

NSX-T Federation Design Content

NSX-T 3.2.x

NSX-T Federation

You can find the most up-to-date technical documentation on the VMware website at: <https://docs.vmware.com/>

NSX-T Federation Design

3401 Hillview
Ave. Palo Alto,
CA 94304
www.vmware.com

Copyright © 2022 VMware, Inc. All rights reserved. Copyright and trademark information.

Contents

Contents	4
Preface	5
Intended Audience	5
VMware Technical Publications Glossary	5
Introduction	6
NSX-T Architecture Components	7
Global Manager	7
Global Manager-Standby	8
Local Manager	9
Objects	9
Management Plane	11
GM Cluster Deployment Models	12
NSX-T GM-Active with VMs Deployed in 3 Different Locations	12
NSX-T GM-Active and GM-Standby	13
Communication Flows for Global Manager and Local Manager	16
GM-Active to GM-Standby Communication Flow	17
GM to LM Communication Flow	18
LM to LM Communication Flow	25
Federation Regions	30
Data Plane	32
Networking in NSX-T Federation	34
Global Manager Network Services	34
Routing Protocols	44
Security in NSX-T Federation	55
Global Manager Security Services	56
NSX-T Federation Limitations	62

Preface

This section describes NSX-T Federation. For more information about design decisions,

please refer to the VMware Validated Design

<https://www.vmware.com/support/pubs/vmware-validated-design-pubs.html>.

VMware NSX-T is designed to address application frameworks and architectures that have heterogeneous endpoints and technology stacks. In addition to vSphere, these environments may include other hypervisors, containers, bare metal operating systems, and public clouds. NSX-T allows IT and development teams to choose the technologies best suited for their particular applications. NSX-T is also designed for management, operations, and consumption by development organizations in addition to IT.

Intended Audience

This information is intended for anyone who wants to install, upgrade, or use NSX-T Federation.

VMware Technical Publications Glossary

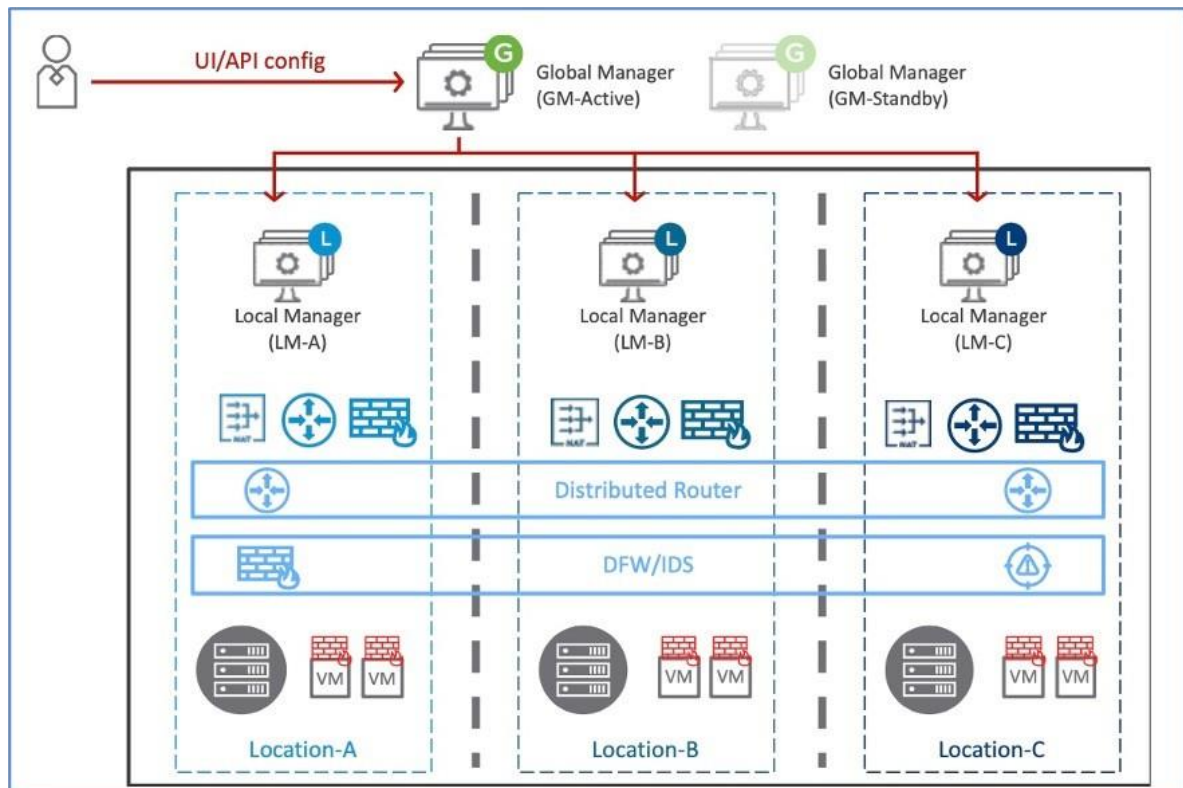
VMware Technical Publications provides a glossary of terms that might be unfamiliar to you. For definitions of terms as they are used in VMware technical documentation, go to <http://www.vmware.com/support/pubs>.

Introduction

With NSX-T Federation, you can manage multiple NSX-T Data Center environments using an intuitive user interface, with a single pane of glass view. You can create gateways and segments that span one or more locations and configure and enforce firewall rules consistently across locations.

NSX-T uses one central NSX-T Global Manager Cluster (GM) that offers central network and security services configuration for all locations:

- There is one NSX-T Manager Cluster per location, which we call the Local Manager (LM) that manages Transport Nodes (hypervisor and Edge nodes) for that location.
- The GM pushes the network and security configuration to the different LMs to implement locally.



It's simple, you install Global Manager, add locations, and configure networking and security from the Global Manager.

NSX-T Architecture Components

In this section, we describe the major components of the NSX-T architecture. This chapter includes the following topics:

- Global Manager
- Global Manager-Standby
- Local Manager
- Objects
- Management Plane

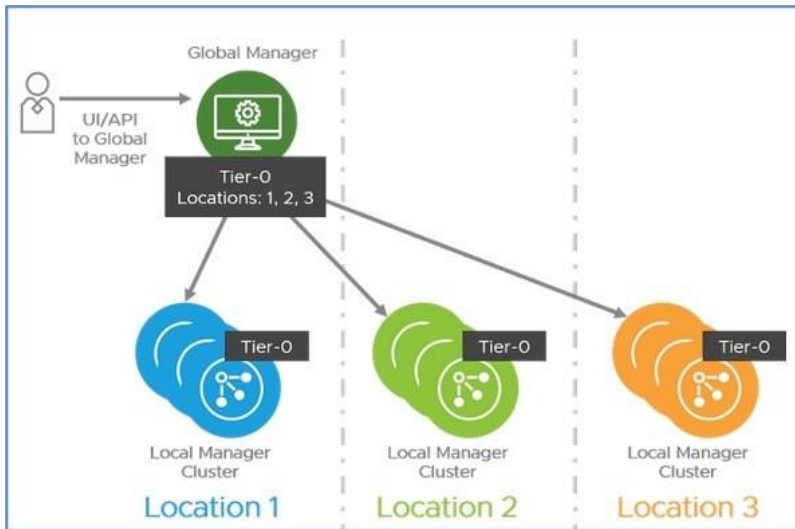
Global Manager

In NSX-T Federation, the Global Manager (GM) offers central network and security services for all locations. You make configuration changes on the Global Manager. The changes are synced with the relevant Local Managers, which also sync some information with one another.

The configurations you create on the Global Manager are read-only on the Local Managers. Configurations on the Local Managers are NOT synced with the Global Manager; instead, the Global Manager syncs a configuration with a Local Manager only if the configuration is relevant to that location.

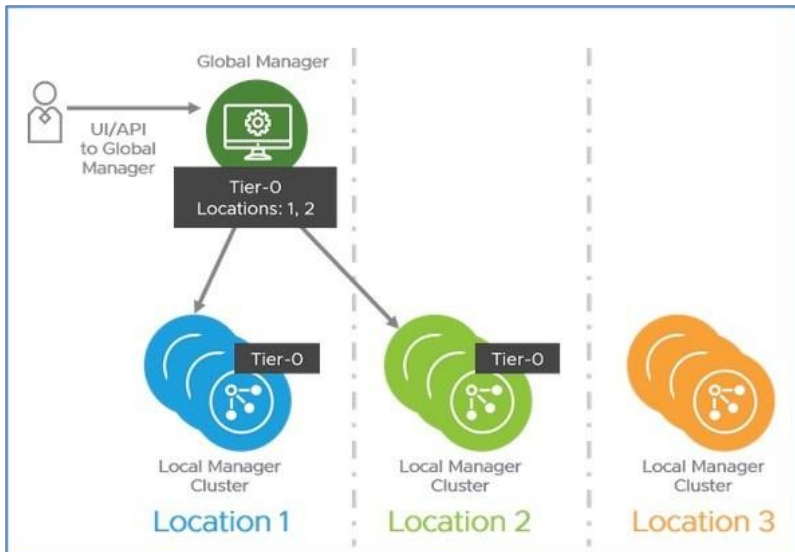
Note You can create objects on a Local Manager; however, those objects are not synced with the Global Manager or any other Local Manager.

Let's say you create a tier-0 gateway and add it to Location 1, Location 2, and Location 3. The configuration is synced with all three Local Managers, as shown in Figure 2.



If you create the Tier-0 gateway and add it only to Location 1 and Location 2, the configuration is not synced with Location 3.

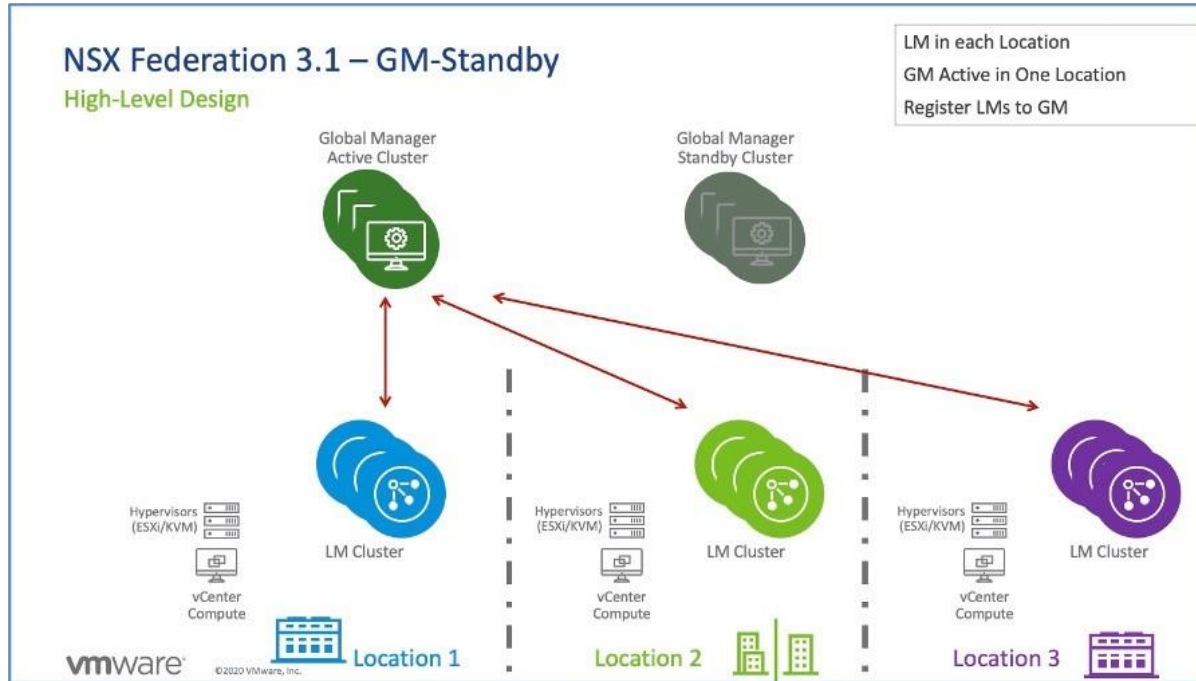
Figure 2-1. Three Locations with 1 Not Synced



Global Manager-Standby

Global Manager-Standby (GM-Standby) is synchronized with Global Manager. Every configuration is synced; however, GM-Standby remains in standby mode in case the GM goes down, as shown in Figure 4. If this occurs, you go into the UI and specify that GM-Standby become Active. More about GM-Standby is in the Disaster Recovery section.

Figure 2-2. Global Manager-Standby



Local Manager

Local Managers reside in each location and, depending on your configurations, sync back to the Global Manager. They do not sync their configurations TO the Global Manager. The Global Manager validates changes against the Global Manager configuration only. When a Local Manager receives a configuration from the Global Manager, the configuration is realized in the fabric nodes of that Local Manager. During this realization, errors or conflicts might be detected.

Objects

There are two types of objects in NSX-T Federation:

- Global Objects: Objects created from the Global Manager.
- Local Objects: Objects created from the Local Manager.

Objects you create from the Global Manager are global objects; however, their span might not include all available locations. The Global Manager displays only global objects but does not automatically receive the status of objects.

On a Local Manager, you only see local objects and any global objects that you've applied to that location. The Local Managers display the status of

both global and local objects.

Management Plane

The Management Plane provides a single API entry point into the system, persists user configuration, handles user queries, and performs operational tasks on all management, control, and data planes in the system. It is the one and only source of truth for the logical, configured system. The active Global Manager, GM-Standby, and one Local Manager per location reside on the Management Plane. You make changes in the Management Plane via a RESTful API or the NSX-T UI.

GM Cluster Deployment Models

Each GM Cluster is composed of three NSX-T Manager VMs of type NSX Global Manager. The LM Cluster is composed of three NSX-T Manager VMs of type NSX Manager. As explained in the VMware NSX-T Reference Design Guide, the NSX-T Manager virtual machines (VMs) members of a cluster (GM or LM) can be the same subnet or different subnets. Also, the maximum latency between any of the NSX-T Manager VMs is 10 milliseconds. To operate, The NSX-T Manager cluster can handle the loss of one of its NSX-T Manager VMs.

