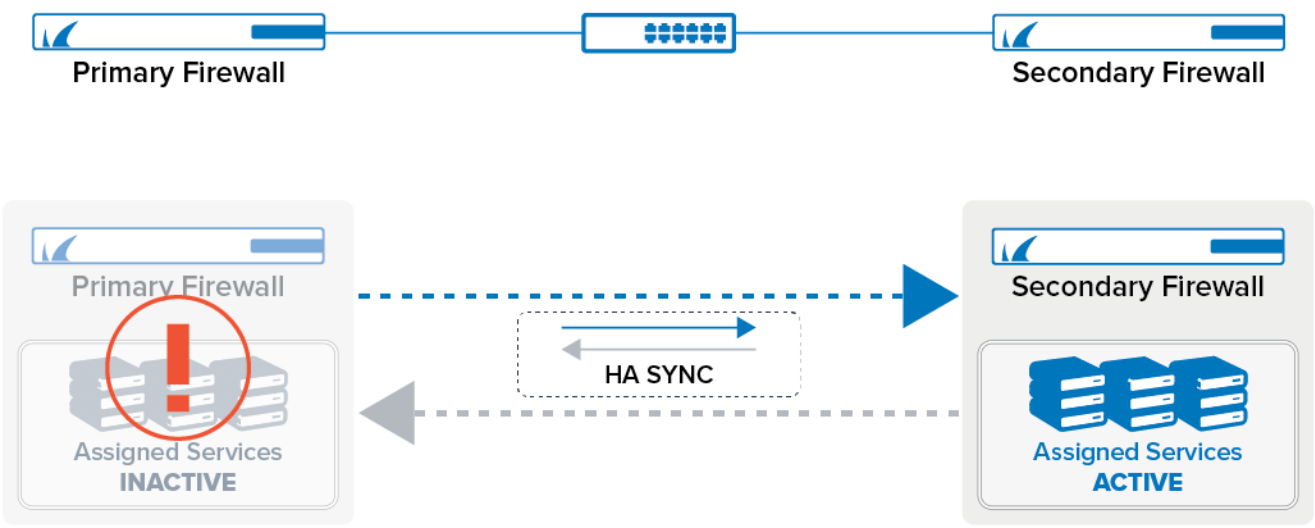
High Availability

https://campus.barracuda.com/doc/96026361/

High Availability ensures that the services running on the Barracuda CloudGen Firewall are always available even if one unit is unavailable due to maintenance or a hardware fault. Reliable HA depends on the correct configuration of the surrounding switches and routers. Especially important is the ARP cache time or ARP timeout which must be set to a value between 30 and 60 seconds. When one or more services fail over to the secondary firewall, the MAC addresses associated with the service IPs also change. The MAC address is immediately sent out via gratuitous or unsolicited ARP requests, updating the MAC address table or ARP cache of the connected switches and routers. If the lifetime of the ARP timeout of the switch is set to be longer, for example, 300 seconds, the secondary unit would not be reachable for up to 5 minutes, because the ARP cache would not be updated for that time. Longer timeouts also increase the number of ARP requests sent out by the firewall, increasing the load on the switch.



Management of an HA-Pair

When configuring an HA-pair, the secondary firewall receives the configuration from the primary firewall. For this, the Network node in the configuration tree is replaced by another Network node that accepts the input of the Management IP of the secondary firewall. With the exception of the Management IP, all other configuration data is mirrored to the secondary firewall.

The secondary firewall receives its device name from the primary firewall. In addition to that, '-HA' is appended in order to be able to distinguish the secondary from the primary firewall:

Example Name of Primary Firewall	Example Name of Secondary Firewall

MyFirewall	MyFirewall-HA

For the purpose of better manageability, only the primary firewall will be visible in the configuration tree. For a better overview, however, both the primary and the secondary firewall will be visible in the Control Center's Status Map as soon as their configuration is completed.

Requirements and Limitations for High Availability

- Both units must use the same platform: You cannot mix virtual and physical appliances.
- Both units must be the same model. Using different revisions of the same hardware appliance is possible.
- Both units must have the same license- and subscription status. Running a box in demo mode causes HA synchronization to fail.
- If you are running an HA setup with different appliance revisions, ensure that both physical ports of the private uplink are using identical port labels. Otherwise, HA synchronization may fail.
- Both Management IPs for the HA pair must be in the same network.
- Latency on the HA sync connection may not exceed 80 ms.

Standalone High Availability

For a standalone HA cluster, the primary unit downloads the licenses for both units, and when the secondary unit is joined to the HA cluster, the license for the secondary unit is transferred over. The licenses are bound to the MAC addresses of the primary and secondary units.

