulates services that would have had hard isolation in their native form. An approximation to this requirement is sufficient in most cases.

Thus, for example, using FlexE or a channelized sub-interface together with packet scheduling as interface slicing, optionally along with the slicing of node resources (Network Processor Unit (NPU), etc.), a type of hard isolation can be provided that is adequate for many VPN+ applications. Other applications may be either satisfied with a classical VPN with or without reserved bandwidth, or may need dedicated point to point fiber. The needs of each application must be quantified in order to provide an economic solution that satisfies those needs without over-engineering.

This spectrum of isolation is shown below:



At one end of the above figure, we have traditional statistical multiplexing technologies that support VPNs. This is a service type that has served the industry well and will continue to do so. At the opposite end of the spectrum we have the absolute isolation provided by traditional networks. The goal of enhanced VPN is pragmatic isolation. This is isolation that is better than is obtainable from pure statistical multiplexing, more cost effective and flexible than a dedicated network, but which is a practical solution that is good enough for the majority of applications.

3.2. Performance Guarantee

There are several kinds of performance guarantees, including guaranteed maximum packet loss, guaranteed maximum delay and guaranteed delay variation.

Guaranteed maximum packet loss is a common parameter, and is usually addressed by setting the packet priorities, queue size and discard policy. However this becomes more difficult when the requirement is combine with the latency requirement. The limiting case is zero congestion loss, and that is the goal of the Deterministic Networking work that the IETF [DETNET] and IEEE [TSN] are pursuing. In modern optical networks, loss due to transmission errors is already approaches zero, but there are the possibilities of failure of the interface or the fiber itself. This can only be addressed by some form of packet duplication and transmission over diverse paths.

Guaranteed maximum latency is required in a number of applications particularly real-time control applications and some types of virtual

reality applications. The work of the IETF Deterministic Networking (DetNet) Working Group [DETNET] is relevant; however the scope needs to be extended to methods of enhancing the underlay to better support the delay guarantee, and to integrate these enhancements with the overall service provision.

Guaranteed maximum delay variation is a service that may also be needed. [I-D.ietf-detnet-use-cases] calls up a number of cases where this is needed, for example electrical utilities have an operational need for this. Time transfer is one example of a service that needs this, although it is in the nature of time that the service might be delivered by the underlay as a shared service and not provided through different virtual networks. Alternatively a dedicated virtual network may be used to provide this as a shared service.

This suggests that a spectrum of service guarantee be considered when deploying an enhanced VPN. As a guide to understanding the design requirements we can consider four types:

- Best effort
- Assured bandwidth
- Guaranteed latency
- Enhanced delivery

Best effort is the service that current VPNs provide. An assured bandwidth service is one in which the bandwidth over some period of time is assured. The instantaneous bandwidth is however, not necessarily assured, depending on the technique used. Providing assured bandwidth to VPNs, for example by using TE-LSPs, is not widely deployed at least partially due to scalability concerns. Guaranteed latency and enhanced delivery are not yet integrated with VPNs.

A guaranteed latency service has a latency upper bound provided by the network. Assuring the upper bound is more important than achieving the minimum latency.

In Section 3.1 we considered the work of the IEEE Time Sensitive Networking (TSN) project [TSN] and the work of the IETF DetNet Working group [DETNET] in the context of isolation. The TSN and Detnet work is of greater relevance in assuring end-to-end packet latency. It is also of importance in considering enhanced delivery.

An enhanced delivery service is one in which the network (at layer 3) attempts to deliver the packet through multiple paths in the hope of eliminating packet loss due to equipment or media failures.

It is these last two characteristics that an enhanced VPN adds to a VPN service.

Flex Ethernet [FLEXE] is a useful underlay to provide these guarantees. This is a method of providing time-slot based channelization over an Ethernet bearer. Such channels are fully isolated from other channels running over the same Ethernet bearer. As noted elsewhere this produces hard isolation but makes the reclamation of unused bandwidth more difficult.

These approaches can be used in tandem. It is possible to use FlexE to provide tenant isolation, and then to use the TSN/Detnet approach to provide a performance guarantee inside the a slice or tenant VPN.

3.3. Integration

A solution to the enhanced VPN problem has to provide seamless integration of both overlay VPN and the underlay network resource. This needs be done in a flexible and scalable way so that it can be widely deployed in operator networks to support a reasonable number of enhanced VPN customers.

Taking mobile networks and in particular 5G into consideration, the integration of network and the service functions is a likely requirement. The work in IETF SFC working group [SFC] provides a foundation for this integration.