There are two modes of NSX-T Federation deployment:

- NSX-T GM-Active with VMs deployed in 3 different locations
- NSX-T GM-Active and GM-Standby

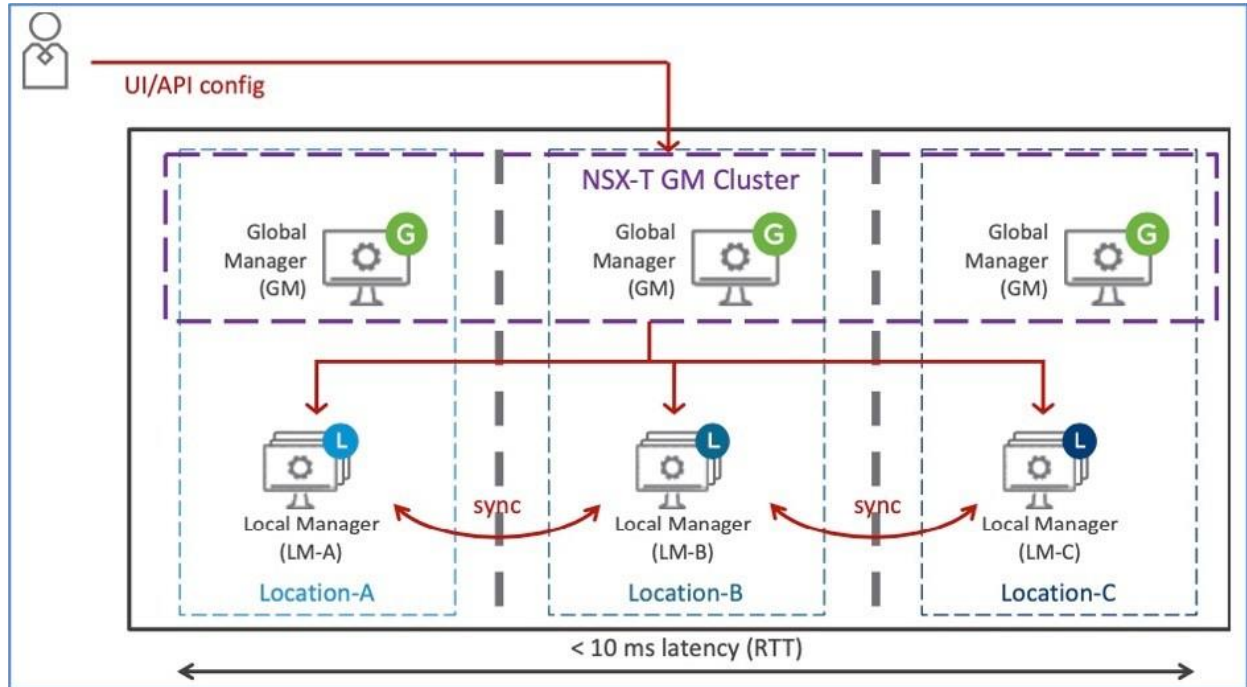
This chapter includes the following topics:

- [NSX-T GM-Active with VMs Deployed in 3 Different Locations](#)
- [NSX-T GM-Active and GM-Standby](#)

NSX-T GM-Active with VMs Deployed in 3 Different Locations

The typical use cases are different buildings in a metropolitan region, where each building requires a dedicated, full operational Management Plane in case of other location failures. And at the same time, a central configuration is required for ease of operation.

Figure 3-1. Buildings in Metropolitan Region (<10ms latency)



It is important to highlight in this NSX-T GM Cluster deployment that the loss of one location does not stop the GM Management Plane service because the GM cluster still has two valid members. Additionally, if you lose one location, the LM Management Plane service does not stop on the locations either because the LM cluster still has two valid members.

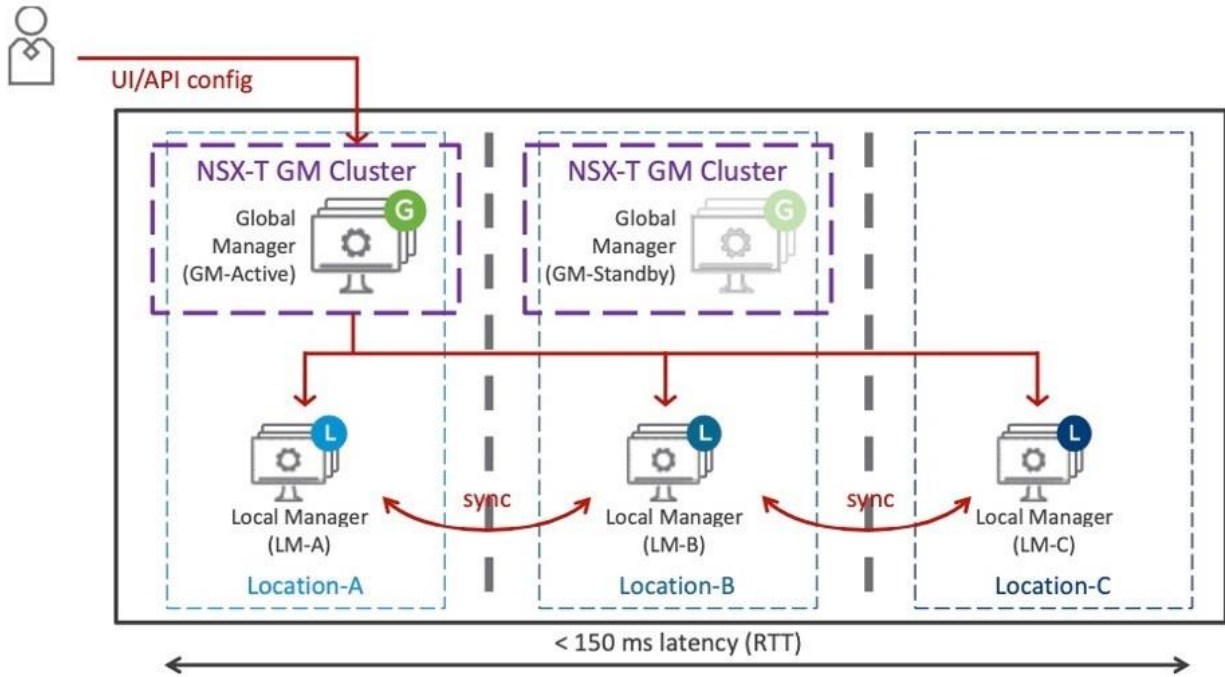
Note For a use case with three locations or more, latency (RTT) below 10ms between each, and no congestion between those locations, we recommend that you have each location with its own LM cluster and one GM NSX-T VM in three locations.

NSX-T GM-Active and GM-Standby

The most common type of deployment is to have GM Active in one location and the GM standby cluster in another location. If you lose GM Active, you can recover by making GM Standby active. Local services are not affected. For this use case with two locations, latency (RTT) above 10ms across the locations, we recommend that you have all three NSX-T GM VMs in one single location.

The typical use cases would be two data centers only or data centers in large distance regions, as show in Figure 6.

Figure 3-2. Two Locations Only and/or Data Centers far apart (>10ms latency)



If you lose the location that hosts GM-Active, it does not stop the GM Management Plane service until you make GM-Standby Active; however, the LM Management Plane service does not stop on the different locations. We discuss recovering the GM Management Plane more in the Disaster Recovery section.

Design Decision ID	Design Decision	Design Justification	Design Implication
NSXT-SDDC-FED-001	Deploy GM Cluster across different location.	If number of locations is more than 3 and latency (RTT) is below 10 ms it is recommended to deploy GM cluster across those locations. No standby GM cluster is required	Low RTT latency implies geographically close locations
NSXT-SDDC-FED-002	Deploy GM cluster and Standby GM	If number of locations is two OR latency is more than 10ms Standby GM cluster deployment is recommended.	Additional VM capacity will be required to host Standby GM cluster nodes.

Design Decision ID	Design Decision	Design Justification	Design Implication
NSXT-SDDC-FED-003	Use LM Cluster VIP, FQDN Cluster VIP or external Load Balancer VIP for GM to LM communication	For GM-LM proper communication	
NSXT-SDDC-FED-004	For best scale and performance, we recommend you have the Section Span, Source/ Destination, and Applied To matching the need.	With these settings, rules are only pushed to relevant LMs and to all relevant VMs.	

Communication Flows for Global Manager and Local Manager

There are three types of federated communication flows:

- GM-Active to GM-Standby
- GM-LM
- LM-LM

NAT is not supported for these communications, and if there is a firewall between GM and LMs, specific ports must be open. Refer to <https://ports.vmware.com/home/NSX-T-Data-Center> .

The GM to GM, GM to LM, and LM to LM management plane and control plane synchronization is offered by the Async Replicator (AR) service. AR runs on port 1236 opened by the Application Proxy (APH) Service. The APH connectivity is established between each:

- GM-Active VMs to GM-Standby VMs.
- GM-Active and Standby VMs to every locations LM VMs.
- Each location LM VM to every other location LM VMs.

There is no APH connectivity between GMs inside a GM Cluster; likewise, no APH connectivity between LMs inside a LM Cluster.

The GM-Active gets NSX objects' status from registered LM IP/FQDN via 443. That is why registration with the LM Cluster VIP is strongly recommended for that channel availability, even in the case of on LM VM loss.

The GM only stores configuration information (management plane) and not realization information (controller plane).

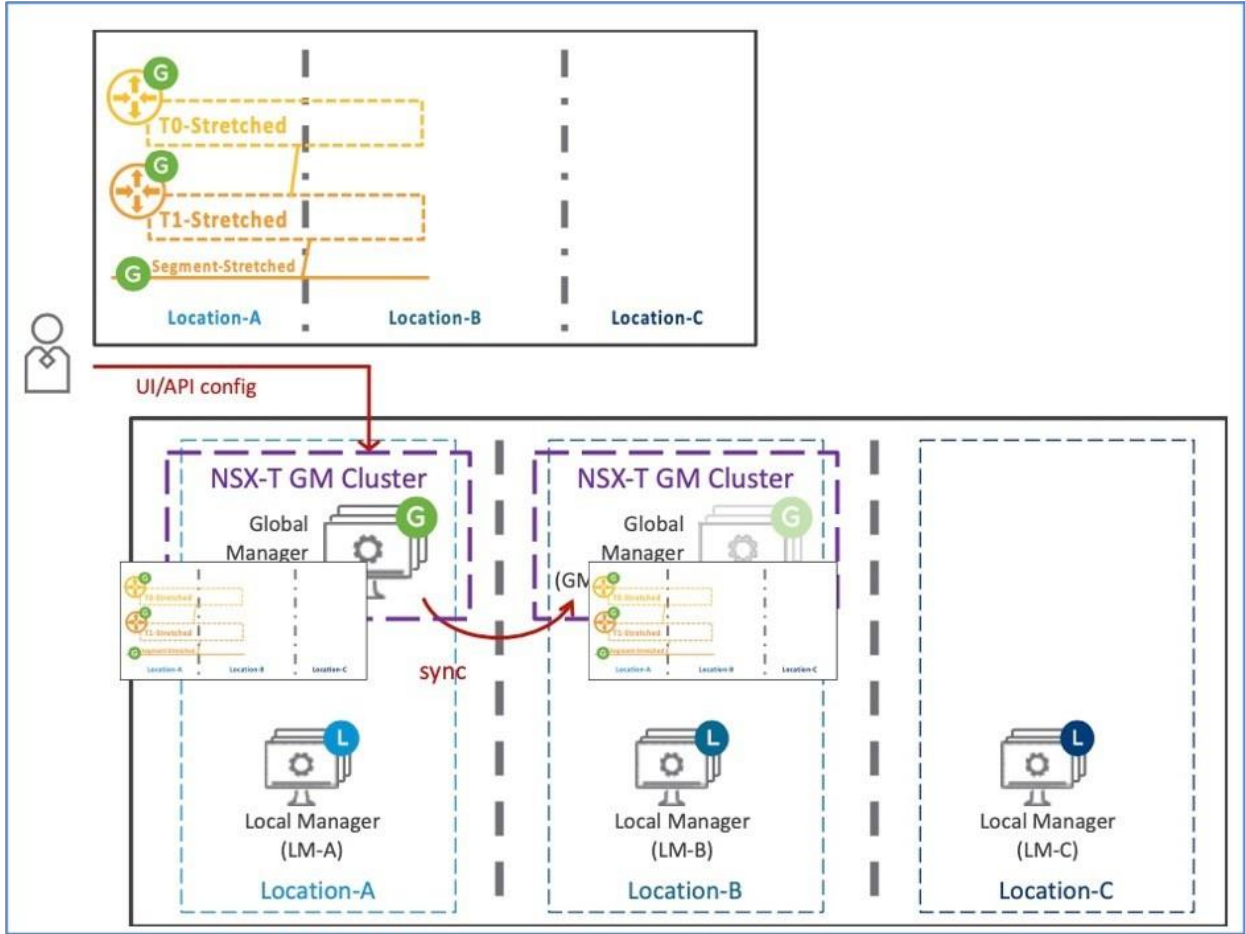
This chapter includes the following topics:

- GM-Active to GM-Standby Communication Flow
- GM to LM Communication Flow
- LM to LM Communication Flow

GM-Active to GM-Standby Communication Flow

GM-Active synchronizes all received configurations to its GM-Standby, as shown in Figure 7.

Figure 4-1. GM-Active to GM-Standby Communication Flow



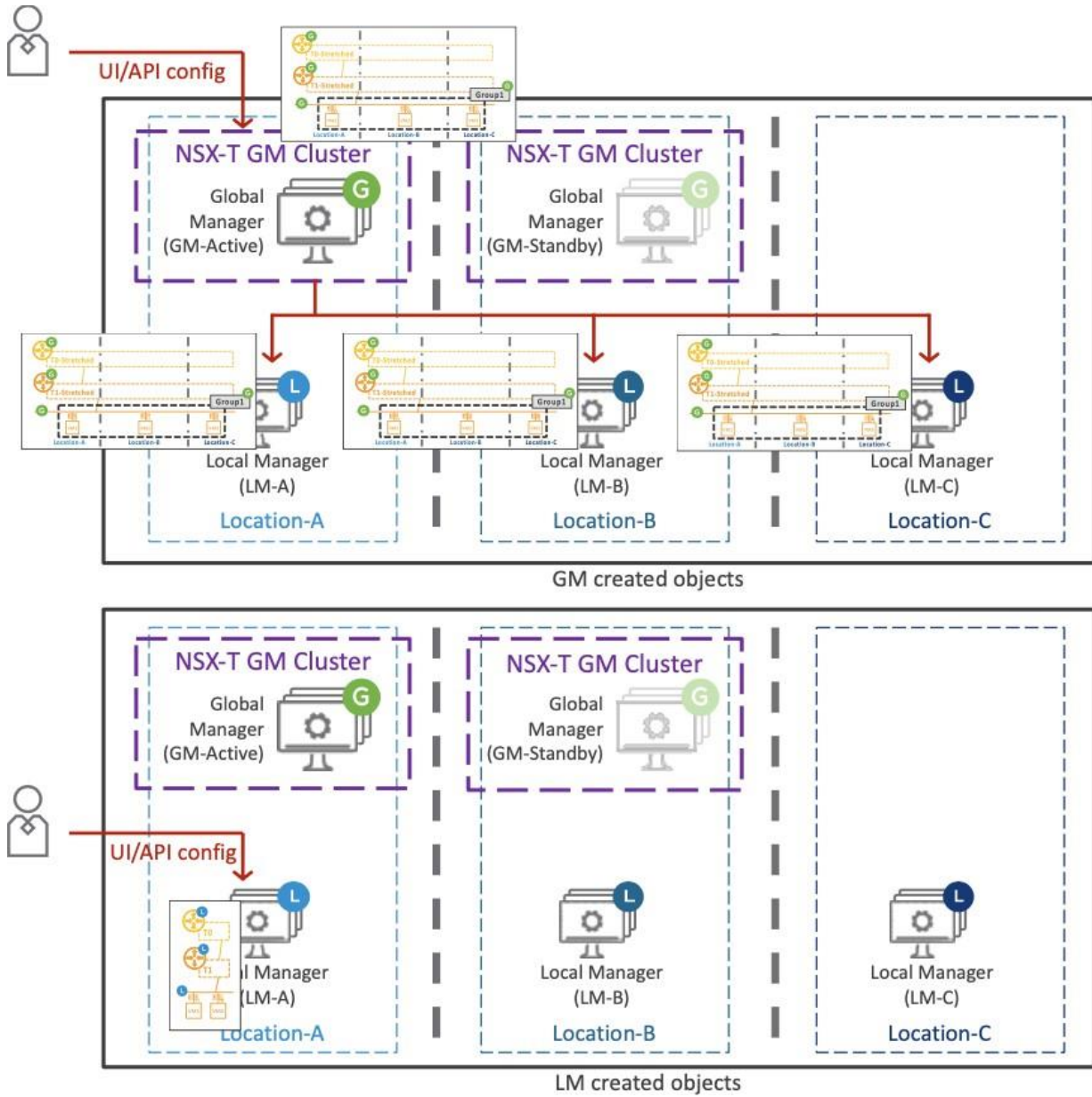
This Figure shows an example of network configuration completed on GM-Active. GM-Active synchronizes this configuration to GM-Standby. The same synchronization occurs for any other configuration, such as security configuration, Principal Identity users, or vIDM users.

