Edge scale out with SDDC Multi-Edge in VMware Cloud on AWS

VMware Architecture

Table of contents

Edge scale out with SDDC Multi-Edge in VMware Cloud on AWS

Introduction

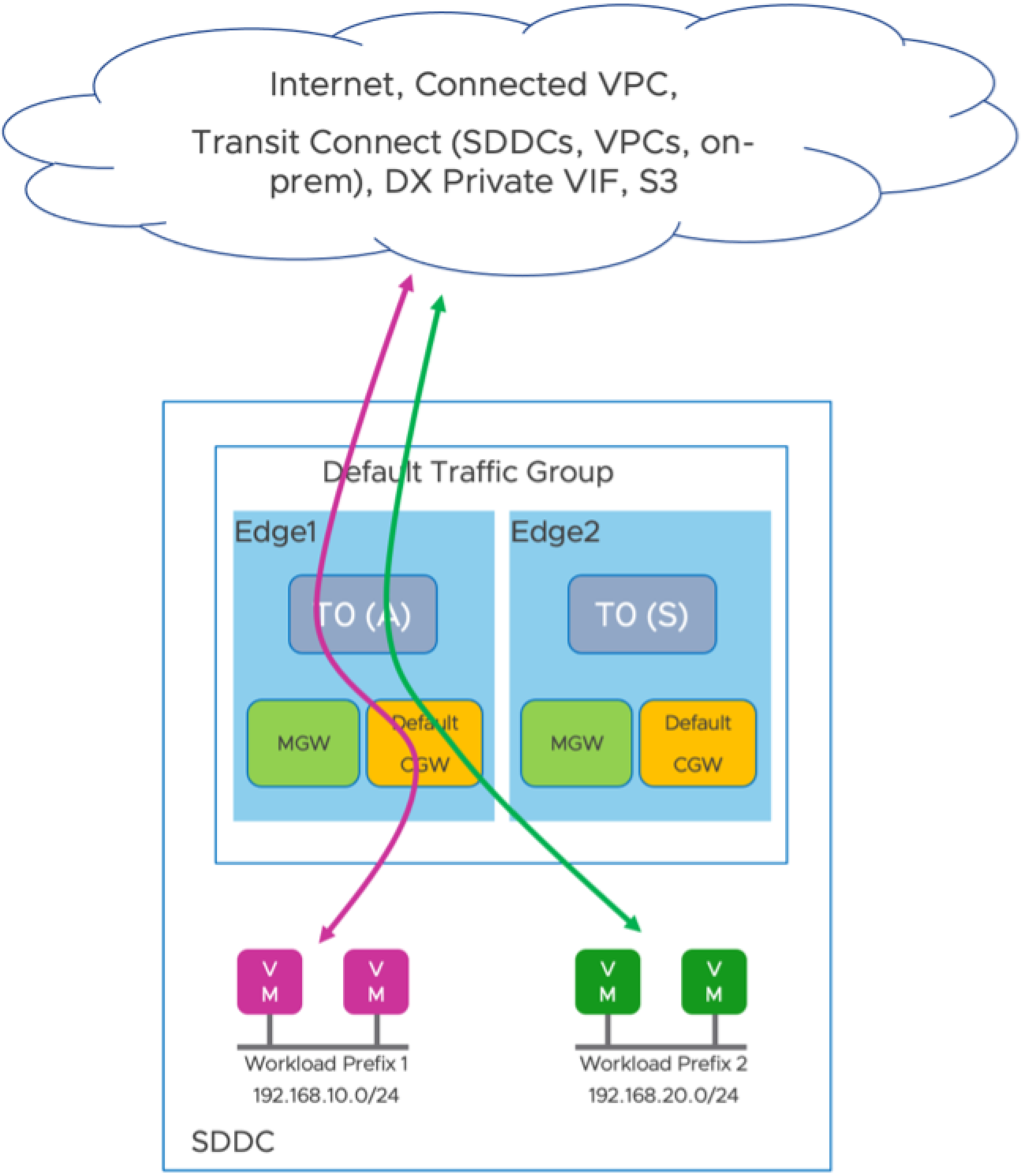
In VMware Cloud on AWS deployments, a SDDC is deployed with a pair of NSX Edge VMs that run in active/standby mode. The active Edge appliance provides the platform on which the default Tier0 (T0) and Tier1 (T1) routers run. North-south traffic for the SDDC workloads goes through the default T0 router. Depending on the type of host the Edge is deployed on, the throughput of the interface on the host can become a bottleneck when consistently high throughput for north-south traffic is required. The SDDC Multi-Edge feature that is supported in VMware Cloud on AWS since version 1.12 can address these concerns.

