Citrix Federated Authentication Service Scalability

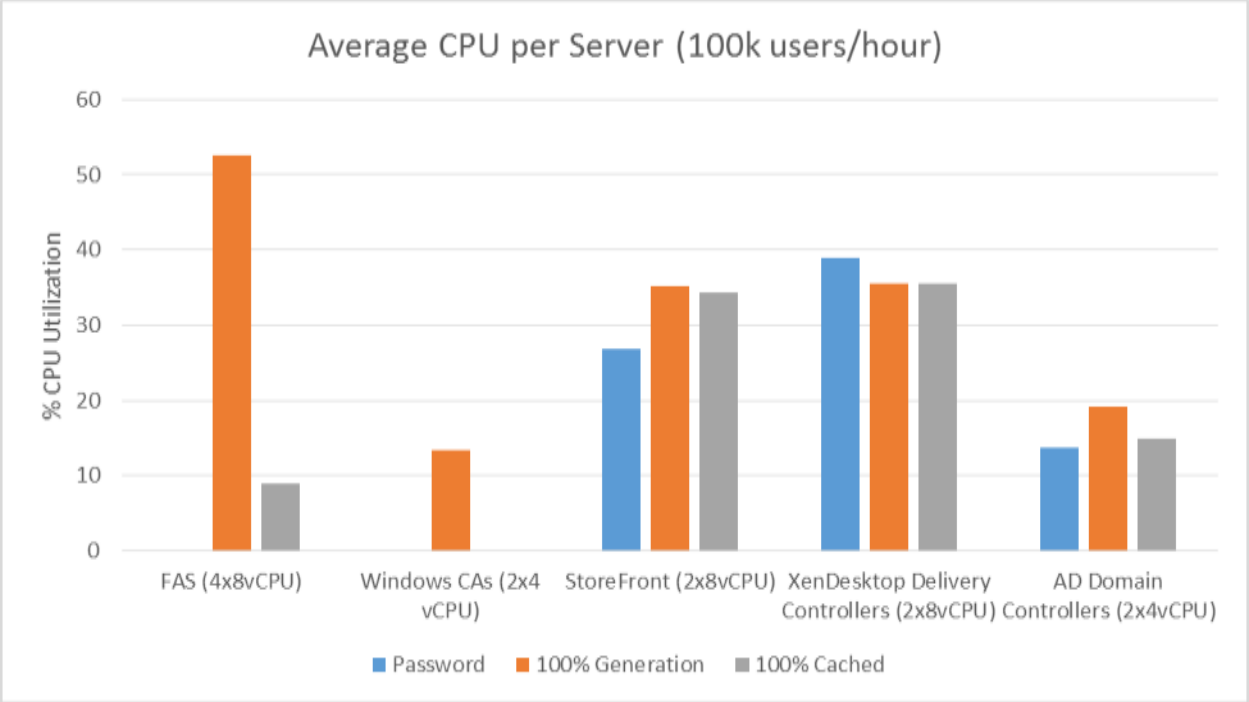
Introduction

The Citrix Federated Authentication Service (FAS) generates digital certificates signed by a Microsoft Windows Certificate Authority server to be used for secure logons. StoreFront then uses the certificates to authenticate users to a Virtual Delivery Agent (VDA) instead of using a password for authentication.

This article discusses how the Federated Authentication Service can be deployed to support large environments. We feature a 100,000-user setup.

Summary & Primary Recommendations

The following graph summarizes the CPU usage required in our test environment to support 100,000 users logging on within one hour:



The quantity and CPU configurations of the non-FAS virtual machines are provided for reference and may not match what is required for your environment.

- Four 8-vCPU Federated Authentication Servers were used at ~50% capacity. This provided N+2 redundancy to maintain close to the original performance level in case two of the four servers failed or became unavailable.

- At least N+1 FAS servers are recommended if the Federated Authentication Service is in use andrequired forlogons. The extra servers require excess CPU capacity to generate certificates in place of the servers that failed.
- FAS certificates are not shared among FAS servers. If a server fails, some organizations may accept degraded performance, in which case higher average resource usage may be acceptable.
- The default certificate lifetime is one week (seven days), although renewals will occur after the halfway point (3.5 days).
- Citrix StoreFront servers process an extra load if FAS is enabled.
- The Windows Certificate Authority only sees a significant load when certificate generation is required for users. You can reduce the load on the servers by generating and replacing certificates at intervals.
- Active Directory Domain Controllers see an extra load because of FAS certificate generation, but this load is comparable to other bulk Windows certificate generation operations.
- Be sure to monitor the Windows event log on the FAS server for logon audit information.
- Windows Certificate Authorities archive all FAS-issued certificates by default. Be sure to follow certificate management practices for your organization, and monitor the available disk space on these servers.