Note vIDM configurations are currently not synchronized to GM-Standby; however, vIDM users are.

GM to LM Communication Flow

In both GM cluster deployment models, network and security is centrally configured and managed through the GM, which pushes the relevant network and security objects to the different LMs.

Figure 4-2. GM to LM Communication Flow



In Figure 8, there is one Tier-0 + Tier-1 + Segment stretched between Location-A and Location-B, configured on GM-Active. GM-Active pushes those GM objects to Location-A LM + Location-B LM. This configuration is not pushed to Location-C LM because none of those objects are relevant to Location-C.

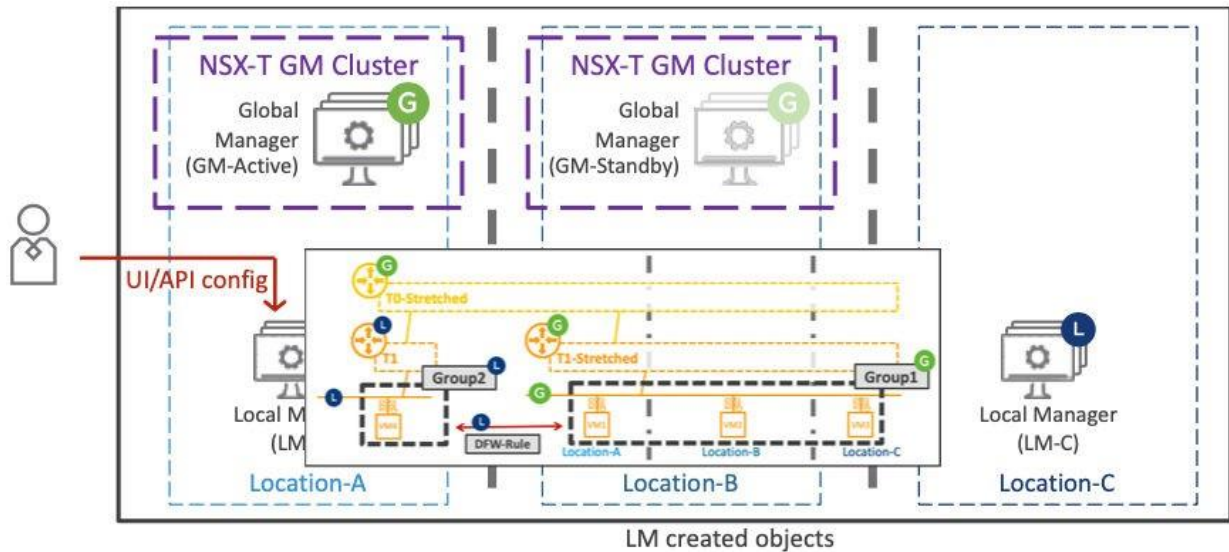
The GM objects that LM receives are tagged as GM are in read-only mode in the LM. Only very specific fields of some specific objects can be edited on LM:

- Tier-0 + BGP and Segment Ports configurations, where the LM admin can quickly make changes in an emergency.

- GM-Segment-Ports, where LM can edit TAG. The goal is to allow orchestration tools, such as vRA, to deploy applications on existing GM-Segments and add TAG on those created LM-Segment-Ports. Then if GM Security Groups are based on Segment-Port-TAGs, those applications automatically receive their appropriate security with no extra configuration.

Each LM can directly receive network and security configurations. LM direct configuration is useful for specific network or security features currently not supported from GM and/or to accept orchestration tools not enhanced to talk to the GM yet, as shown in the next Figure.

Figure 4-3. Direct LM Configuration



Notice that one Tier-1 with LB Service is directly configured on the LM. The LM objects can consume GM objects, such as the LM-T1 that has one-arm attached to a GM-segment.

Note When you configure directly on the LM, those objects are unknown to the GM. You manage those LM objects directly on the LM.

However, it is also possible, in some cases to link those LM objects to GM objects. Following is a list of supported LM Network and Security features configured from LM after registered by GM and the ability to link those LM objects with GM objects.

LM configuration	Object	Support (Yes / No)
Networking		
	LM T1	<ul style="list-style-type: none"> LM config consuming LM objects (LM T1 connected to LM T0) LM config consuming GM objects (LM T1 connected to GM T0)
	LM Segment	<ul style="list-style-type: none"> LM config consuming LM objects (LM_Segment connected to LM T1) No LM config consuming GM objects (LM_Segment connected to GM T1), ability to ability create/update a Segment Port on GM Segment
	LM L2-Bridge	<ul style="list-style-type: none"> LM config consuming LM objects (LM Segment) No LM config consuming GM objects (LM_Segment)
	LM Edge NAT	<ul style="list-style-type: none"> LM config consuming LM objects (LM_NAT on LM T0/T1) No LM config consuming GM objects (LM NAT on GM T1)
	LM LB	<ul style="list-style-type: none"> LM config consuming LM objects (LM_LB on LM T1) No LM config consuming GM objects (LM LB on GM T1)
	LM VPN	<ul style="list-style-type: none"> LM config consuming LM objects (LM_VPN on LM T0/T1) No LM config consuming GM objects (LM_VPN on GM T0/T1)
Security		
	LM Group	<ul style="list-style-type: none"> LM config consuming LM objects (LM_Group with LM_Members) No LM config consuming GM objects (LM_Group with GM_Members), only exception LM Group with Static Member = GM_Segment
	LM DFW	<ul style="list-style-type: none"> LM config consuming LM objects (LM_DFW on LM_Group) LM config consuming GM objects (LM_DFW on GM_Group)
	LM Firewall IPFIX	<ul style="list-style-type: none"> LM config consuming LM objects (LM_FWIPFIX on LM_Group) No LM config consuming GM objects (LM FWIPFIX on GM_Group)

	LM GFWW	<ul style="list-style-type: none"> LM config consuming LM objects (LM GFWW on LM_T0/T1) LM config consuming GM objects (LM_GFWW on GM_T0/T1)
	LM IDFW	<ul style="list-style-type: none"> LM config consuming LM objects (LM IDFW on LM_Group) LM config consuming GM objects (LM_GFWW on GM_Group)
	LM Security Profile (Session Time/DNS/Flood)	<ul style="list-style-type: none"> LM config consuming LM objects (LM_SecProf on LM_Group) No LM config consuming GM objects (LM_SecProf on GM_Group)
	LM IDS/IPS	<ul style="list-style-type: none"> LM config consuming LM objects (LM_IDS/IPS with LM_Group) No LM config consuming GM objects (LM_IDS/IPS with Source=GM_Group), only exception LM_IDS/IPS with "LM_Group = Global_Segment Member" + "Service = GM_Service"
	LM Network Introspection*	<ul style="list-style-type: none"> See below*
	LM Endpoint Protection*	<ul style="list-style-type: none"> See below*
	LM Segment Security	<ul style="list-style-type: none"> LM config consuming LM objects (LM_Segment with LM_SegSec) LM config consuming GM objects (LM_Segment with GM_SegSec)
Monitoring		
	LM Switch IPFIX	<ul style="list-style-type: none"> LM config consuming LM objects (LM_SwitchIPFIX on LM_Segment / LM_Groups / etc) No LM config consuming GM objects (LM_SwitchIPFIX on GM_Segment / GM_Groups / etc)
	LM Port Mirroring	<ul style="list-style-type: none"> LM config consuming LM objects (LM_PortMirror on LM_Segment / LM_Groups / etc) No LM config consuming GM objects (LM_PortMirror on GM_Segment / GM_Groups / etc)

* Starting with NSX-T 3.2.0, LM Network Introspection (Host-based and Cluster-based deployments) and Endpoint Protection configuration are supported with some limitations.

Network Introspection Host-based and Cluster-based and Endpoint Protection

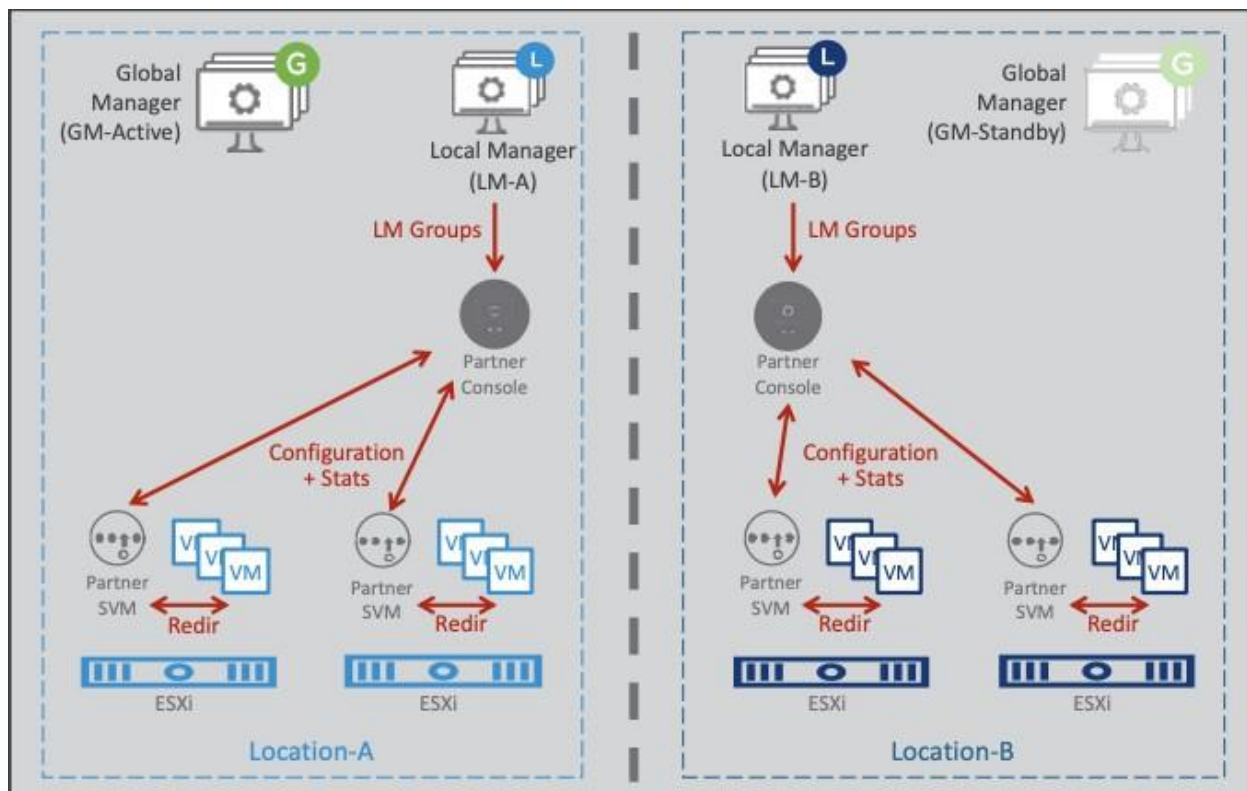
Network Introspection (Host-based and Cluster-based deployments), previously named Service Insertion are supported in NSX-T 3.2.0. Here are the limitations:

- Limitations on NSX side:
 - Use of LM_Segment is required (no GM_Segment)
 - Use of LM_Groups is required (no GM_Group because the Partner Console does not collect/receive the GM Groups)
 - No Network Introspection on Edges T0/T1
- Limitations on Partner side, such as Palo Alto, CheckPoint, Fortinet, Netscout, etc:
 - Partner must validate NSX-T Federation support on their side.

Endpoint Protection, previously named Guest Introspection is supported starting with NSX-T

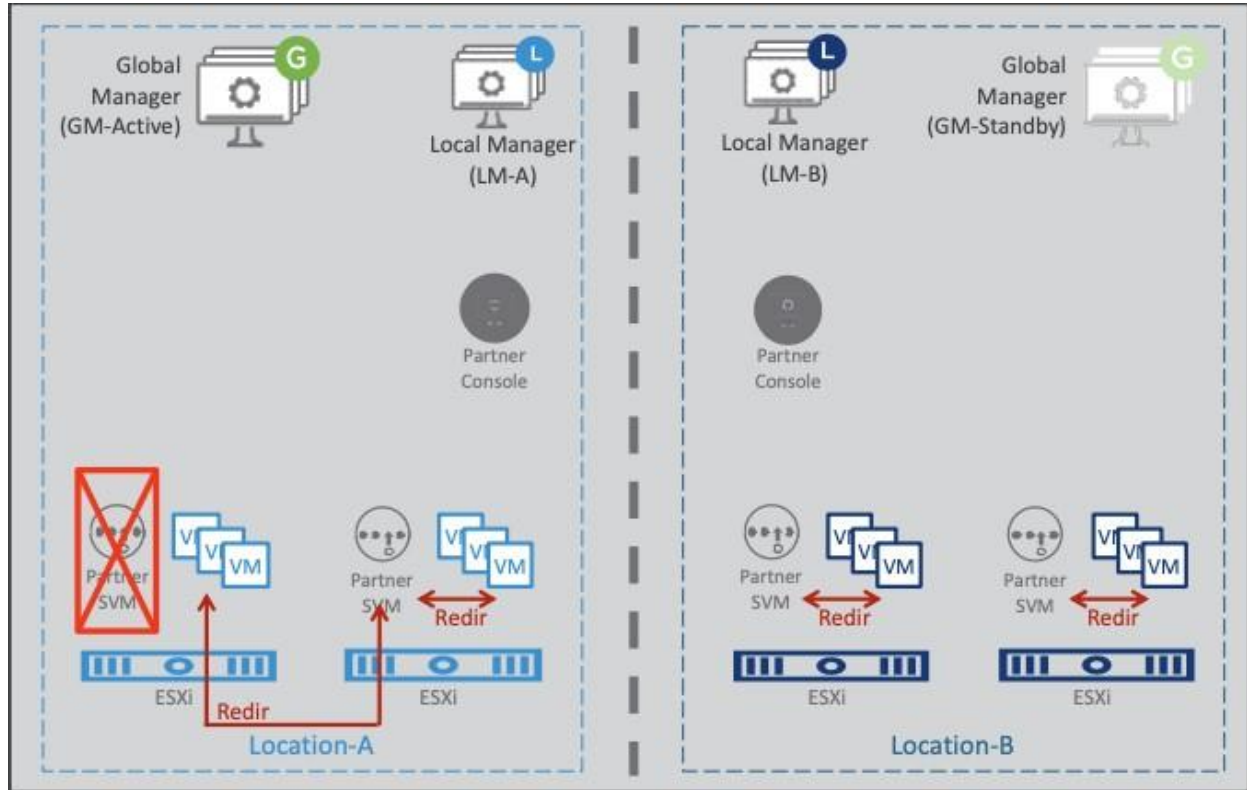
3.2.0. Keep the following points in mind.

