Global Knowledge®

Expert Reference Series of White Papers

Dynamic Multipoint Virtual Private Network (DMVPN)

Dynamic Multipoint Virtual Private Network (DMVPN)

Billl Treneer, CCSI, CCNP, CCDP, CWLSS, CCNA, CCDA, CompTIA Security+, and CompTIA Network+

Introduction

Before implementing DMVPN as a hub and spoke solution, or streaming multicast with a Dynamic Multipoint Virtual Private Network (DMVPN), an explanation of DMVPN may be in order for many of us trying to implement this solution. All examples of VPNs in this paper cross the public Internet. DMVPNs could be used with other networks like Multiprotocol Label Switching (MPLS), but streaming multicast is accomplished quite well using "Default" and "Data" Multicast Distribution Trees (MDTs) with MPLS.

DMVPN

A DMVPN is not a protocol so there are no configuration commands that trigger it like "ip dmvpn xxxx." A DMVPN is instead a network design. This design allows remote sites/spokes in a "Hub and Spoke" or "Star" VPN router topology to connect to each other directly without sending the traffic/data packets through the Hub. In other words, it is one hop rather than two hops which is sometimes called a hairpin turn. Most of this paper will describe Phase 2 DMVPN design. Phase 3 is also available and the differences are explained at the end of this paper.

The DMVPN design is made up of the following technologies, which will be explained separately:

1. Multipoint Generic Routing Encapsulation (mGRE)
2. Next Hop Resolution Protocol (NHRP)
3. Routing protocol—EIGRP is often mentioned as a good choice
4. IP sec encryption

Background

Next Hop Resolution Protocol (NHRP) was originally used in non-broadcast multi-access networks (NBMA), like Frame-Relay and Asynchronous Transfer Mode (ATM). Devices/routers connected to an NBMA network typically are all on the same IPv4 subnet. Broadcasts and multicasts do not reach all devices like they do on an Ethernet network because legacy NBMA networks are usually Layer 2 WAN implementations with no routing inside the WAN. Without routing inside the NBMA network, spoke routers have to go through the hub router to get to another spoke as described above. This limitation could cause a bandwidth bottleneck at the hub router. One solution is to put the routers in a full mesh topology, but spoke routers would need an extensive configuration and the expense of extra virtual circuits to reach each other in one hop (see Figure 1).

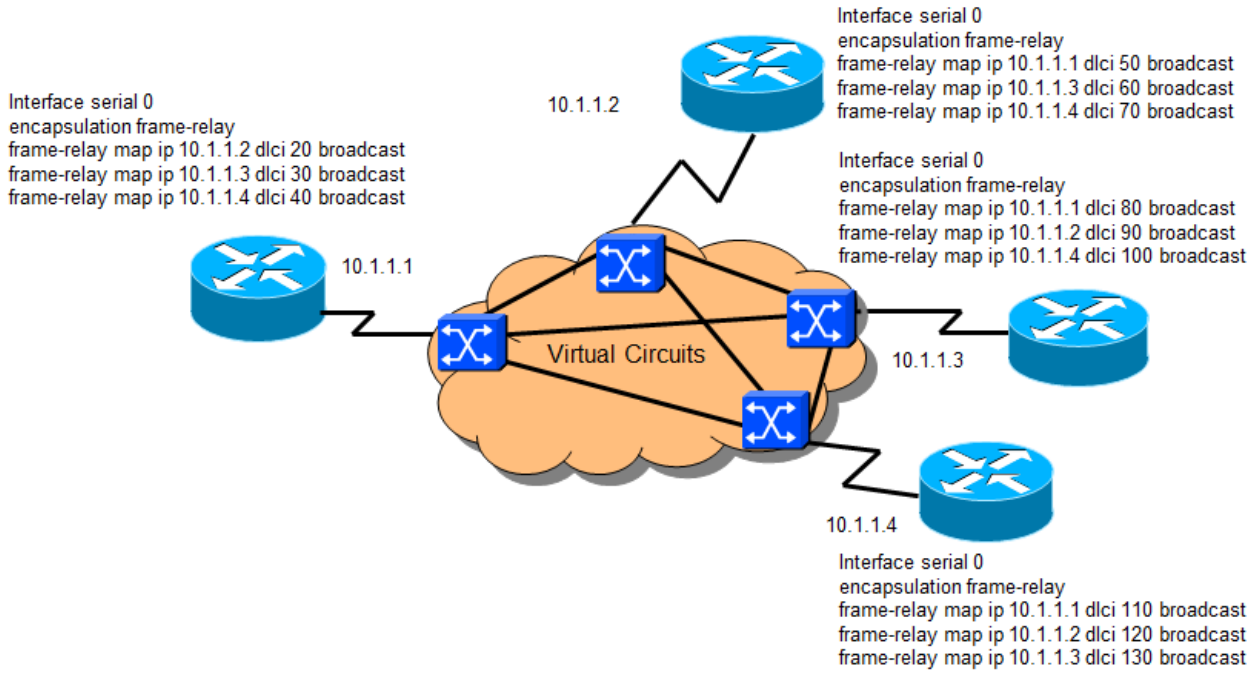


Figure 1

Like any NBMA configuration there needs to be a mapping either statically, like shown here, or dynamically of the IPv4 next hop address to the NBMA next hop address. The NBMA next hop address in this example is a Frame-Relay DLCI number. The above example is with only four locations. A topology of twenty routers would need a total of 190 virtual circuits and 380 map statements. A more scalable, less expensive solution is needed that also does not cause a bandwidth bottleneck at the hub router.

The solution is to have a Layer 3 inexpensive WAN like the public Internet. The Internet has routers inside the backbone that can make routing decisions. The 190 virtual circuit problems go away as Internet routers have a path to every destination. Also, the Internet routers are fully meshed or at least heavily partially meshed so two remote spoke routers likely have a better path to each other than going through a hub router. The problem is the routers connected to the Internet are likely not all on the same subnet so they do not form an NBMA and NHRP registration and discovery would not take place. Therefore, for spoke routers to connect directly to other spoke routers and provide the security necessary on the public Internet, some form of VPN tunnel configuration is needed. The 380 map statements problem would still exist as 380 VPN configurations (see Figure 2).

