Data Center Interconnections: Technical Implementations

By Jaroslav Rajić

3. 12. 2009

The July 2009 article "Data Center Interconnect" discussed data center interconnection requirements on the application level. This article continues that discussion by elaborating on the technical implementation of various interconnection technologies.

This article outlines various ways in which access technologies can be used for data center interconnection, tailored to specific customer requirements: Layer 2 or Layer 3, low latency vs. flexibility etc.

Introduction

The previous article introduced the need for secondary data centers and disaster recovery. We also mentioned how load sharing can be performed on the application level, so that you can provide service without interruption. In this article, we will focus on the back-end connection between data centers, which is required for proper operation of a secondary data center.

Interconnecting Technologies: Overview

The interconnecting technologies are divided into Layer 3 and Layer 2 services. The choice is based on application-level requirements. If the application requires Layer 3 (IP) connectivity, you can use GRE over IPSec or MPLS VPN. When an application requires Layer 2 visibility between the data centers, you need to bridge two Layer 2 domains over a Layer 3 network. Protocols offering this capability include L2TP over IPSec, point-to-point Ethernet over MPLS, VPLS, and so on. If you need to transport protocols such as FC between the locations, you need to establish a connection capable of carrying any protocol directly, such as using Dense-Wavelength Division Multiplexing (DWDM) or similar.

Layer 3 Interconnections

IP-based Internet access is the most common need for most customers. The VPN topology depends on whether the customer is using IP service or MPLS VPN service. For IP service, an IP-based VPN needs to be established to provide routing of packets and privacy. For MPLS VPN service, the provider provides routing and privacy through MPLS.

Layer 3 MPLS VPN

IP MPLS VPN offers clear benefits over other technologies. Traffic routing efficiency is accomplished by optimum routing in the service provider network core. The service provider gives the customer a specified level of service (through a service level agreement – SLA), providing sufficient bandwidth and controlling delay by using MPLS Traffic Engineering.

Another advantage is seamless integration with customer routing: routes from the customer routing protocol are redistributed into multiprotocol BGP (MP-BGP) and delivered automatically to all of the customer's VPN sites, where redistribution into the client routing protocol occurs again.

By contrast with other tunneling solutions, MPLS VPN has the advantage of low packet overhead. MPLS labels usually require only 12 bytes of overhead (for a label stack of three labels), whereas an IPSec/GRE solution can take

up to 94 bytes, minimizing the customer's maximum transmission unit (MTU) and decreasing performance for applications that use all available payload bytes – file copying etc.

MPLS in the provider's core network offers the advantage that the service provider's core routers don't need to know all the customer routes (and possibly no Internet routes), leading to optimal performance for VPN traffic.

MPLS VPN provides good integration with customer internal gateway protocols (IGPs) such as OSPF, EIGRP and RIP. Particularly in OSPF, special care is taken to carry the OSPF route type in the MP-BGP VPNv4-type route. This feature of MPLS VPN (called OSPF Superbackbone) allows for correct redistribution of OSPF routes between the sites, making them of type "Inter-Area" rather than "External."

Figure 1: MPLS VPN label encapsulation. Label L1 is the path label (provided by LDP), label V1 is the VPN label (provided by BGP)

GRE over IPSec over the Internet

If the customer doesn't want MPLS VPN, another option is to set up a Layer 3 VPN using tunneling. This way, the customer overlaps the service provider's Layer 3 network with a (relatively fixed) topology of tunnels, built using the Generic Routing Encapsulation (GRE) and IPSec protocols. This approach uses GRE and IPSec to encapsulate packets. The GRE encapsulation allows the customer to carry routing protocol traffic (multicasts) in order to establish an overlapping topology over a generic (service provider's) IP network, used to transport the packets between locations. As the transport is performed over a public network, IPSec is used to establish a secured communications channel and to guarantee data integrity.

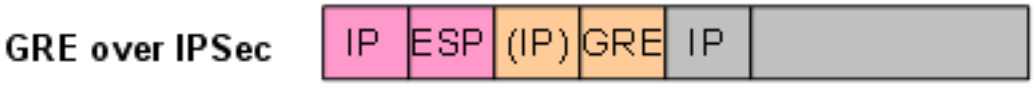
There are several GRE/IPSec designs: point-to-point, hub-and-spoke, partial mesh etc. Each customer establishes the best possible topology

according to its own traffic needs, using the routing protocol to maintain connectivity.