Scaling Citrix Services for Large Environments

FAS Servers

A single FAS server has been tested to support hosting 100,000 users. Depending on connection rates, a single FAS server may become a significant bottleneck.

To support 100,000 users logging on within an hour, we recommend four virtual servers for FAS. The following VM and Hypervisor host specification was used:

FAS Server Virtual Machine

- CPU: 8vCPU
- Memory: 8 GB
- Operating System: Windows 2012 R2
- Physical Hardware: HP Gen8 blade with 2 Intel® Xeon® E5-2640 processors @ 2.5 GHz
- Dedicated CPU cores allocated to the FAS server

A full description of the test setup can be found in Appendix A.

This overprovisioning of resources allows 100,000 users to log in within an hour, even if new user certificates have to be generated for every user. There is also spare generation and user-handling capacity in case two of the FAS servers become unavailable.

If no Federated Authentication Service servers are available, users cannot log on to StoreFront while FAS is enabled. It is therefore critical to ensure that at least one FAS server is always available.

Smaller environments can scale down these recommendations linearly (four 4vCPU servers or two 8vCPU servers for 50,000 users/hour, and so on), but should be configured in at least an N+1 configuration for high availability. When sizing the environment, consider the peak logon rate, including all disaster recovery scenarios.

StoreFront

For StoreFront, the expected CPU load using FAS is 30% higher relative to the same StoreFront environment not using the FAS service.

Delivery Controllers

Delivery Controllers pass FAS-related data through but do not interact directly with it. As such, the use of the Federated Authentication Service does not impact the scalability of Delivery Controllers.

Impacts on Other Services

Active Directory Domain Controllers

Using FAS slightly increases the load XenDesktop places on Windows Active Directory Domain controllers when FAS certificate generation has to occur relative to password-only logons. Preliminary testing suggests that this impact is similar to the load required to do bulk generation of other certificates in an AD environment.

This factor relates purely to the load placed on the AD controllers by the XenDesktop environment to the exclusion of all other Active Directory usage. If any of the XenDesktop users utilizing FAS can employ cached certificates without renewing them, the extra load placed on AD to generate certificates is proportionally decreased.

Microsoft Windows Event Logging

FAS logs a significant number of events. These events help track certificate generation and usage. As a result, the Windows event logs on FAS servers can be verbose.

Ensure that these event logs do not overflow before logon information is lost. In addition, you may want to have a centralized event log management system track FAS logon activity.

Microsoft Windows Certificate Authority

With two 4 vCPU Windows Certificate Authorities (CA), each handling two FAS servers, each Certificate authority saw approximately a 13% CPU load in our 100k user/1-hour certificate generation test.

This load is significantly lower than the CPU load placed on the FAS servers themselves; but may still be noticeable at peak hours.

We recommend that a Windows CA handling 100,000 users during an hour have at least four CPU cores available to process signing requests. Having four cores allows the CA to handle signing certs for all these users within an hour if necessary.

Having fewer certificate authorities than Federated Authentication Service instances is acceptable. However at least one certificate authority must always be available.

Windows Certificate Authority Archiving

Windows Certificate Authorities (CA) archive every issued certificate by default. This archive policy means the Windows CAs retain copies of all current and expired FAS certificates until the certificates are manually deleted. This archiving policy can be turned off, but the appropriate departments and authorities within your organization must approve.

Testing with a Windows 2012R2 Certificate Authority service suggests that for every 100,000 certificates (2048-bit) to be stored, up to 3.0 GB of disk space is required to support the CA transaction log and EDB database file.

Take care to ensure that all Windows CAs used by FAS are regularly backed up and do not run out of disk space. In addition to past certificates, Windows CAs keep an activity transaction log which may be truncated by performing a backup. Be sure to follow all the appropriate certificate management policies required by your organization.