3.4. Dynamic Configuration

New enhanced VPNs need to be created, modified, and removed from the network according to service demand. An enhanced VPN that requires hard isolation must not be disrupted by the instantiation or modification of another enhanced VPN. Determining whether modification of an enhanced VPN can be disruptive to that VPN, and in particular the traffic in flight will be disrupted can be a difficult problem.

The data plane aspects of this problem are discussed further in Section 5.

The control-plane and management-plane aspects of this (particularly garbage collection) are for further study.

Dynamic changes both to the VPN and to the underlay transport network need to be managed in a seamless way in order to avoid disruption to sensitive services.

In addition to non-disruptively managing the network as a result of gross change such as the inclusion of a new VPN endpoint or a change to a link, VPN traffic might need to be moved as a result of traffic volume changes.

3.5. Customized Control

In some cases it is desirable that an enhanced VPN has a customized control plane, so that the tenant of the enhanced VPN can have some control to some of the resources and functions allocated to this VPN. Each enhanced VPN may have its own dedicated controller, it may be provided with an interface to a control plane that is shared with a set of other tenants, or it may be provided with an interface to the control plane of the underlay provided by the underlay network operator.

Further detail on this requirement will be provided in a future version of the draft.

3.6. Applicability

The technologies described in this document should be applicable to a number types of VPN services such as:

- Layer 2 point to point services such as pseudowires [RFC3985]
- Layer 2 VPNs [RFC4664]
- Ethernet VPNs [RFC7209]
- Layer 3 VPNs [RFC4364], [RFC2764]

Where such VPN types need enhanced isolation and delivery characteristics the technology described here can be used to provide an underlay with the required enhanced performance.

4. Architecture of Enhanced VPN

A number of enhanced VPN services will typically be provided by a common network infrastructure. Each enhanced VPN consists of both the overlay and a specific set of dedicated network resources and functions allocated in the underlay to satisfy the needs of the VPN tenant. The integration between overlay and various underlay

resources ensures the isolation between different enhanced VPNs, and achieves the guaranteed performance for different services.

An enhanced VPN needs to be designed with consideration given to:

- A suitable enhanced data plane
- A control plane to create enhanced VPN, making use of the data plane isolation and guarantee techniques
- A management plane for enhanced VPN service life-cycle management

These required characteristics are expanded below:

- Enhanced data plane
  - Provides the required resource isolation capability
  - Provides the required packet latency and jitter characteristics
  - Provides the required packet loss characteristics
  - Provides the mechanism to identify network slice and the associated resources
- Control plane
  - Collect the underlying network topology and resources available and export this to other nodes and/or the centralized controller as required.
  - Create the required set of virtual topologies with the resource and properties needed by the enhanced VPN services that are assigned to it.
  - Determine the risk of SLA violation and take appropriate avoiding action
  - Determine the right balance of per-packet and per-node state according to the needs of enhanced VPN service to scale to the required size
- Management plane
  - Provides the life-cycle management (creation, modification, decommissioning) of enhanced VPN

* Provide a interface between the enhanced VPN provider and the enhanced VPN clients such that some of the operation requests can be met without interfering other enhanced VPN clients.

This document will focus on the data plane and control plane of the enhanced VPN. The details of the management plane is outside the scope of this document.

4.1. Layered Architecture

The layered architecture of enhanced VPN is shown in Figure 1.

