**Lync Server 2013 Stress Testing Guide**



Lync Server 2013 Stress Testing Guide

February 18, 2014

Version 1.0

Prepared by

**Lync Server 2013 Virtualization Team**

Contributors

**Agus Rachman, Cedric Depaepe, James Hornby, Brandon Bernier, Andrew Sniderman, Seth McClure, Campbell Gunn, Ramon Infante**

**Thanks to Christian Lavista and the Microsoft Technology Center, New York**

Acknowledgments

|  |  |
| --- | --- |
| Name | Position |
| Christian Lavista | Director, Microsoft Technology Center, New York |
| Fabian Labat | MTC Service Engineer (Atos) |
| Tim Myers | Senior Service Engineer |
| Joel Yoker | MCS US Hyper-V Architect |
| Chris Dragich | Senior Content Developer |
| Facilities | Microsoft Technology Center, New York |

Table of Contents

[1 Test Plan Overview 2](#_Toc379985607)

[2 Objectives 2](#_Toc379985608)

[3 Test Approach and Assumptions 2](#_Toc379985609)

[4 Lync Server 2013 Stress and Performance Tool Overview 3](#_Toc379985610)

[5 Test Environment Specification 4](#_Toc379985611)

[5.1 Example Scenario 5](#_Toc379985612)

[6 Preparing for a Lync Server Stress Test Exercise 7](#_Toc379985613)

[6.1 Provisioning 8](#_Toc379985614)

[6.2 Load Configuration 13](#_Toc379985615)

[6.2.1 Load Configuration: Common Configuration 14](#_Toc379985616)

[6.2.2 Load Configuration: General Scenarios 16](#_Toc379985617)

[6.2.3 Load Configuration: Voice Scenarios 20](#_Toc379985618)

[6.2.4 Load Configuration: Reach 25](#_Toc379985619)

[6.2.5 Load Configuration: Mobility 27](#_Toc379985620)

[6.2.6 Load Configuration: Summary 29](#_Toc379985621)

[6.2.7 Load Configuration: Placement and Fixup 32](#_Toc379985622)

[6.3 Run Client Scripts Customization 36](#_Toc379985623)

[6.4 Performance Monitoring Preparation 38](#_Toc379985624)

[7 Executing a Lync Server Stress Test Exercise 41](#_Toc379985625)

[7.1 Initialization 42](#_Toc379985626)

[7.1.1 Control Client Users Initialization 43](#_Toc379985627)

[7.2 Test Runs 43](#_Toc379985628)

[8 Analyzing Stress Test Results 44](#_Toc379985629)

[8.1 System Performance: CPU Utilization 45](#_Toc379985630)

[8.2 Role Performance: User Services Queue and Sproc Latency 47](#_Toc379985631)

[8.3 Verdict 48](#_Toc379985632)

[9 Summary 48](#_Toc379985633)

1. Test Plan Overview

To assess and validate the capacity planning requirements of a Lync Server 2013 deployment, we recommend that you perform a stress testing exercise using The Lync Server 2013 Stress and Performance Tool (LSS).

LSS is a toolset that allows organizations planning to deploy Lync Server 2013 to perform a stress test of a design that will be deployed on a target deployment infrastructure (server, storage, and network) based on their specific usage pattern. The LSS can simulate all Lync Server workloads or just specific workloads, such as instant messaging (IM) and presence, application sharing, audio conferencing, and mobility, with a particular scale target.

This test plan will take you through the steps required to plan and conduct a Lync Server 2013 stress testing exercise, and it will help you analyze the results in order to conclude whether the initial capacity plan is sound or requires adjustment.

1. Objectives

The objectives of the test plan are:

* Identify the activities required to prepare for and conduct testing
* Identify configuration controls and metrics

1. Test Approach and Assumptions

A stress testing exercise using LSS must be performed against the same target deployment infrastructure (server, storage, and network) that you intend to use for production implementation, and it must be performed in an isolated lab environment.

Since LSS provides only the tools to provision users and simulate user loads, your stress testing exercise must also involve the proper design and deployment of a Lync Server 2013 implementation with its supporting infrastructure (such as Active Directory, Certificate Services, Office Web Apps Server, load balancer, and reverse proxy).

A stress testing exercise using LSS is typically performed right after the completion of the technical design and before taking the design into pilot. The design, with its supporting infrastructure, should be deployed in the lab environment for functional and nonfunctional testing first. You can then perform stress testing to validate capacity planning.

1. Lync Server 2013 Stress and Performance Tool Overview

For details about LSS, see <http://technet.microsoft.com/en-us/library/jj945609.aspx>.

In summary, LSS provides the following:

* User Provisioning Tool (UserProvisioningTool.exe) to perform the following tasks:
  + Create LSS Active Directory user objects. The user naming convention can be customized as needed, and the user objects index will be appended at the end of the user name.
  + Enable LSS Active Directory user objects as Lync-enabled user.
  + (Optional) Enable Enterprise Voice for the LSS Lync-enabled users with the defined phone area code prefix as a Line URI value.
  + Modify contact lists of each individual LSS Lync-enabled users based on the defined parameters (average contacts per user, average contact groups per user, and federated/cross pool contacts percentage).
  + Create LSS Active Directory distribution group objects and add LSS Lync-enabled users as members of those distribution groups based on the defined parameters.
  + Generate CSV files to be utilized as input files for the E-911 Location Database, covering wireless access point, subnet, port, and switch network elements, and an XML file to be used when simulating LIS user loads. The process of populating the E-911 Location Database must be handled separately using the appropriate Lync Management Shell cmdlets.
* Load Configuration Tool (UserProfileGenerator.exe) to generate stress profiles (in the form of .xml files) and batch files to run the stress test, based on several input parameters covering the number of stress client machines, number of users to be simulated, and workloads to be included as part of the stress testing exercise.
* Lync Performance Tool (LyncPerfTool.exe) that will simulate user loads based on the stress profiles generated by Load Configuration Tool.

In addition, LSS includes all resource files used by Lync Performance Tool when simulating user loads, such as the instant message text, audio files, video files, and PowerPoint files.