Summary and Considerations

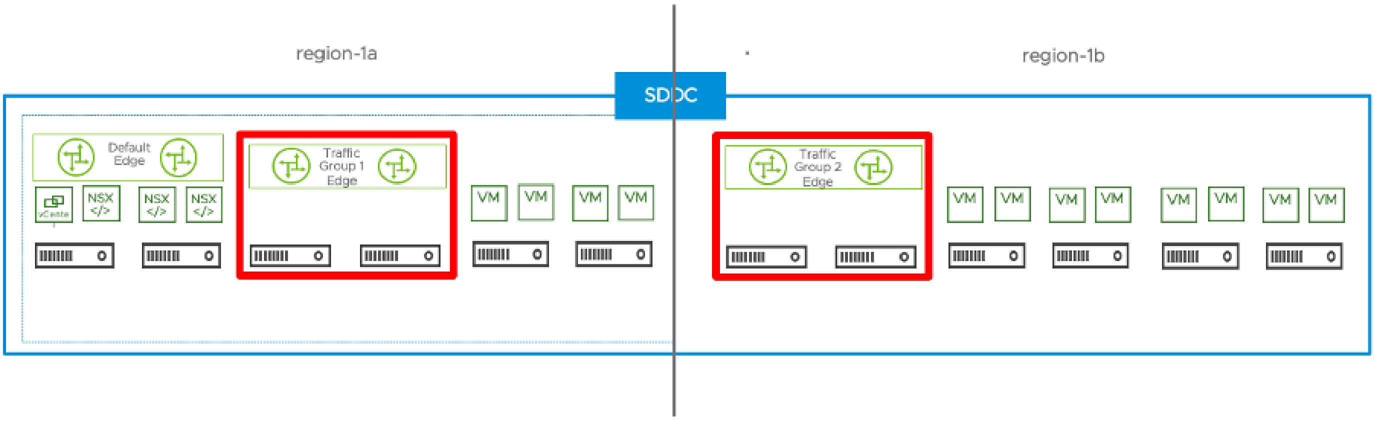
Use Case	Some Large SDDCs have requirements for high north-south throughput and/or traffic isolation requirements. SDDC Multi-Edge could be used to address these use cases.
General Considerations/Recommendations	upsize
Document Reference	demo SDDC Multi-Edge Scaling Out HCX Deployments in a Multi-Edge SDDC VMWare cloud on AWS Management Cluster Planning Understanding VMware Cloud on AWS Network Performance Upsize SDDC Management Appliances
Last Updated	June 2023

SDDC Multi-Edge Overview

As shown in the illustration below, a VMware Cloud on AWS SDDC is deployed with a pair of edge nodes in Active/Standby mode. Traffic flows like vSAN, vMotion and vSphere replication that use vmkernel (vmk) ports do not cross the edge. The edge does not get involved for traffic flows between workloads inside the SDDC. Traffic flows between workload VMs and SDDC Management network will traverse the active edge. Also, all north-south traffic between workload VMs and entities outside SDDC will traverse the active edge. This can be a bottleneck in deployments with high north-south throughput requirements.



With the SDDC Multi-Edge feature also called Edge Scale Out (Active/Standby). You can deploy additional pairs of Edge nodes in sub-clusters called Traffic Groups (TGs). You can enable Multi-Edge feature only on a Large size SDDC. If you have a Medium size SDDC, you must upsize the SDDC to Large. Also note that you will need additional hosts in the management cluster to deploy these additional edge nodes in TGs.