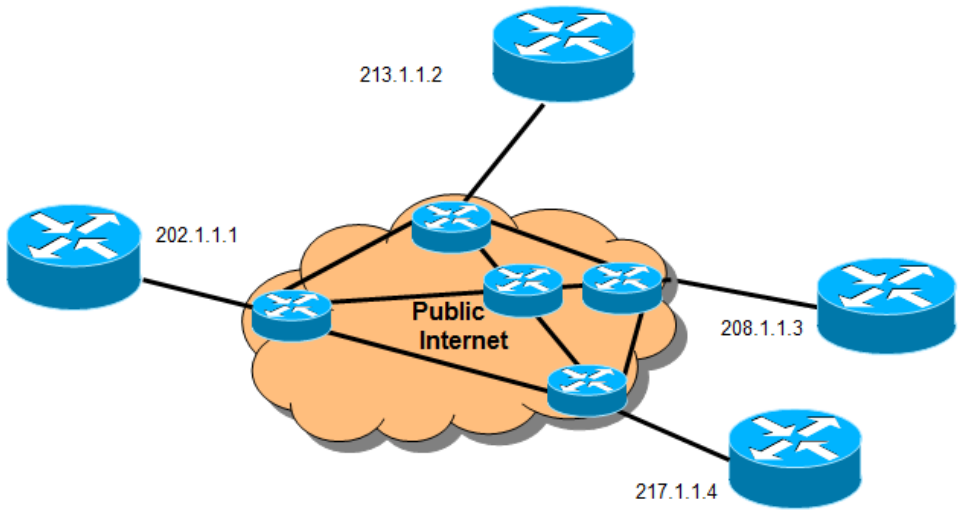


Figure 2

To get the advantages of the Layer 3 Internet, GRE or some other tunnels have to be created, then all the routers can be configured on the same subnet. An NBMA is formed with the routers' tunnel interfaces on the same subnet. The routers can take advantage of NHRP registration and discovery with an NBMA. The most efficient way to create the GRE tunnels is with mGRE, so the first step in DMVPN design is to configure mGRE.

1. Multipoint Generic Routing Encapsulation, (mGRE)

Notice the IPv4 GRE tunnel interfaces in Figure 3 are all on the same subnet. mGRE does not allow broadcasts so the use of mGRE creates an NBMA.

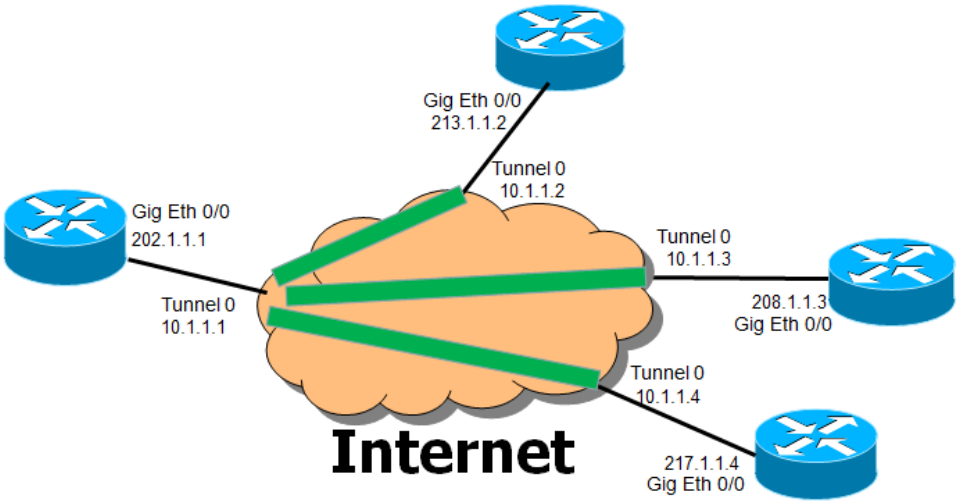


Figure 3

With this NBMA configuration there still needs to be a mapping either statically or dynamically of the IPv4 next hop physical IPv4 address to the NBMA next hop address. The NBMA next hop address in the Figure 3 example is the IPv4 address of Tunnel 0. Another view of making the Internet into an NBMA network is the logical analogy shown below with the small green routers as phantoms (see Figure 4).

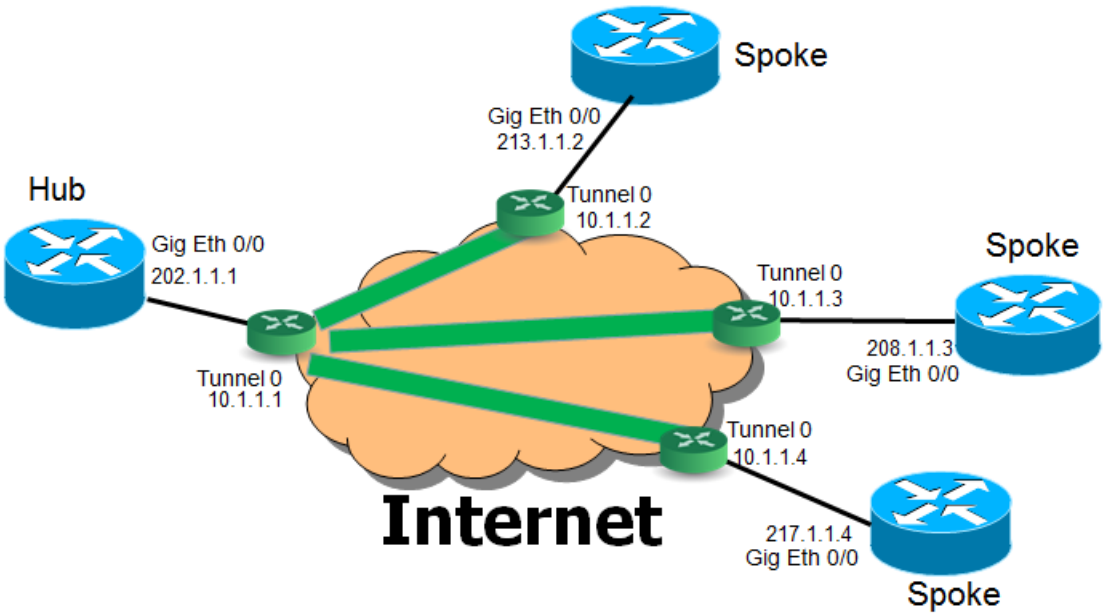


Figure 4: Logical view of mGRE tunnels

The tunnel configuration creates a one-hop logical connection from the hub router to each spoke router. Notice that each router's physical connection is to its local ISP, so physically the routers are on different subnets. Once the tunnel interfaces are configured on each Gigabit Ethernet interface there is a single logical subnet between the hub router and its spokes.