On the Partner side, such as Bitdefender, Trend Micro, etc, the Partner must validate NSX-T Federation support on their side. In the following figure, on the location Management Plane, each Partner Console receives



its local LM Groups and pushes its configuration to the different Host-Based Partner SVM in its location (Partner Console does not receive GM Groups).

On the Data Plane, each ESXi redirects the VM traffic to its hosted Partner SVM. In case of hosted Partner SVM failure in one of the ESXi, that ESXi redirects its VM traffic to another ESXi Partner SVM. That other ESXi will



always be local, as shown in the following figure.

LM to LM Communication Flow

In LM-to-LM communication flows, each LM talks directly to the other LMs and knows about the local and remote members. The GM knows none of this Data Plane information because the Data Plane is not on the GM.

There are two cases where the LM synchronizes information with other LMs:

- Stretched NSX Group
- Stretched Segment

Stretched NSX Group

The following Figure shows one GM-Group1 stretched to all locations it created. Since that GM- Group1 is global, it pushes configurations to all LMs.

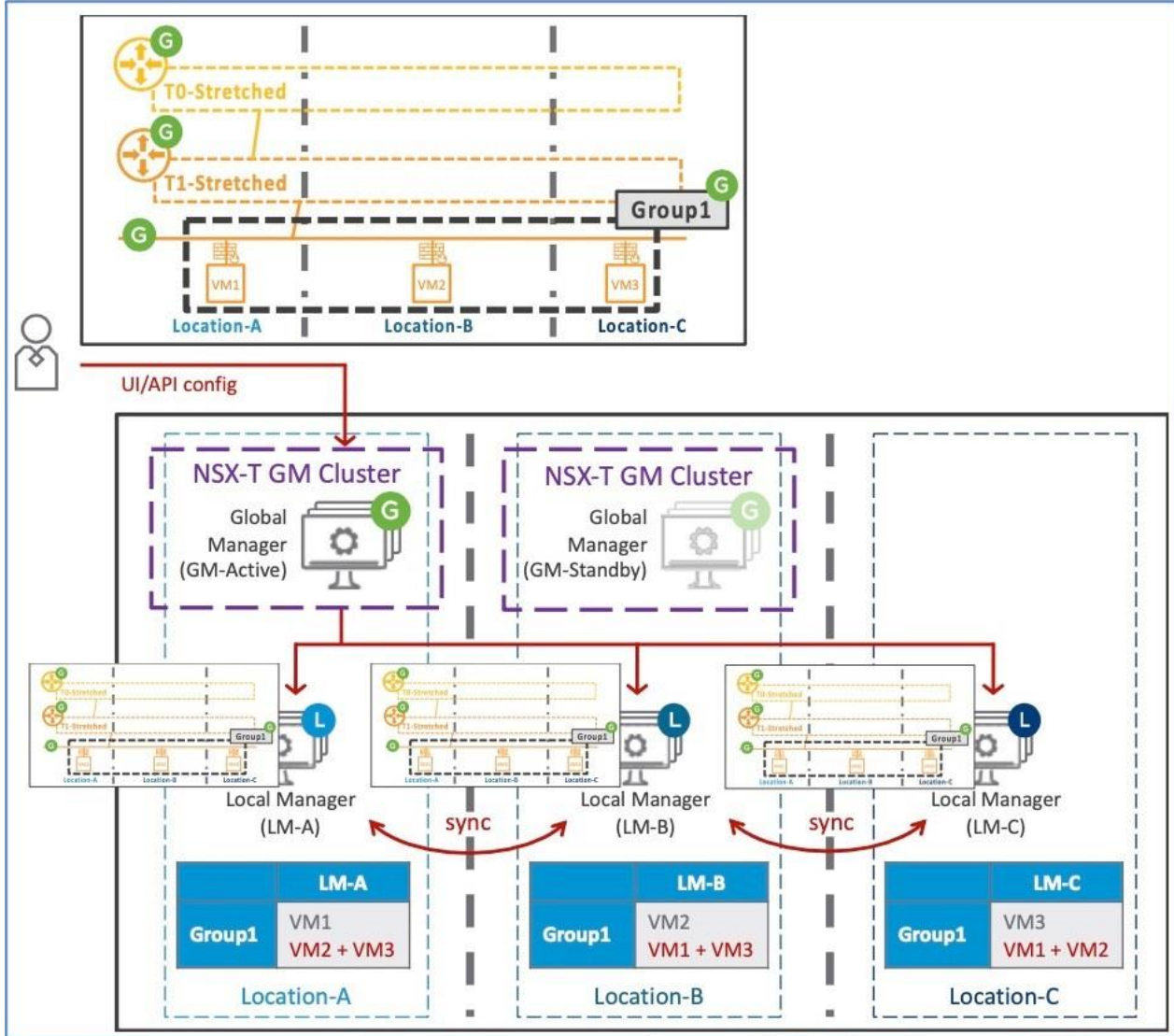
Each LM receives that GM-Group1 and resolves its location membership:

- LM-A=VM1

- LM-B=VM2
- LM-C=VM3

Because that GM-Group1 is global, each LM synchronizes its local membership with the other LMs. At the end of the synchronization, all LMs know about all remote LM memberships.

Figure 4-4. LM to LM Communication Flow Stretched NSX Group



Stretched Segment

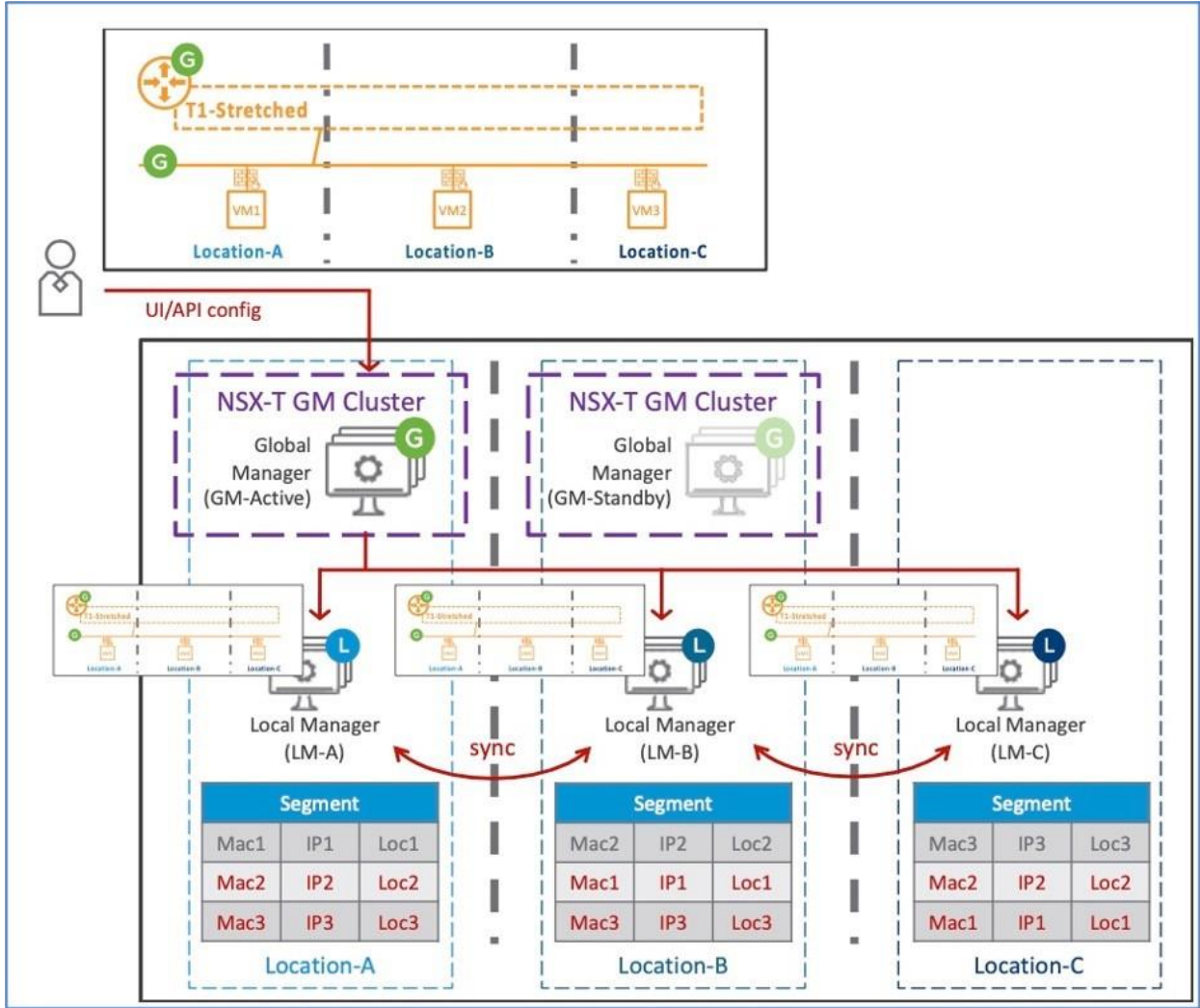
The following Figure shows one GM-Segment stretched to all locations created. Because that GM-Segment is global, it pushes configurations to all LMs.

Each LM receives that GM-Segment and fills up its local segment table:

- LM-A=Mac1/IP1
- LM-B=Mac2/IP2
- LM-C=Mac3/IP3

Because the GM-Segment is global, each LM synchronizes its local segment table with the other LMs. At the end of synchronization, all LMs know about all remote LM segment tables.

Figure 4-5. LM to LM Communication Flow Stretched Segment



Federation Regions

Regions is a new concept in NSX-T 3.1 and are good for security.

When you register a LM, it becomes a region. The GM it is attached to is also a region. By default, the GM creates:

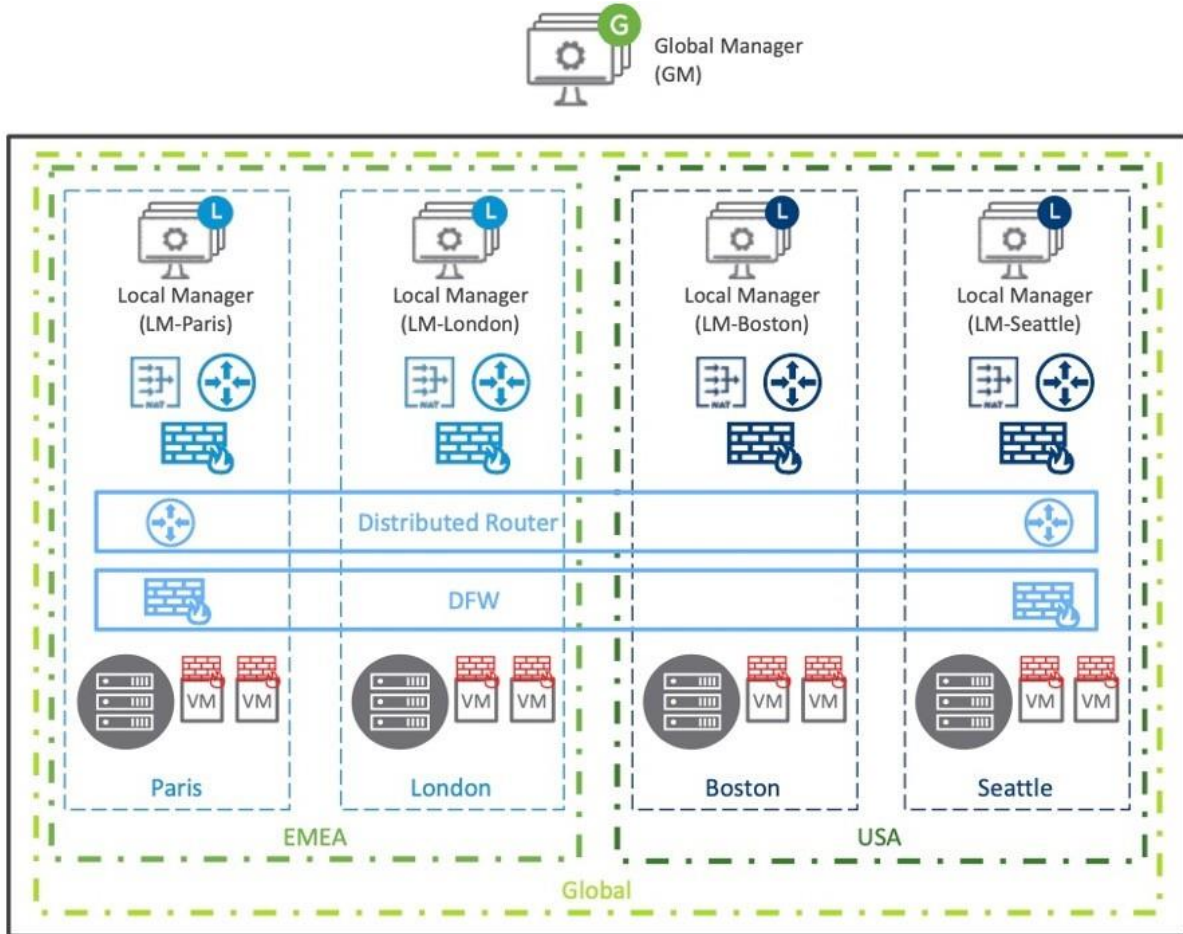
- One Global region that contains all the LMs
- One region per LM that contains only the individual LM.

It is possible to configure extra regions that contain one or multiple LMs. One specific LM can only be in one extra region in addition to its LM region and the Global region.

In the following Figure:

- There are four LMs: LM-Paris, LM-London, LM-Boston, and LM-Seattle.
- The default GM regions are Global, Paris, London, Boston, and Seattle.
- We created two extra regions: EMEA region that contains LM-Paris and LM-London and a USA region that contains LM-Boston and LM-Seattle

Figure 5-1. Example of Federation Regions



Specific regions allow you to create specific security policies to be applied to only those specific regions.

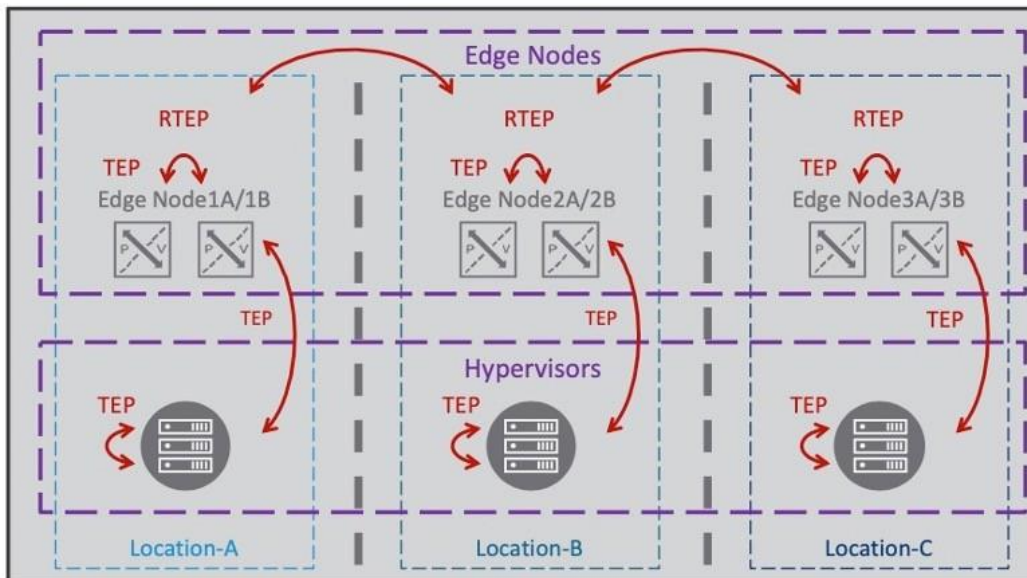
Note Regions do not apply to network constructs (Tier-0, Tier-0, Segment, NAT, GW-FW), where each LM must be individually selected on those stretch network elements.

