PAN-OS HA

Understanding PAN-OS HA states, timers and loops

Palo Alto Networks
232 E. Java Dr.
Sunnyvale, CA 94089
408.738.7700
www.paloaltonetworks.com

Table of Contents

Overview

Service providers and enterprises that deliver revenue-generating and business critical services over the Internet face a myriad of performance and security challenges. However critical those challenges may be, high availability remains the paramount concern. In order to properly perform access control functions, a network firewall must be placed at the single point through which all traffic must pass. Because all traffic must pass through the firewall, it is vital that the traffic flow remains uninterrupted, even in the event of a device or network failure. A well-designed security infrastructure needs to offer high availability tools to create a resilient, scalable, and easy to manage solution.

Palo Alto Networks offers a line of purpose-built security solutions that integrate firewall and VPN functions with a set of high availability (HA) tools to deliver resilient, high performance devices. This technical paper describes the main functionality of PAN-OS high availability

Hardware requirements

Two identical Palo Alto firewalls per cluster (PA 4060, PA 4050, PA 4020, PA 2050, PA 2020, PA 500

Software requirements

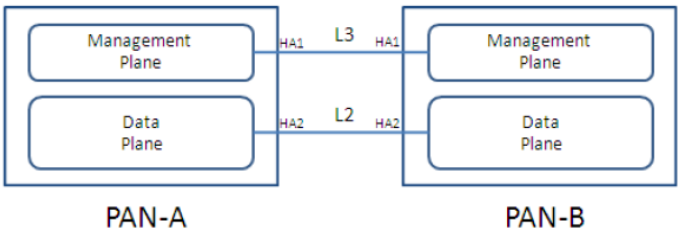
PAN-OS 3.1

Feature description

The fundamental architecture of Palo Alto Networks' next-generation firewalls is provide network security by enabling enterprises to see and control applications, users, and content – not just ports, IP addresses, and packets – using three unique identification technologies: App-ID, User-ID, and Content-ID. These features are complimented by robust hardware which is designed to provide separation of control plane and data plane This separation ensures highly available management, high speed logging and route updates, while the dataplane does all the security processing
PAN-OS HA cluster consists of two identical Palo Alto firewalls with identical software that enforce the same overall security policy and share the same configuration settings. The HA cluster is seen as a single device by both external devices

HA links

The devices in a HA cluster require two dedicated connections to exchange state information and data synchronization. The figure below shows the logical representation of two devices connected in a HA cluster



HA control link

The HA control link also known as the HA1 link is used by HA agent for the devices in HA to communicate with one another. The HA agent uses TCP port 28769 for clear text communication and uses SSH over TCP port 49969 if using encryption. This connection is used to send and receive hellos and HA state information, configuration sync and management plane sync, such as routing, user-id information. Configuration changes to either the active or passive unit are automatically synchronized to the other device over this link
The PA-4000 series of firewalls have dedicated HA1 link. All other platforms require a revenue port to be configured as a HA1 link. HA1 link is a layer3 link requiring an IP address.

HA data link

The HA data link also known as the HA2 link is used to synchronize state, sessions, routing tables, IPSec security associations and ARP tables between devices in a HA cluster. The HA 2 link is a layer 2 link. It uses ethernet as transport with ether type 0x7261. Data flow on the HA2 link is always unidirectional- always from the active device to the passive device
The PA-4000 series of firewalls have dedicated HA2 link. All other platforms require a revenue port to be configured as a HA2 link

HA states

The following are the possible HA system states:

Initial: The transient state of a device when it joins the HA cluster. The device will remain in this state for 60 seconds after bootup unless peer is discovered and negotiation begins. After 60 seconds, the device will become active if HA negotiation has not started.
Active: The state of the device in a HA cluster that all traffic received by the HA cluster.
Passive: The state of the device that monitors the status of the active unit and takes over should the active device fail. Traffic handling interfaces are maintained in a down state.
Suspended: This is an administratively disabled state. In this stage HA cluster member cannot participate in the HA election process and does not sync configurations and sessions
Non-functional: Error sate due to data plane crash or monitor failure

