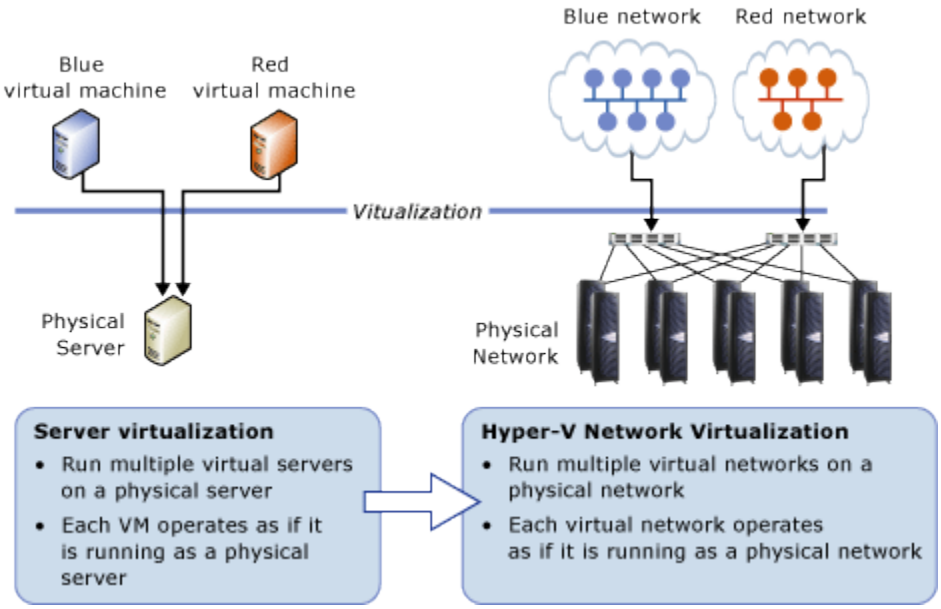
Network Virtualization technical details

Applies To: Windows Server 2012

Server virtualization enables multiple server instances to run concurrently on a single physical host; yet server instances are isolated from each other. Each virtual machine essentially operates as if it is the only server running on the physical computer. Network virtualization provides a similar capability, in which multiple virtual network infrastructures run on the same physical network (potentially with overlapping IP addresses), and each virtual network infrastructure operates as if it is the only virtual network running on the shared network infrastructure. The following figure shows this relationship.