Sample configuration scripts are included (look for the scripts in *%ProgramFiles%\Microsoft Lync Server 2013\LyncStressAndPerfTool\Doc* folder) to set up your Lync environment to support stress testing using LSS. Depending on the user loads that will be simulated, you will need to modify these sample configuration scripts, such as **ResponseGroupConfiguration.ps1** and **LisConfiguration.ps1**.

You can also make the necessary changes in your Lync environment manually. However, keep in mind that LSS requires the test elements to use sequential numbers as an index. For example, if you create multiple Response Group workflows with the naming convention **RgsWorkFlow**, to simulate the test of 20 Response Group workflows, the SIP addresses must be set to **RgsWorkFlow0@contoso.com**, **RgsWorkFlow1@contoso.com**, and so on up to **RgsWorkFlow19@contoso.com**.

1. Test Environment Specification

When you design a Lync Server 2013 implementation, factors such as number of users, functional requirements (such as IM and Presence, conferencing, Enterprise Voice, mobility, external access, etc.), and nonfunctional requirements, such as high availability and site resiliency, influence the Lync Server 2013 topology—such as the required server roles, number of servers, and the supporting infrastructure to support the overall solution.

Stress testing using LSS requires stress client machines that run Lync Performance Tool (LPT) to generate various user workloads. The tool is also capable of simulating a PSTN gateway, and it requires a dedicated stress client machine for this purpose.

Simulation of external access requires separate stress client machines that are placed in front of an Edge server external interface. This ensures that the simulation of media-related workloads traverse Lync Edge Server.

To complement the simulated workloads, consider using control client machines to validate the user experience using an actual Lync client connected against the test environment under load. At a minimum, you must dedicate five control client machines equipped with the proper input devices (microphone and webcam) if you intend to validate a multiview user experience. If applicable, some of the control client machines should be placed on the external network to ensure media-related workloads traverse the Edge Server.

* 1. Example Scenario

The following example of a Lync Server 2013 design supports IM and Presence, conferencing, Enterprise Voice, mobility, and external access. High availability is considered in the design, and the expected number of users to be supported by the design is 4,500. In this example, it is assumed users are making one PSTN call per hour. Based on the supplied information, the topology will consist of:

* An Enterprise Edition pool with three Front End servers with collocated Mediation Server.
* A pair of SQL Server 2012 Standard Edition Back End Servers with a SQL Server 2012 witness.
* An Edge pool with two Edge Servers.
* An Office Web Apps Server farm with two Office Web Apps servers.

The topology also includes:

* Hardware Load Balancer
* Reverse Proxy
* PSTN gateway

To stress test Lync Server using LSS, it is best to use an isolated lab environment. The stress testing lab needs to include:

* Active Directory Domain Services domain controllers
* Active Directory Certificate Services root certification authority

As to planning the stress client machines, we recommend that you simulate no more than 4,500 endpoints per stress test client machine. In this scenario, we are simulating 4,500 users, but using a 40 percent multiple point of presence (MPOP) ratio, giving us 6,300 endpoints. As such, a single stress client machine is not sufficient.

Important

Client hardware requirements for the stress client machine are defined in the following article: <http://technet.microsoft.com/en-us/library/jj945592.aspx>.

While the minimum hardware requirements meet most stress testing scenarios, in certain cases, you need to increase either the CPU, memory, or both of the stress client machine. Monitor the CPU and memory utilizations of the stress test client machine, and increase the resources to ensure proper user loads are generated by the stress test client machines.

Based on these factors, five stress client machines will be deployed in the lab environment.

To plan for the control client machines, the minimum number of five control client machines are used with three control client machines placed in the internal network, and two control client machines are placed in the external network.

The following overall components are required in the isolated lab environment for the purpose of this Lync Server stress testing exercise using LSS.



Figure - Example scenario test environment design

1. Preparing for a Lync Server Stress Test Exercise

The steps to prepare to perform a Lync Server stress test exercise using LSS include:

* Provisioning LSS user accounts, contacts creation, distribution list creation, and Location Information Service (LIS) configuration.
* Defining and generating LPT load configuration.
* Placing and fixing LPT load configuration files.
  1. Provisioning

To perform a Lync Server stress testing exercise, you must capture the following information before running the User Provisioning Tool.

|  |  |  |
| --- | --- | --- |
| **User Creation** | **Value** | **Example** |
| Front-End Pool FQDN |  | lyncpool01.contoso.com |
| User Name Prefix |  | lyncpool01-user |
| Password |  | pass@word1 |
| SIP Domain |  | contoso.com |
| Account Domain |  | contoso.com |
| Organizational Unit |  | lyncpool01-users |
| Phone Area Code |  | +1425 |
| Voice Enabled |  | Enabled |
| Number of Users |  | 4505 |
| Start Index |  | 0 |

|  |  |  |
| --- | --- | --- |
| **Contacts Creation** | **Value** | **Example** |
| Average Contacts per User |  | 50 |
| Fixed |  | Disabled |
| Average Contact Groups per User |  | 5 |
| Federated / Cross Pool Contacts Percentage |  | 0 |
| Federated / Cross Pool User Prefix |  | lyncpool01-user |
| Federated / Cross Pool User SIP Domain |  | fabrikam.com |

|  |  |  |
| --- | --- | --- |
| **Distribution List Creation** | **Value** | **Example** |
| Number of Distribution Lists |  | 13500 |
| Distribution List Prefix |  | lyncpool01-DL |
| Minimum Members in a Dist. List |  | 2 |
| Maximum Members in a Dist. List |  | 500 |

|  |  |  |
| --- | --- | --- |
| **Location Info Service Config** | **Value** | **Example** |
| Number of Addresses |  | 100 |
| Offices Per Address |  | 100 |
| Number of WA Points |  | 100 |
| Number of Subnets |  | 100 |
| Number of Switches |  | 100 |
| Number of Ports |  | 100 |
| **Civic Address Details** |  |  |
| Company Name |  | Contoso Corporation |
| Street Name |  | 163rd |
| Street Name Suffix |  | Ave |
| Post Directional |  | NE |
| City |  | Bellevue |
| State |  | WA |
| Zip Code |  | 99234 |
| Country |  | US |