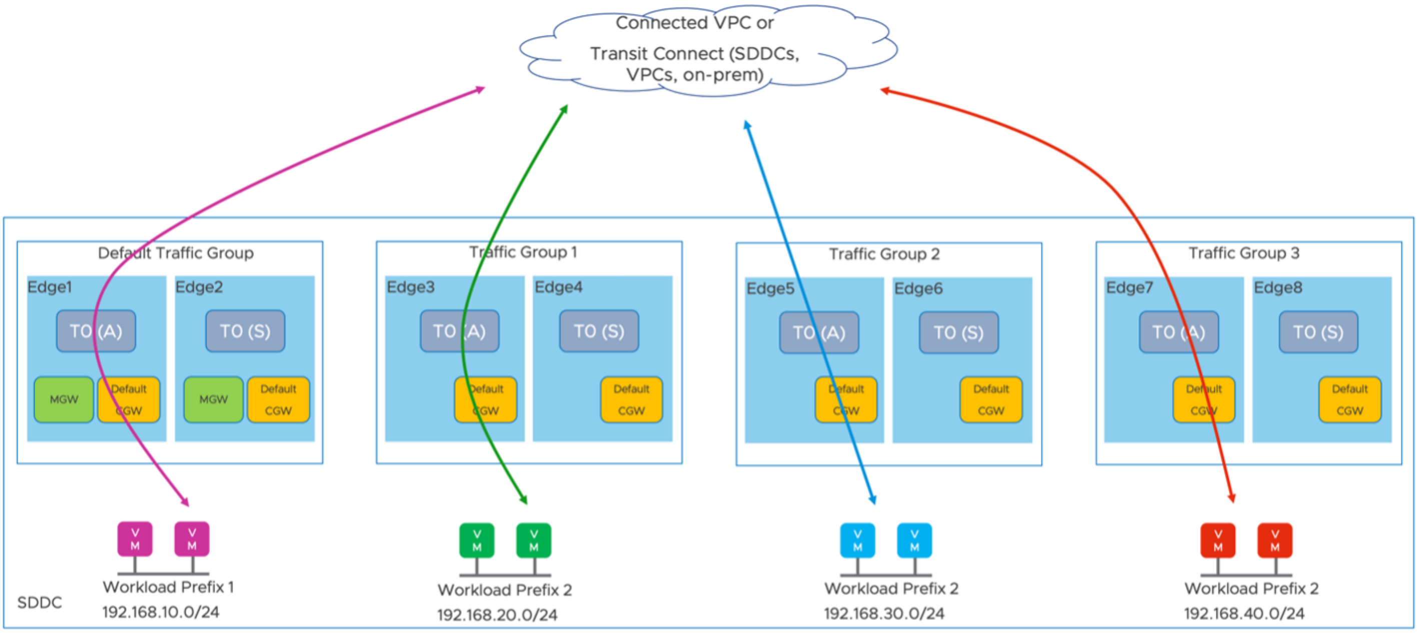


The Edge nodes in the TGs are deployed in Active/Standby mode in each TG. Normally, packets are routed based on destination IP address. The SDDC Multi-Edge feature uses source-based routing (source IP address) for traffic from Virtual Machines (VMs) and destination-based routing for traffic destined to the VMs. You can distinguish traffic based on the source subnet prefixes and create prefix lists. The prefix list can include subnets or individual IP addresses from the workload subnets. You should manually map prefix lists to a TG that should handle traffic for those source prefixes. This will automatically program the AWS routing tables to direct the return traffic to the same TG Edges to ensure symmetric routing.