To achieve high performance, this technology requires special hardware, such as VPN acceleration modules to encrypt and decrypt traffic at near-wire speeds. By contrast, MPLS VPN does not provide encryption, but does provide customer traffic separation at true wire speed.

TIP: Check other NIL IP Corner articles and presentations and related Cisco whitepapers to learn more about secure IPSec-based VPNs.

Figure 2: GRE over IPSec encapsulation, IPSec in Tunnel mode.



Layer 2 Interconnections

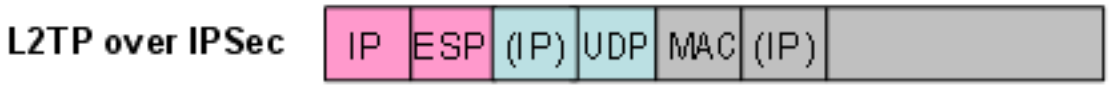
Layer 3 VPN technologies are usually sufficient to satisfy most common VPN requirements. However, if a design can't be implemented using Layer 3 VPNs, Layer 2 VPNs are an option to overcome problems. Layer 2 interconnections between sites over long distances are necessary only if some application the customer is running requires Layer 2 visibility between the sites. The most common of these applications are databases, running on clustered servers that need to be in the same VLAN in order to exchange keepalives for load distribution and high availability.

Layer 2 Tunneling Protocol

Layer 2 Tunneling Protocol (L2TP) can be used to carry Layer 2 traffic, using Internet Protocol (IP) for encapsulation and traffic delivery. The traffic thus becomes Layer 3 and can travel over any IP-based network, such as the Internet. The same security concerns apply as when using GRE, because L2TP does not encrypt the traffic, which is transported over a public network. Just as when using GRE and IPSec, you need to be careful about MTU implications caused by overhead. L2TP can provide a point-to-point extension of a Layer 2 network segment.

L2TP packets can be protected using IPSec when they travel across a public IP network, such as the Internet. This is accomplished in a very similar way as in GRE – by routing the L2TP traffic to a (tunnel) interface that provides encryption.

Figure 3: L2TP encapsulation. The Layer 2 frame is encapsulated in an UDP datagram. IPSec is used in Tunnel mode, but depending on the hardware implementation, it can also be in Transport mode



Point-to-Point Ethernet over MPLS

Similarly to when using L2TP, Ethernet over MPLS provides an extension of a Layer 2 domain, but utilizing MPLS as a transport platform for user traffic. A label stack of two labels is used for transport: the first label identifies the provider edge (PE) router, and the inner label identifies the virtual circuit (VC) to which the traffic belongs.

MPLS encapsulation is performed at the PE device; on one port, the device receives "raw" traffic that needs to be encapsulated, and sends out the port over the link to the core using MPLS. The VC labels are exchanged using the Directed LDP protocol, while labels for the IP next-hop are exchanged using regular LDP or BGP.

Keep in mind that an Ethernet over MPLS link is a point-to-point link – it behaves exactly as a point-to-point Ethernet link does between two devices.

To "capture" the traffic, the same mechanisms are used in this system as with L2TP: xconnect. Only the encapsulation is different (MPLS instead of IP).

Figure 4: EoMPLS/VPLS encapsulation. Label L1 is obtained through LDP, label V1 identifies the VC and is exchanged by Directed LDP. The

packet may contain a control word between the labels and the actual frame.

VPLS

The Virtual Private LAN Service (VPLS) is an extension of the Ethernet over Multiprotocol Label Switching (EoMPLS) service, allowing more than two devices to participate in such a "virtualized LAN" over long distances. This design creates a multiple-access connection in a similar fashion to that of an Ethernet LAN, where all segments "see" themselves on Layer 2.

Because VPLS is a multiple-access topology, it presents a few problems: when a frame comes into such a network, to which location does it need to go to arrive at its destination? Switching functions, such as MAC address learning and frame replication, must be performed on the PE devices, accepting the frames.

Another issue is how to set up such a multiple-access network – which sites are members of a particular VPLS domain? With point-to-point EoMPLS, you have no doubt, as the link includes only two sites; with VPLS you need to set up the domain manually or dynamically by using MP-BGP.