Notes:

* You must create the OU (organizational unit) to be used for LSS before starting the user creation process.
* The Lync Server 2013 pool must be operational in order for the user creation process to be completed successfully.
* The amount of time required to create the users depends on the performance of your Active Directory Domain Services Domain Controllers and your Lync Server 2013 Front End Servers.
* Line URI will be configured using 12-digit E.164 format, prefixed with the value entered as **Phone Area Code** and trailing seven digits (for example, if **+1425** is used as **Phone Area Code**, then when voice-enabling 4,500 users, the Line URI will start from: +14250000000, +14250000001, and so on up to +14250004499).
* Make sure the phone number is not in conflict with existing test users or when you create Dial-in Conferencing phone numbers.
* Your number of users needs to reflect the number of users for stress testing and also the number of users for the control client machines.
* **User Provisioning Tool** creates LIS configurations in the form of .csv files that define the subnets, wireless access points, switches, and ports.
* In addition, **Locations0.xml** file will be generated to be used by Lync Performance Tool when simulating LIS user loads.

After all tables are completed, use the **User Provisioning Tool**, enter the values of the captured information into the tool, and run **Create Users, Create Contacts, Create Distribution Lists,** and **Generate LIS Config Files** sequentially.

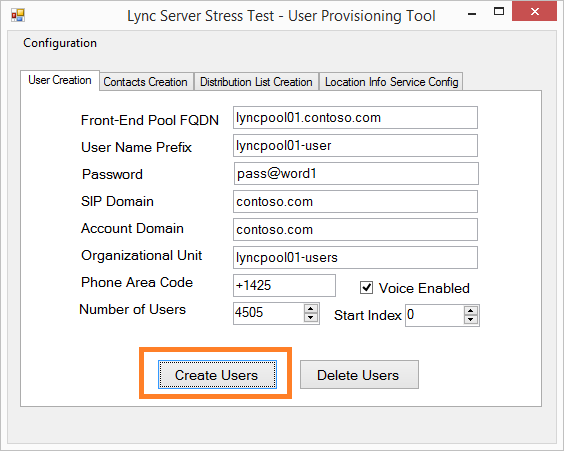


Figure - User Provisioning Tool: User Creation

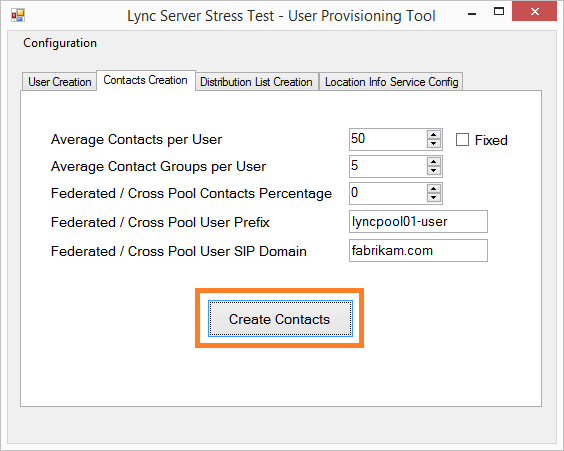


Figure - User Provisioning Tool: Contacts Creation

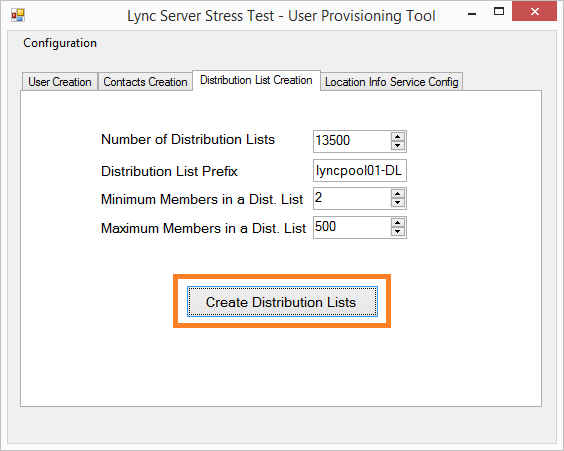


Figure - User Provisioning Tool: Distribution List Creation

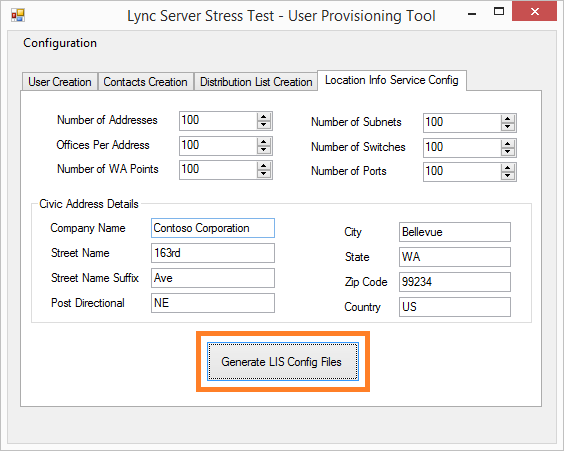


Figure - User Provisioning Tool: Location Info Service Config

Important

The time it takes to complete the whole provisioning process is highly dependent on the performance of your Lync Front End servers and Active Directory Domain Services Domain Controllers.

* 1. Load Configuration

The next step after completing the provisioning process is to create LPT load configurations to simulate the target user loads for stress testing against your Lync Server implementation.

Note

To be able to complete the load configuration step, refer to your Lync Server 2013 topology to supply the required information—such as the Lync Server 2013 pool name and port number, Edge pool name and port number, Internal and External Web Services URL, central site name, etc.

* + 1. Load Configuration: Common Configuration

Use the first tab of the Load Configuration tool to define the number of stress client machines, the user settings that were used during the provisioning step, and the target FQDNs and port numbers of the Front End pool and Edge pool. The values entered in this tab are used to correctly configure the LPT load configuration files.

