Performance Sizing Guide for Client Site Proxy for Standalone Server

Symantec.cloud

Performance Sizing Guide for Client Site Proxy for Standalone Server

Documentation version: 1.0

Legal Notice

Symantec Corporation
350 Ellis Street
Mountain View, CA 94043

http://www.symantec.com

Contacting Technical Support

The Global Client Support Center (GCSC) seeks to provide a consistently high level of service. The team consists of technically-trained client-focused individuals. They respond to your issue with the aim of resolving it within the first contact.

To reduce the time it takes to resolve an issue, before you contact the team refer to the Help on rasing support tickets. The Help explains what information is required for the various types of support issue.

We welcome comments and questions about our services.

Contact GCSC using the following contact details:

Email us at:	support.cloud@symantec.com
Call us on:	EMEA: +44 (0) 870 850 3014 or +44 (0)1452 627766
	US: +1 (866) 807 6047
	Asia Pacific: +852 6902 1130
	Australia: 1 800 088 099
	New Zealand: 0800 449 233
	Hong Kong: 800 901 220
	Singapore: 800 120 4415
	Malaysia: 1 800 807 872
	South Korea: 00798 14 800 6906
Open a support ticket	Log into ClientNet and navigate to **Support > Ticketing**
Visit the Web site	www.symanteccloud.com
Visit the Online Help	Online Help

We recommend that you check ClientNet frequently for maintenance information and to learn what's new. You can also add your mobile number in the **Administration > SMS Alerts** section of ClientNet to receive critical service-related issues by text message.

Contact and escalation details are available in the following PDF: Contact and Escalations document.

Contents

Chapter 1

About the Tests and the Test Environment

This chapter includes the following topics:

About this guide and the CSP for Standalone Server

This aim of this guide is to help you to determine the number of Client Site Proxy (CSP) Standalone servers to deploy in your network environment to support your user base and traffic profile.

The CSP for Standalone Server is the component of the Web Security service that you install on-site on a Microsoft Windows server.

The CSP captures information that is specific to the user's computer that is making requests to the Internet. To achieve this, the CSP authenticates the user making the Web request against the domain controller, captures and encrypts details of the domain name, user name, and local IP address and adds them to the HTTP request as custom HTTP headers. This is information is utilized together with information on the user that is held by the service to apply policy specifically to the user as defined in ClientNet.

For further information about the configuration and deployment of the CSP, see the *Client Site Proxy Administrator Guide*.

About the CSP test environment

We conducted the tests in a controlled environment that simulated high traffic load to test the performance limits of the CSP.

We tested versions 1.0.18 and 1.0.20 of the CSP under two simulated user traffic loads to determine the maximum number of users that could be supported before performance began to degrade.

Note: The performance and capacity of CSP for Standalone Server versions 1.0.18 and 1.0.19 are equvialent.

We used the following system resource configuration to test both versions of the CSP:

Table 1-1 System resource configuration for CSP tests

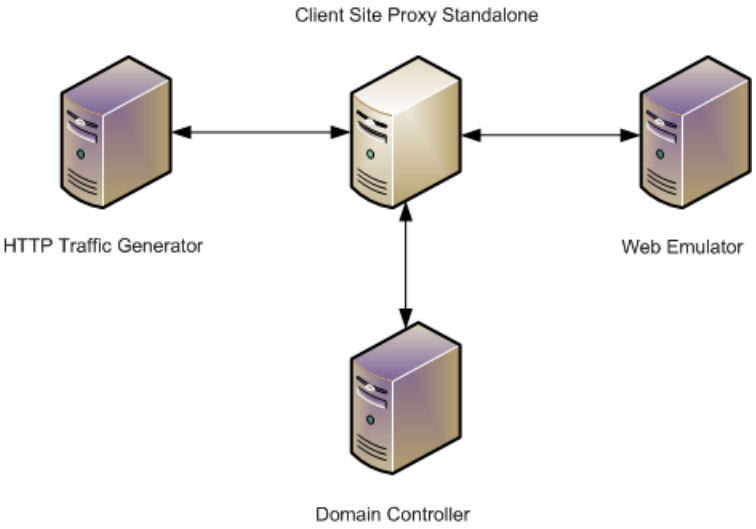
Component	CSP for Standalone Server hardware configuration and version	Domain controller server hardware configuration and version
Processor	1 Dual core 3.0 GHz CPU	2 Quad Core Intel Xeon 2.33 GHz CPUs
Memory	2 GB RAM	4 GB RAM
Disk space	140 GB Ultra fast SCSI	140 GB Ultra fast SCSI
Operating system	Windows 2008 SP2 (32-bit)	Windows 2008 SP2 (32-bit)
NIC	Copper 1 GB/100 MB Ethernet NIC	Copper 1 GB/100 MB Ethernet NIC

Testing the CSP on a virtualized environment was beyond the scope of this guide. You should expect some performance degradation if you choose to deploy the CSP on a virtualized environment when compared to a dedicated system with similar system resources.