For outbound traffic from SDDC, the traffic is source routed based on the source IP address of the traffic and sent to the active edge in the mapped TG. The return traffic will be routed based on the destination IP match in the AWS routing table and directed to the same TG edge.

Subnet prefixes that are not explicitly mapped to a TG will be handled by the default active T0.

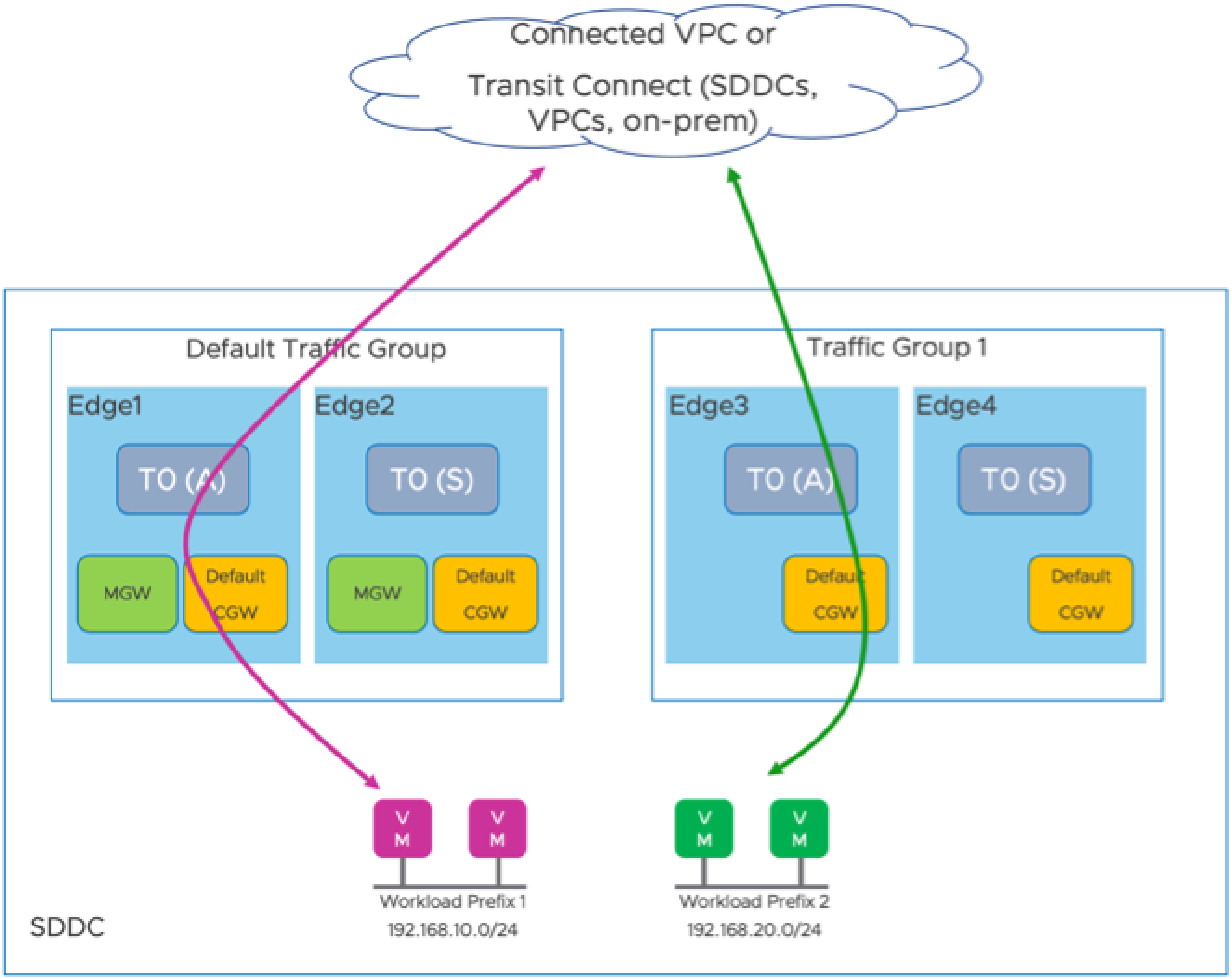
With the SDDC Multi-Edge feature, the firewall rules configured on the gateway firewall will be applied across all TGs. Within each TG, the active TG Edge maintains state information for all north-south traffic crossing that TG. This state information is synchronized between the two edge nodes in each TG. If the active edge node in any TG fails, the standby edge node in that TG can forward traffic without losing state information.