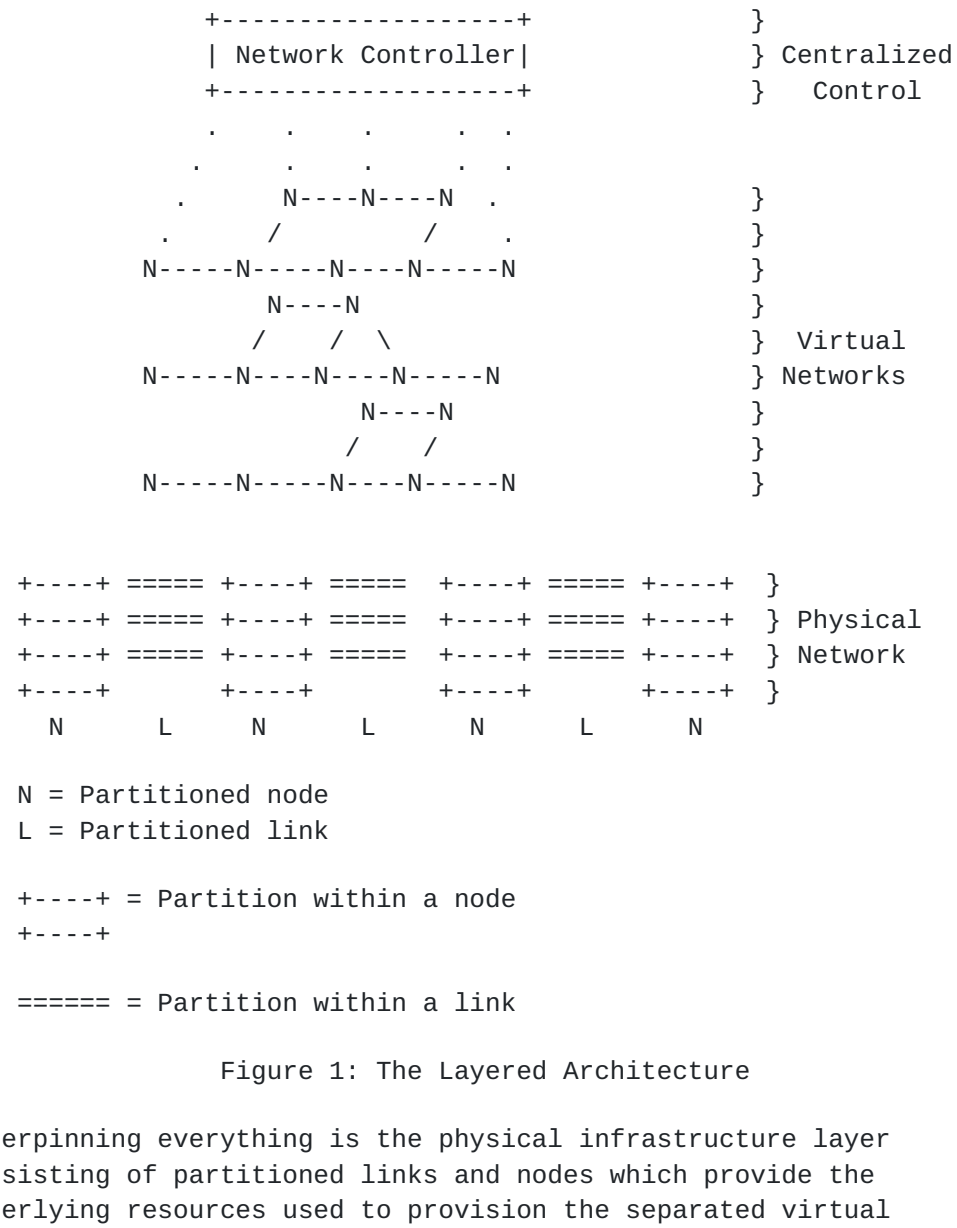


Figure 1: The Layered Architecture

Underpinning everything is the physical infrastructure layer consisting of partitioned links and nodes which provide the underlying resources used to provision the separated virtual networks. Various components and techniques as discussed in

Section 5 can be used to provide the resource partition, such as FlexE, Time Sensitive Networking, Deterministic Networking, etc. These partitions may be physical, or virtual so long as the SLA required by the higher layers is met.

These techniques can be used to provision the virtual networks with dedicated resources that they need. To get the required functionality there needs to be integration between these overlays and the underlay providing the physical resources.

The centralized controller is used to create the virtual networks, to allocate the resources to each virtual network and to provision the enhanced VPN services within the virtual networks. A distributed control plane may also be used for the distribution of the topology and attribute information of the virtual networks.

The creation and allocation process needs to take a holistic view of the needs of all of its tenants, and to partition the resources accordingly. However within a virtual network these resources can if required be managed via a dynamic control plane. This provides the required scalability and isolation.

4.2. Multi-Point to Multi-Point

At a VPN service level, the connectivity are usually multi-point-to-multi-point (MP2MP). For such kind of services, the corresponding underlay is also an abstract MP2MP medium. However when service guarantees are provided, the point-to-point path through the underlay of the enhanced VPN needs to be specifically engineered to meet the required performance guarantees.

4.3. Application Specific Network Types

Although a lot of the traffic that will be carried over the enhanced VPN will likely be IPv4 or IPv6, the design has to be capable of carrying other traffic types, in particular the Ethernet traffic. This is easily accomplished through the various pseudowire (PW) techniques [RFC3985]. Where the underlay is MPLS, Ethernet can be carried over the enhanced VPN encapsulated according to the method specified in [RFC4448]. Where the underlay is IP, Layer Two Tunneling Protocol - Version 3 (L2TPv3) [RFC3931] can be used with Ethernet traffic carried according to [RFC4719]. Encapsulations have been defined for most of the common layer two type for both PW over MPLS and for L2TPv3.

