Astaro Deployment Guide
High Availability Options – Clustering and Hot Standby

Table of Contents

Introduction ................................................................. 2
Active/Passive HA (Hot Standby) ........................................ 2
Active/Active HA (Cluster) .................................................. 2
Astaro’s HA – Act as One .................................................. 2
Deploying a Hot Standby System ........................................ 2
Installation ....................................................................... 3
  Automatic Configuration ...................................................... 3
  Manual Configuration ....................................................... 4
Total Synchronization .......................................................... 4
Failover ........................................................................... 5
Update Process ................................................................. 5
Monitoring ......................................................................... 5
Reporting and Logging ........................................................ 6
Deploying a Cluster System ................................................ 6
Installation ....................................................................... 7
  Initial Setup ................................................................... 7
  Extending a Cluster ........................................................... 7
Integrated Load Balancing .................................................... 8
  Distributed Network Traffic ................................................. 8
HA Roles and Failover ....................................................... 8
Performance Improvements .................................................. 10
Example - Meshed Cluster Setup ......................................... 11
Conclusion ....................................................................... 12
Introduction

The main causes for an Internet security system to fail today are because of a hardware or software failure. To circumvent these cases and ensure your Internet connection stays online, Astaro offers two high-availability (HA) options:

Active/Passive HA (Hot Standby)

The ability of any system to continue providing services after a failure is called failover. In Active/Passive HA this is done by setting up a standby system (slave) which becomes active in case the primary system (master) fails. Active/Passive HA is possible for all Astaro Appliances.

Active/Active HA (Cluster)

You can also use Astaro Gateways to set up an Active/Active HA (also called cluster), which operates by distributing dedicated network traffic to a collection of devices - similar to conventional load-balancing approaches - in order to get optimal resource utilization and decrease computing time. In an Active/Active HA, you are protected against hardware failures on one node by the remaining nodes who automatically take over the workload and/or roles of the failing node. Active/Active HA is possible with up to 10 nodes for all Astaro Appliances.

Astaro's HA – Act as One

Comparing Astaro’s HA architecture to other high availability technologies like external load balancers or IP based NLB systems, many advantages become visible. While the following chapters will detail on how easy it is to setup an Astaro HA system with automatic configuration, it’s worth noting that once up and running, the HA environment acts as only one system.

This is usually also the case for other clustered environments where cluster nodes are efficiently hidden and fully transparent for the client side. However, Astaro takes this concept a step further in representing the cluster to the administrator as only one system, too. This approach saves administrators from maintaining separate machines: Administrator only need to login to the master node - participating nodes are fully manageable from there.

While the whole network traffic is directed through the dedicated master node, totally synchronizing the data between all HA nodes (configuration, activity states, log and monitoring data) makes failover delays insignificantly short. For example, the fact that MAC addresses are shared between the HA nodes, completely masks a failover for client PCs.

Usually HA technologies that are fully integrated into the gateway are coming along with trade-offs in functionality. Astaro however, has build a patent-pending HA technology where limitations don’t exist and all features of an Astaro Gateway are fully supported.
Deploying a Hot Standby System

Installation

The possibility to use a hot standby system for redundancy is the simplest way to protect network environments against hardware failures of a device. This concept usually is used where additional performance is not necessarily required but high availability must be guaranteed.

Hot Standby System

To setup a hot standby system to ensure high availability for an Astaro installation, you have the option to either use automatic or manual configuration.

Automatic Configuration

The automatic configuration feature offers a straightforward way to easily build an Active/Passive HA. Astaro Hardware Appliances always come with automatic HA configuration preconfigured and a dedicated HA interface (eth3) for communication between the two devices. Here are the three steps of the process:

To start the automatic setup you only need to connect two Astaro Gateways of the same type via the HA interface (eth3). Once a link on this interface is detected the gateways will automatically start to configure an HA hot standby environment.

The devices will negotiate the master and slave nodes (typically the system with the longer uptime becomes master), build the configuration and from then on act as one single hot standby system.

Without any required user intervention during this process, the HA hot standby system will go to an active state.
Manual Configuration

The manual configuration of an Active/Passive HA is very simple, too. However, it gives you some more options like which LAN port to use for synchronization, device names, node IDs and encryption key.