Data Plane

The Global Manager works only on the Management Plane, pushing configurations to the different Local Managers. LMs are in charge of the local Management Plane (configuration) and the local Control Plane (Mac/IP/TEP tables).

The Data Plane elements are Edge Nodes and hypervisors in the different locations, as shown in the following Figure.

Figure 6-1. NSX-T Federation Data Plane



North/South traffic is still processed by the Edge Nodes in the different locations. East/West overlay traffic within a location is still processed by the hypervisors using their TEP interfaces. East/West overlay traffic cross-location is not processed between hypervisors' TEP interfaces; instead, the Edge Nodes and their Remote TEP (RTEP) interfaces perform the processing.

NSX-T Federation Edge Nodes have two Overlay interfaces.

Overlay Interface	Description
TEP	<p>For the Overlay communication within a location to other local Edge Nodes and local hypervisors.</p> <p>Each Edge Node can have multiple TEP IP, and fragmentation is not supported on TEP traffic.</p>
RTEP	<p>For the Overlay communication cross-locations to remote Edge Nodes.</p> <p>Each Edge Node can have a single RTEP IP, and fragmentation is supported on RTEP traffic. However, for best performance, we recommend avoiding fragmentation (RTEP MTU greater or equal to 1700).</p>

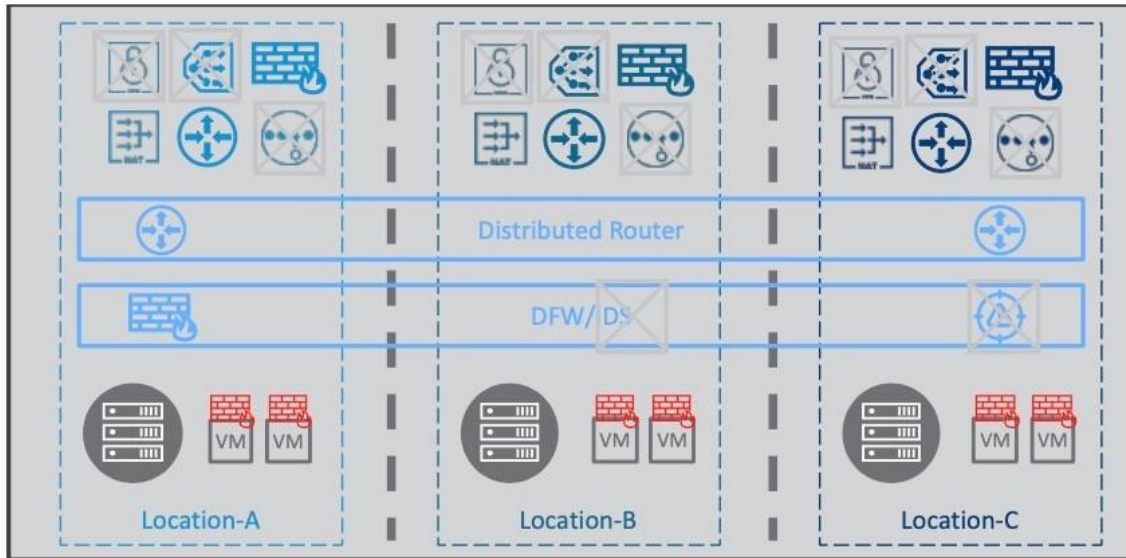
Networking in NSX-T Federation

NSX-T Federation currently supports the following network services:

- Switching: Overlay and VLAN
- IPAM: DHCP Relay, static binging, and DNS
- Routing: NAT and route redistribution
- Routing protocols: BGP and Static

The following figure shows NSX-T Federation Network services.

Figure 7-1. NSX-T Federation Network Services



This chapter includes the following topics:

- Global Manager Network Services
- Routing Protocols

Global Manager Network Services

This section details new options that Federation brings: Network objects span and Tier-0/Tier-1 gateway primary and secondary locations

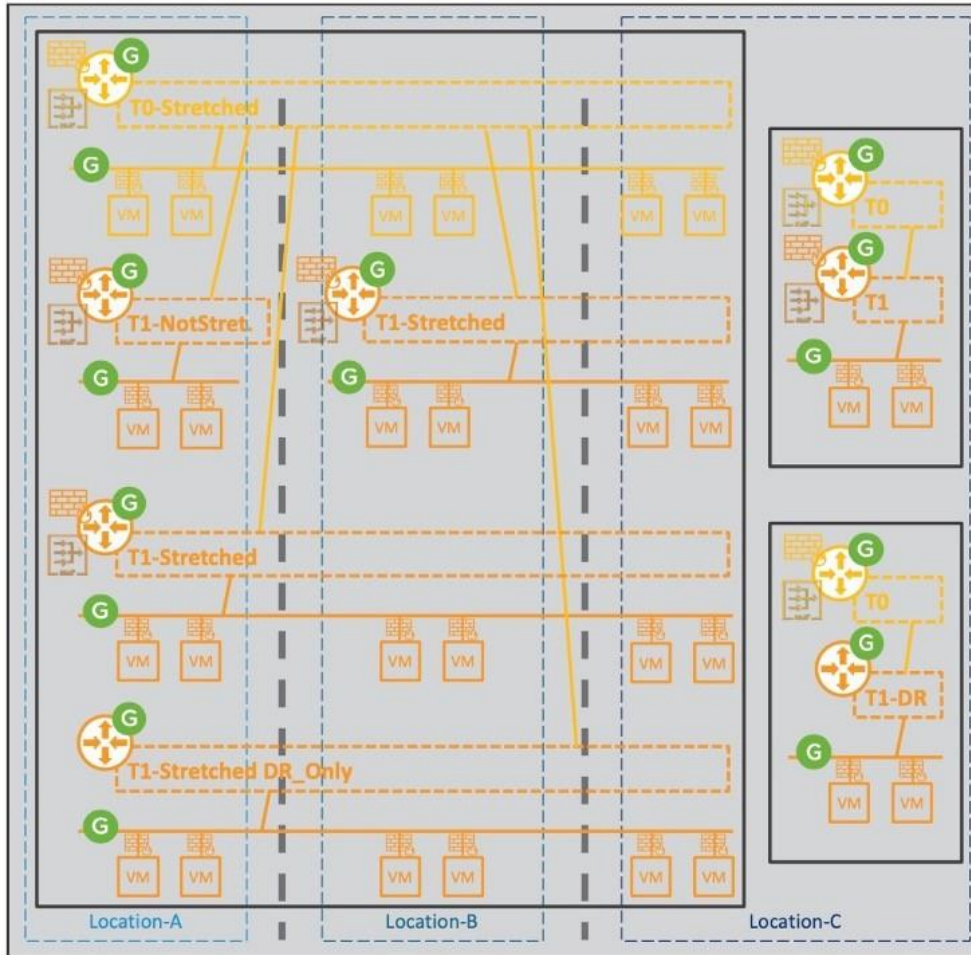
Network Objects Span

GM Tier-0, Tier-1, and Segment-Overlay objects are defined as stretched (multi-locations) or not stretched (single location).

In the following Figure, there are different Tier-0, Tier-1, and Segment-Overlays with different spans. There are a few rules to keep in mind:

- GM Tier-1 DR_Only span equals the attached T0 span
- GM Tier-1 with SR spans is equal or a subset of T0 span
- GM Segment-Overlay span is always equal to its attached Tier-0 and Tier-1 span
- GM Segment-Overlay is realized only when it is attached to Tier-0 or Tier-1

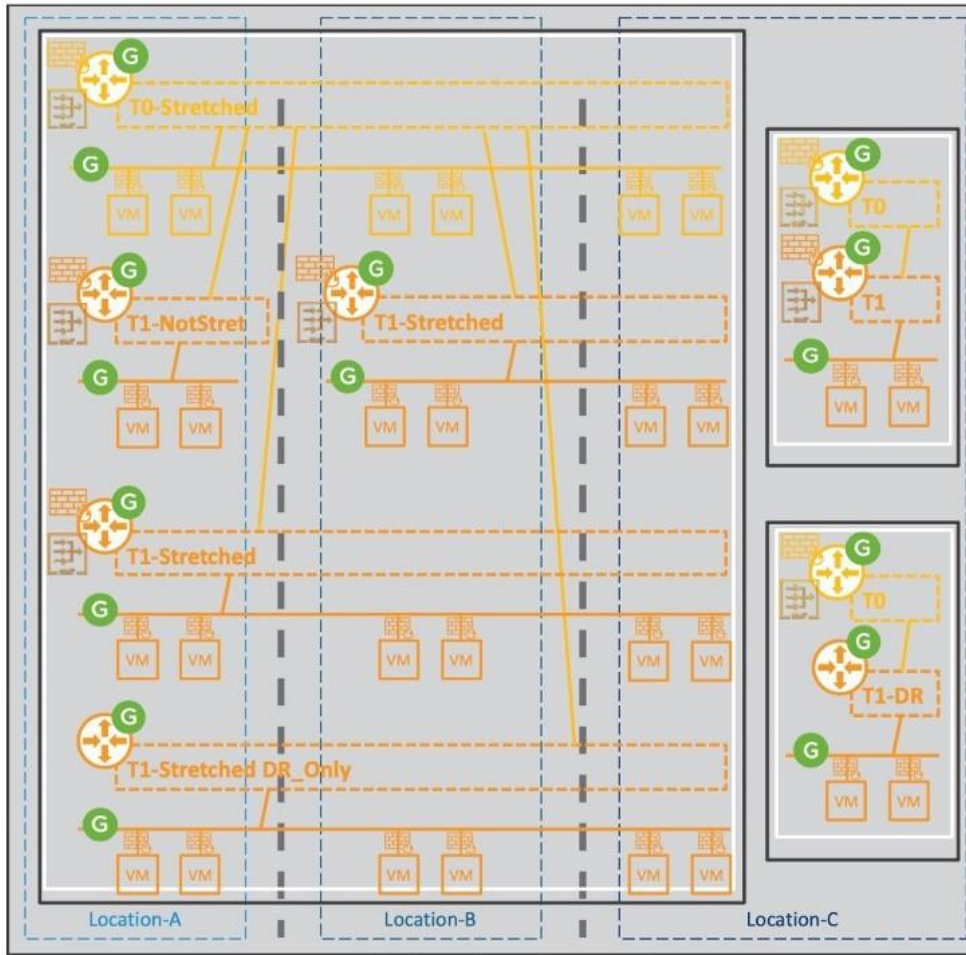
Figure 7-2. Different Supported Tier-0 / Tier-1 / Segments Topologies



GM Segment Configuration Options

GM Segment-Overlay span is always equal to its attached Tier-0 and Tier-1 span. Additionally, GM Segments are realized only when attached to Tier-0 or Tier-1, as shown in the following Figure.

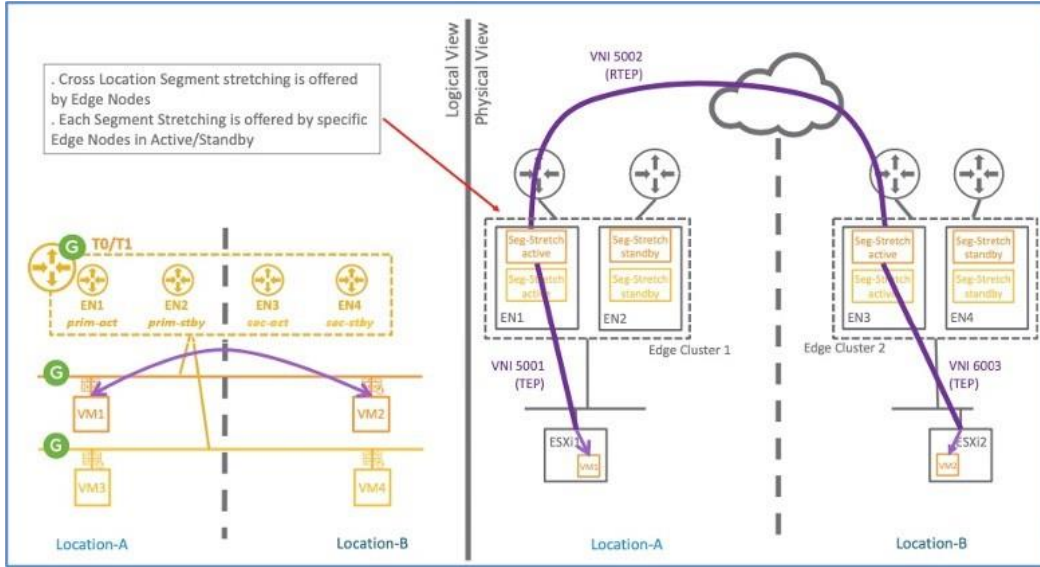
Figure 7-3. NSX-T Federation Segments Topologies



GM Segment Data Plane

The Edge Nodes handle the L2 cross-location traffic to avoid managing many Tunnels/BFD between all hosts cross sites. The following Figure shows the cross-location logical and physical packet walk.

Figure 7-4. Cross-location Logical and Physical Packet Walk



Each stretched segment cross-location communication is offered by different Edge Nodes in Active/Standby mode to offer Edge Node load distribution.

The selection of the pair of Edge Nodes is based on the segment attachment.

Segments	Description
Segments attached to T1_DR	<p>All segments connected to the same T1_DR use the same two Edge Nodes. Some segments are active on one and some segments are active on the other one.</p> <p>The two Edge Nodes selected for the segments of that T1_DR are two Edge Nodes from the Edge Cluster used by the T0 connected to that T1_DR.</p> <p>Different T1_DR connected to the same T0 use different two Edge Nodes if the Edge Cluster that the T0 uses has more than two Edge Nodes.</p>
Segments attached to T1_SR	<p>All segments connected to the same T0 Active/Standby use the same two Edge Nodes. Some segments are active on one and some segments are active on the other one.</p> <p>The two Edge Nodes selected for the segments of that T1_SR are the two Edge Nodes hosting that T1_SR.</p>
Segments on T0 Active/Standby	<p>All segments connected to the same T0 Active/Standby uses the same two Edge Nodes. Some segments are active on one and some segments are active on the other one.</p> <p>The two Edge Nodes selected for the segments of that T0 Active/Standby are the two Edge Nodes hosting that T0_SR.</p>

<p>Segments on T0 Active/Active</p>	<p>All segments connected to the same T0 Active/Active use the same two Edge Nodes. Some segments active on one and some segments active on the other one.</p> <p>The two Edge Nodes selected for the segments of that T0 Active/Active are two Edge Nodes out of the Edge Nodes that T0 Active/Active uses.</p>
-------------------------------------	--

Remote TEP (RTEP)

The cross-location communication between Edge Nodes uses RTEP and Geneve encapsulation. RTEP IP is only supported by Edge Nodes. Also, fragmentation on RTEP is allowed, but for best performance, you should avoid fragmentation.