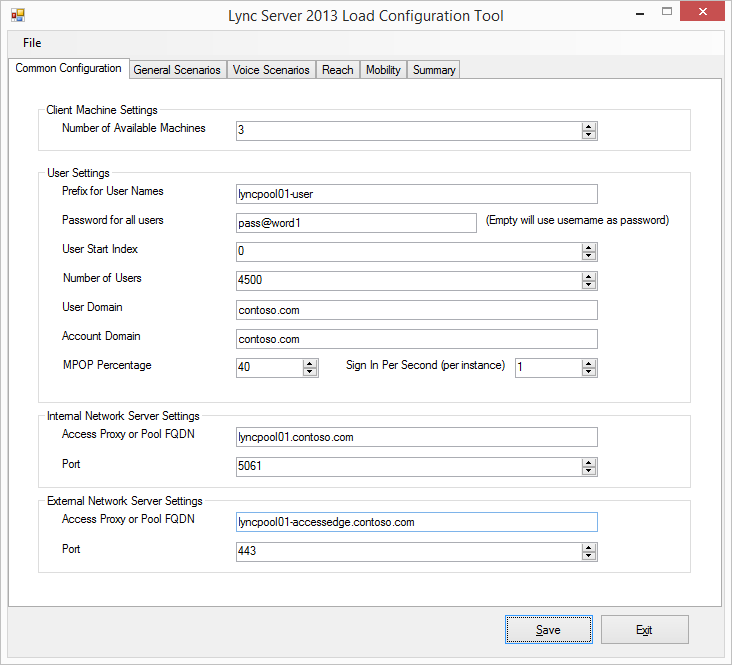


Figure - Load Configuration Tool: Common configuration

Notes:

* We recommend that you use a single stress client machine for no more than 4,500 endpoints, keeping in mind that MPOP percentage is also counted towards that limit. For example, to conduct a stress test exercise for a 4,500-user topology with a 40 percent MPOP ratio, you need at minimum two stress client machines.
* Each PSTN gateway simulator requires a single stress client machine. The number of simultaneous voice calls that can be generated depends on the processor configuration of the PSTN gateway simulator stress client machine. From observation, a quad-core stress client machine can support up to 667 simultaneous voice calls.
* In the example above, one stress client machine is included as PSTN gateway simulator, so in total you will need to prepare three stress clients machines for the purpose of conducting the a Lync stress test exercise using LSS.
* The stress client machines to simulate external access do not need to be included when using this tool. Stress profiles generated by the tool include both internal and external stress profiles, and you must manually copy the stress profiles to the appropriate stress client machines.
* You must specify only the LSS users for the value of **Number of Users**,and do not include the users created for the control client machines.
  + 1. Load Configuration: General Scenarios

The next tab, **General Scenarios**, allows you to choose the type of user loads to be simulated as part of stress testing exercise, covering:

* Instant Messaging and Presence
* Audio Conferencing—while the title includes “audio”, choosing this option will also simulate Video conferencing
* Application Sharing
* Data Collaboration
* Distribution List Expansion
* Address Book Web Query
* Response Group Service
* Location Information Services

Some types of user loads enable you to specify the load level: **Disabled**, **Low**, **Medium**, **High**, **Custom**, and **Enabled**.

**Disabled** and **Enabled** are self-explanatory. The **Low**, **Medium**, and **High** settings determine the amount of conversations occurring per hour for that particular type of user load. For maximum flexibility of configuring the stress test load profile, choose **Custom**, and then configure further using the **Advanced** window of each type of load.

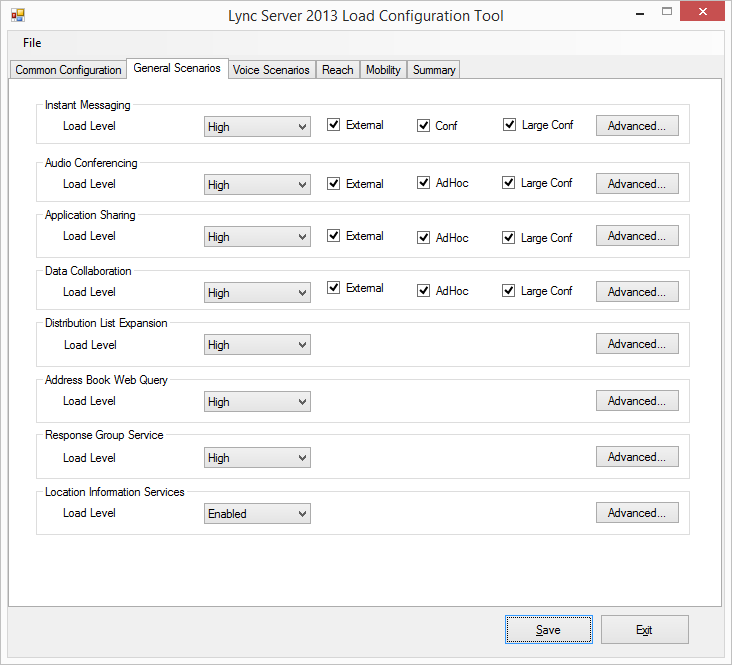


Figure - Load Configuration Tool: General scenarios

Some types of user loads allow you to enable **External**, **Conf**, **AdHoc**, and **Large Conf**.

* **External** Enable this setting to create a stress test configuration for external users signing in through an Edge Server.
* **Conf** Enable this setting to create a stress test configuration that simulates IM conferencing user loads.
* **AdHoc** Enable this setting to create a stress test configuration that simulates ad hoc Application Sharing and Data Collaboration conferencing user loads.
* **Large Conf** Enable this setting to create a stress test configuration that simulates large conferences with up to 250 participants.

Because the **Load Configuration Tool** has no direct relationship with the **User Provisioning Tool**, you must configure advanced options for **Distribution List Expansion** and **Response Group Service** user loads.

In the **Distribution List Expansion** advanced options, modify the **DLXPrefix**, **DLXDomain**, and **DLXIndexHigh** to match the configuration that you used in the User Provisioning Tool.