5. Candidate Technologies

A VPN is a network created by applying a multiplexing technique to the underlying network (the underlay) in order to distinguish the traffic of one VPN from that of another. A VPN path that travels by other than the shortest path through the underlay normally requires state in the underlay to specify that path. State is normally applied to the underlay through the use of the RSVP Signaling protocol, or directly through the use of an SDN controller, although other techniques may emerge as this problem is studied. This state gets harder to manage as the number of VPN paths increases. Furthermore, as we increase the coupling between the underlay and the overlay to support the enhanced VPN service, this state will increase further.

In an enhanced VPN different subsets of the underlay resources are dedicated to different enhanced VPNs. Any enhanced VPN solution thus needs tighter coupling with underlay than is the case with existing VPNs. We cannot for example share the tunnel between enhanced VPNs which require hard isolation.

A number of candidate underlay data plane solutions which can be used provide the required isolation and guarantee are described in following sections.

- FlexE
- Time Sensitive Networking
- Dedicated Queues

We then consider the problem of slice differentiation and resource representation in the network layer. The candidate technologies are:

- MPLS
- MPLS-SR
- Segment Routing over IPv6 (SRv6)
- Deterministic Networking

The considerations about the control plane is also described.

5.1. Underlay Data plane

5.1.1. FlexE

FlexE [FLEXE] is a method of creating a point-to-point Ethernet with a specific fixed bandwidth. FlexE provides the ability to multiplex multiple channels over an Ethernet link in a way that provides hard isolation. FlexE also supports the bonding of multiple links, which can be used to create larger links out of multiple slower links in a more efficient way that traditional link aggregation. FlexE also supports the sub-rating of links, which allows an operator to only use a portion of a link. However it is a only a link level technology. When packets are received by the downstream node, they need to be processed in a way that preserves that isolation in the downstream node. This in turn requires a queuing and forwarding implementation that preserves the end-to-end isolation.

If different FlexE channels are used for different services, then no sharing is possible between the FlexE channels. This in turn means that it may be difficult to dynamically redistribute unused bandwidth to lower priority services. This may increase the cost of providing services on the network. On the other hand, FlexE can be used to provide hard isolation between different tenants on a shared interface. The tenant can then use other methods to manage the relative priority of their own traffic in each FlexE channel.

Methods of dynamically re-sizing FlexE channels and the implication for enhanced VPN are under study.

5.1.2. Dedicated Queues

In order to provide multiple isolated virtual networks for enhanced VPN, the conventional Diff-Serv based queuing system is insufficient, due to the limited number of queues which cannot differentiate between traffic of different enhanced VPNs, and the range of service classes that each need to provide to their tenants. This problem is particularly acute with an MPLS underlay due to the small number of traffic class services available. In order to address this problem and reduce the interference between enhanced VPNs, it is necessary to steer traffic of VPNs to dedicated input and output queues. Routers usually have large amount of queues and sophisticated queuing systems, which could be used or enhanced to provide the levels of isolation required by the applications of enhanced VPN. For example, on one physical interface, the queuing system can provide a set of virtual sub-interfaces, each allocated with dedicated queueing and buffer resources. Sophisticated queuing systems of this type may be used to provide end-to-end virtual isolation between traffic of different enhanced VPNs.

5.1.3. Time Sensitive Networking

Time Sensitive Networking (TSN) [TSN] is an IEEE project that is designing a method of carrying time sensitive information over Ethernet. It introduces the concept of packet scheduling where a high priority packet stream may be given a scheduled time slot thereby guaranteeing that it experiences no queuing delay and hence a reduced latency. However where no scheduled packet arrives its reserved time-slot is handed over to best effort traffic, thereby improving the economics of the network. The mechanisms defined in TSN can be used to meet the requirements of time sensitive services of an enhanced VPN.

Ethernet can be emulated over a Layer 3 network using a pseudowire. However the TSN payload would be opaque to the underlay and thus not treated specifically as time sensitive data. The preferred method of carrying TSN over a layer 3 network is through the use of deterministic networking as explained in the following section of this document.

5.2. Network Layer

5.2.1. Deterministic Networking

Deterministic Networking (DetNet) [I-D.ietf-detnet-architecture] is a technique being developed in the IETF to enhance the ability of layer 3 networks to deliver packets more reliably and with greater control over the delay. The design cannot use re-transmission techniques such as TCP since that can exceed the delay tolerated by the applications. Even the delay improvements that are achieved with SCTP-PR [RFC3758] do not meet the bounds set by application demands. Detnet pre-emptively sends copies of the packet over various paths to minimize the chance of all packets being lost, and trim duplicate packets to prevent excessive flooding of the network and to prevent multiple packets being delivered to the destination. It also seeks to set an upper bound on latency. The goal is not to minimize latency; the optimum upper bound paths may not be the minimum latency paths.