The LM selects the VNI-TEP.

The GM selects the VNI-RTEP and communicates it to each LM. There is no incidence if one LM already uses the same VNI for its TEP because TEP and RTEP VNI are in different zones.

L3 Routing Service

Tier-0 gateways, Tier-1 gateways, and Segments can span one or more locations in the NSX-T Federation environment.

When you plan your network topology, keep these requirements in mind:

- Tier-0 and Tier-1 gateways can have a span of one or more locations.
- The span of a Tier-1 gateway must be equal to, or a subset of, the span of the Tier-0 gateway it is attached to.
- A segment has the same span as the Tier-0 or Tier-1 gateway it is attached to. Isolated segments are not realized until they are connected to a gateway.
- NSX Edge nodes in the Edge Cluster selected on the Global Manager for Tier-0 and Tier-1 gateways must be configured with the Default TZ Overlay.

You can create several topologies to achieve different goals:

- You can create segments and gateways that are specific to a given location. Each site has its own configuration, but you can manage everything from the Global Manager interface.
- You can create segments and gateways that span locations. These stretched networks provide consistent networking across sites.

Routing Options

The following tables provides the stretched Tier-0 and Tier-1 routing options.

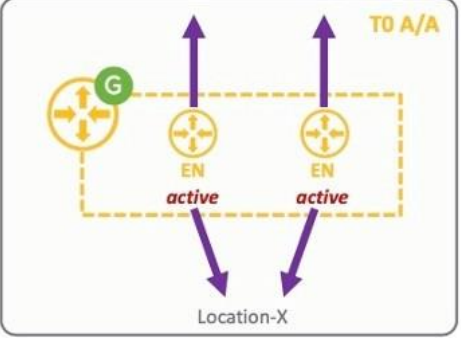
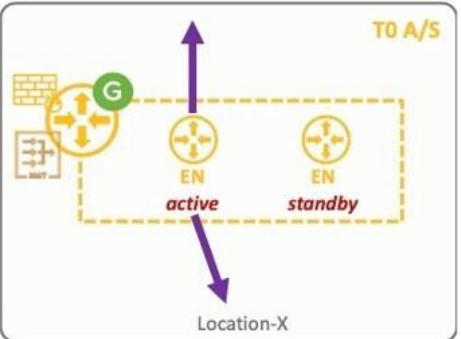
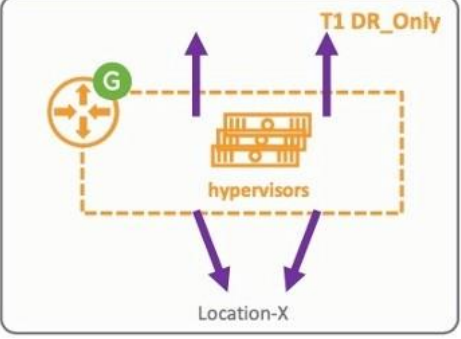
Routing Option	Description
All Primary	This option is only available for Tier-0 without Services. VMs North/South traffic is sent to their default gateway, which is hosted by their local Edge Nodes. Each Edge Node forwards it locally to the fabric.

Primary/Secondary	This option is available for Tier-0 and Tier-1 SR + DR. VMs North/South traffic is sent to their default gateway, which is hosted by their local Edge Node.
T1-Stretched DR_Only	This option is only available for Tier-1 without Services. VMs North/South traffic is sent to their default gateway, which is hosted by their hypervisor. Each hypervisor forwards it locally Tier-0.

Within a location, the Gateways can be in Active/Active or Active/Standby mode. In the specific case of T1-Stretched DR_Only, the router is only distributed and on the hypervisors.

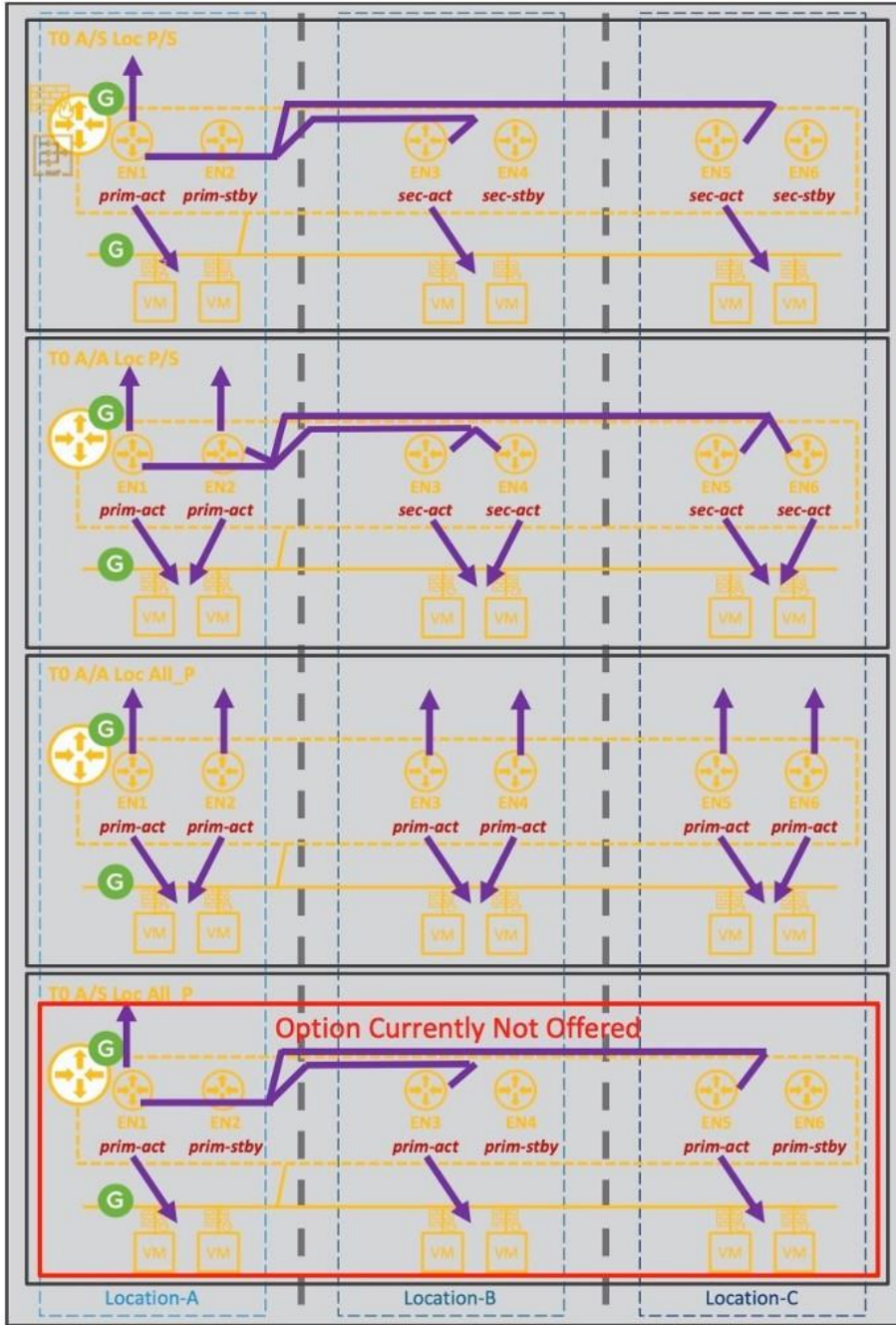
Availability Options

The following table shows the different availability options for Tier-0/Tier-1 within a location.

Availability Option	Description	Example
Active/Active	<p>This option is only available for Tier-0 without Services. Within a location North/South traffic is sent to evenly across all Edge Nodes that are part of the Tier-0.</p> <p>Each Edge Node forwards it locally to the fabric.</p>	
Active/Standby	<p>This option is available for Tier-0 and Tier-1 with Services. Within a location North/South traffic is sent only to the Edge Node hosting the Tier-0 or Tier-1 Active. This Edge Node forwards it locally to the fabric.</p>	
T1 DR_Only	<p>This option is only available for Tier-1 without Services. Within a location North/South traffic is sent directly by all hypervisors.</p> <p>Each hypervisor forwards it locally to the Tier-0.</p>	

All Possible Tier-0 Configuration Options

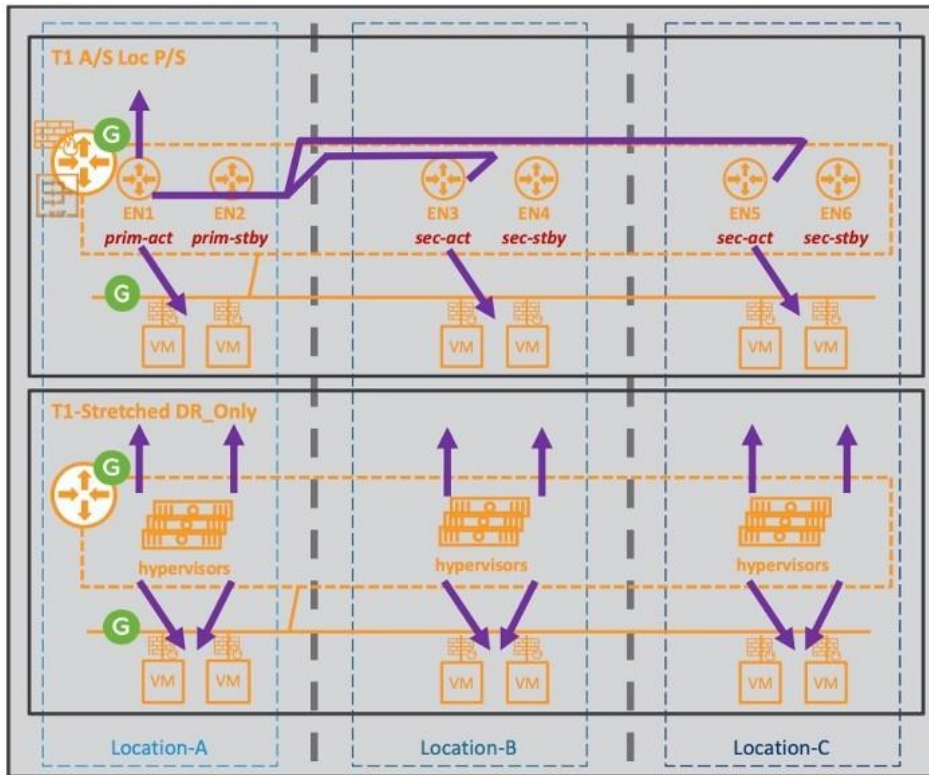
In the following Figure, you can see different stretched Tier-0 configuration options.



Block on the Figure	Option	Description
1st	Active/Standby Location Primary/ Secondary	Available for Tier-0 with Services. VMs North/South traffic sent to their local Edge Node hosting the Tier-0- Active. Those Edge Nodes forward it to the Edge Node hosting the Tier-0 Primary/Active. This Edge Node forwards it locally to the fabric.
2nd	Active/Active Location Primary/ Secondary	Available for Tier-0 without Services. VMs North/South traffic sent to their local Edge Nodes hosting the Tier-0-Active. Those Edge Nodes forward it to the Edge Nodes hosting the Tier-0 Primary/Active. Those Edge Nodes forward it locally to the fabric.
3rd	Active/Active Location All Primary	Available for Tier-0 only without Services. VMs North/ South traffic sent to their local Edge Nodes hosting the Tier-0-Active. Those Edge Nodes forward it locally to the fabric.
4th	Active/Standby Location Primary/ Secondary	This option is not currently offered.

All Possible Tier-1 Configuration Options

Note The following Figure shows the available stretched Tier-1 configuration options.



Block on the Figure	Option	Description
1st	Active/Standby Location Primary/ Secondary	Available for Tier-1 with Services. VMs North/South traffic sent to their local Edge Node hosting the Tier-1-Active. Those Edge Nodes forward it to the Edge Node hosting the Tier-1 Primary/Active. This Edge Node forwards it locally to the fabric.
2nd	T1 DR_Only Location Primary/ Secondary	Only available for Tier-1 without Services. VMs North/South traffic sent directly by all hypervisors. Those hypervisors forward it locally to the Tier-0.

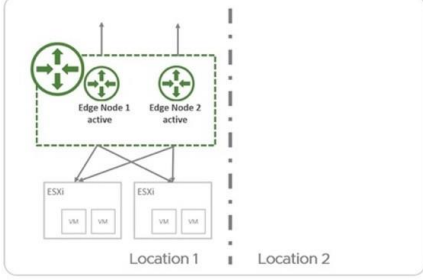
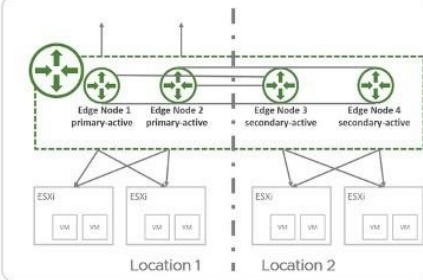
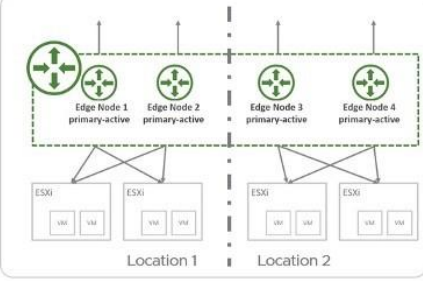
Note Ipv4 and Ipv6 routing are supported.

Routing Protocols

Let's begin with a recap of the various possible topologies with Stretched Tier-0 and Tier-1 gateways.

Tier-0 Gateway

Tier-0 gateways can have one of the following configurations.

Topology	Description	Example
<p>Non-stretched Tier-0 gateway</p>	<p>Create a Tier-0 gateway from Global Manager that spans only one location.</p>	
<p>Stretched Active/Active with primary and secondary locations</p>	<ul style="list-style-type: none"> ■ All Edge nodes are active simultaneously; therefore, Tier-0 cannot run stateful services ■ All traffic enters and leaves through the Edge nodes in the primary location ■ All egress traffic leaves through the primary location ■ If your environment has stateful services, such as external firewall, on the physical network, ensure the return traffic enters via the primary location ■ If you do not have stateful services on your physical network, and you choose to have asymmetric routing, disable Unicast Reverse Path Forwarding (uRPF) on all external Tier-0 interfaces. 	
<p>Stretched Active/Active with all primary locations</p>	<ul style="list-style-type: none"> ■ All Edge nodes are active simultaneously; therefore, tier-0 cannot run stateful services ■ All traffic enters and leaves through Edge nodes in the same location as the workloads ■ Traffic egresses locally from each location ■ Return traffic must enter the same location to allow 	

	<p>stateful services, such as firewall</p>	
<p>Stretched Active/Standby with primary and secondary locations</p>	<ul style="list-style-type: none"> ■ Only one Edge nodes is active at a time; therefore, Tier-0 can run stateful services ■ All traffic enters and leaves through the active Edge node in the primary location ■ NAT, Gateway Firewall, DNS, and DHCP are the supported services 	

Tier-1 Gateway

You can create a tier-1 gateway in NSX-T Federation for distributed routing only. This gateway has the same span as the tier-0 gateway it is linked to.