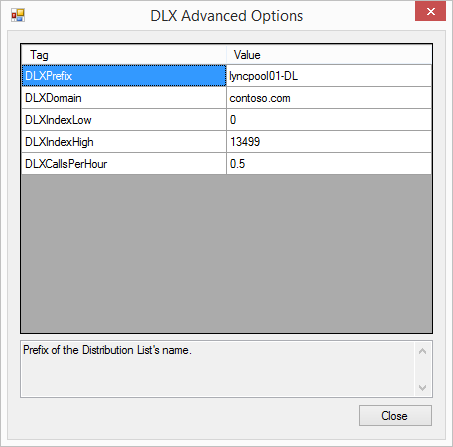


Figure - Load Configuration Tool: General scenarios DLX advanced options

In the **Response Group Service** advanced options, modify **LocationProfile**, **RGSUri**, and **RGSUriSuffixEndIndex** to match your topology (central site name) and Response Group configuration. Modify **NumberOfRGSCallsPerHour** to set the number of RGS Calls per Hour value to be simulated.

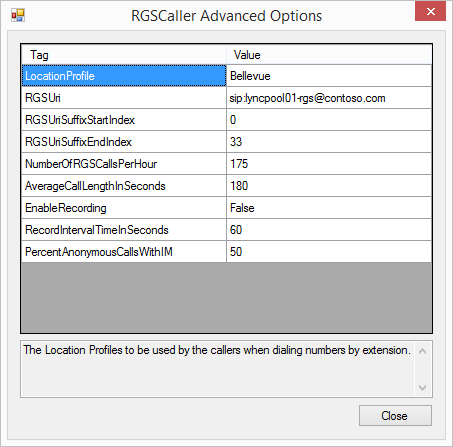


Figure - Load Configuration Tool: General scenarios RGSCaller advanced options

Notes:

* The default values provided by the tool are assumed to be the settings for a pool with 80,000 users. Adjust the number of RGS workflows (RGSUriSuffixEndIndex) and the number of RGS calls per hour accordingly to fit your scenario.
* The Lync Stress Test tool attempts to call RGS using the SIP address defined as the RGSUri value with the index value inserted before the SIP domain notation (for example, **sip:lyncpool01-rgs0@contoso.com**, **sip:lyncpool01-rgs1@contoso.com**, and so on until **sip:lyncpool01-rgs33@contoso.com**). Ensure that you match this behavior when you create RGS workflows for the purpose of RGS user load testing.
* The provided **ResponseGroupConfiguration.ps1** example script can be leveraged as a starting point to create the required RGS agent groups, queues, and workflows. However you must customize it to ensure that you have the right number of RGS workflows, queues, and agent groups, and the right agent memberships within each RGS agent group created by the script.
  + 1. Load Configuration: Voice Scenarios

Use the **Voice Scenarios** tab to choose the type of voice-related user loads to be simulated as part of stress testing exercise, covering:

* Voice over IP (VoIP). This covers UC to UC and UC to PSTN stress tests.
* UC/PSTN Gateway. This refers to the PSTN gateway simulator.
* Conferencing Attendant.
* Call Park Service.

Some types of user loads allow you to specify the load level: **Disabled**, **Low**, **Medium**, **High**, and **Custom**. **Disabled** is self-explanatory. The **Low**, **Medium**, and **High** settings specify the amount of conversations occurring per hour for that particular type of user load. For maximum flexibility of configuring the stress test load profile, choose **Custom**, and configure further using the **Advanced** window of each type of load.

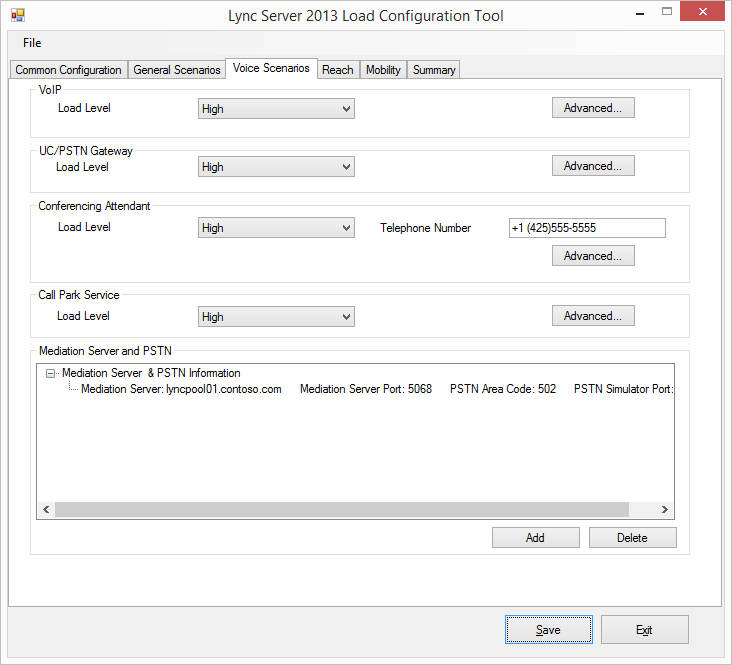


Figure - Load Configuration Tool: Voice scenarios

Because the **Load Configuration Tool** has no direct relationship with the **User Provisioning Tool** and your Lync Server 2013 topology, you must configure advanced options for **VoIP**, **Conferencing Attendant**, and **Call Park Service** user loads to ensure the tests are in-line with your Lync Server 2013 configuration (Enterprise Voice dial plan, Line URI numbering, normalization rules, voice routes, etc.).

In the **VoIP** advanced options, modify **LocationProfile** and **PhoneAreaCode** to match your topology (central site name) and user configuration.

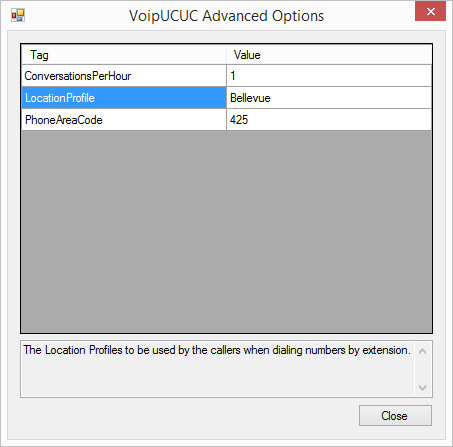


Figure - Load Configuration Tool: Voice scenarios VoipUCUC advanced options

For **Conferencing Attendant** user load configuration, enter the correct telephony number that you have configured for Dial-in Conferencing.

