LARGE SCALE DYNAMIC MULTIPOINT VPN

NOVEMBER 2004
Dynamic Multipoint VPN Facts

- Dynamic Multipoint VPN (DMVPN) can work with static routes but shows its power with routing protocols.
- The routing protocol consumes a lot of CPU with so many neighbors.
- Resource consumption increases with the number of tunnels.
IPsec facts

- IPsec maximum throughput is better with large packets
- On medium and low platforms, CPU is impacted by large SADB
- Cisco recommends that users keep a DMVPN hub within reasonable limits
- Consult your Account Team about platform details
Example – Cisco 7200 Series/VAM2

- The Cisco 7200 Series Router is a popular platform for DMVPN
- It can accept a maximum of 375 tunnels without particular attention (EIGRP)
- In that case, the max throughput would be
  42,000 pps for 64 bytes packets
  22,000 pps for 1400 bytes packets
Scaling the Cisco 7200 Series/VAM2 Further

- If a second mGRE interface is set up on the Cisco 7200 Series Router, it can accept a maximum of 350 tunnels per interface (700 total)
- In that case the max throughput is:
  - 40,000 pps for 64 bytes packets
  - 22,000 pps for 1400 bytes packets
- A third interface does not improve things
Is This Low?

- Yes and no
- The theoretical maximum number of tunnels (Cisco 7200 Series / VAM2) is 5,000 so DMVPN looks bad
- The theoretical max speed is 250Mbps so DMVPN looks the same
- 250Mbps/700 = 350Kbps per spoke
- Not very useful below that throughput anyway
Remarks

• This presentation describes current performance
• Performances change every day and protocols evolve
• Check with your account team to evaluate the best DMVPN platform for your needs
• It is possible to scale DMVPN very high
  Just wait for the next chapter…
Summary on DMVPN Fitness

- If many spokes with very low IPsec throughput, DMVPN may not be a good fit.
- DMVPN starts to become useful at the edge between remote-access and lan-to-lan.
- DMVPN works best for spokes that need statistically constant equal access to central resources.
  Small offices, branch offices, hot-spots, administrations, schools.
- Many existing remote-access or LAN to LAN solutions should actually be DMVPN like networks.
- DMVPN shows a network with integrated security.
APPLICATION TO LARGE SCALE IPSEC
Problem description

- Need to deploy a large DMVPN network
  Any number 700+; tens of thousands allowed
  More than just basic connectivity needed
- Limited to hub and spoke
- Spoke to spoke via the hub is allowed
Requirements

• Constraints
  LAN to LAN
  Dynamic IP addresses

• Solution must:
  Be easy to manage (deployment and monitoring)
  Recover by itself
  Scale to thousands of spokes
  Allow Cisco rich features (ie: Cisco IOS® Intrusion Prevention System (IPS), Cisco IOS Firewall)
Overall Solution

Edge of HQ

Cluster of DMVPN hubs
Aggregates user tunnels
Cluster can be heterogeneous

SLB balances connections
Owns virtual IP address

Spokes (83x)
DMVPN based
Provide QoS
And Firewalling

GRE/IPsec tunnels
IGP + NHRP

No special software needed on PC
IP phones work out of the box
The Load Balancer In General

- Load Balancer owns a Virtual IP Address (VIP)
- When IKE or ESP packets are targeted at the VIP, the LB chooses a hub
- The hub choice is policy (predictor) based:
  - Weighted round-robin
  - Least-connections
- Once a decision is made for a “tunnel”, all subsequent packets go to the same hub (stickyness)
- Once a decision is made for IKE, the same is made for ESP (buddying)
High Level Description

- Spokes think there is a single hub
- They have an NHRP map pointing to the Load Balancer’s Virtual IP Address
- The Load Balancer is configured in forwarding mode (no NAT)
- All the hubs have the same configuration
  - Same Tunnel interface address
  - Same Loopback address (= VIP)
Topography with Addresses

Load Balancer
VIP: 172.17.0.1
(no tunnel)

- Spoke A: 192.168.1.1/24
  - Physical: (dynamic) 172.16.1.1
  - Tunnel0: 10.0.0.11

- Spoke B: 192.168.2.1/24
  - Physical: (dynamic) 172.16.2.1
  - Tunnel0: 10.0.0.12