Upon initial configuration the device with the lowest priority, value close to zero, becomes the active unit (default priority is 100). If two devices have the same priority value, the device with the lowest MAC address becomes the active unit. You can determine whether a better priority number (closer to zero) can initiate a failover by setting the device that you want to be the active unit by configuring preemption on both the devices

Virtual MAC

In layer 3 mode a virtual MAC address is created for the all the configured interfaces. The format of the virtual MAC is 00-1B-17:00: xx: yy where
00-1B-17: vendor ID
00: fixed
xx: HA group ID
yy: interface ID

When a new active device takes over, Gratuitous ARPs are sent from each of the connected interfaces of the new active member to inform the connected Layer-2 switches the new location of the virtual MAC addresses

HA state transition

The firewall can transition HA state for a variety of reasons

- HA agent crashes
- Peer device crashes or power cycles
- Monitored link goes down
- Monitored IP address becomes unreachable

HA agent failure

The HA1 link is critical in having the HA cluster functioning. HA1 link can be either back to back connected or span a L2 switch. It is recommended to have the HA1 links back-to –back connected. HA agent uses this link to communicate with each other maintaining device state. The communication between the devices happens at specific interval referred to as the **Hello-interval.** This is the time in interval in milliseconds to send Hello messages to track the status of the HA agent. Hellos are sent over the HA1 link. The default value of the hello interval is platform dependent.
On the PA500 and PA 2000 series the value is 8000ms and the PA 4000 series use 1000ms.
In the event HA agent crashes, it takes 3 consecutive hellos messages to be missed to realize a failure. On the fourth missed hello HA agent tries to reconnect. Therefore it takes the PA 500 and 2000 series 32 seconds and the PA-4000 series takes 4 seconds to failover. The HA agent can crash if the kernel crashes.
On the PA 500 and 2000 series, the failover time can be decreased by using heart beats. Heartbeats use ICMP pings to monitor peer state. ICMP pings are handled by the kernel. Pings are sent are interval defined as the **heartbeat interval.** The heartbeat-interval is 1000ms for all platforms. It takes 3 successive pings to be missed for HA state transition. Peer device failure can be realized in 4 seconds using heartbeats

HA control link monitoring

HA control link monitoring tracks the state of the HA1 link to see if the peer HA device is down. This will catch a power-cycle, a reboot, or a power down of the peer device. To allow for ignoring flapping of the link that wouldn't necessarily take the HA control connection down, a **monitor hold down** timer for the HA control link monitoring can be configured. The monitor hold down time is configured under the HA1 link. The default value is 3000ms

Link and path monitoring

Link monitoring

The physical state of interfaces can be monitored to trigger a device failure. The interfaces to be monitored are grouped into a link group. A link group can contain one or more physical interfaces that are monitored. A device failure is triggered when any or all of the interfaces in the group fail. The default behavior is failure of any one link in the link group will cause the device to change the HA state to non-functional.

Path monitoring

Path monitoring allows you to monitor the full path through the network to mission-critical IP addresses. ICMP pings are used to verify reachability of the IP address. Pings are sent at an interval of 200ms. .An IP address is considered unreachable when 10 consecutive pings fail. A device failure is triggered when any or all of the IP addresses monitored become unreachable. The default behavior is any one the IP address becoming unreachable will cause the device to change the HA state to non-functional

HA state machine interaction with monitoring

- When a monitored link or path fails, the active device transitions to the non-functional state. The passive device then transitions state to active (this is assuming the passive device has not experienced a failure) In the event the current active device experiences a link failure, it will continue to operate as active device, since the HA peer is in non-functional state because of monitor failure.
- If both the active and passive devices experience link failure, the current active device will continue functioning as the active cluster member
- When a link or a path that is being monitored goes down, and if the peer device is not in passive state and therefore cannot take over as active, then the current active device continues functioning as active even though it is sees either a link or path failure
- If both the active and passive devices experience multiple failures, the device with the least number of failed links or paths will function as the active device.
- If the active device goes into suspended by user intervention, non-functional because of a dataplane down, or be rebooted, then the peer device which is non-functional because of a link or path monitoring failure will leave non-functional state and go into active state.