All systems were set up on Windows 2008; we do not expect the results to vary significantly for other versions of the Windows operating system.

Figure 1-1 shows the test environment. The HTTP traffic generator acts as a client simulator, and requests Web traffic from the Web emulator. The Web emulator acts as a Web server simulator, and serves Web content that is requested by the traffic generator.

Figure 1-1 CSP for Standalone Server test envrionment



We did not include any Web Security infrastructure in the tests as the sizing guidelines are specific to the CSP Standalone component. Our tests were developed to determine the incremental latency that is introduced by the CSP Standalone application, its capacity, and any performance bottlenecks, in an ideal network environment.

It is possible that CSP Standalone application, along with local network limitations and the destination (origin) server, may impact the user throughput and latency levels seen in a real-world environment.

The Web traffic load generation testing tool used a member-space of 10,000 domain users.

About the CSP tests

We analyzed data from our global infrastructure to create realistic per-user traffic profiles.

A traffic profile represents the Web usage profile of users, and includes metrics such as connections per user, average size of requests and number of Web requests per second per connection. We used these profiles to test the limits of the CSP for Standalone Server.

Our tests were designed to determine the achievable throughput of the CSP when using the recommended system resource configuration, and to estimate how many servers may be needed for a given user base.

We measured the incremental latency that the CSP Standalone added when performing proxy authentication against the domain controller. The tests were performed under varying load conditions with different request rates and size of Web content until either the incremental latency increased significantly or the CSP Standalone generated errors indicating that it could no longer successfully process the web requests at the current traffic volume, that is, at the CSP Standalone capacity limit.

To simulate Web traffic, we used Web Polygraph (http://www.web-polygraph.org) Web traffic load generator to send HTTP GET requests through a CSP to a Web emulator that served the Web contents for the requested domains.

HTTP requests were generated at increasing rates until the maximum supportable throughput levels were achieved.

To simulate a real-world deployment, the traffic was generated with requests directed to different Web site addresses served by the Web emulator. Each connection through the CSP Standalone was individually authenticated. Authentication was performed again after every 30 requests per connection. This is a setting within the test harness used to exercise the CSP Standalone's IP authentication credential cache. The related CSP parameter is *authenticate_ip_shortcircuit_ttl*.

For all test scenarios, measurements were taken for throughput levels (in both megabits per second and connections/requests per second) and average latency for Web transactions. We also monitored memory and CPU utilization levels. The measurement of the Web request transaction latency was the primary criteria that we used to determine when the CSP Standalone had reached its supported capacity limit.

We used Windows NTLM authentication protocol. This is used by the CSP to authenticate users making Web requests against the domain controller.

We did not include tests with authenticated HTTPS traffic. The impact to the performance and overall capacity of the CSP due to HTTPS requests versus HTTP requests is expected to be minimal and therefore not material to our recommendations.

All web requests were HTTP GET requests generating a Web page response/download.

Chapter 2

Test Results and Advice on Using this Guide

This chapter includes the following topics:

Results of the CSP tests

Table 2-1 and Table 2-2 show the results for CSP versions 1.0.18 and 1.0.20. These are the figures for network throughput, web request processing rates, and average transaction processing time that you can expect from a single CSP for Standalone Server.

Table 2-1 Test results for CSP for Standalone Server version 1.0.18

	Typical usage profile	High usage profile
Size of Web content (KB)	10	20
Number of requests per second per connection	0.07	0.13
Maximum number of concurrent Web connections per second	1,540	1,530
Throughput (Mbps)	9	33
Incremental processing delay (ms)	7	9
Connections per user ratio	3	3

Table 2-1 Test results for CSP for Standalone Server version 1.0.18 *(continued)*

	Typical usage profile	High usage profile
Maximum number of concurrent users	513	510
Provisioned:concurrent user ratio	2.5	2.5
Maximum number of provisioned users	1,282	1,275