DetNet is based on flows. It currently does not specify the use of underlay topology other than the base topology. To be of use for enhanced VPN, DetNet needs to be integrated with different virtual topologies of enhanced VPNs.

The detailed design that allows the use DetNet in a multi-tenant network, and how to improve the scalability of DetNet in a multi-tenant network are topics for further study.

5.2.2. MPLS Traffic Engineering (MPLS-TE)

MPLS-TE introduces the concept of reserving end-to-end bandwidth for an TE-LSP, which can be used as the underlay of VPNs. It also introduces the concept of non-shortest path routing through the use of the Explicit Route Object [RFC3209]. VPN traffic can be run over dedicated TE-LSPs to provide reserved bandwidth for each specific connection in a VPN. This is not widely deployed in practice due to scaling and management overhead concerns.

5.2.3. Segment Routing

Segment Routing [RFC8402] is a method that prepends instructions to packets at the head-end node and optionally at various points as it passes though the network. These instructions allow the packets to be routed on paths other than the shortest path for various traffic engineering reasons. These paths can be strict or loose paths, depending on the compactness required of the instruction list and the degree of autonomy granted to the network, for example to support Equal Cost Multipath load-balancing (ECMP) [RFC2992].

With SR, a path needs to be dynamically created through a set of segments by simply specifying the Segment Identifiers (SIDs), i.e. instructions rooted at a particular point in the network. Thus if a path is to be provisioned from some ingress point A to some egress point B in the underlay, A is provided with the A..B SID list and instructions on how to identify the packets to which the SID list is to be prepended.

By encoding the state in the packet, as is done in Segment Routing, per-path state is transitioned out of the network.

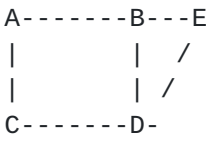


Figure 2: An SR Network Fragment

Consider a further network fragment shown in Figure 2. To send a packet from A to E via B and D: Node A prepends the ordered SID list {B, D, E} to the packet and sends the packet to B. SID list {B, D, E} can be used as a VPN path. Thus, to create a VPN, a set of SID Lists is created and provided to each ingress node of the VPN together with packet selection criteria. In this way it is possible to create a VPN with no state in the core. However this is at the expense of creating a larger packet with possible MTU and hardware restriction limits that need to be overcome.

Note in the above if A and E support multiple VPN an additional VPN identifier will need to be added to the packet, but this is omitted from this text for simplicity.

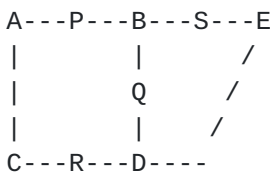


Figure 3: Another SR Network Fragment

Consider a further network fragment shown in Figure 3, and further consider VPN A+D+E (i.e. the VPN connecting together nodes A, D and E with the requires VPN properties.) This requires the nodes to be configured with, or otherwise learn the following path lists to provide complete connectivity:

A has lists: {P, B, Q, D}, {P, B, S, E}
D has lists: {Q, B, P, A}, {E}
E has lists: {S, B, P, A}, {D}

Similarly, to create a new VPN C+D+B the following list are introduced:

C lists: {R, D}, {A, P, B}
D lists: {R, C}, {Q, B}
B lists: {Q, D}, {P, A, C}

Thus VPN C+D+B was created without touching the settings of the core routers, indeed it is possible to add endpoints to the VPNs, and move the paths around simply by providing new lists to the affected endpoints.

However, there are a number of limitations in current SR, which limit its applicability to enhanced VPNs:

- Segments are shared between different VPNs paths

- There is no reservation of bandwidth

- There is limited differentiation in the data plane.

Thus some extensions to SR are needed to provide isolation between different enhanced VPNs. This can be achieved by including a finer granularity of state in the network in anticipation of its future use by authorized services. We therefore need to evaluate the balance

between this additional state and the performance delivered by the network.

With current segment routing, the instructions are used to specify the nodes and links to be traversed. However, in order to achieve the required isolation between different services, new instructions can be created which can be prepended to a packet to steer it through specific network resources and functions.

Traditionally an SR traffic engineered path operates with a granularity of a link with hints about priority provided through the use of the traffic class (TC) field in the header. However to achieve the latency and isolation characteristics that are sought by the enhanced VPN users, steering packets through specific queues and resources will likely be required. The extent to which these needs can be satisfied through existing QoS mechanisms is to be determined. What is clear is that a fine control of which services wait for which, with a fine granularity of queue management policy is needed. Note that the concept of a queue is a useful abstraction for many types of underlay mechanism that may be used to provide enhanced isolation and latency support.