As previously stated multipoint GRE is a major design component of DMVPN. The configurations for mGRE are shown next. The first configuration, Figure 5, shows mGRE at the hub and point-to-point GRE at the spokes. Figure 5 may be a configuration some have seen before, but it will force the tunneled NBMA network into a hub and spoke topology so the spoke routers will not be able to dynamically build a VPN to another spoke router. The second configuration, Figure 6, shows mGRE at the hub and spoke routers, which will allow the spoke routers to dynamically build a VPN to another spoke router. There is still more configuration needed to complete the entire DMVPN design, but Figure 6 shows the critical mGRE part.

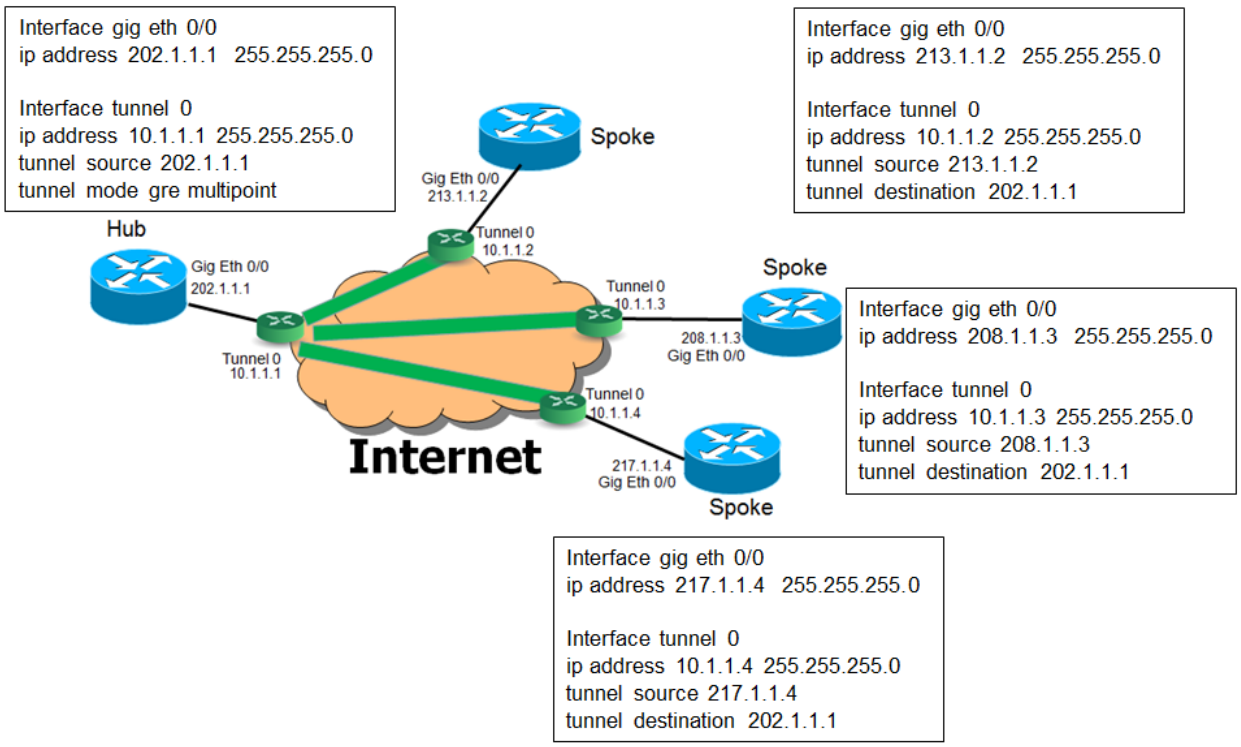


Figure 5: mGRE at Hub and point-to-point GRE at Spokes

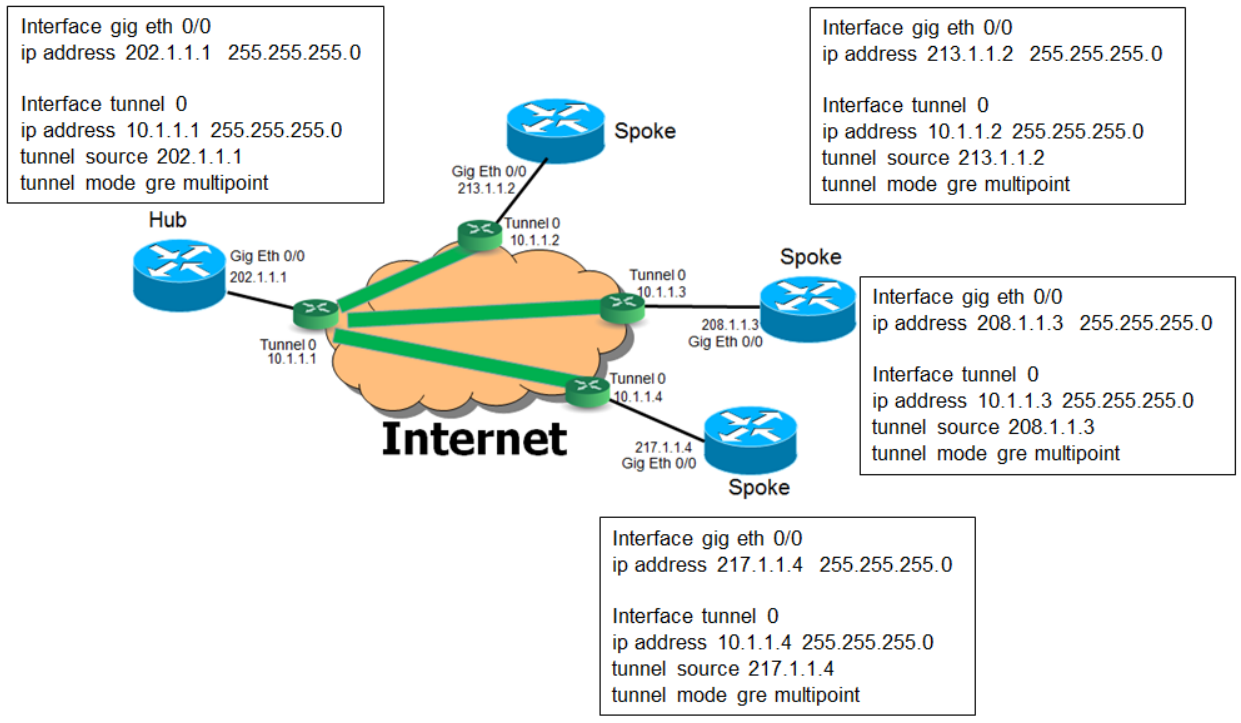


Figure 6: mGRE at Hub and Spokes

It would be a best practice if the physical IPv4 address on interface Gigabit Ethernet 0/0 of the hub router was statically assigned. There will be multiple NHRP commands shown in the next few pages that use this address for mapping. If this address dynamically changed it would create a lot of troubleshooting and reconfiguration. The spoke routers could use dynamic addresses if the tunnel source command was changed to: "tunnel source gig eth 0/0."

The second step of DMVPN design is to link the mGRE configuration to an NHRP configuration.

2. Next Hop Resolution Protocol, (NHRP)

NHRP is a resolution protocol that allows a Next Hop Client (NHC) to dynamically register with Next Hop Servers (NHSs). With the DMVPN design the NHC is the spoke router and the NHS is the hub router. Once all clients are registered, spoke routers can discover other spoke routers within the same NBMA network. With NHRP the above example's 380 VPN statements would be drastically reduced. The additional NHRP configuration statements for registration and discovery are shown in Figure 7. The hub router, NHS, will need the statements on the left of the graphic and each of the spoke routers, NHCs, will need the statements on the right.

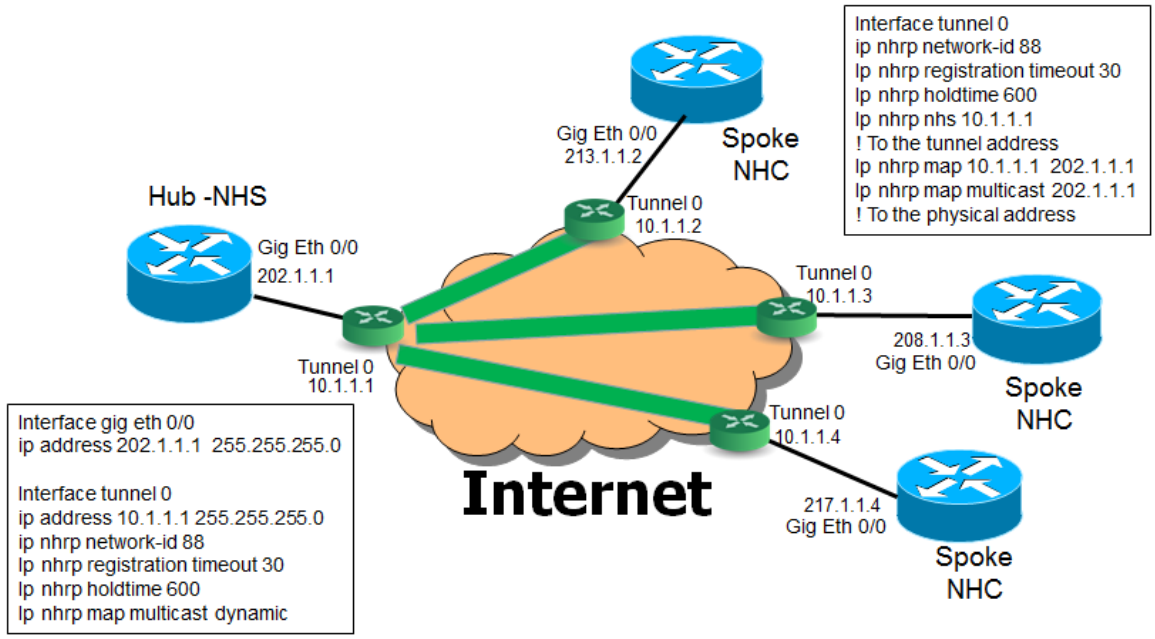


Figure 7: NHRP Configuration Statements

The NHRP commands do the following:

ip nhrp network-id (*number*) -
enables the NHRP on an interface. Use the command in interface configuration mode, in this case the tunnel interface. All NHRP devices, in this case the routers, within one logical NBMA network must be configured with the same network identifier. Number range is 1 to 4294967295.