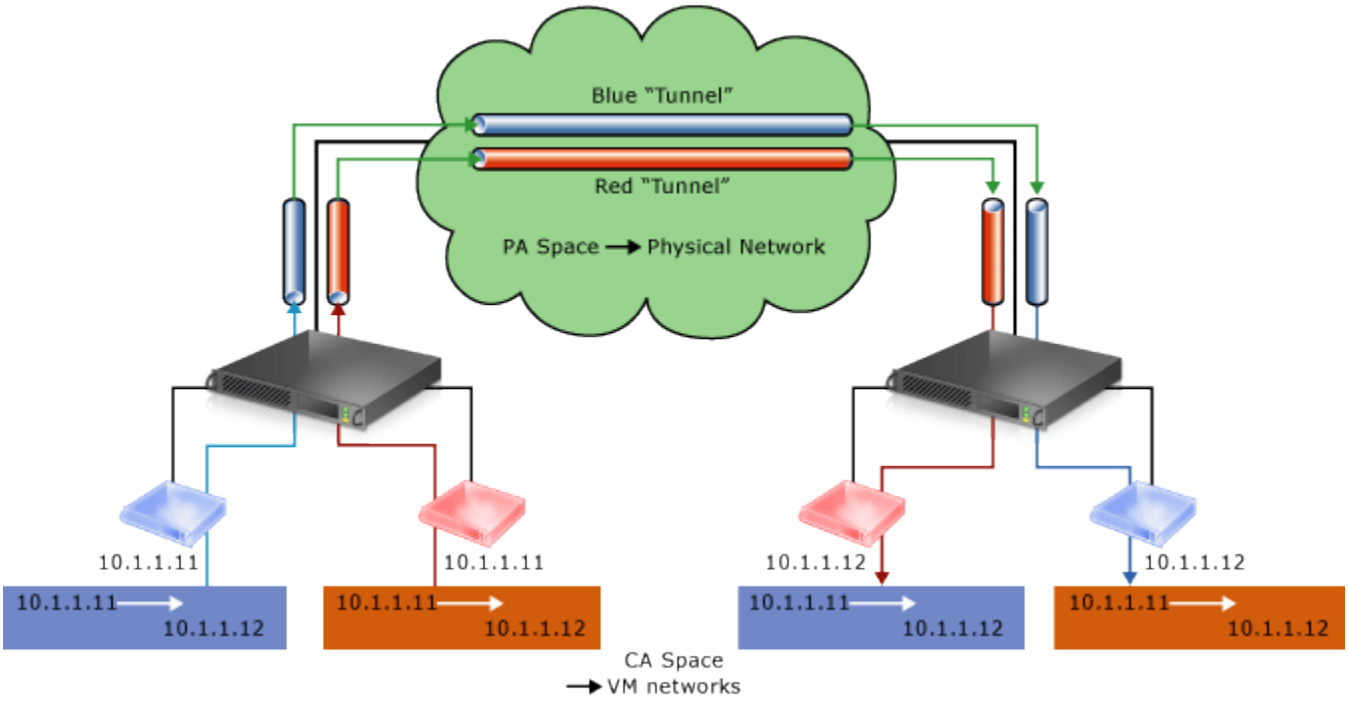
Each virtual network adapter in Hyper-V Network Virtualization is associated with two IP addresses:

- **Customer Address** (CA) The IP address that is assigned by the customer, based on their intranet infrastructure. This address enables the customer to exchange network traffic with the virtual machine as if it had not been moved to a public or private cloud. The CA is visible to the virtual machine and reachable by the customer.

- **Provider Address** (PA) The IP address that is assigned by the hoster or the datacenter administrators based on their physical network infrastructure. The PA appears in the packets on the network that are exchanged with the Hyper-V server that is hosting the virtual machine. The PA is visible on the physical network, but not to the virtual machine.

The CAs maintain the customer's network topology, which is virtualized and decoupled from the actual underlying physical network topology and addresses, as implemented by the PAs. The

following diagram shows the conceptual relationship between virtual machine CAs and network infrastructure PAs as a result of network virtualization.



In the diagram, customer virtual machines are sending data packets in the CA space, which traverse the physical network infrastructure through their own virtual networks, or "tunnels". In the example above, the tunnels can be thought of as "envelopes" around the blue and red data packets with green shipping labels (PA addresses) to be delivered from the source host on the left to the destination host on the right. The key is how the hosts determine the "shipping addresses" (PA's) corresponding to the blue and the red CA's, how the "envelope" is put around the packets, and how the destination hosts can unwrap the packets and deliver to the blue and red destination virtual machines correctly.

This simple analogy highlighted the key aspects of network virtualization:

