PPVPN Working Group
Internet Draft
Expiration Date: February 2002

Heinrich Hummel
Siemens AG

July 2001

Tree/Ring/Meshy VPN tunnel systems
draft-hummel-ppvpn-tunnel-systems-01.txt

Status of this Memo

Abstract

Version 01 is a significant redesign. Exclusive full mesh (EFM) intersite tunneling per VPN is considered as overkill. Options for optimizations are discussed and compared:

a) Shared full mesh (SFM)

b) Exclusive partial mesh per VPN (EPM)

c) Shared partial mesh (SPM)

This draft favors partial mesh (EPM, SPM) but cautions from pursuing SPM (complexity, further study needed).

The draft also contains algorithms for computing minimal/optimal tree/ring/mesh inter-site tunnel topologies.

The establishment of partial mesh MPLS tunnel systems is removed and will be subject for a separate draft.

1 Introduction

This draft deals with full mesh versus partial mesh intersite VPN tunneling.

Full mesh does not mean that there is a tunnel from site S to site D even if no traffic is expected from S to D. However, in case some traffic is expected from S to D, then there is a direct tunnel from S to D.

Partial mesh means a set of tunnels via which all sites are interconnected- either directly or indirectly. In general, a sequence of tunnels is to be passed when data flows from site S to site D.

Exclusive full mesh (EFM) intersite tunneling per VPN is overkill. Optimization options are:
a) Shared full mesh (SFM): Share PE-to-PE tunnels among different VPNs where applicable.
b) Exclusive partial mesh per VPN (EPM): Use PE-to-PE-to-PE resp. CE-to-CE-to-CE tunnel sequences exclusively for the traffic of one particular VPN.
c) Shared partial mesh (SPM): Share PE-to-PE-to-PE tunnel sequences among different VPNs where applicable.

Though all known IP VPN models mention b) as a viable option they rather concentrate on a): Notice the backbone virtual router in the VR-model, notice the BGP-transmitted label in RFC2547bis being called the "second label" - it would be the "third label" if option b) applied.

This draft is focussed on partial mesh and its advantages over full mesh (see section 2 and 3). Therefore it favors options b) and c). However a cautioning warning against c) is appropriate: you may loose something while trying to get everything, and also, it is complex and needs further study.

Section 3 contains algorithms for computing minimal/optimal tree/ring/mesh inter-site tunnel topologies.

The establishment of partial mesh MPLS tunnel systems is removed and will be subject for a separate draft.

2 Savings of tunnels: Minimal partial mesh versus full mesh

Full mesh topologies for interconnecting n sites either require n*(n-1) uni-directional tunnels or n*(n-1)/2 bi-directional tunnels. For comparison, minimal but contiguous topologies require either n uni-directional tunnels which form a ring or n-1 bi-directional tunnels which form a (meshless) tree.

In the first i.e. uni-direction case, the absolute savings are n*(n-1)-n =n*(n-2) tunnels while the relative savings are [n*(n-1)-n]/[n*(n-1)] = 100*(n-2)/(n-1) %.

In the second i.e. bi-direction case, the absolute savings are n*(n-1)/2-(n-1) = (n-1)*(n-2)/2 tunnels while the relative savings are [n*(n-1)/2-(n-1)]/[n*(n-1)/2] = 100*(n-2)/ n %.

The absolute and relative savings S-abs and S-rel in terms of tunnels compared with full mesh is exorbitant:
Uni-directional ring:
n=11; S-abs= 99 tunnels; S-rel =90 %
n=101; S-abs= 9,999 tunnels; S-rel =99 %
n=1001; S-abs= 999,999 tunnels; S-rel =99.9 %

Bi-directional tree:
n=10; S-abs= 36 tunnels; S-rel =80 %
n=100; S-abs= 9,702 tunnels; S-rel =98 %
n=1000; S-abs= 997,002 tunnels; S-rel =99.8 %

3 Partial mesh versus full mesh

Diverse traffic engineering aspects are discussed from the "partial versus full mesh" point of view.

3.1 Admission Control

Exclusive Partial Mesh (EPM):

In the attempt to add a further VPN, EPM enables "VPN Admission Control" as to maintain the promised quality with respect to all previously established VPNs as well as w.r.t. the actual VPN which is about to be established. It may be an iterative process to determine all the tunnels of the partial, VPN-dedicated mesh and also all their routes such that the requested and eventually aggregated traffic bandwidth is reserved on each physical link of each determined tunnel.

Shared Full Mesh (SFM):

The attempt to add a further VPN, may impact some existing (shared) PE-to-PE tunnel and/or require the establishment of some new PE-to-PE tunnel: Hereby the requested bandwidth reservation may either be successful or may fail.

Shared Partial Mesh (SPM):

The attempt to add a further VPN, may impact some existing (shared) PE-to-PE-to-PE tunnel sequence and/or require the establishment of some new PE-to-PE-to-PE tunnel sequence: Hereby the requested bandwidth reservation may either be successful or may fail. (Of course, such a tunnel sequence may also consists of one single tunnel as well).