Table 2-2 Test results for CSP for Standalone Server version 1.0.20

	Typical usage profile	High usage profile
Size of Web content (KB)	10	20
Number of requests per second per connection	0.07	0.13
Maximum number of concurrent Web connections per second	7,200	4,800
Throughput (Mbps)	41	102
Incremental Processing delay (ms)	100	100
Connections per user ratio	3	3
Maximum number of concurrent users	2,400	1,600
Provisioned:concurrent user ratio	2.5	2.5
Maximum number of provisioned users	6,000	4,000

- *Size of Web content* - The average size of Web content through the CSP in KB.
- *Number of requests per second per connection* - The average number of web requests per second for every connection.
- *Maximum number of concurrent users* - The maximum recommended capacity in terms of concurrent user connections that can be supported. Further testing showed that the CSP failed when traffic exceeds the numbers listed above and is therefore not recommended.
- *Throughput* - The maximum raw throughput reached during the tests. The connection limit was reached well before the throughput limit. Consequently, your sizing of the CSP should not be based on throughput.
- *Incremental processing delay* - The incremental latency added by the CSP when operating at the maximum number of concurrent connections.

- *Connections per user ratio* - The ratio of the average number of Web requests per connnections initiated by each individual user. This is is based on a sampling of the connection ratios across a variety of Symantec.cloud customer environments. Customers with higher or lower ratios of connections/user can adjust this number to arrive at a sizing estimate matched to their specific circumstances
- *Maximum number of concurrent users* - The number of active or simultaneous users who are using the CSP. The figure is derived by dividing the maximum number of connections by the connections/user ratio.
- *Provisioned: concurrent user ratio* - The ratio of subscribed or provisioned users to the number of active or concurrent users. The figure is based on a sampling of the user ratios across a variety of Symantec.cloud customers. Customers with higher or lower ratios of active to provisioned users can adjust this number to arrive at a sizing estimate that is matched to their specific circumstances.
- *Maximum number of provisioned users* - This is is derived by multiplying the maximum number of concurrent users by the provisioned: concurrent user ratio.

See "Additional information for CSP for Standalone Server version 1.0.20" on page 15. This section provides information on alternative latency data points and the corresponding number of connections for CSP Standalone version 1.0.20.

Using this guide to plan your CSP deployment

We recommend that you keep in mind the following when planning your CSP deployment:

- You should validate the sizing guidelines and assumptions in terms of how they apply to your own test and production environments before deployment.
- You should test with policies and configurations that are consistent with your specific deployment scenario.
- You should carry out testing with a traffic profile that is consistent with your live production environment and use this profile together with this guide to determine the most suitable sizing estimate.

Understanding your organization's current web traffic will help you to determine the number of CSP servers that are required to stay within the connection, throughput, and response time limits shown in this guide. The Web traffic and number of user connections that needs to be processed in a CSP deployment can often be obtained from the network switch or firewall.

To help you to size the CSP, we offer a tool that enables you to extract various metrics from your proof of concept CSP environment and use them to estimate the number of CSP servers that are required to support your user base.

Click CSP Sizing Tool to download the tool. See the documentation included with the tool for information on installation and usage.

When the user traffic volume is known, your server requirements can be roughly estimated by extrapolating from the testing numbers that are shown in Table 2-3. The estimates assume that:

- The traffic profile is based on typical user usage as described in this guide.
- Load distribution is equal across all servers
- There is no redundancy

Multiple CSP Standalone servers can be deployed to support user numbers in excess of the figures that are shown in Table 2-3. You can distribute user traffic across multiple CSP servers as necessary. It is common for large user locations to deploy the CSP Standalone in a redundant, load-balanced configuration.

Table 2-3 Estimating the number of CSP Standalone servers

Number of provisioned users	Number of dedicated servers required - CSP for Standalone Server version 1.0.18	Number of dedicated servers required - CSP for Standalone Server version 1.0.20
1,000	1	1
2,000	2	1
5,000	4	1
10,000	8	2
20,000	16	4

Appendix A

Additional Information

This appendix includes the following topics:

Additional information for CSP for Standalone Server version 1.0.20

The following tables provide additional information on alternative latency data points and the corresponding number of connections for CSP for Standalone Server version 1.0.20.

Table A-1 Typical usage latency versus connections

Latency	Connections
10	2,408
20	3,838
30	4,286
40	5,072
50	5,588
60	5,818
70	6,368
80	6,740
90	7,118
100	7,208

Table A-2 High usage latency versus connections

Latency	Connections
10	1,718
20	2,640
30	3,338
40	3,832
50	4,042
60	4,208
70	4,374
80	4,540
90	4,688
100	4,816