The tier-1 does not use Edge nodes for routing. All traffic is routed from host transport nodes to the tier-0 gateway. However, to enable cross-location forwarding, the Tier-1 allocates two Edge Nodes from the Edge cluster configured on the linked tier-0 to use for that traffic.

iBGP is internally used to exchange routes with the locations of the stretch Tier-0 gateways. eBGP and/or static routes are used to exchange routes between the Tier-0 gateway and the physical fabric.

Routing Protocol Table

The Tier-0 routes exchange varies based on its configuration and is summarized in the following tables.

T0 A/A Loc_P/S Routing Protocol Table

Protocol	Primary	Secondaries
eBGP receive	EN Primary	EN Secondaries
eBGP advertise	EN Primary: T1 DR_Only routes T1 SR+DR routes <ul style="list-style-type: none"> ■ Stretched Primary Local ■ Stretched Primary Remote ■ Not_Stretch Local ■ Not_Stretch Remote iBGP routes No	EN Secondaries: T1 DR_Only routes <ul style="list-style-type: none"> ■ User should add cost ++ to avoid asymmetric routing T1 SR+DR routes <ul style="list-style-type: none"> ■ Stretched Primary Local ■ Not_Stretch Local ■ User should add cost ++ to avoid asymmetric routing iBGP routes <ul style="list-style-type: none"> ■ No
iBGP receive*	Edge Node Primary: Yes	Edge Node Secondaries: Yes
iBGP advertise*	Edge Node Primary: Yes	Edge Node Secondaries: Yes
FIB	EN Primary: <ul style="list-style-type: none"> ■ Static ■ iBG P eBGP	EN Secondaries: <ul style="list-style-type: none"> ■ Static ■ iBGP ■ eBGP NOTE: Unlike usual routing, if the same route is received from multiple locations, route from EN-Primary is kept instead of

		the eBGP route.
--	--	-----------------

T0 A/A Loc_All_P Routing Table

Protocol	Primary	Secondaries
eBGP receive	Edge Node Primary: Yes	
eBGP advertise	EN Primary: T1 DR_Only routes T1 SR+DR routes <ul style="list-style-type: none"> ■ Stretched Active Local ■ Not_Stretched Local iBGP routes No	
iBGP receive*	Edge Node Primaries	
iBGP advertise*	Edge Node Primaries	
FIB	EN Primary Static	EN Secondaries: Stat ic iBGP eBGP

*iBGP is automatically configured within stretched Tier-0 and cannot be tuned (like set up filters).

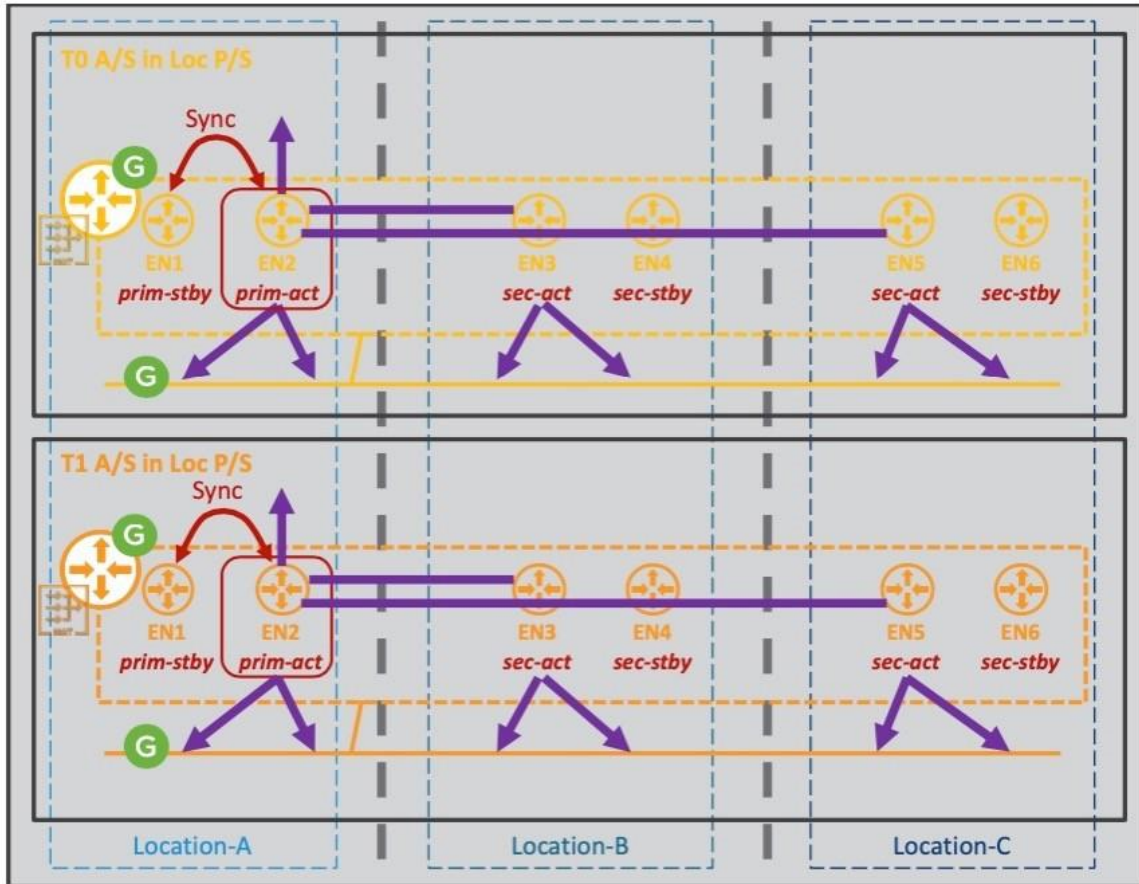
Stateful Network Address Translation (NAT)

NAT is available on Tier-0 and Tier-1 Active/Standby. The NAT function is offered by the Active element.

All NAT sessions are synchronized between the Primary/Active and Primary/Standby. There is no synchronization of NAT sessions to secondary Edge Nodes, as there is no need.

- Stretched Tier-0 and Tier-1, NAT is available on Tier-0 and Tier-1 Active/Standby Location Primary/Secondary. NAT is offered by the Primary/Active element.
- Primary/Active Edge Node failure, there is no data plane impact.

Figure 7-5. NSX Federation NAT Service

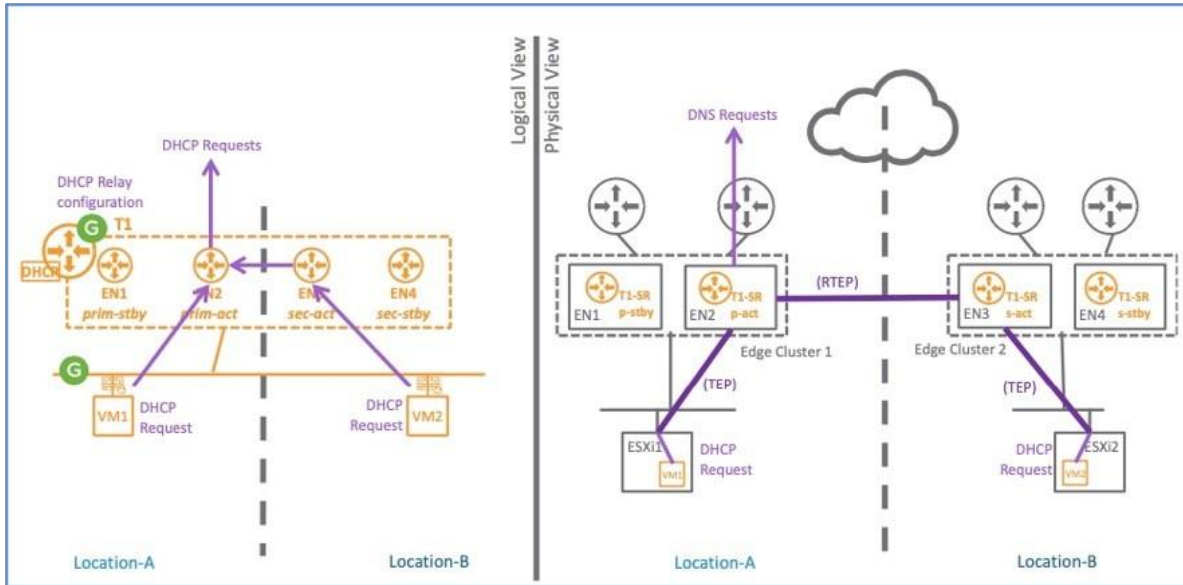


DHCP

DHCP Relay and static binding and DNS are supported from the Global Manager. The NSX DHCP Relay configuration is attached to a Tier-1.

GM DHCP Relay configuration is pushed to the different Local Managers hosting that Tier-1, but only the Edge Node hosting the Tier-1 Primary/Active has the DHCP service running.

Figure 7-6. DHCP Server with DHCP Relay



The Figure above shows the logical and physical packet walk of clients' DHCP requests:

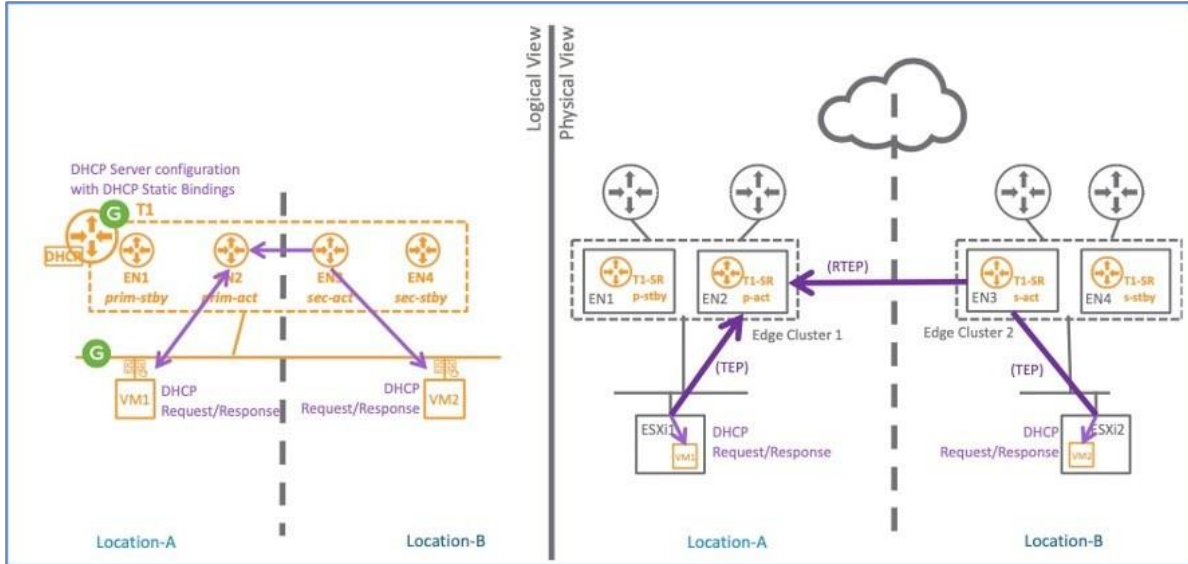
- In both cases, VM1_Location-A and VM2_Location-B, the DHCP response is offered by the Edge Node hosting the Tier-1 Primary/Active (EN2).
- For the VM2_Location-B, the cross-location traffic occurs between the Edge Node hosting Tier-1 Secondary/Active (EN3) and the Tier-1 Primary/Active (EN2).
- The Tier-1 then forwards clients' DHCP requests to its configured external DHCP server.

DHCP Server with DCHP Static Bindings

The NSX DHCP Service configuration is attached to a Tier-1.

GM DHCP Service configuration is pushed to the different LMs hosting that Tier-1, but only the Edge Node hosting the Tier-1 Primary-Active has the DHCP service running.

Figure 7-7. DHCP Server with DHCP Static Bindings



This Figure shows the logical and physical packet walk of clients' DHCP requests and server DHCP response:

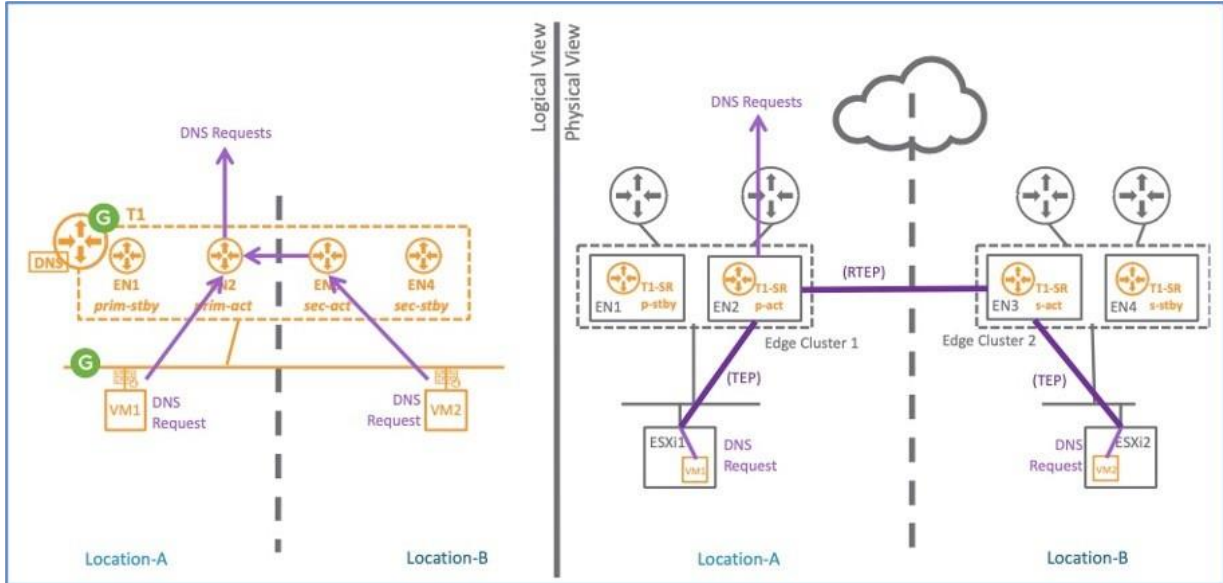
- In both cases, VM1_Location-A and VM2_Location-B, the DHCP response is offered by the Edge Node hosting the Tier-1 Primary/Active (EN2).
- For the VM2_Location-B, the cross-location traffic occurs between the Edge Node hosting Tier-1 Secondary-Active (EN3) and the Tier-1 Primary-Active (EN2).

DNS Service

The NSX DNS Service configuration is attached to a Tier-1.

GM DNS Service configuration is pushed to the different LMs hosting that Tier-1, but only the Edge Node hosting the Tier-1 Primary-Active has the DNS service running.

Figure 7-8. DNS Service



The Figure above shows the logical and physical packet walk of clients DNS requests:

- In both cases VM1_Location-A and VM2_Location-B, the DNS request is forwarded to the Edge Node hosting the Tier-1 Primary-Active (EN2).
- For the VM2_Location-B, the cross-location traffic is done between the Edge Node hosting Tier-1 Secondary-Active (EN3) and the Tier-1 Primary-Active (EN2).