ip nhrp nhs -
specifies the address of one or more NHRP servers. Use the command in interface configuration mode. The command is configured on the tunnel spokes which identifies to what server (NHS) address each client (NHC) goes to register its address with the NHS. (Notice the NHS itself, the hub router, does not have this command, similar to BGP router reflector configuration, if the reader is familiar with BGP.) Normally, NHRP consults the network layer forwarding/routing table to determine how to forward NHRP packets. Legacy ATM designs used a

separate devise as the next hop server which had a next hop forwarding table similar to an ARP table. If an NHS is configured that is not the hub router then the next hop addresses in the next hop table overrides the forwarding path that would otherwise be used for NHRP traffic. That is not the design deployed in this paper. The network layer forwarding/routing table of the hub router is how NHRP packets are forwarded.

ip nhrp map (*NHS ip address*) (*physical ip address of the hub router*) -
statically configures the IP of the NHS to map to the physical address of the hub router. One probably needs to configure at least one static mapping in order to reach the next hop server. Command is needed so NHRP registration and discovery packets needing to reach the NHS are linked to the routing table. If there are multiple NHSs then this command can be repeated.

ip nhrp map multicast (*physical ip address of the hub router*) **-**
usually configured on a spoke router, this command configures the physical address of the hub router used as a destination for broadcast or multicast packets to be sent over a tunnel network. The command is useful for supporting broadcasts or multicasts over a tunnel network when the underlying physical network does not support IP multicast mode. The Internet does not allow broadcasts and multicast on its physical interfaces so routing protocols like OPSF and EIGRP would not be supported unless they went over a tunnel.

ip nhrp map multicast dynamic -
allows NHRP to automatically add routers to the multicast NHRP mappings. Use this command when spoke routers need to initiate mGRE and IPSec tunnels and register their unicast NHRP mappings. This command is needed to enable routing protocols to work over the mGRE and IPSec tunnels because routing protocols often use multicast packets. This command prevents the hub router from needing a separate configuration line for a multicast mapping for each spoke router.