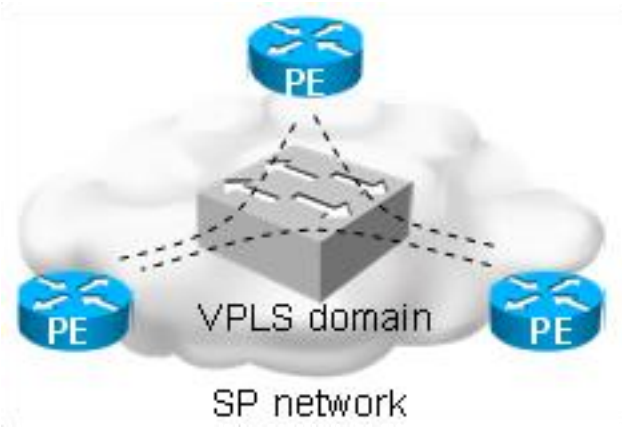
Customers with dual-homed sites that connect to the same VPLS cloud need to be extremely careful when designing the VPLS and when provisioning. Dual-homing to the cloud leaves two options: use the Spanning Tree Protocol (in essence, paying for one link that goes unused, as it will be blocked by STP), or design the WAN network in such a way that Layer 2 loops cannot occur. If a network loop occurs, broadcast storms are highly probable, which can have a severe impact on the service provider network.

The service provider can use MPLS Traffic Engineering to provide bandwidth for these EoMPLS/VPLS links, but needs to employ strict QoS to block these broadcast storms, in order to avoid consuming all available bandwidth. Furthermore, in order to save on backbone link costs, service providers plan their links with overprovisioning in mind, counting on all

customers not to use all their available bandwidth at the same time. A broadcast storm would consume a large portion of available bandwidth (a full portion available to one customer), having an impact on VPLS services for other customers as well.

The bottom line is that such "Layer-2 over MPLS" services should be used with MPLS Traffic Engineering and rigorous QoS measures in the network (traffic shaping/policing).

Figure 5: VPLS network. The SP cloud acts as a multiple access Ethernet link. Traffic between sites is transported using MPLS.



802.1Q and QinQ

802.1QinQ is a pure Layer 2 technology for providers that have a backbone based on Layer 2. It allows them to encapsulate IEEE 802.1Q VLAN tags within another 802.1Q tag. This is a practical approach to support customers with multiple VLANs, by using a single VLAN to carry the customer traffic through the backbone.

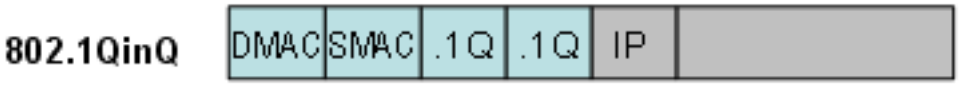
This form of Layer 2 tunneling is used when DSL service providers in some parts of the network share a common Layer 2 infrastructure (such as ATM or a Metro Ethernet backbone). 801.1QinQ is a key technology in the service provider DSL wholesale model, in which all customers of a single DSL provider are put into the same (outer) provider-specific VLAN, but are isolated from each other in various (inner) customer-specific VLANs.

Technically, 801.1QinQ is implemented by inserting an Ethertype field and a VLAN tag after the MAC addresses (becoming the outer, service-provider

tag) and the existing Ethertype and VLAN tag (becoming the inner, customer tag). The frame is switched through the backbone by using the outer VLAN tag, until it reaches the port on the switch/router, which is configured to decapsulate the frame and forward the frame to the customer switching domain.

QinQ can be also used within an enterprise network, to trunk a VLAN that uses incompatible numbering across a larger Layer 2 network. Or it can be used to scale the current VLAN scheme, which is now offering 4096 VLANs (12 bits), up to 16,8 million VLANs (4096 x 4096, 24 bits).

Figure 6: 802.1QinQ frame format. Additional 802.1Q field is inserted for the outer VLAN.



Dark Fiber

Fiber interconnections are suitable for data center interconnections because they are universal and Layer 2 independent. They can carry either Fibre Channel traffic or Gigabit Ethernet. Adequate equipment is needed at the customer side, of course.

The business model for service providers is that they can lend particular fibers to their customers. Service providers are not the only companies in this market – most companies that own physical infrastructure, such as roads, highways, and railways, also own fiber cables installed within these facilities, which cover large distances.

