Network Working Group
Internet-Draft
Intended status: BCP

L. Yong
L. Dunbar
Huawei
T. Herbert
Facebook

Expires: April 2017 October 31, 2016

Interconnecting Network Sites by IP Tunnels
draft-yong-intarea-inter-sites-over-tunnels-00.txt

Abstract

This document specifies use of a set of IP tunnels to interconnect multiple network sites over IP backbone networks. The interconnected network sites form 'virtual' private network(s). The networks at any of those sites can be Layer 2 domains or/and Layer 3 subnets. The IP backbone networks that the tunnels traverse through can be IPv4 or IPv6. This document describes the special property (or features) that those IP tunnels need to have in order to interconnect multiple sites as if those sites are directly connected by wires.

Status of This Document

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 http://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 April 31, 2017.

Table of Contents

1. Introduction

This document specifies use of a set of IP tunnels to interconnect multiple network sites over IP backbone networks. The networks at any of those sites can be Layer 2 domains or Layer 3 subnets; IP backbone networks that tunnels traverse can be IPv4 or IPv6. This document describes the properties (or features) that IP tunnels need to have in order to interconnect multiple sites as if those sites are directly connected by wires.

An IP tunnel in this document is identified by a pair of IP addresses (IPv4 or IPv6) that IP backbone networks can reach or the IP addresses plus a pair of UDP ports; one address per each tunnel endpoint. Tunnel ingress receives network traffic from its site (or access ports) and will encapsulate the traffic with an outer header whose destination and source address are the tunnel's two end points respectively. Tunnel egress receives IP packets from the backbone network, decapsulates the IP packets, and forwards decapsulated traffic toward its site (or an access port).

Section 1.1 describes the motivation of this specification. Section 3 describes the solution architecture for sites interconnection by IP tunnels. Section 4 specifies the sites interconnection requirements for the IP tunnels. The rest of Sections describe Site Interconnection Tunnel (SITE) solution. Section 11 describes operational considerations for sites interconnection over IP tunnels.

Note: The site interconnection policy configuration and the prefix routing within a site and across sites are outside the scope of this document.

1.1. Motivation of sites interconnection by IP tunnels

Tunneling technology has being widely used in IP networks. For examples: transporting packets in one address family (e.g. IPv6) over a network of different address family (e.g. IPv4) [RFC7059]; establishing a direct logical "link" for traffic engineering; traffic security over the Internet (e.g. IPsec tunnel [RFC5996] [RFC3884]); or Location and Identifier Separation application (LISP) [RFC6830].

Provider based L2VPN and L3VPN solutions [RFC4762] [RFC4364] use MPLS LSP tunnel technology to interconnect provider edge devices, i.e., Provider Edge (PE). In these solutions, PEs are part of provider backbone networks that implement the VPN solutions. For a customer to get a provider based VPN service, each customer site

must attach to at least one of the PEs in the provider backbone networks, this attachment is called attachment circuit (AC) [RFC4664] [RFC4364].

Today, a company can use IPSec tunnels [RFC5996][RFC3884] to interconnect its IP subnets at different sites over the Internet if the company has an experienced operator to configure its site networks and the devices that support the IPsec tunnel capability. To achieve this, each site needs to have at least one device attaching to an IP backbone network and have at least one public IP address that the IP backbone networks can reach to. As a result, the interconnected sites form one "virtual" private network among the sites. In this case, an ISP provider (i.e. IP backbone provider) does not provide the sites interconnection service to the company although it carries the IPsec tunnel traffic; in other words, the ISP provider only provides Internet access service to each site. As a result, there is no guaranteed QoS between a pair of sites, which is the difference from provider based VPN solutions.

Sites interconnection by IP tunnels is attractive to some companies that operate at multiple locations. Some companies wish to do it but do not have such an experienced operator to construct/configure site networks and tunnels to achieve this. A vendor may make a product to achieve this and sell to these companies. The product includes a site gateway device or component (called S-GW in this document) and software to allow a customer to specify their site interconnection requirements including sites interconnection policies. The software will manage the configuration of the S-GWs at multiple locations; furthermore the software can automate the processes from client specifying the sites interconnection to the sites interconnection completion. Section 3 highlights this solution architecture.

A SP provider may use this product and integrate it with its provider based VPN solution [RFC4364][RFC4762][RFC7432] to provide agile VPN services to its customers. [Dunbar]

2. Terminology

2.1. Requirements language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2.2. Terms defined in this document

Site: A place that contains switches, routers, services, appliances and these devices are configured to form L2 domain (s) or L3 domain. For examples, an Enterprise company data center, a college campus network center.