- 192.168.128.1/25
- 10.1.0.0/24
- 10.1.1.0/24
Spoke Configuration

- The spoke configuration is the same as with a single hub
- It has an NHRP map
  
ip nhrp map 10.0.0.1 172.17.0.1
Load Balancer

• We will study Cisco IOS Software SLB
  Runs on most Cisco IOS Software platforms, including the Cisco Catalyst® 6500 Series Switch
  Opt for Releases 12.2S or 12.1E
• CSM 3.1 or above should work too but we do not need most of its features (useless)
• Load balancing must be able to do Layer 3 and 4 load balancing
  Upper layers are useless (encrypted)
Cisco IOS Software SLB performances

- Cisco IOS Software SLB on a Cisco Catalyst 6500 Series Switch (MSFC-2)
  - Can manage 1M connections w/ 128MB RAM
  - Can create 20,000 connections per second
  - Switches packets at 10Gbps (64 bytes)

- Cisco IOS Software SLB on a Cisco 7200 Series Router (NPE-400)
  - Can create 5,000 connections per second
  - Switches packets at ½ the Cisco Express Forwarding rate (depending on other features)

- Should not be a bottleneck
Cisco IOS Software SLB cluster definition

ip slb probe PINGREAL ping
  faildetect 2

ip slb serverfarm HUBS
  failaction purge
  probe PINGREAL

! predictor round-robin

real 10.1.0.2
  weight 4
  inservice

real 10.1.0.3
  weight 4
  inservice

Weighted round-robin
This is the default

If all the hubs are equivalent, the weight is the same
Cisco IOS Software SLB VIP definition

ip slb vserver ESPSLB
virtual 172.17.0.1 esp
serverfarm HUBS
sticky 60 group 1
idle 30
inservice

ip slb vserver IKESLB
virtual 172.17.0.1 udp isakmp
serverfarm HUBS
sticky 60 group 1
idle 30
inservice
Monitoring and managing

SLB-7200#sh ip slb connections

<table>
<thead>
<tr>
<th>vserver</th>
<th>prot</th>
<th>client</th>
<th>real</th>
<th>state</th>
<th>nat</th>
</tr>
</thead>
<tbody>
<tr>
<td>IKESLB</td>
<td>UDP</td>
<td>64.103.8.8:500</td>
<td>10.1.0.2</td>
<td>ESTAB</td>
<td>none</td>
</tr>
<tr>
<td>ESPSLB</td>
<td>ESP</td>
<td>217.136.116.189:0</td>
<td>10.1.0.2</td>
<td>ESTAB</td>
<td>none</td>
</tr>
<tr>
<td>IKESLB</td>
<td>UDP</td>
<td>213.224.65.3:500</td>
<td>10.1.0.2</td>
<td>ESTAB</td>
<td>none</td>
</tr>
<tr>
<td>ESPSLB</td>
<td>ESP</td>
<td>80.200.49.217:0</td>
<td>10.1.0.2</td>
<td>ESTAB</td>
<td>none</td>
</tr>
<tr>
<td>ESPSLB</td>
<td>ESP</td>
<td>217.136.132.202:0</td>
<td>10.1.0.3</td>
<td>ESTAB</td>
<td>none</td>
</tr>
</tbody>
</table>

SLB-7200#clear ip slb connections ?

- firewallfarm: Clear connections for a firewallfarm
- serverfarm: Clear connections for a specific serverfarm
- vserver: Clear connections for a specific virtual server

SLB-7200#sh ip slb reals

<table>
<thead>
<tr>
<th>real</th>
<th>farm name</th>
<th>weight</th>
<th>state</th>
<th>conns</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1.0.2</td>
<td>HUBS</td>
<td>4</td>
<td>OPERATIONAL</td>
<td>4</td>
</tr>
<tr>
<td>10.1.0.3</td>
<td>HUBS</td>
<td>4</td>
<td>OPERATIONAL</td>
<td>1</td>
</tr>
</tbody>
</table>
Hub Tunnel configuration

interface Tunnel0
bandwidth 10000
ip address 10.0.0.1 255.255.0.0
no ip redirects
ip mtu 1350
ip nhrp map multicast dynamic
ip nhrp network-id 1
ip nhrp holdtime 3600
no ip split-horizon
no ip mroute-cache
tunnel source Loopback0
tunnel mode gre multipoint
tunnel key 1
tunnel protection ipsec profile tp
end

interface Loopback0
ip address 172.17.0.1 255.255.255.255
end

interface FastEthernet0/0
ip address 10.1.0.{2,3} 255.255.255.0
interface FastEthernet0/1
ip address 10.2.0.{2,3} 255.255.255.0