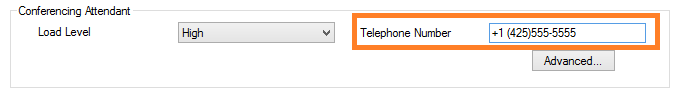


Figure 12 - Load Configuration Tool: Voice scenarios CAA telephone number

In the **Conferencing Attendant** advanced options, make sure **LocationProfile** matches your topology (central site name).

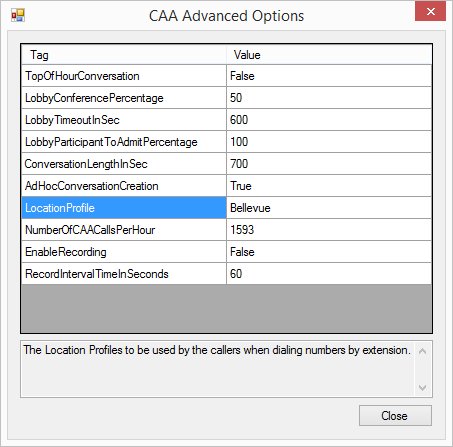


Figure - Load Configuration Tool: Voice scenarios CAA advanced options

In the **Call Park Service** advanced options, modify **PhoneAreaCode** to match your Enterprise Voice configuration.

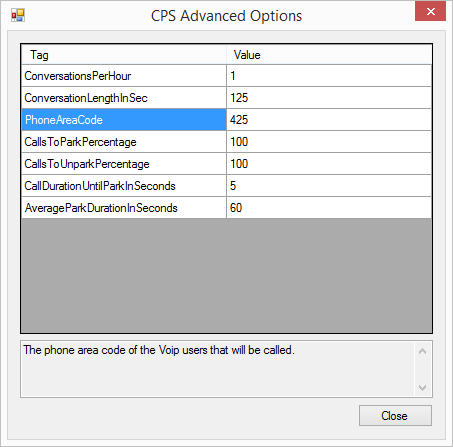


Figure - Load Configuration Tool: Voice scenarios CPS advanced options

After modifying the user load configurations, add **Mediation Server** configuration to ensure the PSTN gateway simulator is targeting the mediation server in your topology.

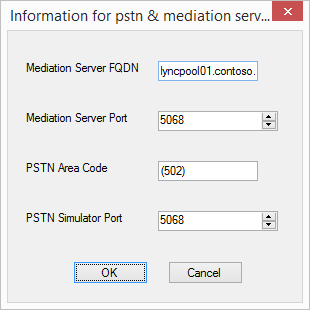


Figure - Load Configuration Tool: Voice scenarios information for PSTN and mediation server

Notes:

* The Load Configuration Tool does not support the configuration of the PSTN gateway simulator using SIP/TLS and SRTP. Make sure you enable TCP support when you configure the mediation server settings on your Lync Server 2013 topology.
* The Load Configuration Tool can only create the load configuration for a single PSTN gateway simulator for each mediation server or pool. To simulate more than 667 simultaneous voice calls, you need to increase the number of processors available to the PSTN gateway stress client machine or prepare additional PSTN gateway stress client machine and use the same load configuration file in each stress client machine. To distribute the load generated by the PSTN gateway simulator, divide the **ConversationPerHour** value evenly between each PSTN gateway stress client machines, and modify the load configuration XML files manually.
  + 1. Load Configuration: Reach

The **Reach** tab allows you to choose the type of Lync Web App-related user loads to be simulated as part of stress testing exercise, covering:

* Application Sharing
* Data Collaboration
* Instant Messaging
* Voice Conferencing

Some types of user loads allow you to specify the load level: **Disabled**, **Low**, **Medium**, **High**, and **Custom**. **Disabled** is self-explanatory. The **Low**, **Medium**, and **High** settings specify the amount of conversations occurring per hour for that particular type of user load. For maximum flexibility of configuring the stress test load profile, choose **Custom**, and configure further using the **Advanced** window of each type of load.