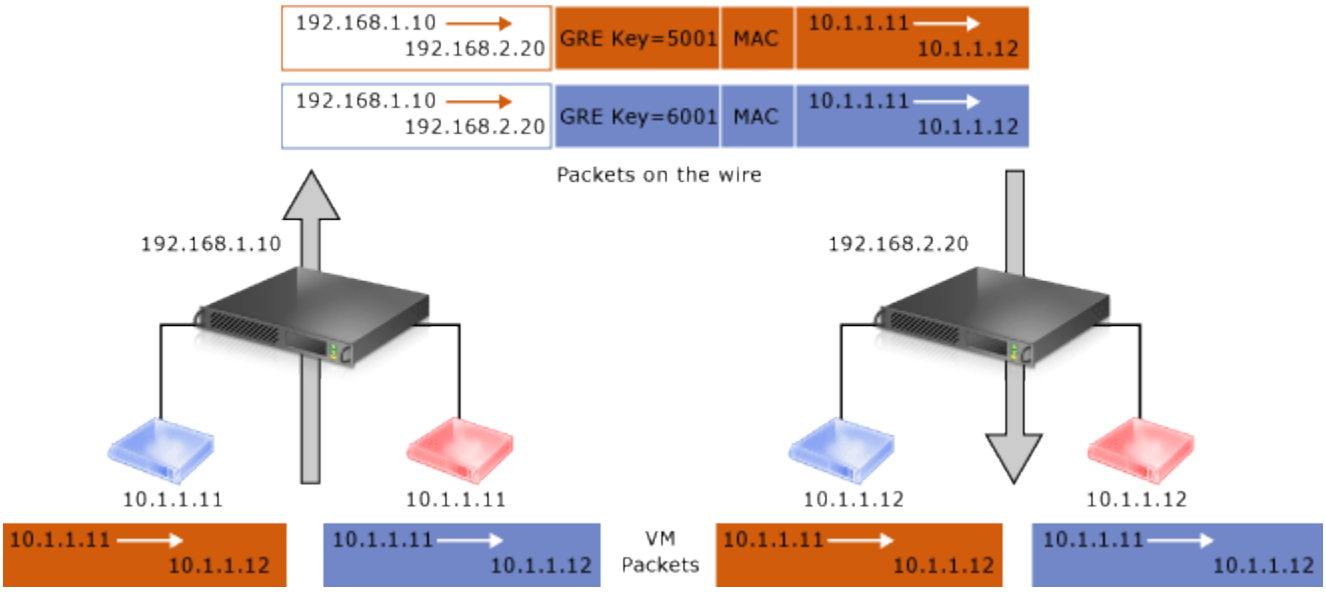
- Each virtual machine CA is mapped to a physical host PA.
- Virtual machines send data packets in the CA spaces, which are put into an "envelope" with a PA source and destination pair based on the mapping.
- The CA-PA mappings must allow the hosts to differentiate packets for different customer virtual machines.

As a result, the mechanism to virtualize the network is to virtualize the network addresses used by the virtual machines. The next section describes the actual mechanisms of address virtualization.

Network virtualization through address virtualization

Hyper-V Network Virtualization supports the following two mechanisms to virtualize the IP address:

Generic Routing Encapsulation The first network virtualization mechanism uses the Generic Routing Encapsulation (NVGRE) as part of the tunnel header. This mode of address virtualization mechanism is intended for the majority of datacenters deploying Hyper-V Network Virtualization. In NVGRE, the virtual machine's packet is encapsulated inside another packet. The header of this new packet has the appropriate source and destination PA IP addresses in addition to the Virtual Subnet ID, which is stored in the Key field of the GRE header, as shown in Figure 3.