SITE: Site Interconnection Tunnel Protocol Solution

More coming

3. Sites interconnection solution architecture

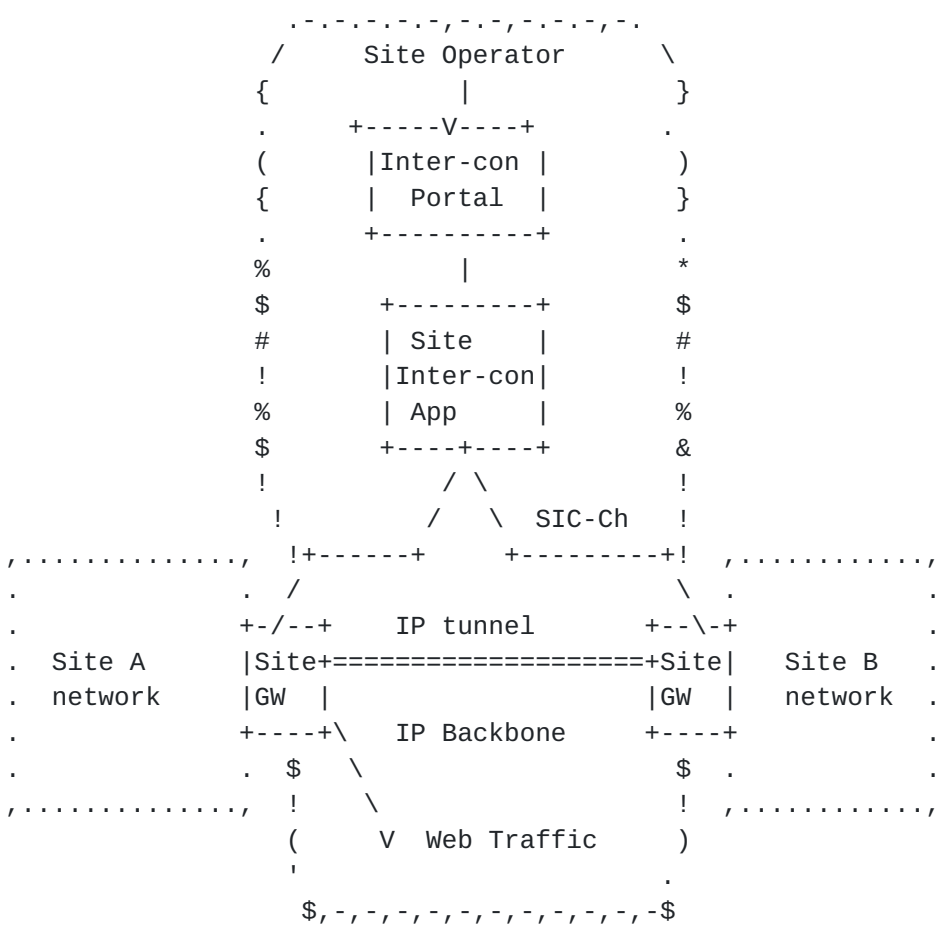

Figure 1 : Sites Interconnection Solution Architecture View

Figure 1 illustrates the solution architecture of sites interconnection. The architecture contains the following components:

- Site Gateway (S-GW): A router and/or switch device/component. It can be configured to perform routing/forwarding function at the site. S-GW attaches to an IP backbone network and support a tunnel capability specified in this document. S-GW can be configured with one IP address to be used as tunnel end point for all tunnels to other sites or be configured with one IP address per a tunnel to a remote site.

- Sites Interconnection Controller (SIC): A software component that controls/manages individual S-GWs via a SIC channel. It receives a sites interconnection request from the SIP; processes it, and sends the configuration data to individual S-GWs via SIC channel.

- Sites Interconnection Portal: A software with GUI interface to allow site operator to specify its sites interconnection requirement including sites interconnection policy. The same portal can be used for the site operator to specify the site Internet access policies.

- SIC Channel: a communication channel between S-GW and SIC over IP network. It could be a TCP application with security capability. The channel is used for SIC to send S-GW configuration data and get the PM data from S-GW. It may be used for dynamic routing purpose among the sites; in this case, SIC gets the prefix information from the S-GW at a site, and send the information to the S-GWs at other sites where the S-GW will distribute the information to the site.

