SGA 4

Introduction

Streaming Gateway Appliance (SGA) 4 simplifies the deployment and management of the SGA by eliminating the need for:

- Public key certificates, as HTTPS is no longer used to communicate between Frame Terminal and the SGA.
- Load balancer, as Frame control plane is responsible for load balancing user sessions across an SGA cluster.

SGA 4 supports:

- Self-service features within Frame Console to:
  - View the status of all SGA clusters and SGA nodes (VMs).
  - Manage the lifecycle of both manually and auto-deployed SGA VMs.
  - Attach (and detach) a Frame account to an SGA cluster.
  - Ability to power on and off individual auto-deployed SGA nodes.
  - Add and delete SGA nodes for a given SGA cluster.
- Customers who require outbound traffic to have a different source public IP address than the inbound public IP address of each SGA VM.
- Customers who require their SGA VMs to have two virtual network interfaces (a virtual network interface for traffic between users on the Internet and the SGA VM and a private virtual network interface for traffic between the SGA VM and the workload VMs).

SGA 4 can only be used with Frame Remoting Protocol (FRP) 8.

Limitations

- SGA 4 image for ESXi will be supported in a future release. SGA 4 is now Generally Available for all other infrastructures.

SGA 4 Clusters and Nodes

SGA 4 introduces two new concepts for customers: SGA Cluster and SGA Node. An SGA Cluster is composed of one or more SGA Nodes, where each node is an SGA 4 VM. Each SGA Cluster is deployed in a specific public cloud region to support Frame Accounts in that region or deployed on-premises to support one or more AHV VLANs. Customers may deploy more than one SGA Cluster. A Frame Account may only be associated to one SGA Cluster.

Once an SGA cluster with one or more nodes is created with at least one node powered on, customers can then in Frame Console:

1. Create a new Frame Account, specifying an existing SGA Cluster.
2. Attach a previously created Frame Account to an existing SGA Cluster (to be supported in a future release).
3. Detach a Frame Account from its SGA Cluster to ensure users can only access the Frame workload VMs in a private networking deployment model (to be supported in a future release).

Network Requirements

When deploying an SGA VM, the customer's network must: (1) allow Internet traffic to reach the SGA VM and (2) from the SGA VM to the network containing the Frame-managed workloads (e.g., Sandbox, test/production pools, Utility Servers). As a best practice, we recommend the SGA VM or VMs (if high availability is required) be deployed in a DMZ (e.g., VPC, VNET, or VLAN) network, separate from the workload VM network.

Customers who have previously configured their network for SGA 3.X will need to allow SGA 4 VMs to initiate outbound HTTPS/Secure WebSocket (tcp/443) connections to cch.console.nutanix.com for communication with Frame control plane.

Customers who are starting with a new network will need to configure their network to satisfy the Frame private networking with SGA 4 network requirements.

Consult Public Cloud with Private Networking and SGA or Nutanix AHV with Private Networking and SGA to ensure that network requirements are satisfied before continuing to SGA 4 installation and configuration.

SGA 4 VM provides Frame Platform its public IP address based on the following:

1. For automated deployments of SGA 4, the public IP address returned from the cloud provider's Instance Metadata Service (IMDS) endpoint.

2. For manual deployments of SGA 4, the public IP address specified by the administrator before the SGA Node is registered to the Frame control plane using the Registration Code.

> **NOTE**
> 1. While each SGA VM must have an associated public IP address, the public IP address does not have to be attached to virtual network interface of the SGA VM itself. Instead, the customer administrator can manually deploy an SGA VM with only a private IP addresses and then configure a NAT rule either on their firewall, web application firewall, or load balancer that maps the inbound public IP address of the SGA VM to the corresponding private IP address on the SGA VM.
>
> 2. SGA 4 no longer requires a corresponding DNS A record for its public IP address; however, customers can create DNS records for their SGA 4 public IP addresses, if desired.
>
> 3. SGA 4 does not support IPv6 addresses.