ip nhrp holdtime (*seconds*) -
changes the number of seconds that NHRP dynamic entries expire. The default is 7,200 seconds (two hours). The NHRP cache can contain static and dynamic entries. The static entries never expire. Cisco recommends that this setting be changed to 600 seconds, or ten minutes, so the holdtime does not have an adverse effect on the NHRP registration and discovery process.

ip nhrp registration timeout (*seconds*) -
sets the time between periodic registration messages. By default this is set to one-third of the value of the NHRP holdtime, or forty minutes. If the defaults of these two settings are not changed and something happens to the mappings on the NHS, the spokes will not attempt to re-register for forty minutes. Thus the maps would work from the spoke to the hub, but not from the hub back to the spoke for forty minutes. A much better practice is the thirty-second re-registration shown in this paper. If a little less overhead is desired, then a sixty-second value may be appropriate. Since the next hop of "Next Hop Registration Protocol" relies on a forwarding/routing table, a routing protocol needs to be configured.

Step three of DMVPN design is to select and configure a routing protocol.

3. Routing Protocols with DMVPNs

Several routing protocols can be used in a DMVPN design, including EIGRP, BGP, OSPF, and RIPv2. EIGRP is often preferred as the dynamic routing protocol because of its conservation of router CPU cycles and network bandwidth, as well as its quick convergence times. EIGRP also provides a range of options for address summarization and default route propagation.

The address space should be summarized as much as possible, and it is a good idea to configure the spokes as EIGRP stub networks although that was not done in this paper. As with all EIGRP networks, the number of neighbors should be limited to ensure the hub router can re-establish communications after a major outage. In

very large EIGRP networks, it may be necessary to adjust the EIGRP hold time to allow the hub more time to recover.

Spoke-to-spoke DMVPN networks present a unique challenge because the spokes cannot directly exchange routing information with one another, even though they are on the same logical subnet. This limitation requires that the hub router advertise subnets from other spokes on the same subnet. This would normally be prevented by split horizon. To fix this use the command:

no ip split-horizon (*eigrp AS*)

at the hub router. In addition, the advertised route must contain the original next hop as learned by the hub router. EIGRP advertises itself as the next hop by default so the command:

no ip next-hop-self (*eigrp AS*)

was added to allow this type of operation.

Note: DMVPN Phase 3 does not need the **no ip next-hop-self (***eigrp AS*) command.

The routing protocol configuration is pretty generic on the spoke side. The hub router needs the above statements on the tunnel interface (see Figure 8).

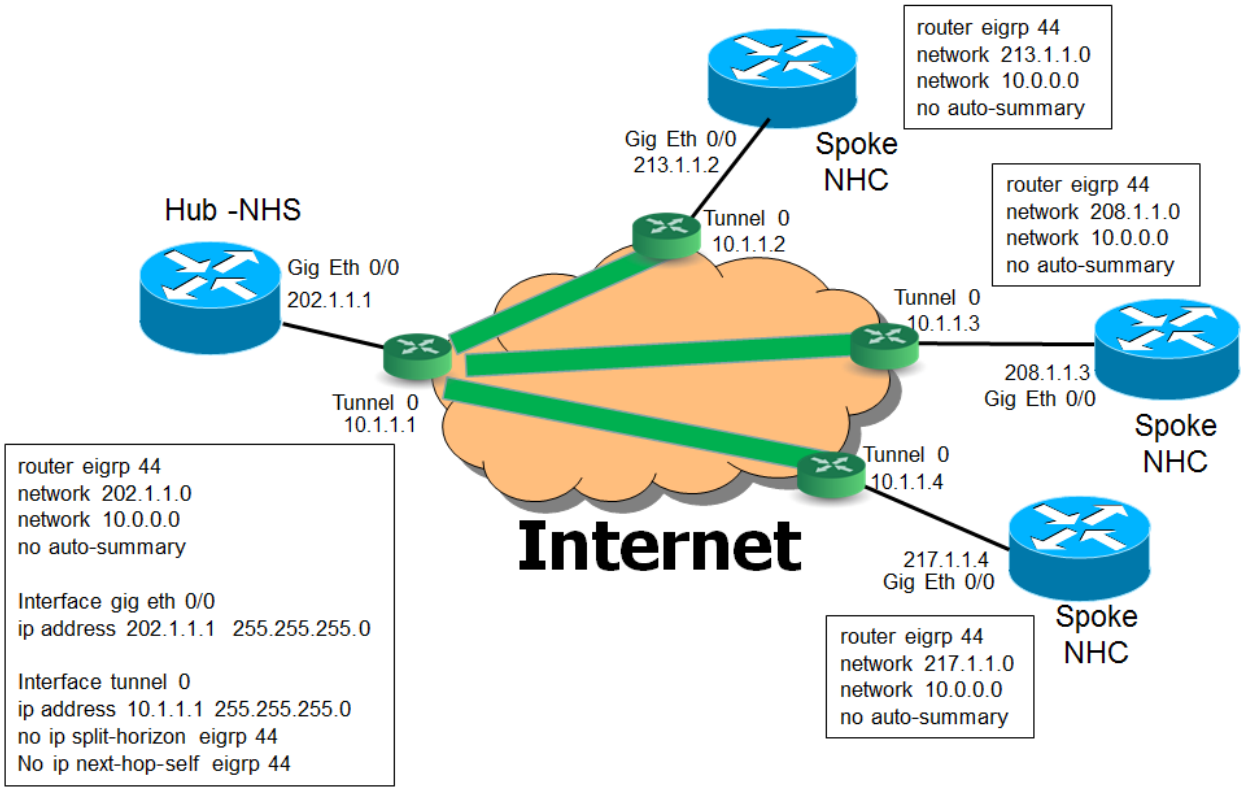


Figure 8: EIGRP Configuration Statements

In this configuration EIGRP is being used. Both the disabling of split horizon and the disabling of next hop marking is required. The reason behind the latter is because, by default, an EIGRP router will advertise routes

with a destination of self (i.e., to get to this network send traffic to me); but in this case EIGRP needs to advertise the original advertising EIGRP router IP address. Without this all of the spokes would still forward traffic through the hub router.