In case of a customer edge (CE) equipment back-to-back connection, the fiber must be delivered to the customer. Splicing might be needed along the path so that a continuous fiber is produced end-to-end.

Figure 7: Back-to-back Dark Fiber connection

One option to better exploit the fiber is to use multiplexing technologies such as Coarse-Wavelength Division Multiplexing (CWDM) and Dense-Wavelength Division Multiplexing (DWDM). CWDM is a simpler technology, allowing you to carry between 4 and 8 distinct data channels over the same fiber by using a different laser beam wavelength for each channel, sending in all the channels through a prism (passive optical element) into the same fiber, and similarly de-multiplexing it at the other end. This approach offers more efficient use than a simple back-to-back connection.

If your requirements are even higher, you can use DWDM, in which you can have up to 80 simultaneous channels by using different wavelengths and active multiplexing equipment, such as the Cisco ONS .

In both cases, these connections can be either native Fibre Channel or Gigabit Ethernet, making them useful for data center interconnections, offering high speed and low latency, with the other device only one Layer 2 hop away.

Figure 8: Dark Fiber interconnection using CWDM or DWDM. Multiple channels can be set up, i.e. one for Gigabit Ethernet, one for Fibre Channel

Fibre Channel interconnects are especially suitable when synchronous replication of storage is required, i.e., when the primary storage (at the primary datacenter) confirms the disk write when data has been written to the primary *and* the secondary storage equipment (at the secondary, distant datacenter). This is to satisfy the most demanding high availability requirements.

Fiber

In addition to dark fiber, service providers might offer fiber connectivity services by using their active equipment, offering the customer only a part

of their link and one or a few wavelengths. In this case, service providers use DWDM equipment for their and customer traffic, and it is up to each particular customer to determine how to connect to this DWDM equipment. Metro Ethernet, dark fiber or CWDM/DWDM equipment can be used at the customer side, with intermediate equipment (linecards) at the service provider.

Figure 9: Fiber interconnection by using one or a few wavelenghts at a service provider

DWDM optical transceivers are available even as SFPs that can be plugged directly into customer switches, routers or FC switches. Bear in mind that such transceivers are very expensive, with prices comparable to a linecard for the Cisco Optical Network System (ONS); however, they do save the provider from having to keep a free slot at the "target" SP Cisco ONS device, and the customer does not need to purchase any DWDM equipment except for the transceiver.

Conclusion

This is an overview of the most popular data center interconnection technologies available today. They are very different in terms of use, equipment used, performance and costs. The most economical variants use the Internet as a carrier service for packet transport (GRE over IPSec, L2TP over IPSec), but are usually limited in terms of performance by the encryption hardware.

More flexible are the MPLS-based solutions (MPLS VPN, EoMPLS, VPLS), which provide traffic separation by using labels and wire-speed performance, but implementation for these solutions is more demanding (CE-PE routing, high-speed access etc.), and SLAs need to be honored by the service provider, leading to additional administrative costs (MPLS Traffic Engineering and QoS).

Finally, we considered pure "Layer 2 and below" technologies such as QinQ, Dark Fiber and Fiber connectivity. Fiber-based connectivity offers the best performance, but is also the most expensive solution, and should be used only when requirements are at maximum demand (such as for synchronous storage replication).

QinQ is a niche technology used mainly by DSL wholesale service providers, allowing these companies to separate traffic in VLANs of different service providers, carrying it over a single Ethernet backbone with additional VLAN tagging.

Table 1: Technology-Requirements Matrix

Technology	Application	Equipment	Advantages/Disadvantages
GRE over IPSec	Connecting a few sites over the Internet	Cisco ISR routers, Cisco ASA devices	+ relatively easy to configure + no costly equipment required - low performance - low scalability
MPLS VPN	Connecting sites over an MPLS VPN service provider	Cisco routers and switches	+ optimum routing through the service provider backbone + service provider participates in routing exchange + wirespeed technology + scalability - high-end solution only
L2TP over IPSec	Connecting Layer 2 domains over a public IP network	Cisco routers and switches	+ simple to deploy, no configuration required at the service provider - high encapsulation overhead - performance limitations of IPSec on routers
EoMPLS	Connecting Layer	Cisco routers	+ Layer 2 interconnect with