Traffic flows that can take advantage of SDDC Multi-Edge feature

For the following traffic flows, the SDDC Multi-Edge feature increases available north-south throughput capacity by 100% for each additional TG

- Traffic going over Transit Connect to/from other SDDCs, Native AWS VPCs and/or On-premises.
- Traffic to/from Connected VPC

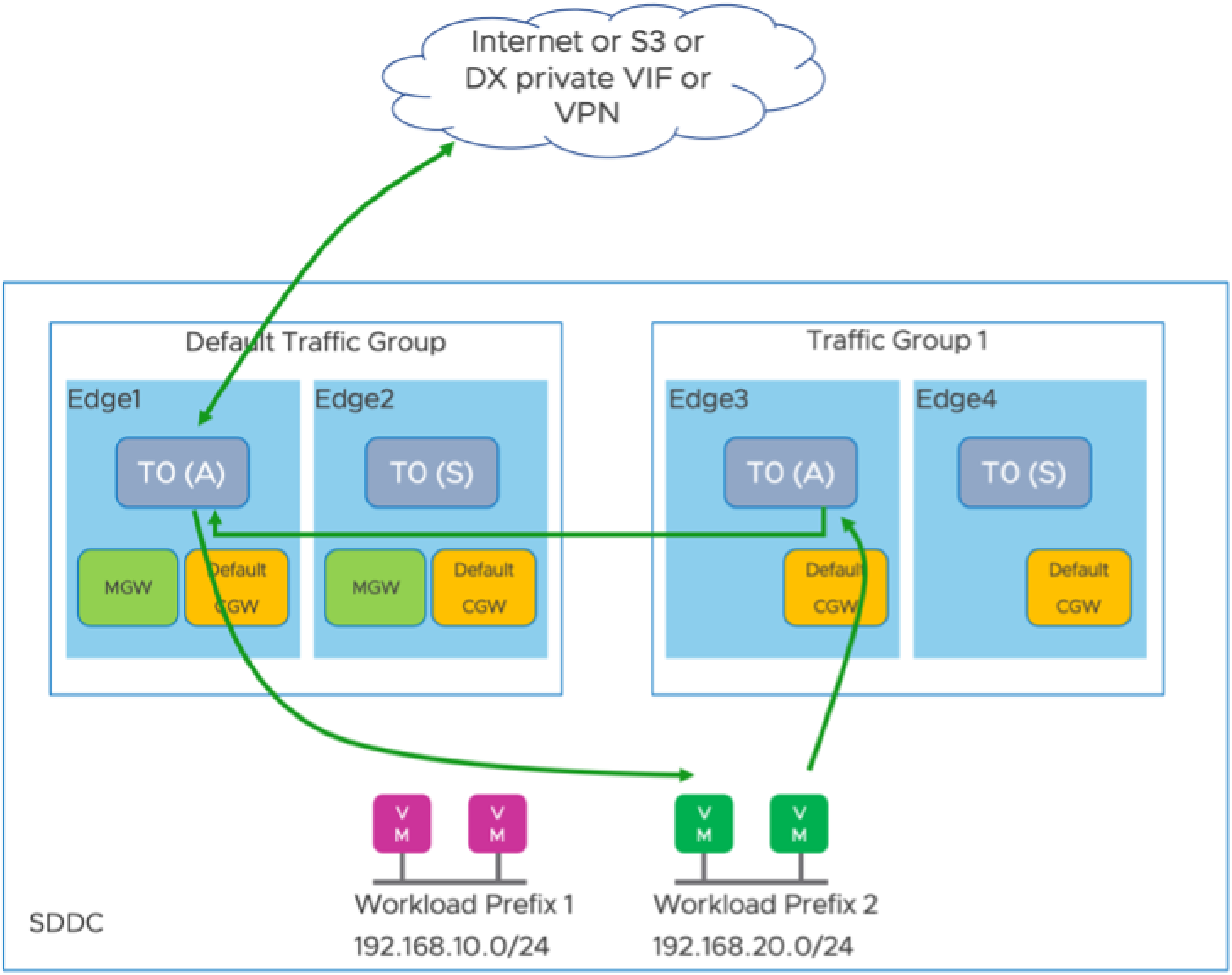


Traffic flows that cannot take advantage of SDDC Multi-Edge feature

Some traffic flows cannot benefit from scaling out to TG edges. This traffic must traverse the active edge in the default traffic group even if the prefix is mapped to a scaled-out TG.

- T0 and T1 VPN traffic, DX (Direct Connect) traffic
- NATted traffic (SNAT/DNAT)
- Traffic to and from the SDDC's native Internet connection
- Traffic to Amazon private or public S3 Gateway endpoint
- Subnet prefixes behind customer configured CGW (Multi-CGW feature)
- Traffic flows that use the Management gateway (MGW)

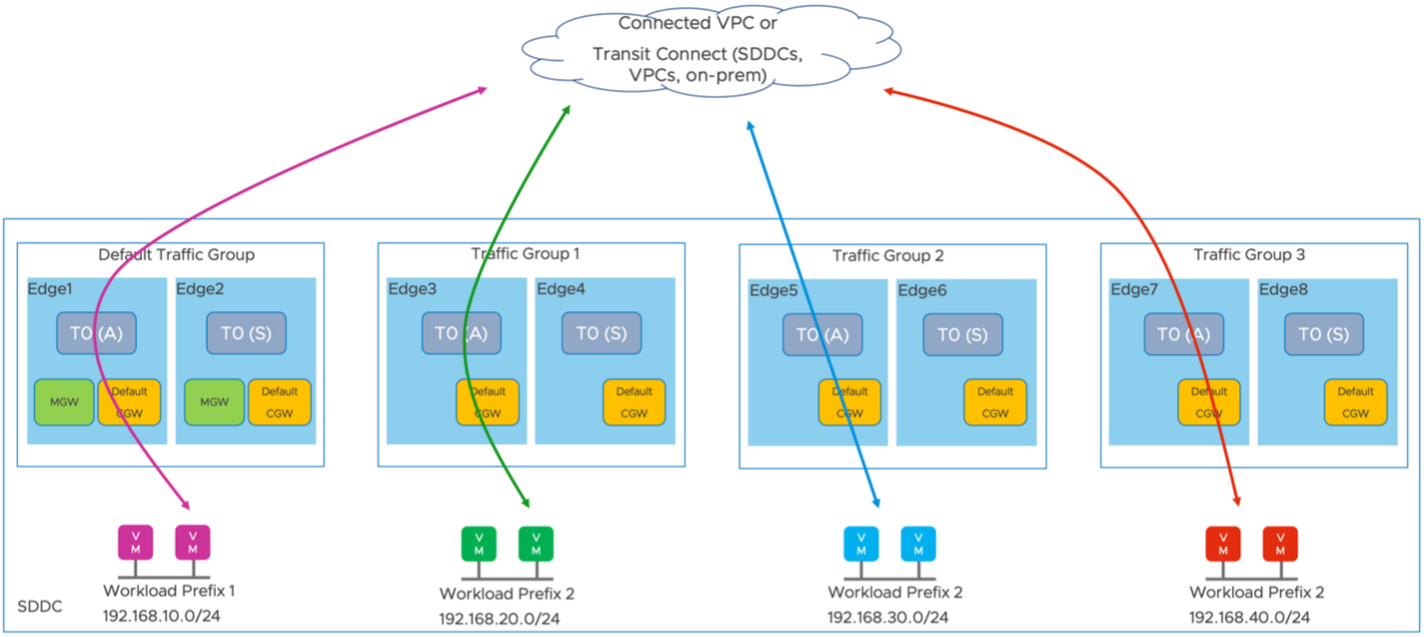
As shown in the illustration below, if the workload prefix is mapped to a TG, the above traffic will first hit the corresponding TG edge T0 and will be redirected to the active T0 in the default traffic group. For these cases the additional edges in TG will not provide any benefit.