- IP tunnel: A tunnel exists between a pair of S-GWs. The tunnel end point as an interface on a S-GW receives network traffic and encapsulate the traffic with an outer header whose Destination and Source address are the tunnel's two end points respectively. Tunnel egress receives IP packets from the IP backbone network, decapsulates the IP packets, and forwards decapsulated packets toward its site (or one access port(s)). IP tunnel is unidirectional, two sites interconnection requires two tunnels, one for each direction. Two tunnels may traverse different paths in backbone network.

In this architecture, each site uses its S-GW to attach to IP backbone network, which implies that the site traffic may go to the Internet via the S-GW. For security and performance reason, it is RECOMMENDED to use different public IP addresses for the Internet access and the tunnel end point of sites interconnection. It is possible to use more than one S-GW at a site for sites interconnection.

In this architecture, a site may have L2 domains, L3 subnets, or TRILL networks. However, the interconnected site networks MUST have the same type (see Section 3.1).

Note: This document will not specify the whole solution for this architecture. It only specifies the tunnel end point functions at a S-GW and tunnel end point configuration at the S-GW. Other functions in this solution architecture will be addressed in other documents.

4. Key properties of sites interconnection over IP tunnels

This section only lists the requirements for sites interconnection over IP tunnels. The entire solution architecture requirements are beyond the scope of this document.

4.1. Network sites interconnection properties

1. A site in this context may have a single network or multiple networks. For single network case, the network may be an L2 domain, L3 subnets, or a TRILL network. For the case of multiple networks, each network may be an L2 domain or L3 subnets and individual network traffic is segregated within the site including on the S-GW. Sites interconnection MUST NOT interconnect the networks at two sites that have different type. For example, one is L3 subnet, another is L2 domain. In the case of multiple networks at a site, sites interconnection MUST support individual networks at the site to be connected to the same type of networks that reside in either a remote site or different remote sites.

2. The S-GW MUST support site Internet access. The traffic to the Internet and the traffic to a remote site SHOULD be segregated by different S-GW physical ports or different IP addresses to avoid DDoS attack [RFC4732].

3. Tunnels for sites interconnection MUST secure site traffic over the IP backbone network.

4. Tunnels for sites interconnection MUST segregate the site traffic from different networks.

5. Sites interconnection topology may be mesh or hub-spoke.

6. Tunnel for sites interconnection MUST be able to detect congestion on the path of IP backbone network and MUST stop some site traffic based on the operator specified policy. (stop some traffic on both directions or congestion direction?). Early dropping traffic will help site applications to take an action. S-GW MUST reports the congestion status to site interconnection app (SIA).

7. A site can be configured with two S-GWs for redundant and/or transport capacity. The remote S-GW MUST be able to support load balance tunnel traffic. A S-GW at a site MUST NOT forward incoming traffic from IP backbone to another S-GW at the site, which creates the loop.

8. Site network traffic may be unicast, mcast, or bcast(L2) traffic. Sites interconnection solution MUST be able to carry unicast, mcast, bcast traffic over tunnels.

9. Site network traffic may be control plane packets such as IBGP (ISIS?) or control packets such as ICMP, ARP. Tunnel SHOULD be able to carry control plane or control packets according to operator specified site interconnection policy. In default, a tunnel MUST NOT carry control plane packets over the tunnel.

10. Other?

4.2. Tunnel transport properties for sites interconnection

This section lists tunnel transport requirements over IP backbone networks for sites interconnection application.

1. Tunnel type (tunnel mode, or transport mode, or both)

2. IP backbone network can be IPv4 or IPv6 network. A tunnel MUST be able to run over an IPv4 or IPv6 network.

3. Tunnel solution MUST meet IP [RFC791] [RFC2460] and UDP application requirements [RFC5405bis]

4. Tunnel MTU and Fragmentation [JOE]. Tunnel SHOULD avoid tunnel traffic to be fragmented on IP backbone networks.

5. Tunnel solution supports configuration option to turn off traffic encryption function.

6. Tunnel solution supports congestion control upon detecting congestion condition on the tunnel path (use circuit breaker, packet drop/report at ingress, notification to source).

7. Tunnel solution MUST map site traffic QoS to proper DSCP value on tunnel IP packets based on operation specified policy.

8. Tunnel solution SHOULD be able to monitor the inter-sites connectivity and report the status.

9. Tunnel solution SHOULD support some tools for sites interconnection operator such as tunnel path trace, ping, tunnel's throughput test, etc.

10. ICMP handling (from Internet or from remote tunnel end-point).

11. Others, Hop count for tunnel traffic, middle box considerations, etc.

5. Site traffic encapsulation

GRE-in-UDP w/DTLS [GRE-in-UDP] or GUE [GUE]: the reason is:

- It can run over the Internet (Ipv4 and Ipv6)
- It runs as UDP application, ECMP advantage, and middle box traversal benefit.
- Encapsulate different types of traffic
- Ability to support fragmentation
- Security/encryption capability
- Support traffic segregation

Like to hear other's opinions on encapsulation protocol choices.