From the perspective of the control plane, and from the perspective of the segment routing, the method of steering a packet to a queue that provides the required properties is an abstraction that hides the details of the underlying implementation. How the queue satisfies the requirement is implementation specific and is transparent to the control plane and data plane mechanisms used. Thus, for example, a FlexE channel, or a time sensitive networking packet scheduling slot are abstracted to the same concept and bound to the data plane in a common manner.

We can also introduce such fine grained packet steering by specifying the queues through an SR instruction list. Thus new SR instructions may be created to specify not only which resources are traversed, but in some cases how they are traversed. For example, it may be possible to specify not only the queue to be used but the policy to be applied when enqueuing and dequeuing.

This concept can be further generalized, since as well as queuing to the output port of a router, it is possible to queue to any resource, for example:

- A network processor unit (NPU)
- A central processing unit (CPU) Core
- A Look-up engine

Both SR-MPLS and SRv6 are candidate network layer technologies for enhanced VPN. In some cases they can be supported by DetNet to meet the packet loss, delay and jitter requirement of particular service. However, currently the "pure" IP variant of DetNet [I-D.ietf-detnet-dp-sol-ip] does not support the Packet Replication, Elimination, and Re-ordering (PREOF) [I-D.ietf-detnet-architecture] functions. How to provide the DetNet enhanced delivery in an SRv6 environment needs further study.

5.3. Control Plane

Enhanced VPN would likely be based on a hybrid control mechanism, which takes advantage of the logically centralized controller for on-demand provisioning and global optimization, whilst still relies on distributed control plane to provide scalability, high reliability, fast reaction, automatic failure recovery etc. Extension and optimization to the distributed control plane is needed to support the enhanced properties of VPN+.

RSVP-TE provides the signaling mechanism of establishing a TE-LSP with end-to-end resource reservation. It can be used to bind the VPN to specific network resource allocated within the underlay, but there are the above mentioned scalability concerns.

SR does not have the capability of signaling the resource reservation along the path, nor do its currently specified distributed link state routing protocols. On the other hand, the SR approach provides a way of efficiently binding the network underlay and the enhanced VPN overlay, as it reduce the amount of state to be maintained in the network. An SR-based approach with per-slice resource reservation can easily create dedicated SR network slices, and the VPN can be bound to a particular SR network slice. A centralized controller can perform resource planning and reservation from the controller's point of view, but this cannot ensure resource reservation is actually done in the network nodes. Thus, if a distributed control plane is needed, either in place of an SDN controller or as an assistant to it, the design of the control system needs to ensure that resources are uniquely allocated in the network nodes for the correct service, and not allocated to multiple services causing unintended resource conflict.

Abstraction and Control of Traffic Engineered Networks (ACTN) [I-D.ietf-teas-actn-framework] specifies the SDN based architecture for the control of TE networks. The ACTN approach can be applicable to the provisioning of enhanced VPN service. The details are described in [I-D.lee-rtgwg-actn-applicability-enhanced-vpn].

6. Scalability Considerations

Enhanced VPN provides the performance guaranteed services in packet networks, with the cost of introducing necessary additional states into the network. There are at least three ways of adding the state needed for VPN+:

- Introduce the complete state into the packet, as is done in SR. This allows the controller to specify the detailed series of forwarding and processing instructions for the packet as it transits the network. The cost of this is an increase in the packet header size. The cost is also that systems will have capabilities enabled in case they are called upon by a service. This is a type of latent state, and increases as we more precisely specify the path and resources that need to be exclusively available to a VPN.

- Introduce the state to the network. This is normally done by creating a path using RSVP-TE, which can be extended to introduce any element that needs to be specified along the path, for example explicitly specifying queuing policy. It is of course possible to use other methods to introduce path state, such as via a Software Defined Network (SDN) controller, or possibly by modifying a routing protocol. With this approach there is state per path per path characteristic that needs to be maintained over its life-cycle. This is more state than is needed using SR, but the packet are shorter.

- Provide a hybrid approach based on using binding SIDs to create path fragments, and bind them together with SR.

Dynamic creation of a VPN path using SR requires less state maintenance in the network core at the expense of larger VPN headers on the packet. The packet size can be lower if a form of loose source routing is used (using a few nodal SIDs), and it will be lower if no specific functions or resource on the routers are specified. Reducing the state in the network is important to enhanced VPN, as it requires the overlay to be more closely integrated with the underlay than with traditional VPNs. This tighter coupling would normally mean that more state needed to be created and maintained in the network, as the state about fine granularity processing would need to be loaded and maintained in the routers. However, a segment routed approach allows much of this state to be spread amongst the network ingress nodes, and transiently carried in the packets as SIDs.