3.2 VPN-specific Traffic Policing

Only EPM allows for VPN-specific traffic policing.

3.3 Multiple routes

The immense tunnel savings in case of MINIMAL partial mesh as shown in section 2 allow some luxury, i.e a few more tunnels, so that services like traffic balancing, fast path restoration or traffic-type/Qos-specific traffic multiplexing can be supported. The resulting number of tunnels would still be from the order of n and would still yield a partial mesh.

3.3.1 Fast path protection

In the full mesh case, though there are n*(n-1) uni-dir. tunnels, no fast path protection mechanism can be provided for some traffic stream in case one of their links would break. However, if you have only 2*n uni-dir.tunnels that form two inversive rings, then an alternative route between any pair of sites would always be available in case one of their links is broken. To match this capability the full mesh needs to be doubled i.e. needs 2*n*(n-1) tunnels.

3.3.2 Traffic-type/Qos-specific traffic multiplexing

Assume there should be x completely different tunnels resp. tunnel sequences for any site-to-site communication, e.g. one tunnel per traffic-type (voice, data,...) resp. per QoS-class. In the full mesh case, x*n*(n-1) tunnels were required. In the partial mesh case, x*n tunnels were required.

3.3.3 Traffic balancing

Assume there should be x alternate routes for any site-to-site communication for reasons of traffic balancing. In a full mesh x*n*(n-1) tunnels were required. In a partial mesh not more than x*n tunnels were required. There may be even less, if the alternate routes may have partially shared route sections.

3.4 Customer-based VPN and Layer-2 provider-provisioned VPN

Customer-based VPNs, whereby the service provider is completely unaware of what is the purpose of any traversing CE-to-CE tunnels, can only be optimized by EPM, i.e not by SFM and not by SPM.

The same is true for provider-provisioned CE-to-CE tunnels (i.e. for Layer-2 provider-provisioned VPN).

3.5 VPN Multicast

In this section VPN multicast is evaluated from the resource taking prospective.

Shared and Exclusive Full Mesh (SFM and EM): A general goal of multicast is to employ a (tree-like) delivery channel as to avoid multiple transmission over the same physical links. This goal is not supported at all in case of full mesh models like EFM and SFM. Indeed, at the source-PE the multicast data must be forwarded as often as there are destination-PEs. None of these flows is ever branched on its way to its destination-PE.

Solution favored by E.Rosen in [2]: That solution essentially favors one multicast tree per VPN or one per suitable set of VPNs (Multicast domain). All multicast applications within the VPN shall forward the multicast packets along that tree to ALL respective PEs - even in case there are no receivers on some of the sites (so that the packet has to be discarded). This solution requires P-routers' involvement and multicast trees in additon (!!!) to the tunnels for point-to-point traffic.

Exclusive Partial Mesh (EPM): The exclusive partial mesh which is already provided for point-to-point traffic can be utilized also for multicast. The P-router stays completely VPN-unaware.

4 Computing partial mesh topologies

The intersite tunnels may be CE-to-CE (L2vpn, customer-based VPNs) or PE-to-PE (RFC2547bis, Virtual Router).

Tunnels may have associated weight values:

a) Weight= number of hops

b) Weight= 1/traffic load between its endpoint nodes

c) Weight= Price taken from a Least-Cost-Routing table, structured according to service providers, time-of-day/day-of-week time zones, distance zones, direction of tunnel establishment, anticipated tunnel lifetime, etc.

VPN tunnel computation according to 4.1 and 4.2 may either be done in awareness of the carrier network's topology (weight according to a) or b) may apply) or without that awareness (weight according to c) may apply). A customer who computes a VPN tunnel system based on weight definition c) may order the resulting VPN tunnel system at the right service provider(s), e.g. by post-mail, or order it based on user signalling (i.e. the service provider becomes aware of what will be established), or establish it all by himself based on user signalling such that the service provider is not informed about the purpose of the established tunnels. All respective signalling details should be described and provided by the IETF.

4.1 Computing Minimal Tree VPN tunnel systems

The minimal number of bi-directional tunnels for interconnecting n sites is n-1. In a full-meshed VPN tunnel system n sites will be interconnected by n*(n-1)/2 bi-directional tunnels. Among them we select those n-1 tunnels of smallest weight which contiguously interconnect all n sites. They will form a tree-like topology.

In case of uni-directional tunnels the full mesh configuration will require n*(n-1) tunnels. Among them we need to select those n-1 pairs of inverse tunnels whose weights (for the total pair) are smallest and which still form a contiguous topology.

The following algorithm selects n-1 bi-directional tunnels which form a tree-like topology. Minor modification were only necessary as to build an algorithm which selected n-1 pairs of inverse, uni-directional tunnels that form a tree-like topology as well.

Start of algorithm:
We have sorted all n*(n-1)/2 candidate tunnels in ascending order according to their weights and have all of them marked REGARD rather than DISREGARD.