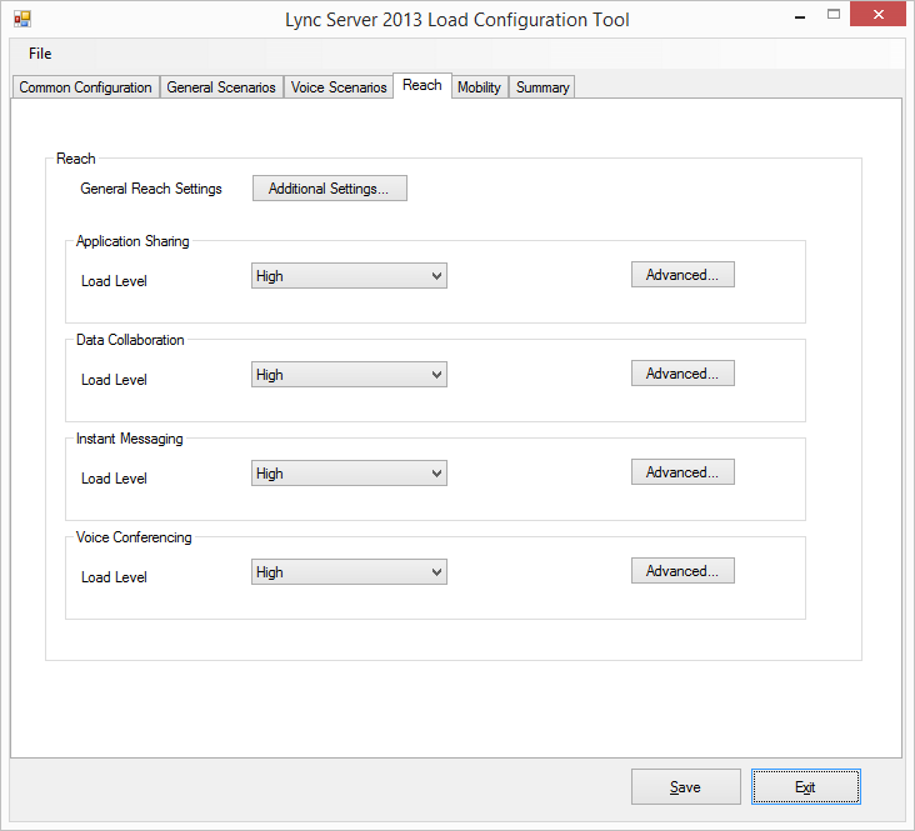


Figure - Load Configuration Tool: Reach

The most important step when setting up load profiles for Lync Web App-related user loads is configuring the **General Reach Settings**. Use your Front End internal Web Services FQDN as **ReachTargetServerUri** value.

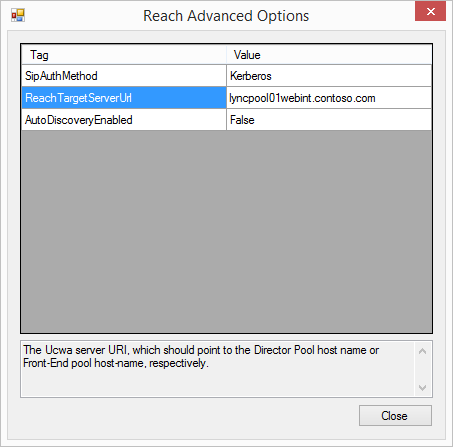


Figure - Load Configuration Tool: Reach advanced options

* + 1. Load Configuration: Mobility

Use the **Mobility** tab to configure the UWCA-based user loads that are to be simulated by the Lync Performance Tool:

* Lync Mobile App Instant Messaging and Presence
* Lync Mobile App Peer-to-Peer Audio/Video calls

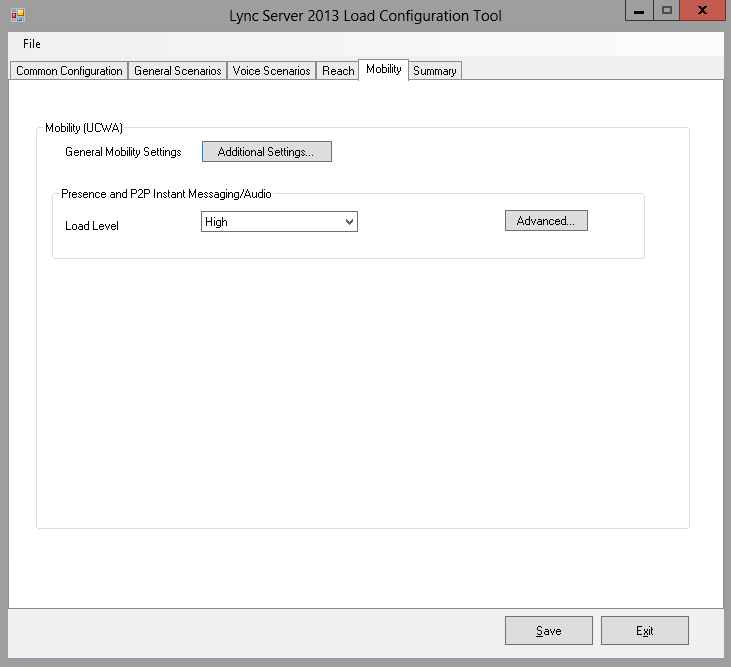


Figure - Load Configuration Tool: Mobility

Some types of user loads allow you to specify the load level: **Disabled**, **Low**, **Medium**, **High**, and **Custom**. **Disabled** is self-explanatory. The **Low**, **Medium**, and **High** settings specify the number of conversations occurring per hour for that particular type of user load. For maximum flexibility of configuring the stress test load profile, choose **Custom**, and configure further using the **Advanced** window of each load type.

The most important step when setting up load profiles for UCWA user loads is configuring the **General Mobility Settings**. Use your Front End External Web Services FQDN and port number as **UcwaTargetServerUri** and **UcwaTargetServerPort** values.

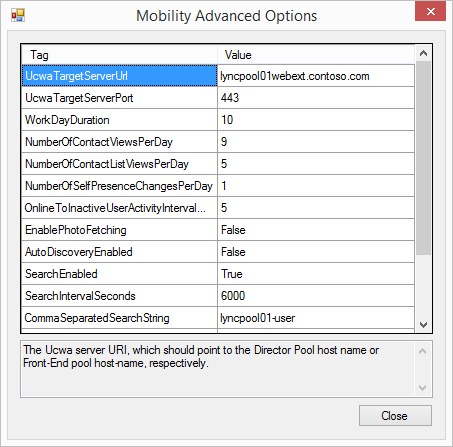


Figure - Load Configuration Tool: Mobility advanced options

Notes:

* With the **SearchEnabled** value set to **True**, you must set a value for **CommaSeparatedSearchString**. The default is blank. The value of **CommaSeparatedSearchString** must match the prefix of the accounts that have been provisioned.
* If your LSS lab does not include connectivity to push notification services, ensure that the **PnchUserRanges** value is not configured. The mobility stress test will fail to start if this parameter is defined while push notification services is not enabled in the topology.
  + 1. Load Configuration: Summary

After you go through all of the previous tabs, the **Summary** tab provides you with the list of user load profiles to be generated based on your input.