The Tier-1 then forwards clients' DNS requests to its configured external DNS server.

Security in NSX-T Federation

NSX Federation security provides the following benefits:

- Consistent security policy across your deployments
- Effective disaster recovery that ensures continuity of established security framework
- Extension of network and security framework to another location if you run out of compute resources in one location

NSX-T Federation currently supports the following security services:

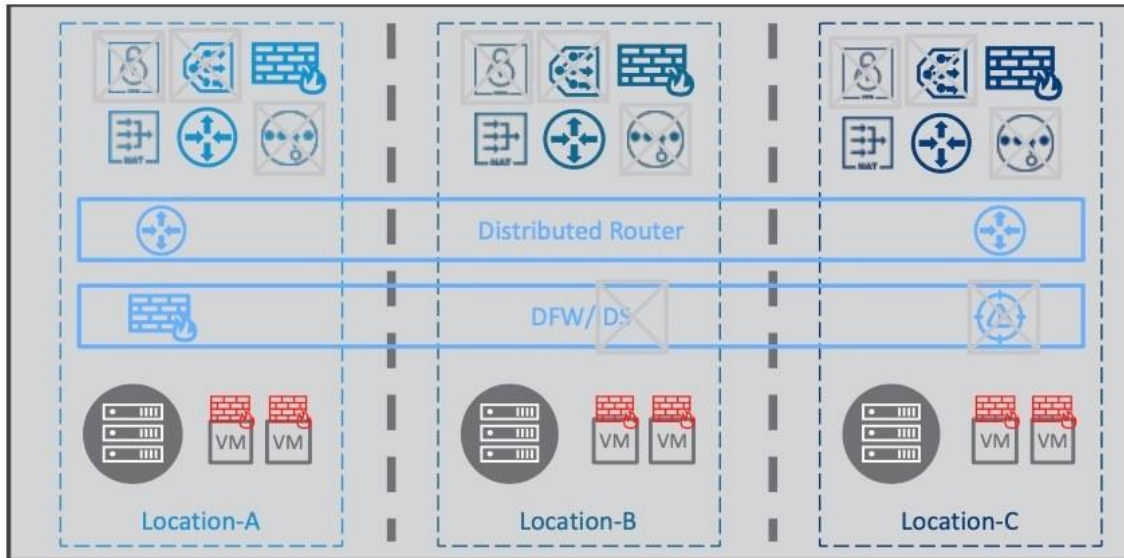
- Distributed Firewall
- Gateway Firewall

Federation supports both firewalls with L7 App ID context support. NSX-T Federation does NOT support the following security services:

- Time-based Firewall
- Identity Firewall
- Distributed IDS
- FQDN filtering
- Network Introspection
- Endpoint Protection

The following Figure shows NSX-T Security services.

Figure 8-1. NSX-T Federation Security Services



This chapter includes the following topics:

- Global Manager Security Services

Global Manager Security Services

In this section, we detail all of the security services supported within Federation.

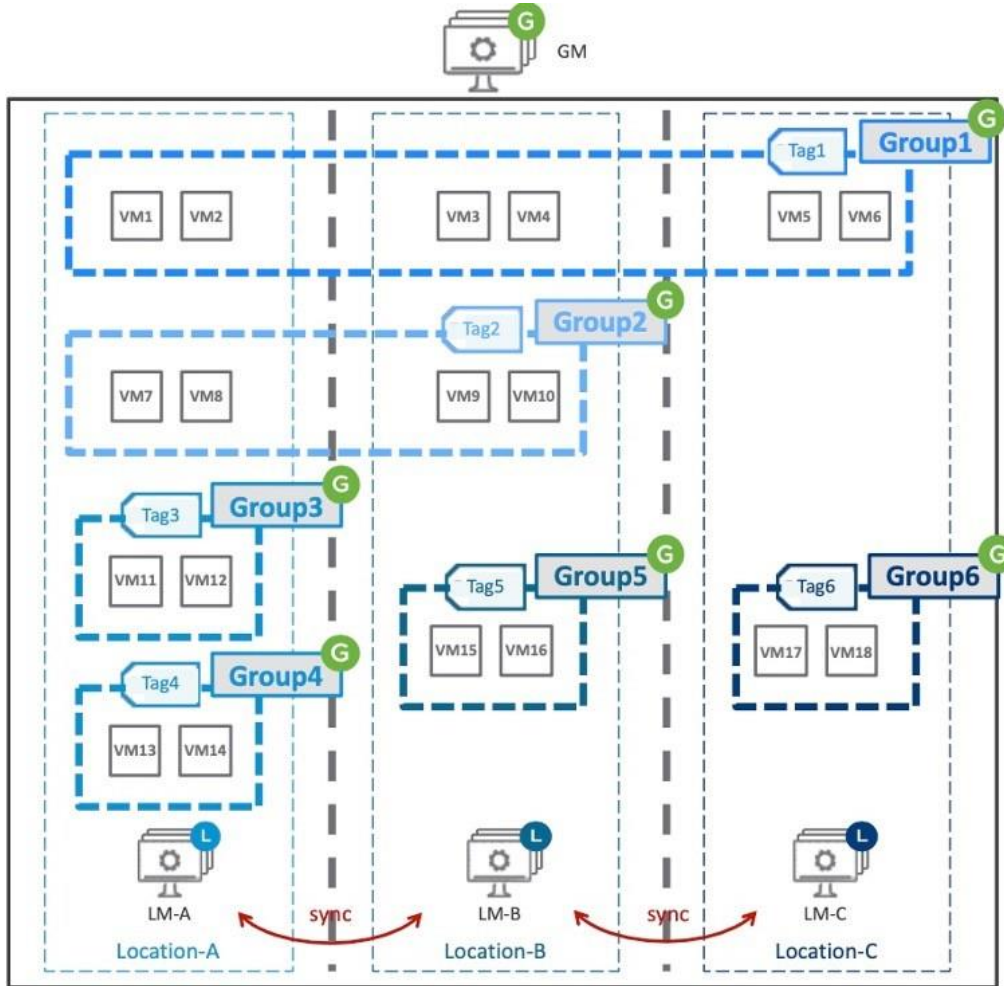
Global Manager Groups

GM Group configuration is very similar to LM Group configuration. Federation brings a new option for Groups through the Region feature.

You can define GM Groups as:

- Global: All locations
- Regional: Multiple locations
- Local: Single location

Figure 8-2. NSX-T Federation Group Span



In the Figure above:

- The Group1 span is Global and is pushed to all LMs (LM-A +LM-B+LM-C).
- The Group2 span is Regional (LM-A + LM-B).
- The Group3 + Group4 + Group5 span are Local, where each is in a specific LM location.

The membership of each group can be static or dynamic. Figure 65 shows dynamic membership with groups' membership based on VM tags.

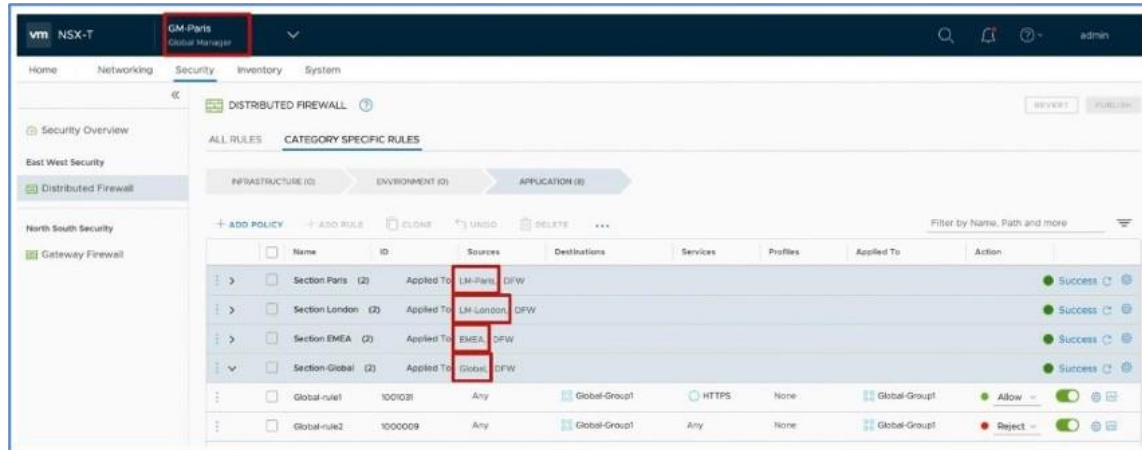
For each Global or Regional Group, each LM synchronizes its local members with the other LMs in the Group span.

Global Manager Distributed Firewall

GM DFW configuration is very similar to LM DFW configuration. DFW Sections use Regions, which you can consider as the span of the DFW Section. It's called DFW Policy in the UI.

You can define GM DFW Sections as Global, Regional, or Local. The DFW rules within a section have the same span as its DFW Section.

Figure 8-3. Distributed Firewall Span



All other GM DFW configuration options are the same as LM DFW options.

Each section has an "Applied-To" to define the scope of that section (which VMs will receive those section rules). If no specific "Applied-To" is configured at the section level, an "Applied-To" can be defined at the rules level to define the scope of that rule.

There are a few constraints in creating GM DFW Sections and Rules:

- GM can create the firewall in all categories except Ethernet and Emergency. Only the Local Manager can create firewalls in Ethernet and Emergency categories.
- GM cannot see or edit the LM Default Layer3 Section in Application.
- GM DFW Rules:
 - In DFW Section Local: Must have Source or Destination equals to ANY or Groups that belong to the same location.
 - In DFW Section Region: Must have Source or Destination equals to ANY or Groups that belong to the same region or location within that region.
 - In DFW Section Global: Source or Destination can be anything.

The created GM DFW Sections and Rules are pushed to the LM associated to the span of the GM DFW Section. For instance, the DFW Sections Global are pushed to all LMs, and the DFW Sections Paris are pushed only to LM-Paris.

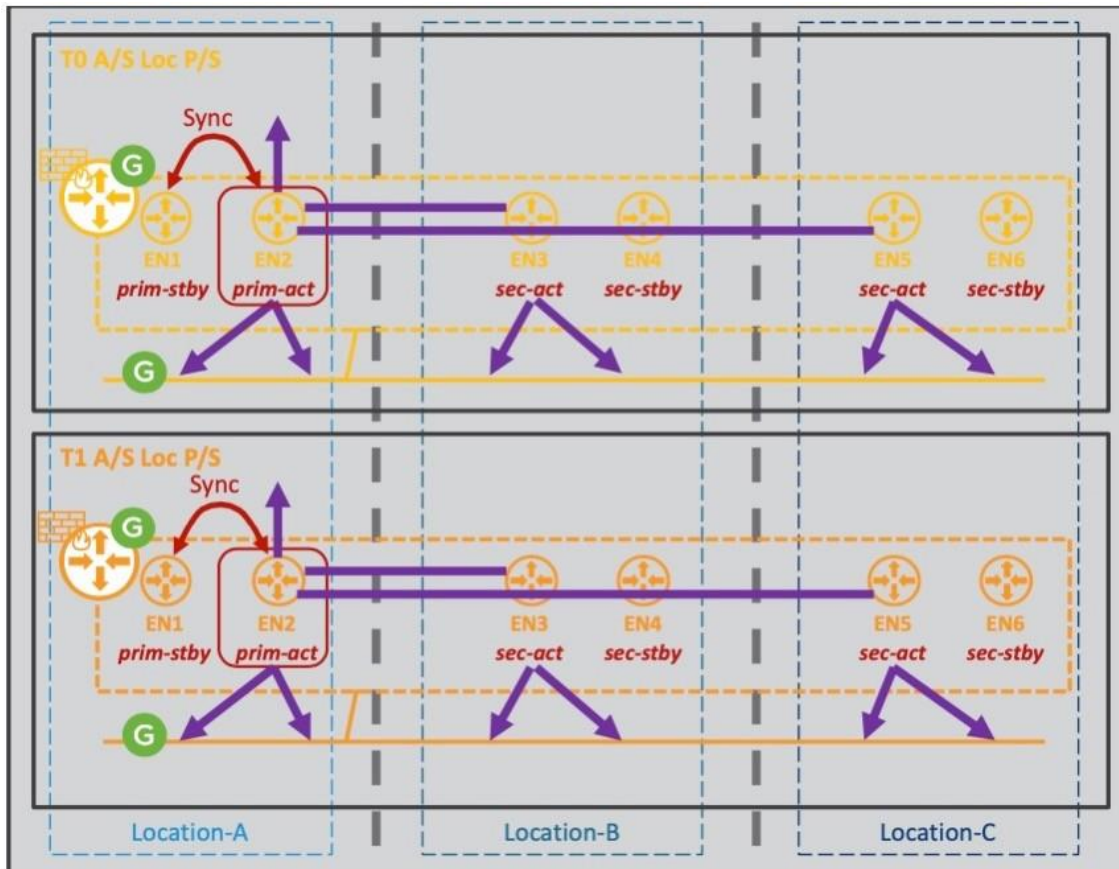
LM receives the GM DFW Sections in each category (Infrastructure, Environment, Application) and always places them on top of its LM DFW Sections within each category. To have an LM DFW rule above any GM DFW Section, the LM can create DFW Rules under the Emergency category".

Global Manager Gateway Stateful Firewall

Gateway Firewall is available on Tier-0 and Tier-1 Active/Standby. The Active element offers the Gateway Firewall function:

- In a stretched Tier-0 and Tier-1, Gateway Firewall is available on Tier-0 and Tier-1 Active/ Standby, Location Primary/Secondary. The Primary/Active element offers the Gateway Firewall function.
- All Gateway Firewall sessions are synchronized between the Primary/Active and Primary/ Standby.
- If there is a Primary/Active Edge Node failure (EN2), there is no data plane impact.
- We cover loss of location and data plane recover in the Data Plane Recovery section
- Region construct is only used for DFW. The span of Tier-0 and Tier-1 is defined by its selected locations.
- Groups used in Gateway Firewall Rules must have the same span as the Gateway.

Figure 8-4. Gateway Firewall Service



NSX-T Federation Limitations

Management

- GM with vIDM: GM-Active does not synchronize vIDM configuration to GM-Standby
- Port Mirroring

Networking

- All Networking features are supported from GM, except:
 - L2 Bridge
 - DHCP dynamic binding
 - Routing VRF and EVPN
 - Layer4+ services Load Balancing and VPN
 - OSPF
- List of supported LM Network features configured from LM after registered by GM in the GM to LM Communication Flow section.
- No Tier-0/Tier-1 Automatic Disaster Recovery
 - Network DR requires GM Tier-0/Tier-1 primary location configuration change.

Security

- Stretched Groups based on Segment Ports or Segment Ports Tags do not support VM cold vMotion / SRM across locations.
- All Security features are supported from GM, except:
 - Time-Based Firewall
 - Identity Firewall
 - Distributed IDS
 - Network Introspection
 - Endpoint Protection

- Malware Prevention
- Network Detection and Response
- Distributed Security for vCenter VDS Port Group (using GM dynamic group membership based on LM VDS Port Group Tags)

- List of supported LM Security features configured from LM after registered by GM in the GM to LM Communication Flow section.