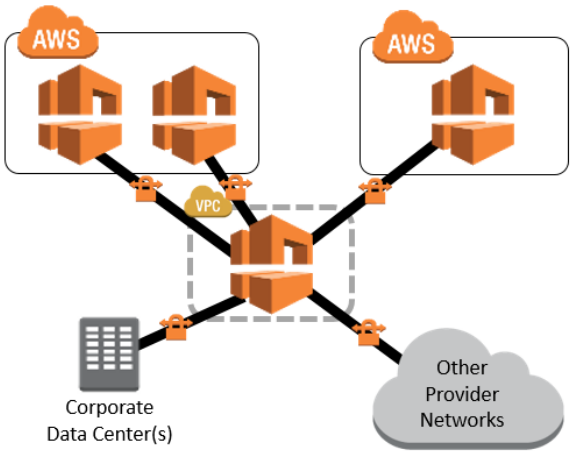
TRANSIT VPC

"How do I build a global transit network on AWS?"

Overview

Amazon Virtual Private Cloud (Amazon VPC) provides customers with the ability to create as many virtual networks as they need, as well as different options for connecting those networks to each other and to non-AWS infrastructure. One common strategy for connecting multiple, geographically disperse VPCs and remote networks is to create a transit VPC that serves as a global network transit center. A transit VPC simplifies network management and minimizes the number of connections required to connect multiple VPCs and remote networks. This design can save time and effort and also reduce costs, as it is implemented virtually without the traditional expense of establishing a physical presence in a colocation transit hub or deploying physical network gear.

This document provides an overview of a transit VPC solution that assumes a typical hub-and-spoke network topology where remote VPCs access each other and remote networks through the transit VPC, as depicted in the diagram to the right. The following sections describe key considerations and recommendations for building a transit VPC and assume basic knowledge of highly available remote-network connectivity,[1] IPsec VPNs, network addressing, subnetting, and routing.



General Best Practices

When creating transit networks, there are some universal network-design principles to consider. For example, the transit network will become a critical component of your network backbone, so choose network vendor products you are familiar with and comfortable supporting. With this in mind, consider the following AWS remote-connectivity best practices:

- Create a dedicated VPC solely for containing your transit network infrastructure. This greatly simplifies routing and failover configurations compared to a single shared services VPC[2] that combines transit network instances with other shared service infrastructure.
- Leverage VPC peering when possible for network connectivity between VPCs to reduce the amount of traffic that must traverse the transit network. This will reduce transit network contention and latency, which can improve application performance.
- Implement non-overlapping network ranges for your private networks to simplify the ability to route between remote networks. Although a transit network can be an excellent place to implement NAT rules to compensate for overlapping networks, this adds additional complexity to the network design.
- Leverage multiple dynamically routed, rather than statically routed, connections to the transit VPC. This allows the transit network infrastructure to automatically fail over between available connections as necessary, creating a highly available, resilient, and more scalable network.

Application on AWS

The following sections provide a high-level overview for creating a dedicated transit VPC to directly route network traffic regardless of where each network is physically located. This approach creates a transitive network using host-based VPN appliances on Amazon Elastic Compute Cloud (Amazon EC2) instances in a dedicated VPC. AWS highly recommends leveraging virtual network appliances from the AWS Marketplace[3] to significantly reduce the level of effort to establish and maintain these VPN connections.

[1] See the *Resources* section for relevant Solution Briefs.
[2] For more information on shared services VPCs, see the Multiple VPC VPN Connection Sharing Solution Brief
[3] For recommended products, search AWS Marketplace for one the following terms: Cisco CSR 1000V, Fortinet FortiGate, Palo Alto Networks, Sophos UTM, Vyatta

A transit VPC is applicable to customers with the following use case/requirements:

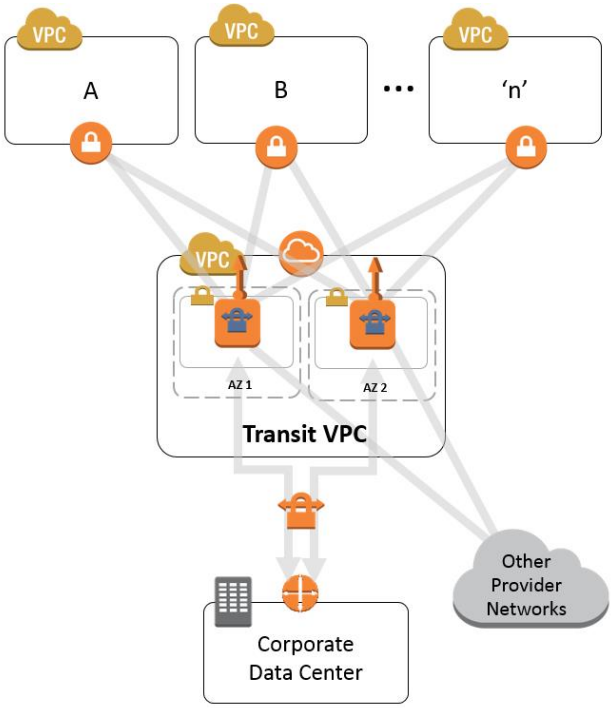
- AWS resources in spoke VPCs need access to a wide variety of on-premises or remote infrastructure
- Spoke VPCs are located in different AWS regions
- Complex network-routing is required to implement a hybrid network architecture
- Security or compliance programs require additional network-based monitoring or filtering between resources in different networks (e.g., Network Intrusion Detection Systems or next-generation firewalls)
- The use of AWS network providers and partner products would reduce high colocation or other physical transit network costs

Configuration Details

This design deploys VPN appliances into a dedicated transit VPC. VPN appliances should be deployed into separate Availability Zones for maximum availability. Spoke VPCs are connected to the transit network through dynamically routed VPN connections between their virtual private gateways (VGWs) and the network appliances. This design uses VPN connections from spoke VPCs, rather than VPC peering to enable routing between any connected network, including external networks or VPCs in other AWS regions. This also allows spoke VPC resources to leverage VGW capabilites for routing and failover in order to maintain highly available network connections to the transit VPC network appliances. Remote networks also connect to the transit VPN appliances using redundant, dynamically routed VPN connections. Once connected, leverage dynamic routing protocols to automatically route traffic around potential network failures as well as to propogate network routes to remote networks.

Note that in the diagram to the right, all communication with the VPN appliances (including the VPN connection between the corporate data center and other provider networks and the transit VPC) uses the transit VPC Internet gateway and Elastic IP addresses. In addition to using dynamically routed connections, AWS highly recommends the use of Auto Recovery for EC2 to protect instances in the transit VPC.

Along with providing direct network routing between VPCs and on-premises networks, this design also enables the transit VPC to implement more complex routing rules, such as network address translation between overlapping network ranges, or to add additional network-level packet filtering or inspection.



Considerations

This design supports any IP-based connectivity requirements between Amazon VPCs and remote resources with minimal on-premises network changes. It also provides an opportunity to select products available on the AWS Marketplace that integrate seamlessly with AWS-provided VPN connections, without the need to deploy these products into existing data centers. However, it does require the customer to configure and manage the EC2-based VPN instances deployed in the transit VPC. This will result in additional EC2 and, potentially, third-party license charges. Also, be aware that this design will generate additional data-transfer charges for traffic traversing the transit VPC: data is charged when it is sent from a spoke VPC to the transit VPC, and again from the transit VPC to the on-premises network.

AWS provides a transit VPC reference implementation for deploying a fully automated Cisco-based transit VPC in minutes. This solution actively monitors a customer's environment for specifically tagged VGWs to automatically join to the transit network. It supports VPCs located in multiple AWS regions and in different AWS accounts. See the implementation guide for detailed information.

Resources

Transit Network VPC (Cisco CSR) Implementation Guide	http://docs.aws.amazon.com/solutions/latest/cisco-based-transit-vpc AWS-provided, automated solution using Cisco Cloud Services Router (CSR) 1000V
Amazon VPC Documentation	https://aws.amazon.com/documentation/vpc/ AWS webpage with links to VPC technical documentation
Auto Recovery for EC2	http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-recover.html EC2 product documentation describing the Auto Recovery for EC2 feature
Auto Scaling Website	https://aws.amazon.com/autoscaling/
Multiple VPC VPN Connection Sharing	http://d0.awsstatic.com/aws-answers/AWS_Multiple_VPC_VPN_Connection_Sharing.pdf AWS Solution Brief describing options for connecting multiple VPCs to on-premises networks using a single VPN connection
Single Data Center HA Network Connectivity	https://d0.awsstatic.com/aws-answers/AWS_Single_Data_Center_HA_Network_Connectivity.pdf AWS Solution Brief describing options for creating highly available connections from a single data center to AWS
Multiple Data Center HA Network Connectivity	https://d0.awsstatic.com/aws-answers/AWS_Multiple_Data_Center_HA_Network_Connectivity.pdf AWS Solution Brief describing options for creating highly available connections from multiple data centers to AWS