Configuring OSPF over a DMVPN network has some of the same limitations as OSPF over other types of networks. Historically, a single OSPF area should not contain more than fifty routers, and there should not be more than three areas on a router. Although current routers have stronger processors, the additional overhead of encryption and NHRP negates much of this. For this reason, the fifty-router limit per area should be observed. In addition, because only the hub is in direct communications with all of the branches, it must be configured as the designated router (DR) on the DMVPN subnet. There is not typically a backup designated router (BDR).

The mGRE tunnel on the hub router must be configured as an OSPF broadcast network to allow the selection of a DR. Each spoke router is configured with an OSPF priority of 0 to prevent a spoke from becoming the DR. The tunnel IP MTU must match on all GRE interfaces that are OSPF-adjacent. In addition, OSPF areas running over DMVPN should be stubby or totally stubby areas to reduce LSA flooding over the WAN.

mGRE, NHRP, and EIGRP are in place so the spoke routers are now ready to learn the addresses of the other spoke routers even if you scale the number of spokes to a very large number. The process is described below (see Figure 9).

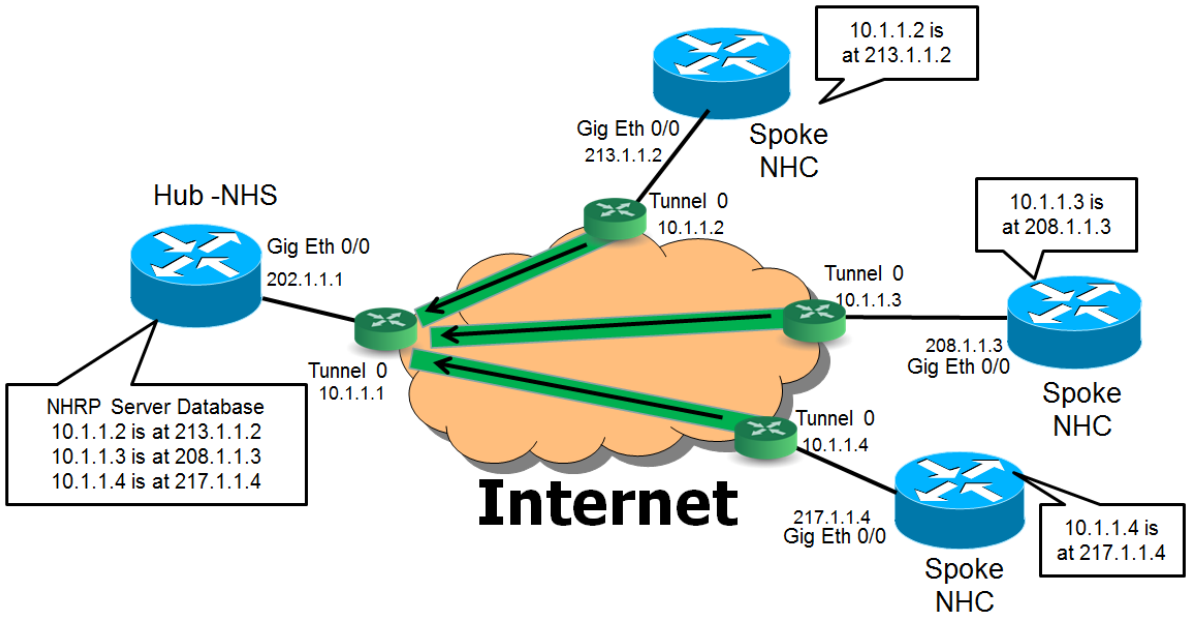


Figure 9: NHCs Register with NHS

As Figure 9 shows, the spoke routers/NHCs register their tunnel address to physical address mapping with the NHS. The NHS creates an NHRP table. Now if an NHC wants to discover the address of another NHC it queries the NHS as show in Figure 10. Once the address is discovered the NHC builds an mGRE tunnel to the other NHC (see Figure 10).

