High Availability Designs

Introduction

SecureAuth is a browser-based Identity Provider (IdP) that provides multi-factor authentication, federation and Single Sign-on (SSO). As such, it is deployed with the same design as other highly-available web services.

There are two basic models for deploying a highly available web service: a **centralized cluster** and a **distributed set of nodes**.

The central cluster approach load-balances a number of servers at a single site such as a central data-center. Typically, this central site will be the primary data center for an enterprise, and will be built with redundancy in place for Internet access, power and other infrastructure needs, as well as having multiple servers in place. Tools such as load-balancers, hardware or software, prevent the outage of a web-site when one of the servers fail. An alternative method is to use virtualization tools, such as those offered by VMWare, to provide for a highly-available system.

A distributed high-availability model makes use of multiple site locations such as several data-centers. Systems, such as a Global Server Load-Balancer (GSLB) a DNS entry with multiple A-records, can be used to return the IP address for the multiple locations to the client/browser. Third party GSLB systems can provide additional functionality when returning the address for the service, such as checking the health and availability of a server before returning the address to a client or checking the geographic location of a client and returning the address of a 'closer' server 'higher' in the list.

Central cluster

The centralized-cluster design creates a single highly-available deployment of SecureAuth IdP servers.

SecureAuth IdP can be deployed at a central site to provide high-availability (HA) and for load-balancing (LB).

This design centralizes SecureAuth IdP services at a single location designed for High-Availability (HA) and with performance in mind. All services and remote locations would utilize the central SecureAuth IdP deployment for identity services, Multi-Factor Authentication, Single Sign-on (SSO) and other functions provided as SecureAuth IdP features.

Deploying multiple servers behind a load-balancing appliance, or utilizing a software load-balancing tool, prevents the outage of a server from affecting the availability of the service (i.e., SecureAuth IdP authentication, federation and SSO). Multiple servers deployed for load-balancing increase the maximum load of the overall system by distributing sessions across multiple servers. In this way, a load-balanced system can support the number of users a single system can service multiplied by the number of servers deployed in the load-balanced pool of servers. That is, if a single server could handle 500 authentications per minute, two servers behind a load-balancer could scale to 1,000 per minute. However, if an enterprise were to anticipate a peak volume of 1,000 authentications per minute, and desired to provide redundancy in the case of a node failure, three servers would be deployed so that in the event of a single system failure the anticipated peak load would still be handled.

The image below depicts a centralized and load-balanced deployment of SecureAuth IdP:
Clients coming across the Internet would all access SecureAuth IdP through the load-balancer, as could VPNs. At an enterprise location that is connected to the datacenter via a Wide Area Network (WAN), such as MPLS, the clients would authenticate across the WAN and through the load-balancer. In distributed VPNs, where a VPN exists at satellite locations it is more common that the SecureAuth IdP authentication or enrollment would occur across the WAN rather than across the Internet, as the WAN is considered an internal network. In the image above red lines depict Internet traffic and blue lines depict WAN traffic.

Distributed / multi-site deployment

SecureAuth IdP can be deployed at multiple sites to provide high-availability (HA), for performance considerations or for load-balancing (LB) geographically.

Various techniques, such as Global Server Load Balancing (GSLB) or returning multiple DNS A-records, can be used to provide a form of load balancing and high-availability. Disaster Recovery (DR) datacenters operating as a secondary site can have fail-over provided using BGP or other techniques.

In a multi-site design, SecureAuth IdP can be deployed as a single server or in a group, with HA/LB provided by an external system or via software (software-based LB is only recommended for customers with experience using the feature due to specific network configuration requirements).
A variation of the multi-site deployment model is multiple locations with a service, such as a VPN, at each location. Often each location is accessed by a unique URL. An example is an enterprise with a Los Angeles datacenter and a New York datacenter, with SSL VPN access at each site. The Los Angeles VPN may have a URL such as LA-VPN.domain-name.com, and the New York location could have the URL NY-VPN.domain-name.com. In this way the end-user makes the selection for the VPN to access. Their choice may be based on their location, which would improve the performance of the VPN by reducing the distance across the Internet between their client and the gateway. Another option would be to choose the VPN that is closer to the resource that the end-user would like to access; if the application or file a user wants to access is in L.A. they may achieve better performance across the Internet to the L.A. VPN, than across the enterprise's Wide Area Network (WAN) between the N.Y. and L.A. locations. Either model can be used by an enterprise based on the design and infrastructure of their network.

The image below depicts two sites with SecureAuth IdP in a multi-site deployment:

Clients resolve the DNS addresses for the URL for SecureAuth IdP through the DNS settings configured in their local system, typically their Internet Service Provider's (ISP) DNS servers. The ISP's DNS server will look up the authority for the domain from the root DNS servers. The authority for the DNS domain could be a GSLB system, or it could be a DNS server configured with multiple A-records (IP addresses) for the domain name in the URL. The client would then access SecureAuth IdP using the IP address(es) provided during DNS name resolution, and each SecureAuth IdP deployment would be equally able to service the request.

The multi-site deployment with unique names (URLs) for each site works similarly. The client attempts to access the location, such as Site1.domain-name.com, and DNS resolves to the site-1 IP address. Clients accessing Site2 would resolve the IP address for site-2. Again, SecureAuth IdP, in each location, would be equally able to service the client's identity request.