Internet or S3 or DX private VIF or VPN
Default Traffic Group
Edge1
T0 (A)
MGW
Default CGW
Edge2
T0 (S)
MGW
Default CGW
Traffic Group 1
Edge3
T0 (A)
Default CGW
Edge4
T0 (S)
Default CGW
V M
V M
V M
V M
SDDC
Workload Prefix 1
192.168.10.0/24
Workload Prefix 2
192.168.20.0/24

SDDC Multi-Edge Use Cases

Increase throughput to address high north-south throughput requirements

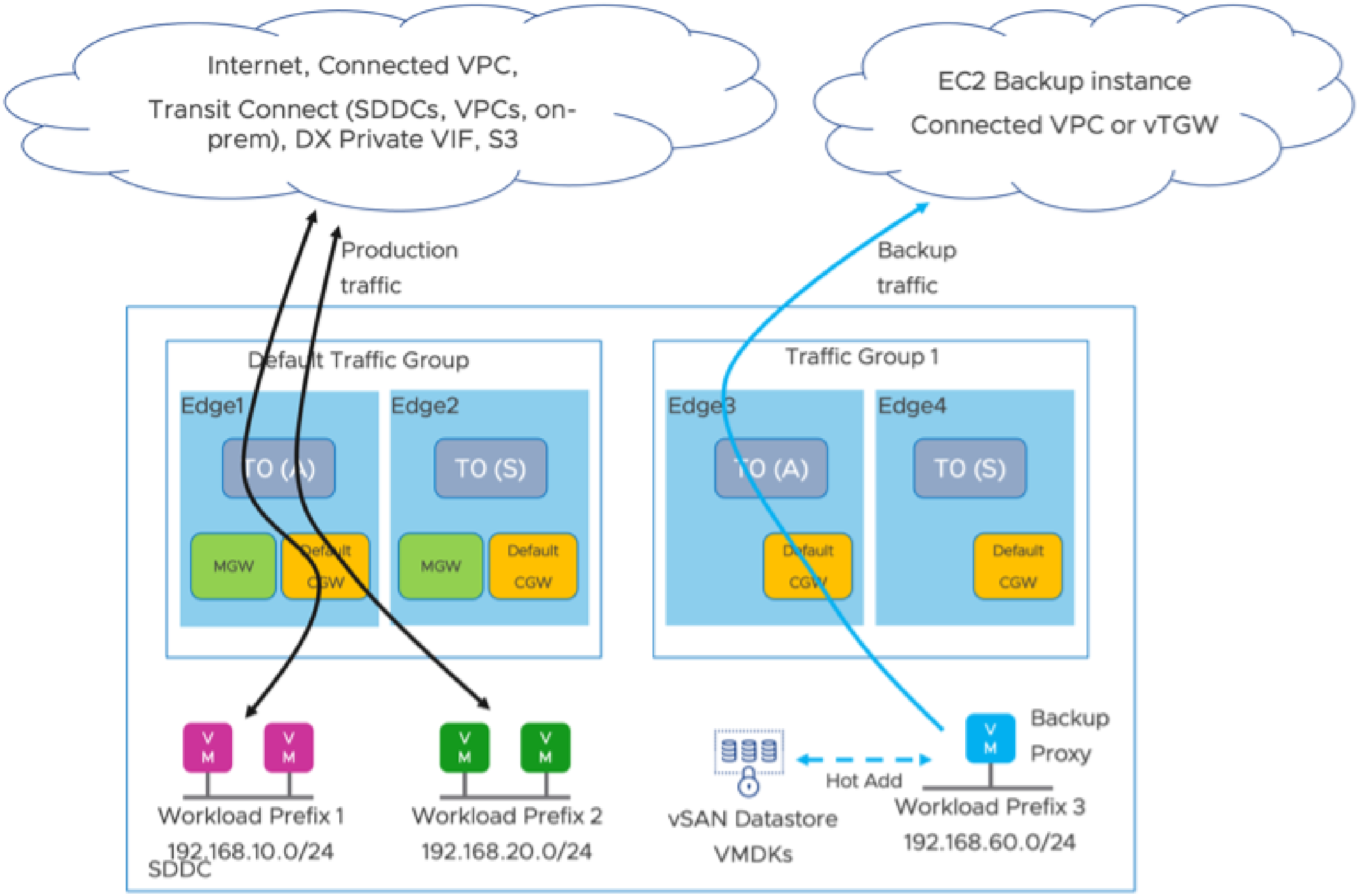
The single active Edge can become a bottleneck in the deployments where there is a requirement for very high north-south throughput that a single Edge node cannot support. The SDDC Multi-Edge feature will enable you to add additional TGs with edge nodes resulting in increased north-south throughput for the SDDC. You can pin different north-south traffic flows across these additional edge nodes to address the north-south throughput requirements of the SDDC.