More information can be found at https://technet.microsoft.com/en-us/library/cc731183(v=ws.11).aspx

Failover/Redundancy

Take care to ensure that at least one FAS server is available at all times. If no FAS server is reachable by a FAS-enabled StoreFront server, then users cannot log on or start applications.

Place a FAS server in maintenance mode using the *Set-FasServer* PowerShell command to indicate to StoreFront that another FAS server should be used for new logons. While in maintenance mode, the FAS server continues to support in-session certificates for currently logged-on users, but does not issue certificates or provide cached certificates for new logons.

At least one Windows certificate authority must be reachable to each FAS server at all times for the FAS server to issue new certificates. If no Windows CA is available, existing FAS certificates can continue to be used, but new users or users with expiring certificates cannot log on.

A FAS server can be configured to use more than one Certificate Authority for a certificate definition using PowerShell as described in the product documentation. It then fails over in case the first cannot be reached.

Certificate Expiration

Introduction

The "Validity period" set for the Citrix_SmartcardLogon certificate template (or its chosen replacement) controls how long user certificates are considered valid. When a user who has a previously generated certificate logs on after their certificate's half-life has been reached, the FAS server requests a new certificate for that user.

This early renewal helps ensure that the certificate does not expire while a user is logged on if in-session certificate use is enabled for that user.

Only the FAS server handling a user's current logon request issues an initial or updated certificate. If the user has a certificate on a failover or other alternate server, that certificate is not renewed and may expire if left untouched.

Adjusting the Validity Period

If CPU availability is at a premium and your organization's security policies permit, the validity period can be extended. To extend the validity period, edit the appropriate certificate template in the Windows Certificate Templates MMC snap-in. The extended validity period allows cached certificates to be used for a longer period.

The default validity period is one week (seven days), with certificate replacement the first time the user logs on after the halfway point (3.5 days).

Alternatively, more CPU may be required if the validity period must be shortened. The validity period must be at least 30 minutes to account for clock skew and the Windows CA issuing a certificate valid slightly in the past. Do not use such a short time setting if in-session certificates are enabled.

Certificate Pre-generation

Determining if Pre-generation is Necessary

Citrix believes that most customers using software-only encryption do ***not*** require certificate pre-generation. It takes less than one second for the Federated Authentication Service to generate a certificate for a single user on a modern Intel® processor. Certificate generation for multiple simultaneous logons can be parallelized across multiple CPU cores.

In "Scaling the FAS Architecture" in this document, the benchmarks mentioned used the default software cryptographic provider for FAS and did not use any pre-generation technique.

Customers using older hardware, a Trusted Platform Module (TPM), or certain Hardware Security Modules (HSM) may encounter longer delays, and might want to generate certificates in advance. Some customers may want to use pre-generation to stagger the generation and expiration dates and times of certificates to ensure fail over availability and a more balanced renewal rate.

The product documentation includes instructions to pre-generate user certificates.