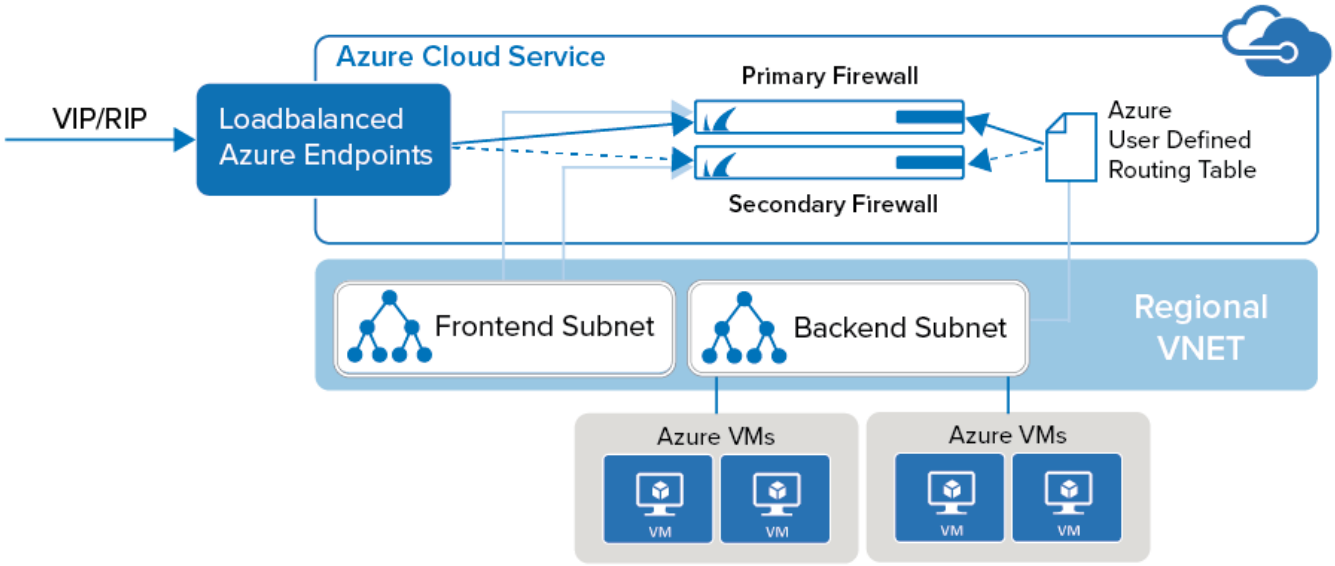
For more information, see How to Set Up a High Availability Cluster.

High Availability for Managed Firewalls

You can also configure an HA cluster with two managed firewalls as primary and secondary box in the cluster on the Firewall Control Center. The box level configuration and licensing of the firewall are completely separate.

For more information, see How to Set Up a Managed High Availability Cluster from Scratch and How to Set Up a Managed High Availability Cluster from Two Stand-Alone Firewalls.

High Availability in Azure



To be able to run a high availability cluster for firewalls in Microsoft Azure, a special setup is required to integrate into the Azure networking environment.

For more information, see High Availability in Azure.

High Availability with a Private Uplink

When configuring HA without a private uplink, the switch both firewalls are connected to represents a single point of failure. If traffic is not forwarded by the switch, the HA sync breaks because the primary and secondary units cannot establish a connection. To always have a reliable connection, you can configure a private uplink.

For this, one network interface must be dedicated for HA purposes. It is recommended to directly connect the two firewalls and use a /30 subnet for the uplink. You can configure the firewall to use either just the MIP or also the secondary network interface so that both HA partners can redundantly verify their availability. If you configure a second uplink interface, the IP address of that interface will

be associated with the interface and its configured MIP on the same firewall.

For more information, see How to Configure a Private Uplink for a High Availability Cluster.

In Depth: Transparent Failover Procedure and Limitations

An HA system can be used for load balancing to exploit all features that are available through the CloudGen Firewall architecture. Use transparent failover to synchronize the forward packet sessions (inbound and outbound TCP, UDP, ICMP-Echo, and OTHER-IP-Protocols) of the Firewall server between the two HA partners. Transparent failover is enabled by default and is set per access rule.

Unsynchronized Components

The following information is not HA-synced:

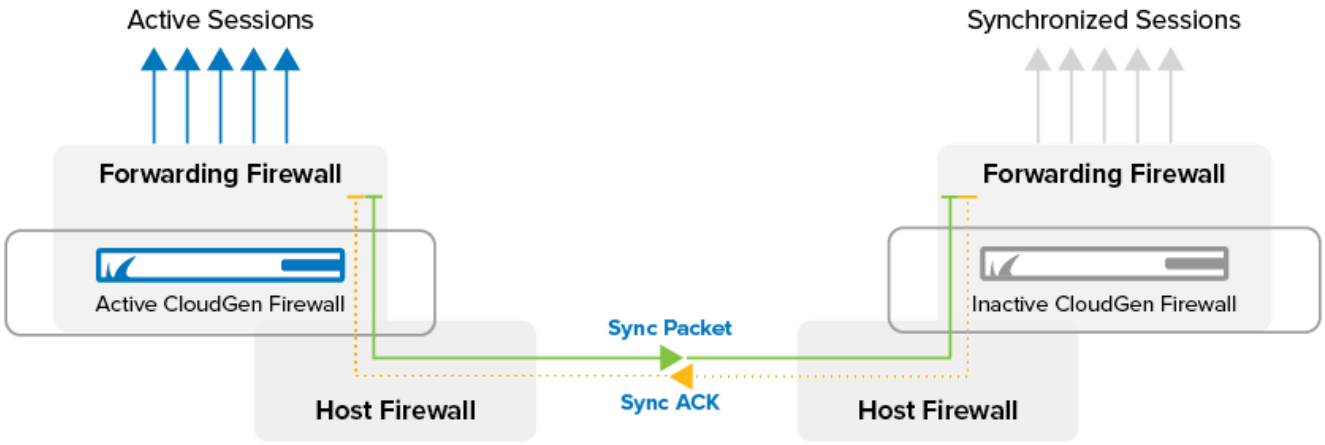
Module or Component	Unsynchronized Sub-Components
Firewall	No Transparent Failover HA Sync for the Following Connections: • Sessions using Dynamic IPs: **(xDSL, DHCP, Wireless WAN)** • Sessions using **Box IP Addresses**, **Additional Local IPs** • SSL-Decryption Sessions • FTP Sessions • HTTP Sessions with **Archive Content Scan** • Generic TCP-Proxy Sessions • RAW-TCP • WANOPT Sessions • Sessions excluded from HA synchronization via **Advanced Rule Settings** in the matching access rule No Data Synchronization for Content in: • Firewall History • Firewall Monitor-Data • ATP-Scan Queue Limited Functionality for Synchronized Sessions after an HA-Takeover: • Application/Protocol/Content information • IPS
VPN Service	• IPSec tunnels
Access Control Service	All
Eventing	All
Logging	All
Statistics	All

Home Directories (Admins)	All
SMS Messages	All

Synchronizing Procedure

Synchronization can be carried out via dedicated HA uplink and/or the LAN connection. Synchronization traffic is transmitted by AES-encrypted UDP packets, so-called sync packets, on port 689. The AES keys are created by using the BOX RSA Keys and renewed every 60 seconds.

Only a small amount of synchronization traffic is necessary for synchronizing via LAN connection. Sync traffic is kept at a minimum by synchronizing only sessions and not each packet. Due to the characteristics of the TCP protocol (SYN, SYN-ACK, ...), only existing established TCP connections are synchronized. When the synchronization takes place during the TCP handshake, the handshake must be repeated.



The synchronizing procedure takes place immediately (if possible). If synchronization packets are lost, up to 70 sessions per second are synchronized.

Depending on the system availability, the behavior differs:

- If the partner unit is inactive/rebooted – Sometimes, the backup unit might not be available and, therefore, does not respond to the sync packets (for example, for maintenance reasons). In this case, the active unit stops synchronizing. As soon as the partner unit reappears, the active unit checks whether the other one was rebooted or has an obsolete session state and re-synchronizes all necessary sessions.
- If the active unit reboots without a takeover – The **Firmware Restart** button was clicked. The ACPF sessions and sockets are gone, but the unit is not rebooted physically. In this case, the partner unit recognizes that its session state is obsolete and removes all synchronized sessions.

Takeover Procedure

When the primary, active HA unit does not respond to the heartbeat (Control UDP 801), a takeover is initiated after a 10-15 second delay. This delay is necessary to account for potentially low network performance.

> Services are unavailable during the takeover procedure.

When the primary unit stays inactive, the synchronized sessions on the second unit are activated and all connections are available again. The backup unit does not have the current TCP sequence numbers. In case of a takeover, the sequence number is not checked for correctness. As soon as the connection has traffic, the sequence number is known to the former backup unit, and the sequence number check can be performed again. The missing sequence number on the backup unit also results from the fact that TCP connections that were taken over but have since had no traffic cannot be reset in a clean way. Terminating the session via the **Terminate Session** button removes the connection but does not send a TCP Reset (TCP-RST) signal.

If the connection between the HA partners gets interrupted, the primary and secondary systems activate their services at the same time. When the connection becomes active again the primary system immediately shuts down its services. This procedure ensures that only one HA partner is in operational mode while the other one is in standby mode.

Configuration

In each access rule, you can set the **Transparent Failover active/inactive** setting to define whether sessions matching this rule are synchronized. For more information, see Advanced Access Rule Settings.

Figures

1. ha_system_80.png
2. ha_setup_azure_80.png
3. ha_uplink03.png
4. synack.png