These approaches are for further study.

6.1. Maximum Stack Depth of SR

One of the challenges with SR is the stack depth that nodes are able to impose on packets [I-D.ietf-isis-segment-routing-msd]. This leads to a difficult balance between adding state to the network and minimizing stack depth, or minimizing state and increasing the stack depth.

6.2. RSVP Scalability

The traditional method of creating a resource allocated path through an MPLS network is to use the RSVP protocol. However there have been concerns that this requires significant continuous state maintenance in the network. There are ongoing works to improve the scalability of RSVP-TE LSPs in the control plane [RFC8370].

There is also concern at the scalability of the forwarder footprint of RSVP as the number of paths through an LSR grows [I-D.sitaraman-mpls-rsvp-shared-labels] proposes to address this by employing SR within a tunnel established by RSVP-TE.

7. OAM Considerations

A study of OAM in SR networks has been documented in [RFC8403].

The enhanced VPN OAM design needs to consider the following requirements:

- Instrumentation of the underlay so that the network operator can be sure that the resources committed to a tenant are operating correctly and delivering the required performance.

- Instrumentation of the overlay by the tenant. This is likely to be transparent to the network operator and to use existing methods. Particular consideration needs to be given to the need to verify the isolation and the various committed performance characteristics.

- Instrumentation of the overlay by the network provider to proactively demonstrate that the committed performance is being delivered. This needs to be done in a non-intrusive manner, particularly when the tenant is deploying a performance sensitive application

- Verification of the conformity of the path to the service requirement. This may need to be done as part of a commissioning test.

These issues will be discussed in a future version of this document.

8. Enhanced Resiliency

Each enhanced VPN has a life-cycle, and needs modification during deployment as the needs of its tenant change. Additionally, as the network as a whole evolves, there will need to be garbage collection performed to consolidate resources into usable quanta.

Systems in which the path is imposed such as SR, or some form of explicit routing tend to do well in these applications, because it is possible to perform an atomic transition from one path to another. This is a single action by the head-end changes the path without the need for coordinated action by the routers along the path. However, implementations and the monitoring protocols need to make sure that the new path is up and meet the required SLA before traffic is transitioned to it. It is possible for deadlocks arise as a result of the network becoming fragmented over time, such that it is impossible to create a new path or modify a existing path without impacting the SLA of other paths. Resolution of this situation is as much a commercial issue as it is a technical issue and is outside the scope of this document.

There are however two manifestations of the latency problem that are for further study in any of these approaches:

- The problem of packets overtaking one and other if a path latency reduces during a transition.

- The problem of the latency transient in either direction as a path migrates.

There is also the matter of what happens during failure in the underlay infrastructure. Fast reroute is one approach, but that still produces a transient loss with a normal goal of rectifying this within 50ms [RFC5654] . An alternative is some form of N+1 delivery such as has been used for many years to support protection from service disruption. This may be taken to a different level using the techniques proposed by the IETF deterministic network work with multiple in-network replication and the culling of later packets [I-D.ietf-detnet-architecture].

In addition to the approach used to protect high priority packets, consideration has to be given to the impact of best effort traffic on the high priority packets during a transient. Specifically if a conventional re-convergence process is used there will inevitably be micro-loops and whilst some form of explicit routing will protect the high priority traffic, lower priority traffic on best effort shortest

paths will micro-loop without the use of a loop prevention technology. To provide the highest quality of service to high priority traffic, either this traffic must be shielded from the micro-loops, or micro-loops must be prevented.

9. Security Considerations

All types of virtual network require special consideration to be given to the isolation between the tenants. In this regard enhanced VPNs neither introduce, no experience a greater security risk than another VPN of the same base type. However, in an enhanced virtual network service the isolation requirement needs to be considered. If a service requires a specific latency then it can be damaged by simply delaying the packet through the activities of another tenant. In a network with virtual functions, depriving a function used by another tenant of compute resources can be just as damaging as delaying transmission of a packet in the network. The measures to address these dynamic security risks must be specified as part to the specific solution.

10. IANA Considerations

There are no requested IANA actions.

11. References

11.1. Normative References

[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>.

11.2. Informative References

[DETNET] "Deterministic Networking", March , <https://datatracker.ietf.org/wg/detnet/about/>.

[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-07 (work in progress), August 2018.

[I-D.ietf-detnet-dp-sol-ip]
Korhonen, J. and B. Varga, "DetNet IP Data Plane Encapsulation", draft-ietf-detnet-dp-sol-ip-00 (work in progress), July 2018.