Appendix A: Test Configuration

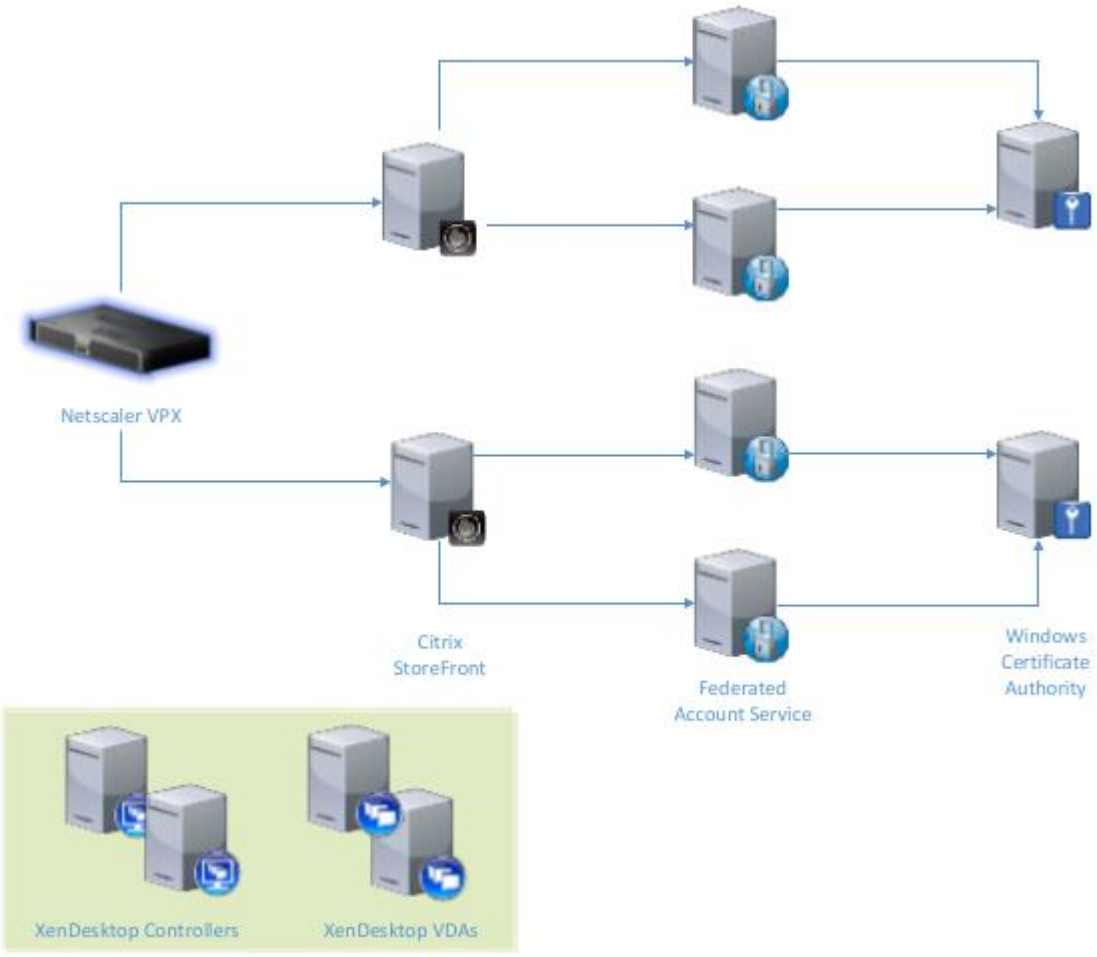
NOTE: This appendix describes the test setup used to perform logon scalability testing with the Federated Authentication Service. Depending on your environment and needs, a different setup may be required.

NOTE: Although test was performed using XenDesktop 7.9 and Microsoft Windows 2012 R2 with the hardware described below, measurements and outcomes are likely to apply to later versions of similar simulated environments.

All servers except the physical hardware, SQL Server, and Storage Controller were hosted on virtual machines. All Windows-based virtual machines and the SQL server ran Microsoft Windows 2012 R2.

Virtual machines were allocated across physical servers so every virtual CPU core was backed by a physical CPU core and there was no CPU contention.

Hyperthreading was enabled but not required to avoid CPU contention.



Authentication workflow used for 100,000-user testing

- 24 HP Generation 8 blade servers acting as hypervisor hosts, each containing:
  - 128 GB RAM
  - 2 Intel® Xeon® E5-2640 processors @ 2.5 GHz
  - 2 500 GB internal hard drives in a RAID 1 array
- 1 Microsoft SQL Server 2014 server
  - 64 GB RAM
  - 2 Intel® Xeon® E5-2640 processors @ 2.0 GHz
  - 2 TB local hard drive array
- Windows 2012 R2 Active Directory environment
  - 2 Domain Controllers (8 GB RAM, 4 vCPU, 40 GB disk)
  - 2 Certification Authorities (8 GB RAM, 4 vCPU, 80 GB disk)
- 1 NetScaler VPX 11.0 65.72.nc (2 GB RAM, 2 vCPU) configured as an HTTP load balancer
- XenDesktop 7.9 setup including:
  - 2 StoreFront Servers (8 GB RAM, 8 vCPU, 40 GB disk)
  - 2 Delivery Controllers (8 GB RAM, 8 vCPU, 40 GB disk)
  - 4 Federated Authentication Service Servers (8 GB RAM, 8 vCPU, 40 GB disk)
- 10 GB Ethernet Network
- NetApp FAS3270 Storage Controller
  - 2 TB NFS storage volume shared to all hypervisors
- Citrix proprietary user logon simulation framework