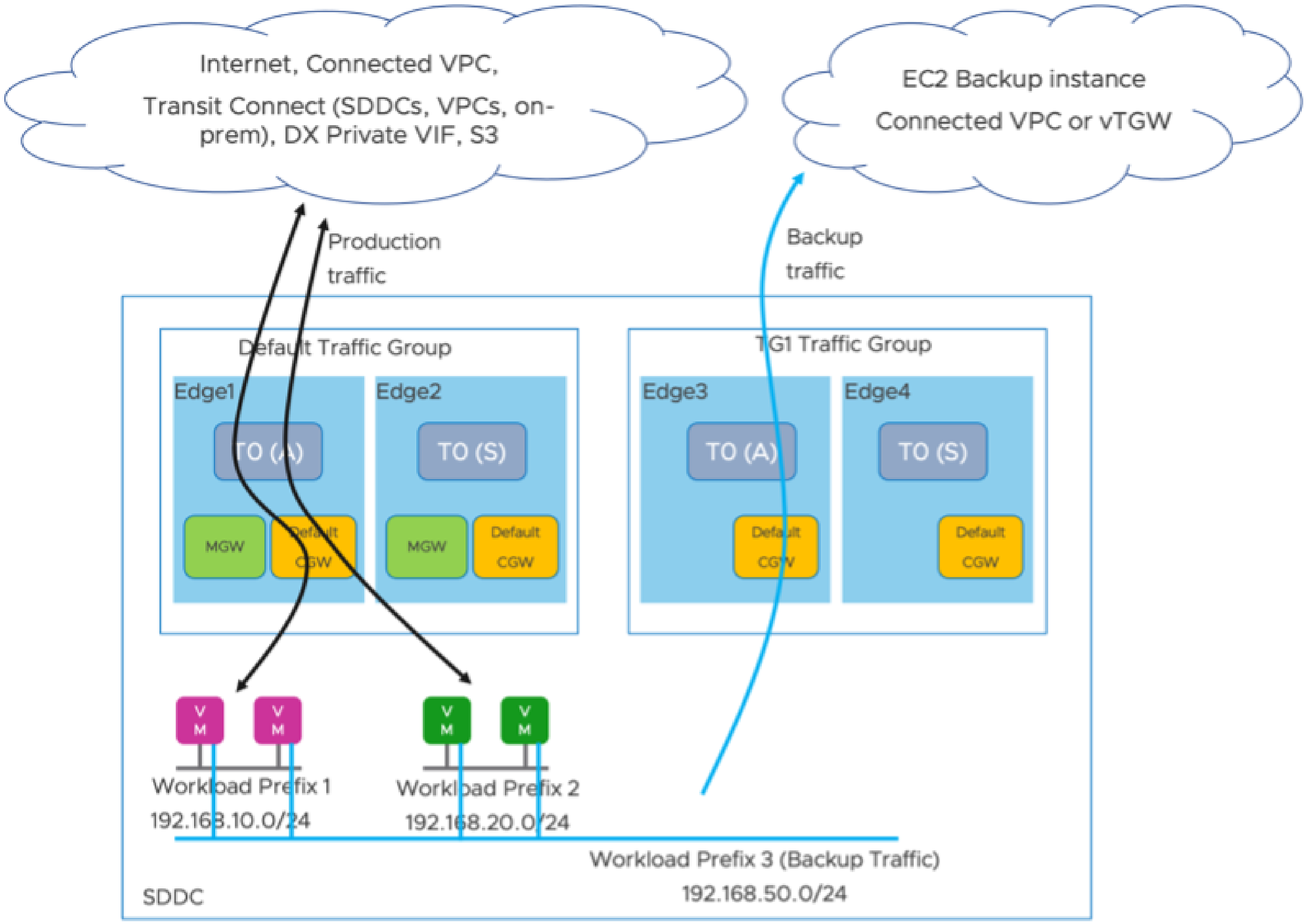
Scale out or Isolate backup/restore traffic

Typically, backup or restore traffic between SDDC and entities external to SDDC introduces a burst of traffic increasing the requirement for north-south throughput during backup and restore operations. This can negatively affect the performance of production traffic. With SDDC Multi-Edge feature, you can map the backup or restore traffic to a TG to isolate this traffic and prevent it from affecting the production traffic.

- Backup using a proxy VM. Typically, a proxy VM in the cluster hot-adds the vmdks and sends them to an EC2 instance located in the Connected VPC. The proxy VM prefix can be mapped to a scaled-out TG to isolate the backup traffic.



- For in-guest backup like SQL server backup, the workload can send backup traffic directly to an EC2 instance in Connected VPC or another VPC via vTGW. In this case, you can add a second NIC to the workload for backup traffic. The IP address of the second NIC can be mapped to a scaled-out TG to isolate this backup traffic. A static route in the guest OS will be required to ensure traffic destined for the EC2 backup server is sent out the correct NIC.