	2 domains over an MPLS network	and switches	high performance (wire speed) - QoS greedy
VPLS	Connecting more than 2 Layer 2 domains over an MPLS network	Cisco routers and switches	+ multiple-access topology - MAC address learning at the PE, frame replication +/- manual or automatic configuration of the switching domain - very careful provisioning needed (QoS, Spanning Tree / switching loops)
802.1QinQ	Carrying Layer 2 traffic over another Layer 2 domain; DSL wholesale, etc.	Cisco routers and switches	+ increases scalability limit of 4096 VLANs + segregating 802.1Q encapsulated traffic from different ISPs, using the same L2 network - application-specific; working Ethernet domain required
Dark Fiber	Providing customers with fabric end-to-end for storage replication etc.	Cisco routers and switches with dedicated SFP transceivers, Cisco ONS	+ all available fiber bandwidth for all possible uses - equipment very costly for full exploit - cable very costly to own/rent end-to-end (construction works)
Fiber	Providing customers with limited-bandwidth channel for DC interconnection	Cisco routers and switches with dedicated SFP transceivers, DWDM transceivers,	+ compromise between costs of owning dark fiber end-to-end solution and provider-based solution + flexibility of solution in terms of price and protocol

	Cisco ONS	(Ethernet, FC etc.) - very costly equipment (Cisco ONS) usually required at CE and PE

NIL – More Than Just a Training Company

NIL Learning delivers the leading-edge Cisco training to IT professionals and companies around the globe. Through field-proven experts — each both active engineer and instructor — NIL Learning enhances the standard learning curriculum with real-life experience and helps clients to maximize their training investment.

NIL Learning is part of NIL, a leading global IT solutions provider. Since 1992, NIL has been at the forefront of advanced contributors to strategic partner Cisco's technologies, learning curriculum and value-added solutions deployed to clients around the globe. Today, NIL has earned the highest certifications offered by Cisco, VMware, EMC, HP, IBM, Microsoft, F5, Jive, MobileIron, RSA, VCE and others. Their portfolio of solutions consists of managed services, professional services and learning services.

NIL is headquartered in Slovenia, with regional offices in Croatia, Serbia, Saudi Arabia, the U.S., Turkey, South Africa, Morocco, Nigeria, Kenya and Botswana.

Why learn at NIL LEARNING?

- All NIL LEARNING instructors are **field-proven experts** - each both active engineer and instructor.
- **75% of NIL LEARNING engineers hold CCSI** certifica- tions, and **18 have already achieved the respected CCIE rank.**

- NIL LEARNING **enhances the standard learning curriculum** with real-life experience and helps clients to maximize their training investment.
- NIL has been a Cisco Training Partner for many years; it became a **Cisco Learning Partner in 1993**, and has been a **Cisco Gold Partner since 1995**.
- NIL was awarded the **Cisco Most Business Relevant Learning Partner in MEA in 2010 and the most innova- tive learning partner in MEA.**
- NIL received the **Innovation Award for its Technology Led Training and its extensive contribution to Cisco learning solutions** at the Cisco EMEAR Learning Partner Summit in 2012.
- NIL received the **Innovation Award for its Technology Led Training and Advanced Engineer Program** at the Cisco Global Learning Partner Summit in 2013.
- NIL LEARNING runs a **centralized training schedule across the whole EMEAR region.**

More Info

Slovenia

T: +386 1 4746 500

E: sales-support@nil.com

Botswana

T: +267 318 1684

E: training@it-iq.bw

Croatia

T: +385 (0)51 583 255

E: info-nilcroatia@nil.com

Kenya

T: +27 (0)11 575 4637

E: mea_sales@nil.com

Morocco

T: +212(0) 660 808 394

E: info-nilmorocco@nil.com

Nigeria

T: +27 (0)11 575 4637

E: mea_sales@nil.com

Saudi Arabia

T: +966 1 465 4641

E: info.nilme@nil.com

Serbia

T: +381 11 2282 818

E: info-nilserbia@nil.co.rs

South Africa

T: +27 (0)11 575 4637

E: mea_sales@nil.com

Turkey

T: +902 123 81 8639

E: info-nilturkey@nil.com

USA

T: +1 612 886 3900

E: info-nilusa@nil.com

www.learning.nil.com