Once this data is entered on the first node you only need to connect the slave node to the master node and either take the same configuration steps as for the first node or use the automatic configuration feature to add the second node automatically.

Total Synchronization

Astaro Gateways working in either high availability mode, continuously exchange data over the HA interface to stay totally synchronized. Synchronization only is effected between the master and slave node. Worker nodes are not part of the synchronization process. The synchronization is key when a hardware failure occurs and the surviving node takes over at the exact point where the failing devices quit.

In this way for example, an IPsec VPN tunnel will keep working during a hardware failure even without the need to reconnect.

The following list includes the data that is synchronized between the master and the slave node of a HA system:

- Complete ASG Configuration
- E-mail queue (e-mails older than five minutes)
- E-mail quarantine
- Reporting data
- Update packages
- System time
- Firewall state
- IPsec state
- DHCP Leases
Failover

All nodes in a HA system monitor each other by means of a heartbeat signal, a periodically sent multicast UDP packet used to check if the other node is still alive. If the master fails to send this packet due to a technical error, the node will be declared dead and the slave node will immediately take over operation.

The time for a takeover is so short, that at most you will lose for example one ping packet – a delay that usually remains unnoticed by any client or application.

Update Process

Each time new updates are available a special process will be started to ensure uninterrupted operation throughout the whole update process and safe transition to new versions of firmware and patches on all nodes of the system.

Whenever an update is started, first the slave node and half of the worker nodes (in clusters) are updated. Only if this initial update was successful, a failover will be initiated and the former master node together with eventually remaining worker nodes will be updated. The two-step approach is necessary to maintain a minimum downtime and a maximum working performance (for example in clustered environments). Also, a complete switch to a new version is required for consistency reasons so that all active systems in an HA scenario are guaranteed to run the same software version.
Monitoring

As soon as you have setup your HA system it will be visible from outside the HA compound only as one device. However, to keep track of the HA status you will find dedicated monitoring options in WebAdmin – as well as in Astaro Command Center if you choose to manage your Astaro Gateways with this management platform.

WebAdmin shows basic HA status information within the dashboard as well as detailed information on the high availability tabs. At any time you will find information about:

- Role of each HA node
- Status of each HA node
- Last status change of each HA node
- Real-time hardware resource usage for CPU, RAM and disk space

Resource Monitoring in a Cluster

Reporting and Logging

All reporting data is consolidated on the master node and is synchronized to the other cluster nodes at intervals of five minutes. In the case of a failover, you will therefore lose not more than five minutes of reporting data.

However, there is a distinction in the data collection process. The graphs displayed in the hardware reporting tabs always represent the data of the node currently being master whereas accounting information on network, web and mail usage represents data that was collected by all nodes involved.

Logging data is also synchronized to the master node (and replicated to all nodes in intervals) via Syslog to ensure that all relevant data is available at any time you login to the Web Admin.
Deploying a Cluster System

Installation

The cluster functionality of Astaro Gateways adds scalability to the high availability feature. Hardware resources on additional HA systems in this setup can effectively be used to enhance performance and capacities of your Astaro environment. How the load is shared and what the performance improvements are will be discussed in the chapter “Performance Improvements” below.

Initial Setup

To build a cluster of two or more Astaro devices you need to first identify and manually configure the first system in the Active/Active HA environment. Besides choosing “Cluster” as an operation mode, the configuration is very similar to the manual configuration of an Active/Passive HA system: You need to provide the same configuration information and also have the same options to add additional cluster nodes manually or automatically.

Once two or more systems become active in a cluster, they automatically share and balance the workload between each other and provide fault tolerance to ensure high availability.
Extending a Cluster

After some time operating the cluster, performance of the system may decrease. This happens usually either by growing amounts of users the system must support, or use of new features, for example by adding optional subscriptions. Astaro clusters are highly scalable and adding additional nodes to a cluster is done in minutes: In Web Admin you only need to set the high availability status to “Automatic configuration” and select the appropriate synchronization interface.

After applying these changes, the new node will automatically be added to the cluster as an additional worker node.