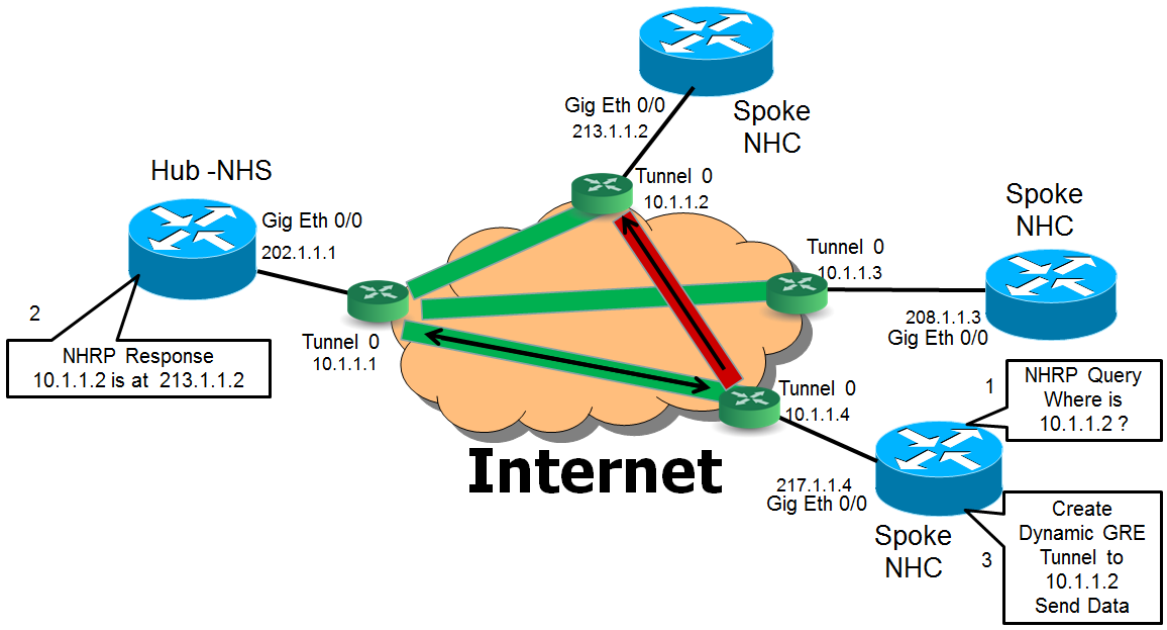


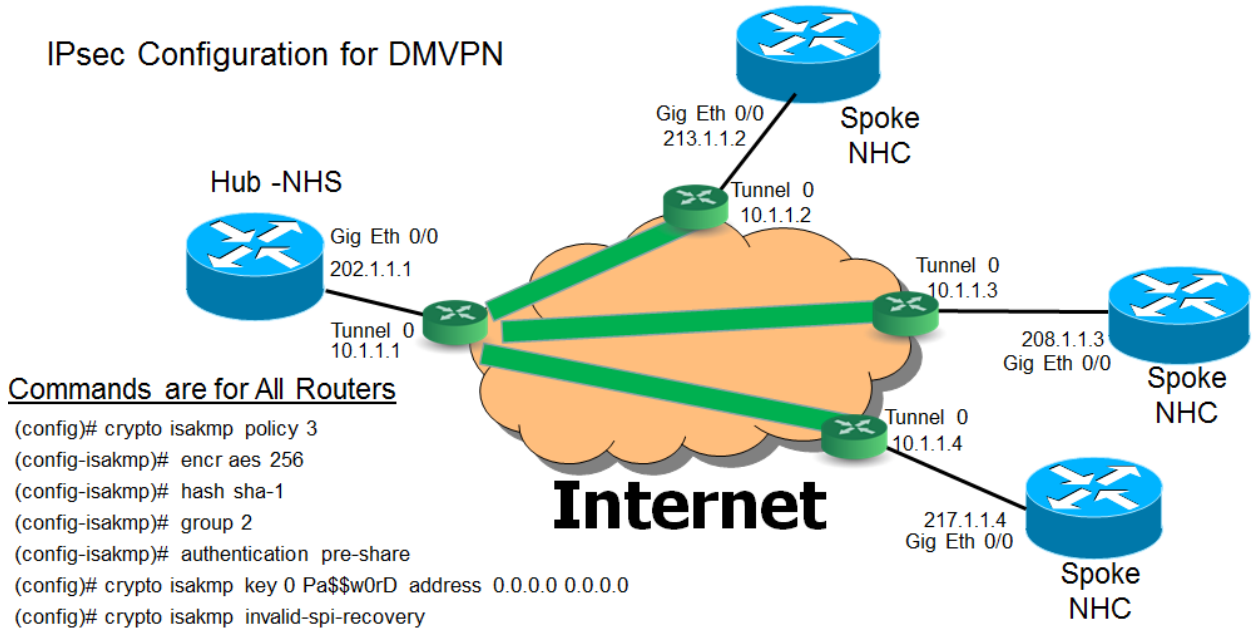
Figure 10: NHC Builds Tunnel to Other NHC

The DMVPN design with mGRE, NHRP, and EIGRP in place is functional, but still not secure as it travels over the public Internet. A crypto process is still needed.

Step four of DMVPN design is to configure IPsec.

4. IPsec Encryption Configuration

This paper's example travels over the Internet so IPsec encryption is necessary (see Figure 11).



Commands are for All Routers

```
(config)# crypto isakmp policy 3
(config-isakmp)# encr aes 256
(config-isakmp)# hash sha-1
(config-isakmp)# group 2
(config-isakmp)# authentication pre-share
(config)# crypto isakmp key 0 Pa$$w0rD address 0.0.0.0 0.0.0.0
(config)# crypto isakmp invalid-spi-recovery
(config)# crypto isakmp keepalive 10

(config)# crypto ipsec security-association lifetime kilobytes 110592000
(config)# crypto ipsec security-association replay disable
(config)# crypto ipsec transform-set ExampleDMVPN esp-aes 256 esp-sha-hmac
(config)# crypto ipsec df-bit clear
(config)# crypto ipsec profile DMVPN-ID88
(ipsec-profile)# set transform-set ExampleDMVPN

(config)# Interface tunnel 0
(config-if)# tunnel protection ipsec profile DMVPN-ID88
```

Figure 11: IPsec Configuration

The isakmp commands for this configuration are described here first.

The Internet Key Exchange (IKE) policy is defined with the command:

crypto isakmp policy *priority*

The priority numbers uniquely identify the IKE policy and assign a priority to the policy. Use an integer from 1 to 10,000, with 1 being the highest priority and 10,000 the lowest. This command invokes the Internet Security Association Key Management Protocol (ISAKMP) policy configuration (config-isakmp) command mode.

While in the ISAKMP policy configuration command mode, some of the commands for which you can specify parameters are as follows:

- authentication ; default = RSA signatures
- encryption (IKE policy) ; default = 56-bit DES-CBC
- group (IKE policy) ; default = 768-bit Diffie-Hellman
- hash (IKE policy) ; default = SHA-1
- lifetime (IKE policy) ; default = 86,400 seconds (one day)

If you do not specify any given parameter, the default value will be used for that parameter.

The crypto isakmp key command is used to configure a preshared authentication key:

crypto isakmp key *enc-type-digit keystring* { **address** *peer-address [mask]* }

enc-type-digit: Specifies whether the password to be used is encrypted or unencrypted.

- 0—specifies that an unencrypted password follows
- 6—specifies that an encrypted password follows

The crypto isakmp invalid-spi-recovery command initiates the IKE security association (SA) to notify the receiving IP Security (IPSec) peer that there is an "Invalid SPI" error:

crypto isakmp invalid-spi-recovery

This command allows the router to be configured so that when an invalid security parameter index error (shown as "Invalid SPI") occurs, an IKE SA is initiated to recognize the "Invalid SPI" situation. The IKE module then sends an "Invalid Error" message to the packet-receiving peer so that resynchronization of the security association databases (SADBs) of the two peers can be attempted. As soon as the SADBs are resynchronized, packets are no longer dropped.

The crypto isakmp keepalive command allows the router to send dead peer detection (DPD) keepalive messages to the peer:

crypto isakmp keepalive *seconds [retry-seconds]* [**periodic**]

When the periodic keyword is used, this argument is the number of seconds between DPD messages; the range is from ten to 3,600 seconds. (Optional) Number of seconds between DPD retry messages if the DPD retry message is missed by the peer; the range is from two to sixty seconds. Once one DPD message is missed by the peer, the router moves to a more aggressive state and sends the DPD retry message at the faster retry interval, which is the number of seconds between DPD retries if the DPD message is missed by the peer. The default DPD retry message is sent every two seconds. Five aggressive DPD retry messages can be missed before the tunnel is marked as down.

The IP sec commands for this configuration are described next.