Must be same on all
Mask is /32

Mask allows $2^{16}-2$ nodes
Routing protocols

- HQ Speaks EIGRP 2
  - Redistribute EIGRP 2 into BGP (summary)
  - Redistribute floating static (Null0) into EIGRP2
  - Redistribute EIGRP 1 into BGP (with filtering)
  - Redistribute BGP (summarized) into EIGRP 1

- Spokes are EIGRP 1 stubs
  - They speak to hubs thru GRE/IPsec tunnel
Hub Routing protocol configuration

```
router eigrp 1
  redistribute bgp 1 metric 1 0 255 20 1400
  network 10.0.0.0 0.0.255.255
  default-metric 64 2000 255 1 1400
  no auto-summary
router bgp 1
  bgp router-id 10.2.0.{2,3}
  bgp log-neighbor-changes
  neighbor 10.0.0.1 remote-as 1

  address-family ipv4
  redistribute eigrp 1 route-map <IGPREDIST>
  neighbor 10.2.0.1 activate
  neighbor 10.2.0.1 next-hop-self
  no auto-summary
  no synchronization
  bgp redistribute-internal
  exit-address-family
```
router bgp 1
  no synchronization
  bgp log-neighbor-changes
  aggregate-address 10.0.0.0 255.0.0.0 summary-only
  aggregate-address 192.168.0.0 255.255.0.0 summary-only
  redistribute eigrp 2
  neighbor HUB peer-group
  neighbor HUB remote-as 1
  neighbor HUB next-hop-self
  neighbor 10.0.0.2 peer-group HUB
  neighbor 10.0.0.3 peer-group HUB
  no auto-summary
Edge router EIGRP configuration

- EIGRP 2 attracts spoke subnets to the edge router
- Floating static route to Null0 discards packets to unconnected spokes

```plaintext
ip route 192.168.0.0 255.255.255.127 Null0 254

router eigrp 2
    redistribute static
    network 192.168.1.0 0.0.0.128
    no auto-summary
    no eigrp log-neighbor-changes
```
Packet Flow
Result

- Tunnels reconnect automatically
- Working sessions are not lost
- QoS allocates bandwidth to voice
- All other features are available
- No need to touch the hubs while adding a spoke
- New hubs can be added/removed on the fly
- Simple to deploy
- Leverages monitoring infrastructure (interfaces, CDP)
Support

- Each feature has plenty of nerd knobs for tuning
- Each feature has advanced debugging capabilities
- Each feature can be troubleshooting independently
IGP choices

- **BGP** between Hubs and Edge is good due to number of prefixes and flexibility
- Scaling the IGP between hubs and spokes is the hardest part
- A distance vector is recommended
- EIGRP shows best results so far but ODR is under test (lightweight)
Positioning

• The main advantages of the solution are:
  - Virtually **limitless** scaling
  - Can be deployed in zero touch with ISC and Intelligent Engine
  - Automatic load management
  - Load balancing **AND** resilience
  - Multiply performances by number of hubs (creation rate, speed, max SA’s)
  - No forklift when upgrading
  - Resilience in **N+1**
Improvements

• It is possible to collapse the Load balancer and the edge router (hubs in lollipop)

• If the load balancer is a Cisco Catalyst 6500 Series Switch, this is even recommended as Layer 3 switching will accelerate spoke to spoke traffic
Routing Protocols

- HQ Speaks EIGRP 2
- Redistribute EIGRP 2 into BGP (summary)
- Redistribute floating static (Null0) into EIGRP2
- Redistribute EIGRP 1 into BGP (with filtering)
- Redistribute BGP (summarized) into EIGRP 1

Spokes are EIGRP 1 stubs
They speak to hubs thru GRE/IPsec tunnel