[I-D.ietf-detnet-use-cases]
Grossman, E., "Deterministic Networking Use Cases", draft-ietf-detnet-use-cases-17 (work in progress), June 2018.

[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-13 (work in progress), July 2018.

[I-D.ietf-teas-actn-framework]
Ceccarelli, D. and Y. Lee, "Framework for Abstraction and Control of Traffic Engineered Networks", draft-ietf-teas-actn-framework-15 (work in progress), May 2018.

[I-D.lee-rtgwg-actn-applicability-enhanced-vpn]
King, D., Lee, Y., Tantsura, J., Wu, Q., and D. Ceccarelli, "Applicability of Abstraction and Control of Traffic Engineered Networks (ACTN) to Enhanced VPN", draft-lee-rtgwg-actn-applicability-enhanced-vpn-03 (work in progress), July 2018.

[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-03 (work in progress), December 2017.

[RFC2764] Gleeson, B., Lin, A., Heinanen, J., Armitage, G., and A. Malis, "A Framework for IP Based Virtual Private Networks", RFC 2764, DOI 10.17487/RFC2764, February 2000, <https://www.rfc-editor.org/info/rfc2764>.

[RFC2992] Hopps, C., "Analysis of an Equal-Cost Multi-Path Algorithm", RFC 2992, DOI 10.17487/RFC2992, November 2000, <https://www.rfc-editor.org/info/rfc2992>.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <https://www.rfc-editor.org/info/rfc3209>.

[RFC3758] Stewart, R., Ramalho, M., Xie, Q., Tuexen, M., and P. Conrad, "Stream Control Transmission Protocol (SCTP) Partial Reliability Extension", RFC 3758, DOI 10.17487/RFC3758, May 2004, <https://www.rfc-editor.org/info/rfc3758>.

[RFC3931] Lau, J., Ed., Townsley, M., Ed., and I. Goyret, Ed., "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, DOI 10.17487/RFC3931, March 2005, <https://www.rfc-editor.org/info/rfc3931>.

[RFC3985] Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, DOI 10.17487/RFC3985, March 2005, <https://www.rfc-editor.org/info/rfc3985>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <https://www.rfc-editor.org/info/rfc4364>.

[RFC4448] Martini, L., Ed., Rosen, E., El-Aawar, N., and G. Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks", RFC 4448, DOI 10.17487/RFC4448, April 2006, <https://www.rfc-editor.org/info/rfc4448>.

[RFC4664] Andersson, L., Ed. and E. Rosen, Ed., "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, DOI 10.17487/RFC4664, September 2006, <https://www.rfc-editor.org/info/rfc4664>.

[RFC4719] Aggarwal, R., Ed., Townsley, M., Ed., and M. Dos Santos, Ed., "Transport of Ethernet Frames over Layer 2 Tunneling Protocol Version 3 (L2TPv3)", RFC 4719, DOI 10.17487/RFC4719, November 2006, <https://www.rfc-editor.org/info/rfc4719>.

[RFC5654] Niven-Jenkins, B., Ed., Brungard, D., Ed., Betts, M., Ed., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, DOI 10.17487/RFC5654, September 2009, <https://www.rfc-editor.org/info/rfc5654>.

[RFC7209] Sajassi, A., Aggarwal, R., Uttaro, J., Bitar, N., Henderickx, W., and A. Isaac, "Requirements for Ethernet VPN (EVPN)", RFC 7209, DOI 10.17487/RFC7209, May 2014, <https://www.rfc-editor.org/info/rfc7209>.

[RFC8370] Beeram, V., Ed., Minei, I., Shakir, R., Pacella, D., and T. Saad, "Techniques to Improve the Scalability of RSVP-TE Deployments", RFC 8370, DOI 10.17487/RFC8370, May 2018, <https://www.rfc-editor.org/info/rfc8370>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8403] Geib, R., Ed., Filsfils, C., Pignataro, C., Ed., and N. Kumar, "A Scalable and Topology-Aware MPLS Data-Plane Monitoring System", RFC 8403, DOI 10.17487/RFC8403, July 2018, <https://www.rfc-editor.org/info/rfc8403>.

[SFC] "Deterministic Networking", March , <https://datatracker.ietf.org/wg/sfc/about>.

[TSN] "Time-Sensitive Networking", March , <https://1.ieee802.org/tsn/>.

Authors' Addresses

Jie Dong
Huawei

Email: jie.dong@huawei.com

Stewart Bryant
Huawei

Email: stewart.bryant@gmail.com

Zhenqiang Li
China Mobile

Email: lizhenqiang@chinamobile.com

Takuya Miyasaka
KDDI Corporation

Email: ta-miyasaka@kddi.com