High Availability

With SGA 4, Frame control plane will handle load balancing user session requests across the available SGA nodes in the SGA cluster. A load balancer is no longer needed to perform the load balancing function.

High Availability SGA 4 Architecture (FRP8)

High Availability SGA 4 Architecture (FRP8)

Typical FRP8 Workflow

Frame users log in to the Frame Platform and are directed to their Launchpad. When a user clicks the desktop or an application icon in their Launchpad, Frame Platform provides the user's browser with the public IP address of the SGA VM associated with the Frame account.

The user's browser or Frame App begins communicating directly with the specific SGA VM using the provided public IP address using (or). The SGA VM validates the session start request and then forwards the session start request to the user's assigned Frame workload VM using . The Frame Agent on the workload VM validates the session start request and begins the Frame session video/audio stream. FRP8 traffic flows back from the Frame Agent on the workload VM through the SGA VM to the user's browser or Frame App.

Internal Access to SGA-enabled Workloads

SGA 4 also supports the scenario where end users within the private network access the workload VMs of an SGA-enabled Frame account while users on the Internet are accessing workload VMs trough the SGA.

During the WebRTC Interactive Connectivity Establishment (ICE) candidate exchange between user and workload VM, FRP8 will test all ICE candidate pairs and determine the best ICE candidate pair to use. If WebRTC verifies that the user and workload VM can communicate over an internal network path, then the FRP8 stream will use that internal network path.

For internal access by users to the workload VMs of an SGA-enabled Frame account, ensure that the users within their private network can route their traffic to the workload VMs in the private network following the private networking requirements for private networking (public cloud) or private networking (AHV) between the end user and workload VMs.

Multi-Frame Account Support

An SGA 4 cluster can be configured for one or more Frame accounts. If there are Frame accounts in different regions or data centers, we recommend you deploy SGA 4 clusters in each of those different regions or data centers to minimize unnecessary network latency.

Security

SGA 4 appliances use Ubuntu 22.04.3 LTS, hardened using CIS Level 1 Server profile (https://ubuntu.com/security/certifications/docs/usg/cis/compliance). SGA administrators can only access the SGA VM command line only through the infrastructure console. SSH is disabled.

The following ports are bound on the SGA VMs:

- 3478 – (udp/tcp) for FRP8
- 4369 – restricted to localhost requests only by SGA component
- 53 – (udp/tcp) restricted to localhost requests only for Ubuntu systemd-resolve service (DNS)

When a user connects to an SGA 4 node, SGA validates the user session request by confirming the validity of the request with the Frame control plane, before connecting the user with the assigned workload VM.

All communication between the SGA 4 VM and the Frame control plane is conducted using a Secure WebSocket (WSS) connection. The WSS connection is initiated by the SGA 4 VM using HTTPS. During the registration process, the SGA 4 VM will authenticate itself to the Frame control plane using a registration code generated by the control plane (and manually entered by the customer administrator for manually deployed SGA VMs) and provide the SGA 4 VM-specific metadata (UUID, SGA public-private key pair, SGA VM public IP address). Once the Secure WebSocket connection is established, the Frame control plane can communicate with the SGA VM to broker new user sessions, facilitate FRP8 WebRTC negotiation, and monitor the availability of the SGA VM.

The public/private key pair is used by the SGA VM to authenticate itself to Frame control plane each time the SGA VM needs to establish a Secure WebSocket connection to the Frame control plane. The private key is used to sign the initial HTTPS GET request by the SGA VM and the digital signature is sent as one of the HTTPS headers, including the timestamp, UUID, and nonce. The control plane validates the digital signature using the SGA VM public key before agreeing to switch to a Secure WebSocket for bidirectional communication.

Revision #7
Created 1 October 2025 04:47:48
Updated 13 January 2026 11:39:00 by Nikola Savic