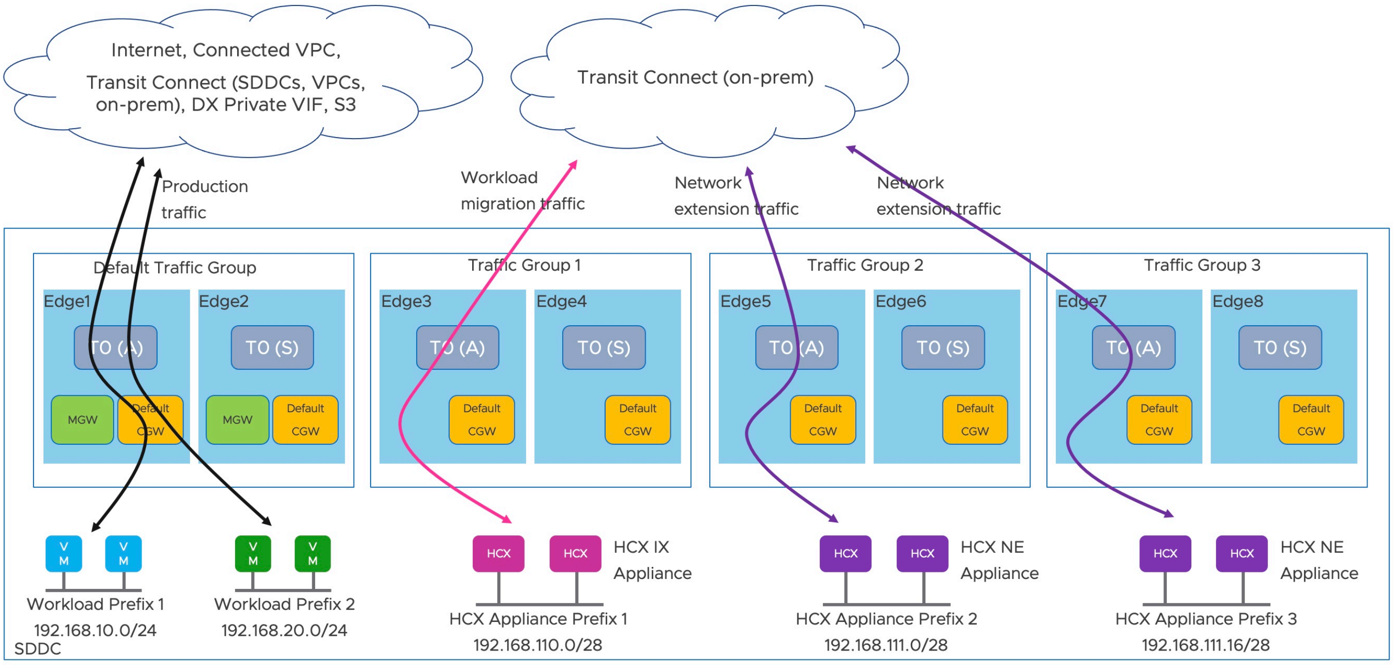
- In some cases, S3 object store is used directly to backup and restore data. For example, Dell EMC PowerProtect DD Virtual Edition (DDVE). The DDVE virtual appliance is deployed inside the SDDC and S3 object store is used to backup and restore data. In these cases, you cannot take advantage of the SDDC Multi-Edge feature to scale out the backup/restore traffic. The traffic between the appliance and S3 gateway endpoint must traverse the default active T0. To address this, you can deploy DDVE as an EC2 instance in Connected VPC. This use case will be equivalent to the previous scenario, and you can take advantage of the SDDC Multi-Edge feature.

Scale out HCX workload migration and network extension traffic

By default, HCX uses the management uplink interface for all uplink traffic. You can override the default Uplink Network for the destination site with a specific TG, and isolate traffic for the HCX service. HCX manager provides workflows to map HCX IX (Migration) traffic and NE (Network Extension) traffic to TGs.

You can isolate migration and NE traffic by creating one Service Mesh for Bulk Migration and one Service Mesh for Network Extension. Within each specific Service Mesh, you can configure a unique TG for the Uplink Network. This solution requires SDDCs to be configured with VMware Transit Connect (vTGW).

- HCX-IX migration service can only use one TG
- HCX-NE can take advantage of multiple TGs


Internet, Connected VPC,
Transit Connect (SDDCs, VPCs, on-prem), DX Private VIF, S3
Transit Connect (on-prem)
Production traffic
Workload migration traffic
Network extension traffic
Network extension traffic
Default Traffic Group
Edge1
Edge2
T0 (A)
T0 (S)
MGW
Default CGW
Traffic Group 1
Edge3
Edge4
Traffic Group 2
Edge5
Edge6
Traffic Group 3
Edge7
Edge8
VM
HCX
HCX IX Appliance
HCX NE Appliance
Workload Prefix 1
192.168.10.0/24
Workload Prefix 2
192.168.20.0/24
HCX Appliance Prefix 1
192.168.110.0/28
HCX Appliance Prefix 2
192.168.111.0/28
HCX Appliance Prefix 3
192.168.111.16/28
SDDC

Summary

VMware Cloud on AWS is an integrated cloud offering jointly developed by Amazon Web Services (AWS) and VMware. You can deliver a highly scalable and secure service by migrating and extending your on-premises VMware vSphere-based environments to the AWS Cloud.

VMware Cloud on AWS can support deployments of varying sizes and scale to address the requirements of the deployment. The SDDC Multi-Edge feature is another example of a feature that enables you to plan and scale your SDDC based on the north-south throughput or traffic isolation requirements of your applications.

Authors and Contributors

- Ron Fuller, Technical Product Management, VMware
- Michael Kolos, Product Solutions Architect, VMware
- Suresh Palguna Krishnan, Product Management, VMware