Integrated Load Balancing

In a clustered Astaro HA environment the load is evenly distributed between the participating cluster nodes. Astaro’s patent pending load balancing technology automatically makes use of all hardware resources in a cluster, whereas the master node takes the responsibility to distribute the load between the slave and worker nodes. Even new worker nodes are immediately assimilated and the master node seamlessly adds the additional resources to the cluster. Additionally, it is worth noting that when building a cluster with more than three nodes, the master is dedicated to act solely as the central load balancing and synchronization system and does not process any traffic data itself anymore.

Distributed Network Traffic

The following list includes the types of network traffic that is distributed to the cluster nodes:

- IPSec
- IPS
- FTP
- HTTP
- POP3
- SMTP
HA Roles and Failover

Each node within the cluster can assume one of the following roles:

- **Master**: The primary system in a standby/cluster setup and responsible for synchronizing and distribution of data within the HA system.
- **Slave**: The standby system in a standby/cluster setup which takes over operations if the master fails.
- **Worker**: A simple node in a cluster setup, responsible for data processing only.

All nodes monitor each other by means of a heartbeat signal, a periodically sent multicast UDP packet used to check if the other nodes are still alive. If any node fails to send this packet due to a technical error, the node will be declared dead. Depending on the role the failed node had assumed, the configuration of the setup changes as follows:

![Failover in Astaro HA systems](image-url)
Performance Improvements

One of the main benefits of clustering Astaro gateways are the scalability and performance improvements you can achieve. By simply adding additional nodes to an existing cluster, adding hardware resources to an existing environment is no longer a major configuration effort. In a cluster you can add up to 10 nodes on the fly, nearly linearly increasing the performance of the overall system.

The matrix below shows how additional nodes increase overall performance in a cluster. Note, that between three and four nodes the performance increase is lower due to the fact that with adding the 4th node, the master node will be dedicated to handle load balancing and synchronization tasks, leaving the workload to the slave and worker nodes. Also, in sizing your clustered environment you have to observe a 1 Gbps limit of the HA interface for the sum of all balanced network traffic (IPSec, IPS, FTP, HTTP, POP3 und SMTP).
Example - Meshed Cluster Setup

With Astaro Security Gateways you can build high availability internet access solutions in a meshed cluster setup. Redundancy here is not only given within the cluster but can be extended to the WAN and LAN side of your network without any additional special devices such as external load-balancers or special switches.

A meshed cluster setup as shown in the diagram above is easily configured:

1. Add the ASG units to your local network. To achieve redundancy, connect the units to your local network by connecting two ports to separate LAN switches.
2. Prepare the cluster by connecting the eth3 interfaces of all ASG units forming the cluster to a network switch.
3. Connect the ASG units to at least two uplink switches or routers and configure Astaro uplink load-balancing on the designated master node to create a redundant WAN connection. With the uplink load-balancing feature you also gain automatic network traffic load-balancing by using both WAN connections.
4. To setup a redundant LAN connection, on the designated master node, create a link aggregation group. Link aggregation, which is also known as "port trunking" or "NIC bonding", allows you to merge multiple Ethernet network ports into one virtual interface. The merged ports appear as a single IP address to your system and provide basic failover and fault tolerance by redundancy in the event of a failing switch. This failover is completely transparent to the system using the connection.
5. Configure the cluster.

Configure the cluster as described in the chapter "Cluster Setup" above.
This meshed setup protects against many different problems that might occur in your environment, like:

- General WAN problems at your provider side (unless you use the same one for both connections).
- Cabling and hardware issues on your WAN side (i.e. network routers).
- Hardware issues on one of the Astaro cluster nodes.
- Cabling and hardware issues on your LAN side (i.e. network switches).

Conclusion

With the two different high availability options, Astaro provides redundancy and adds scalability possibilities to your network security environment. Without the need for additional hardware investments like external load-balancers or high availability switches these HA solutions are very affordable.

Whereas most of the scenarios described in this guide where based on Astaro hardware appliances all HA options are also available with the Astaro software and virtual appliances offering you even more flexibility to design a secure and highly available network infrastructure.