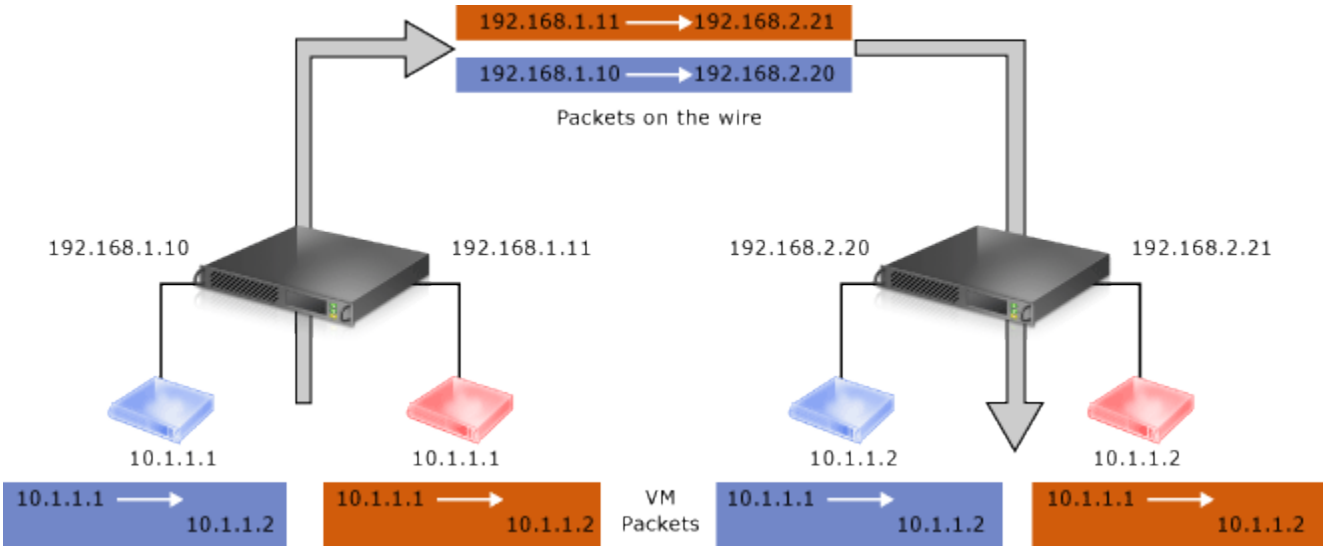
The Virtual Subnet ID included with the GRE header allows hosts to identify the customer virtual machine for any given packet, even though the PA's and the CA's on the packets may overlap. This allows all virtual machines on the same host to share a single PA, as shown in Figure 3.

Sharing the PA has a big impact on network scalability. The number of IP and MAC addresses that need to be learned by the network infrastructure can be substantially reduced. For instance, if every end host has an average of 30 virtual machines, the number of IP and MAC addresses that need to be learned by the networking infrastructure is reduced by a factor of 30.The embedded Virtual Subnet IDs in the packets also enable easy correlation of packets to the actual customers.

With Windows Server 2012, Hyper-V Network Virtualization fully supports NVGRE out of the box; it does NOT require upgrading or purchasing new network hardware such as NICs (Network Adapters), switches, or routers. This is because the NVGRE packet on the wire is a regular IP packet in the PA space, which is compatible with today's network infrastructure.

Windows Server 2012 made working with standards a high priority. Along with key industry partners (Arista, Broadcom, Dell, Emulex, Hewlett Packard, and Intel) Microsoft published a draft RFC that describes the use of Generic Routing Encapsulation (GRE), which is an existing IETF standard, as an encapsulation protocol for network virtualization. For more information, see the following Internet Draft: Network Virtualization using Generic Routing Encapsulation. As NVGRE-aware becomes commercially available the benefits of NVGRE will become even greater.

IP Rewrite The second virtualization mechanism supported by Hyper-V Network Virtualization is IP address rewrite. In this mode, the source and the destination CA IP addresses are rewritten with the corresponding PA addresses as the packets leave the end host. Similarly, when virtual subnet packets enter the end host, the PA IP addresses are rewritten with appropriate CA addresses before being delivered to the virtual machines. Figure 4 shows the virtual machine packets using IP address rewrite. The key difference between IP Rewrite and NVGRE, in addition to the packet format, is that IP Rewrite mode requires a unique PA for each virtual machine CA in order to differentiate virtual machines from different customers using overlapping IP addresses. So IP Rewrite requires one PA per virtual machine CA IP addresses, whereas NVGRE only needs one PA per host.



IP Rewrite is targeted at special scenarios where the virtual machine workloads require or consume very high bandwidth throughput (~10Gbps) on today's existing hardware. Existing network hardware offload technologies such as Large Send Offload (LSO) and Virtual Machine Queue (VMQ) work as expected on existing NICs. These offloads provide significant benefit for network intensive scenarios particularly in a 10 GbE environment. Additionally, multipath routing in the switches (for example, ECMP, Equal-Cost, Multi-Path) continues to work as expected.