The link state will be monitored even when the device is in non-functional state. This will help device transition from non-functional to passive state when the link connectivity is restored.
The link group configuration is not synchronized to the peer device.

Preemption loop

When preemption is configured and the higher priority (value close to zero) device has a link or path monitoring configured it goes into non-functional state when there is a link or path failure. The device will remain in the non-functional state for monitor-fail-hold-time after which it will transition to passive state. Because of preemption, the device will wait for period equal to the sum of passive hold time and preemption hold time and take over as active device. At this point, if monitoring fails again, the device gets into a loop to repeat the active ->non-functional ->passive->active transitions. This state transitions are referred to as flaps. The device will remain in the suspended state even if the link or path connectivity is restored. The default number of flaps is 3. A value of “0” means infinite flaps. The maximum number of flaps defined will have to happen within 15 minutes after which the device enters suspended state. Once the device enters the suspended state, it requires user intervention to transition to functional state. This is accomplished by using the operational command
“request high-availability state functional “

Non-functional loop

A non-functional loop is when both devices in an HA pair have link or path monitoring failures that are not detectible while in non-functional state. This happens when the link state on passive device is set to shutdown in layer 3 mode. The link state on the passive device is always shutdown in vwire and layer2 deployments. If device in HA cluster starts in active state, detects a link or path down and it changes state to non-functional. The peer device at this time will go active. The non-functional device will remain in this state for monitor-fail-holddown time and change state to passive. The active device upon seeing the peer device as passive will change to non-functional because of the link failure. At this point, if monitoring fails again, the device gets into a loop to repeat the active ->non-functional ->passive->active transitions.
This state transitions are referred to as flaps. The device will remain in the suspended state even if the link or path connectivity is restored. The default number of flaps is 3. A value of "0" means infinite flaps. The maximum number of flaps defined will have to happen within 15 minutes after which the device enters suspended state. Once the device enters the suspended state, it requires user intervention to transition to functional state. This is accomplished by using the operational command **"request high-availability state functional "**

HA parameters

Hello-interval: It is the time interval in milliseconds to send Hello messages to track the status of the HA agent. Hellos are sent over the HA1 link. The default value of the hello interval is platform dependent.
PA500 and PA 2000 series uses 8000ms and PA 4000 series use 1000ms.
Heartbeat-interval: It is the time interval in milliseconds to send ICMP ping to the HA peer over the control link. The peer kernel directly responds to the pings. The default value of the heartbeat interval for all platforms is 1000ms
Monitor- hold-time: It is the hold time in milliseconds to allow HA1 link flapping, default 3000ms
Monitor-fail-hold-time: The time to wait in non-functional state due to a monitor failure, before changing state to passive to check the state of monitored object. Default value is 1 minute This timer is applicable for the following cases

- Layer 3 mode with link state on passive device set to shutdown
- HA deployment with layer2
- HA deployment with vwire

Passive-hold-time: It is the time in milliseconds the device waits to change state change from passive to active. Default value for all platforms is 2000ms
Preemption-hold-time: It is the time in minutes a high priority (value closes to zero) device remains in the passive state, the active device relinquishes the active state. Default value is 1 min. If the two devices are configured with different preemption hold-down timeouts, then the preemption hold-down timeout configured on the higher priority (lower value) device is used. The preemption hold-down timeout on the lower priority device is ignored
Flap max: Maximum number of HA state changes to non-funcational before entering suspended state

Configuring HA

PaloAlto networks firewall can configured in Active Passive HA in all of the operating modes – vwire, layer 2 and Layer3. For details on configuring HA refer to the document at https://live.paloaltonetworks.com/docs/DOC-1349

Summary

Palo Alto Networks next-generation firewalls are based on a unique Single Pass Parallel Processing (SP3) Architecture – which enables high-throughput, low-latency network security, even while incorporating unprecedented features and technology. Palo Alto Networks solves the performance problems that plague today's security infrastructure with the SP3 architecture, which combines two complementary components-Single Pass software, Parallel Processing hardware. The results is the perfect mix of raw throughput, transaction processing and network security that today's high performance networks require.