6. Tunneling multicast and broadcast traffic

7. Tunnel transport over IP

7.1. Tunnel transport mode

Tunnel mode or transport mode or both.[RFC3884]

7.2. MTU and fragmentation

Intarea-tunnels draft or GUE-extension draft

7.3. Checksum

Tunnel solution MUST implement the checksum specification for default GRE-in-UDP tunnel in [GRE-in-UDP].

7.4. Congestion management

More than likely, a site operator has no way to control transport path resource in IP backbone networks for sites interconnection. Tunnel packets are treated as regular IP packets and traverse a path in IP backbone networks that other IP application packets traverse as well. Congestion may happen due to "ship-in-night" situation. IP backbone network uses explicitly congestion notification (ECN) [RFC6040] to indicate IP applications about the network congestion.

Upon backbone path congestion, a tunnel for the site interconnection MUST stop some traffic based on operator's interconnection policy. This section specifies the tunnel congestion control mechanism.

7.4.1. Congestion detection

7.4.2. Congestion notification

7.4.3. Tunnel ingress traffic control

7.5. Tunnel traffic hop count and DSCP value setting

7.6. Middle box considerations

8. Sites interconnection security

For site traffic security, tunnel MUST use DTLS to encrypt site traffic, i.e., use GRE-in-UDP w/ DTLS. This feature MAY be turned off by a site operator if the site network traffic is already encrypted.

A site operator SHOULD request a security service from IP backbone provider to prevent DDoS traffic to reach the tunnel end point at a S-GW of the sites.

9. Tunnel configuration

10. Tunnel tools

11. Operational considerations

12. IANA considerations

13. Security considerations

14. References

14.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC2119, March 1997.

[RFC791] DARPA, "INTERNET PROTOCOL", RFC791, September, 1981.

[RFC792] Postel, J., "Internet Control Message Protocol", STD 5, RFC792, September 1981.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC5405bis] Eggert, L., "Unicast UDP Usage Guideline for Application Designers", draft-ietf-tsvwg-rfc5405bis, work in progress.

[RFC6040] Briscoe, B., "Tunneling of Explicit Congestion Notification", RFC6040, November 2010.

[GRE-in-UDP] Yong, L., et al, "GRE-in-UDP Encapsulation", draft-ietf-tsvwg-gre-in-udp-encap-19, work in progress.

[JOE] Touch, J. and Townsley, M., "IP Tunnels in the Internet Architecture", draft-ietf-intarea-tunnels-03, work in progress.

14.2. Informative Reference

[RFC3884] Touch, J., Eggert, L., and Wang, Y., "Use of Ipsec Transport Mode for Dynamic Routing", September 2004.

[RFC4364] Rosen, E. and Rekhter, Y., "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC4364, February 2006.

[RFC4664] Andersson, L. and Rosen, E., "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC4664, September 2006.

[RFC4732] Handley, M. and Rescorla, E., "Internet Denial-of-Service Considerations" RFC4732, November 2006.

[RFC4762] Lasserre, M. and Kompella, V.,"Virtual Private LAN Service (VPLS) using Label Distribution Protocol (LDP) Signaling", RFC4762, January 2007.

[RFC5996] Kaufman, C., et al, "Internet Key Exchange Protocol Version 2 (IKEv2)", September 2010.

[RFC6830] Farinacci, D., et al, "The Locator/IP Separation Protocol (LISP)", RFC6830, January 2013.

[RFC7059] Steffann, S., et al, "A comparison of IPv6-over-IPv4 Tunnel Mechanism", RFC7059, November 2013

[RFC7432] Sajassi, A., et al, "BGP MPLS-Based Ethernet VPN", RFC7432, February 2015.

[Dunbar] Dunbar, L., Yong, L., Song, X., "Client Defined Private Networks laid over Thin CEPs", draft-dunbar-interarea-private-networks-over-thinCPE, work in progress.

[GUE] Herbert, T., Yong, L., Zia, O, "Generic UPD Encapsulation", draft-ietf-nvo3-gue-05, work in progress.

15. Authors' Addresses

Lucy Yong
Huawei Technologies

Email: lucy.yong@huawei.com

Linda Dunbar
Huawei Technologies

Email: linda.dunbar@huawei.com

Tom Herbert

Facebook

Emails: tom@herbertland.com