Changing the global lifetime values when negotiating IPsec security associations requires the command:

crypto ipsec security-association lifetime { seconds *seconds* | kilobytes *kilobytes* }

The defaults are {3,600 seconds = one hour} and {4,608,000 kilobytes which is about one hour if the throughput is 10 megabits per second}. (4,608,000,000 bytes X 8 bits/byte ÷ 10,000,000 bits/seconds = 36,86 seconds = 1 hour). The lifetime in this example above is 110,592,000 kilobytes or twenty-four hours. There are two lifetimes: a "timed" lifetime and a "traffic-volume" lifetime. The security association expires after the first of these lifetimes is reached.

The crypto ipsec security-association replay disable command disables anti-replay checking.

crypto ipsec security-association replay disable

There are multiple transforms that can be set as shown by the two stated above in this example. To define an IPsec transform set, an acceptable combination of security protocols and algorithms uses the crypto ipsec transform-set command.

crypto ipsec transform-set *transform-set-name transform1 [transform2] [transform3] [transform4]*

One may want to use the clear setting for the DF bit when encapsulating IPSec traffic in tunnel mode. Then packets larger than the available maximum transmission unit (MTU) size can be sent, or if you are unsure what the available MTU size is, the packets will still get through. To set the DF bit for the encapsulating header in tunnel mode to all interfaces, use the crypto ipsec df-bit command in global configuration mode.

crypto ipsec df-bit [clear | set | copy] The default setting is copy.

An IPsec profile gathers the IPsec policy settings into a single profile that can be used in other parts of the Cisco IOS configuration. The IPsec profile shares most of the same commands with the crypto map configuration, but only a subset of the commands are valid in an IPsec profile.

crypto ipsec profile *name*

After enabling this command, the only parameter that must be defined under the profile is the transform set via the set transform-set command.

set-transform-set —Specifies a list of transform sets in order of priority.

Other optional commands may be enabled, but they are often not needed. The crypto IP sec profile in tunnel protection mode is then applied on the tunnels interfaces of each router.

DMVPN Has Been Updated with Phase 3:

Phase 3 DMVPN design is more scalable than Phase 2 and has some configuration changes. Phase 3 reduces latency during DMVPN setup for direct spoke-to-spoke connections and improves resilience in case of hub failures and allows for hierarchical hub design. Here is a list of the differences between Phase 2 and Phase 3 DMVPN.

Forwarding in Phase 2 relies exclusively on IP routing table (RIB). Whatever IP next hop is in the routing table is copied to the forwarding table (FIB). In **Phase 3,** there's an NHRP redirect cache in every router. FIB entries are copied from the routing table, but the next hop in the FIB table doesn't necessarily reflect the actual next hop. The next hop may be overridden by a dynamic NHRP entry. This functionality allows direct spoke-to-spoke traffic even if the spokes have no more than a default route toward the hub router. This is accomplished in Phase 3 by the use of "**NHRP Shortcut Switching"** in the Cisco Express Forwarding output switching path. For each data packet that is forwarded out the multipoint Generic Routing Encapsulation (mGRE) interface, NHRP performs a lookup in its mapping table to find an entry for the destination IP address of the data packet. If there is one, it overrides the adjacency determined by Cisco Express Forwarding during the FIB/Adjacency lookup as described before. The next hop router that appears in the FIB table or NHRP cache isn't used unless there's an already established IPsec session with that next hop router. Otherwise, the packet is sent toward the hub router.

The Phase 3 DMVPN configuration changes are needed to enable NHRP shortcut switching. All spokes need to have the commands:

ip nhrp shortcut
ip nhrp redirect

added to their tunnel interfaces. For the hubs use only **ip nhrp redirect**.

For EIGRP, in the hub side only, remove:

no ip next-hop-self eigrp *eigrp AS#*

from the hub tunnel configuration. Leave:

no ip split-horizon eigrp *eigrp AS#*

in the hub tunnel configuration. Add as needed for the summary of the spoke subnets:

ip summary-address eigrp *eigrp AS#* **(**summary address summary mask**)**

For OSPF, for all hubs and spokes, change from:

ip ospf network broadcast

to:

ip ospf network point-multipoint.

Conclusion

Once Multipoint Generic Routing Encapsulation, Next Hop Resolution Protocol, a routing protocol, and IP sec encryption are configured a DMVPN network design is complete. This design allows remote sites/spokes in a "Hub and Spoke" or "Star" VPN router topology to connect to each other directly without sending the traffic/data packets through the Hub.

Learn More

Learn more about how you can improve productivity, enhance efficiency, and sharpen your competitive edge through training.

IINS 2.0 - Implementing Cisco IOS Network Security

ASA e-Camp v2.0 (FIREWALL 2.0 + VPN 2.0)

FIREWALL 2.0 - Deploying Cisco ASA Firewall Solutions

IPS - Implementing Cisco Intrusion Prevention System v7.0

SASAA - Implementing Advanced Cisco ASA Security v1.2

SASAC - Implementing Core Cisco ASA Security v1.0

SSFAMP - Securing Cisco Networks with Sourcefire FireAMP Endpoints

SSFAMP - Securing Cisco Networks with Sourcefire FireAMP Endpoints

SSFRULES - Securing Cisco Networks with Snort® Rule Writing Best Practices

SSFSNORT - Securing Cisco Networks with Open Source Snort®

VPN 2.0 - Deploying Cisco ASA VPN Solutions

Visit **www.globalknowledge.com** or call **1-800-COURSES (1-800-268-7737)** to speak with a Global Knowledge training advisor.

Resources

1. Cisco IOS IP Addressing Services Command Reference, www.cisco.com
2. Migrating from Dynamic Multipoint VPN Phase 2 to Phase 3: Why and How to Migrate to the Next Phase, www.cisco.com
3. Cisco IOS Software Releases 12.4 Mainline, Configuring NHRP, www.cisco.com
4. Introduction to Multipoint GRE and NHRP by Sean Wilkins, October 3, 2012, www.sr-wconsulting.com
5. DMVPN Phase 3 by Petr Lapukhov
6. Spoke-to-spoke IP multicast over DMVPN? By Ivan Pepelnjak, Wednesday, April 20, 2011, http://blog.ipspace.net
7. GNS3 Lab–Introduction to DMVPN by Peterson Amar Apr 28 2014 linkedin.com/in/petersonamar

About the Author

Bill Treneer has taught Cisco courses for 17 years.