We start with an empty tunnel system and with n separated subnetworks, each of which initially contains just one site=node without any tunnel link. In each of the following iteration step we will combine any two so far separate subnetworks by a particular interconnecting tunnel (which will be added to the tunnel system) and repeat the step n-1 times as to get one contiguous subnetwork and consequently one contiguous tunnel system.

Iteration step:
Take the tunnel of smallest weight which is marked REGARD and add it to the tunnel system. Mark this tunnel to DISREGARD. Build the combined subnetwork out of those two subnetworks A and B which are interconnected by the taken tunnel. Mark DISREGARD all tunnels which start at a node in A and end at a node in B.

Result: We have determined a Minimal Tree VPN tunnel system, which interconnects all n sites by the least number of shortest tunnels.

4.2 Computing Optimal Ring VPN tunnel systems

The following algorithms A and B select n bi-directional tunnels which form a ring while their weight sum is fairly minimal.
Analogous tasks: Select n uni-directional tunnels resp. n pairs of inverse, uni-directional tunnels which form a ring as well.

Note that both algorithms are of order n, and not of order n!, with respect to the number of iterations.

Goal of algorithm A: Form a ring such that as often as possible two nodes become immediate neighbors if the respective interconnecting tunnel has a smallest weight; furthermore such, that as often as possible two nodes become 2nd (3rd, 4th,..) degree neighbors, if the respective interconnecting tunnel string has a smallest weight sum.

Algorithm A:

Start with an empty tunnel system and n tunnel strings Si .Each Si has weight Wi =0, contains just 1 node which is both Left- and Right tunnel string border node.

Build n-2 times combined tunnel strings Si&k, by concatenating the

strings Si and Sk by their shortest interconnecting tunnel. In each step, search for those two tunnel strings Si and Sk , which yield the smallest weightsum: Wi&k =Wi +Wk + Weight of connecting Tunnel. Hereby determine out of (maximal) 4 candidates that tunnel of smallest Weight; and also the two new border nodes of tunnel string Si&k.

Add the shorter final tunnel pair to the tunnel system as to close the ring.

Algorithm B: Add step by step one further node (site) to the ring which initially contains any two nodes. Each time insert the new,arbitrarily chosen node Z between those nodes X and Y for which is true: ABS (weight for tunnel X-Z + weight for tunnel Z-Y - weight for tunnel X-Y) is smallest.

There are certainly more such algorithms, see also [5, 6].

4.3 Forming VPN tunnel systems of arbitrarily meshed graph

VPN tunnel systems of arbitrarily meshed graph may be formed e.g. based on pure configuration or based on some algorithms like the following algorithm:

Spend a direct tunnel between any two nodes, if the traffic, exchanged between them, exceeds some limit. As a result you may get m tunnel-and-node fragments, 1 <= m <= n. Some of them may form some meshes, or may form some trees, or may still be isolated nodes (i.e. without any tunnel).

Interconnect these fragments by a minimal tree topology (apply algorithm A) or by a minimal ring topology (apply algorithm B). Hereby, each fragment is an initial "node" according to the algorithm description.

5 Summary and proposals

The draft clearly shows the value/advantages of partial mesh VPN tunneling. The editors of the ppvpn-documents may take this into consideration.

6 Intellectual Property Considerations

This proposal in is full conformity with [RFC-2026].

Siemens may have patent rights on technology described in this document which employees of Siemens contribute for use in IETF standards discussions. In relation to any IETF standard incorporating any such technology, Siemens hereby agrees to license on fair, reasonable and non-discriminatory terms, based on reciprocity, any patent claims it owns covering such technology, to the extent such technology is essential to comply with such standard.

7 References

[1] Eric Rosen (Cisco) RFC 2547bis,
draft-rosen-rfc2547bis-03.txt

[2] Eric Rosen (Cisco) : Multicast in MPLS/BGP VPNs
draft-rosen-vpn-mcast-00.txt

[3] RFC2917: A Core MPLS IP VPN Architecture

[4] Kampella (Juniper Networks): MPLS-based Layer 2 VPNs
draft-kompella-mpls-l2vpn-02.txt

[5] Walid Ben-Ameur and Bernard Liau, Computing Internet routing metrics; Annals of telecommunications (April 2001)

[6] Walid Ben-Ameur, Nicolas Michel and Bernard Liau, Routing Strategie for IP networks; Telektronikk magazine, 2001

[7] Tissa Senevirathne (Force10) : Use of Partial meshed tunnels to achieve forwarding behavior of full meshed tunnels
draft-tsenevir-l2vpn-pmesh-00.txt

[8] Tissa Senevirathne(Nortel),Waldemar Augustyn (Nortel), Pascal Menezes (TeraBeam):
A Framework for Virtual Metropolitan Internetworks (VMI)
draft-senevirathne-vmi-frame-01.txt

8 Author's Address

Heinrich Hummel
Siemens AG
Hofmannstrasse 51
81379 Munich, Germany
Tel: +49 89 722 32057

Email: heinrich.hummel@icn.siemens.de