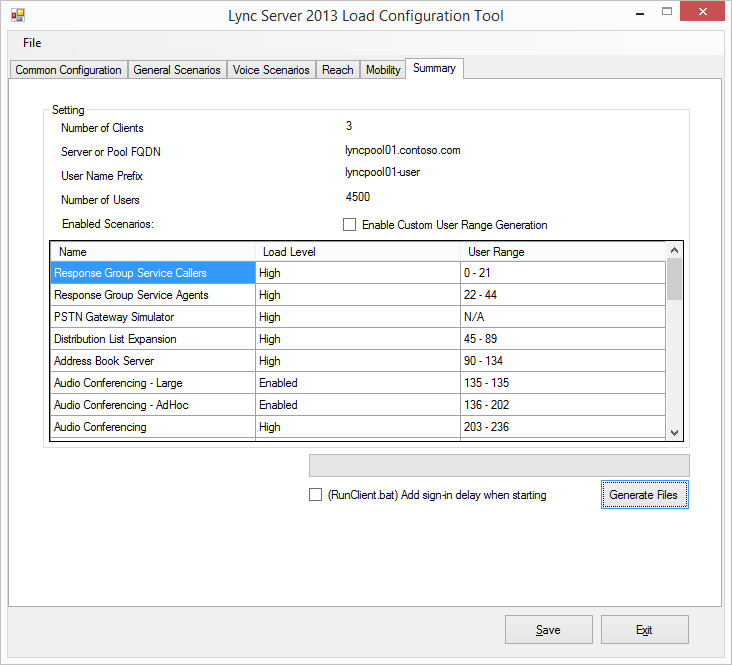


Figure - Load Configuration Tool: Summary

Click **Generate Files** to create the LPT load configuration files. The main folder where the LPT load configuration files are placed uses a date- and time-reference folder name. Multiple subfolders will be created based on the number of stress test client machines that have been defined previously.

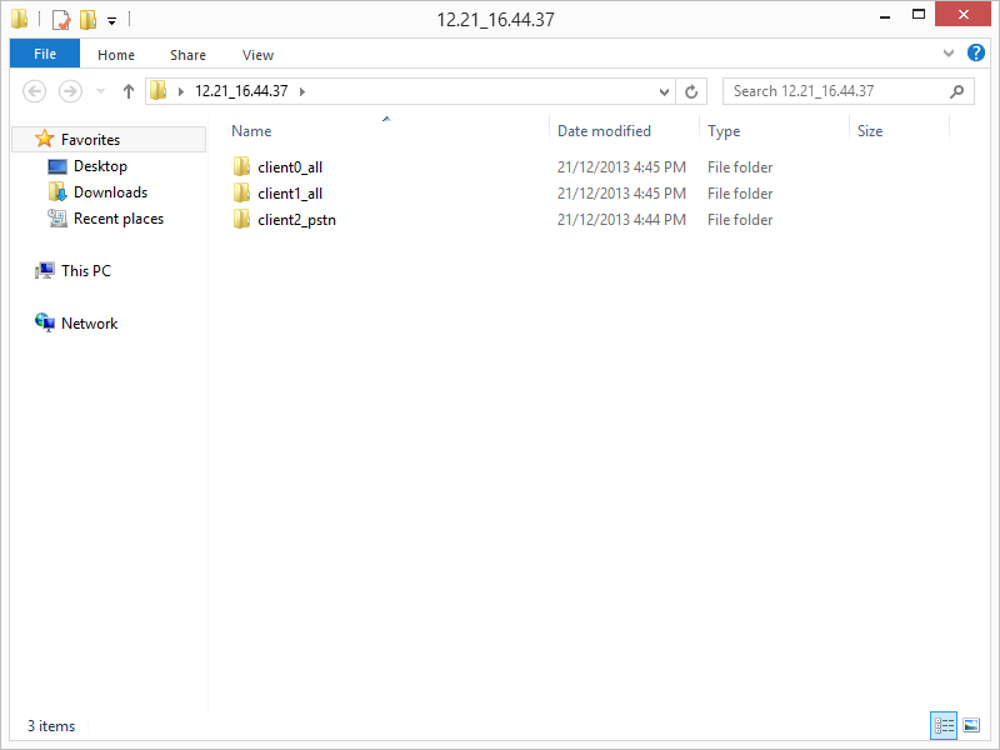


Figure 21 - Folder structure of LPT load configuration files generated by the load configuration tool

Within each client folder, except for the folder containing PSTN gateway simulator load configuration file, you will find multiple LPT load configuration files (in XML format) and a batch file to run LPT against the configuration files.

The batch file also registers LPT counters. This is a requirement before you can use LPT to simulate user loads.

Note

You must manually copy the client folders to each stress test client machine and ensure that LSS is installed in all stress test client machines.

* + 1. Load Configuration: Placement and Fixup

When you simulate all available stress test user loads, there are 32 types of LPT load configuration files.

Types of LPT Load Configuration Files

|  |  |  |
| --- | --- | --- |
| **Number** | **LPT load configuration file type** | **LPT load configuration file naming convention** |
| 1 | Response Group Service Callers | RGSCaller\_clientx.xml |
| 2 | Response Group Service Agents | RGSAgent\_clientx.xml |
| 3 | Distribution List Expansion | DL\_clientx.xml |
| 4 | Address Book Server | ABSQuery\_clientx.xml |
| 5 | Audio Conferencing - Large | AvLarge\_clientx.xml |
| 6 | Audio Conferencing - AdHoc | AvAdHoc\_clientx.xml |
| 7 | Audio Conferencing | AV\_clientx.xml |
| 8 | Audio Conferencing - External | AvExternal\_clientx.xml |
| 9 | Conferencing Attendant | CAA\_clientx.xml |
| 10 | VoIP (UC To UC) | VoipUCUC\_clientx.xml |
| 11 | VoIP (PSTN to UC and vice versa) | VoipPSTNUC\_clientx.xml |
| 12 | Application Sharing - Large | AsLarge\_clientx.xml |
| 13 | Application Sharing - Ad-Hoc | AsAdHoc\_clientx.xml |
| 14 | Application Sharing - External | AsExternal\_clientx.xml |
| 15 | Application Sharing - Sharer | ASSharer\_clientx.xml |
| 16 | Application Sharing - Viewer | ASViewer\_clientx.xml |
| 17 | Location Information Service | LIS\_clientx.xml |
| 18 | Call Parking Service | CPS\_clientx.xml |
| 19 | Data Collaboration - Large | DataCollabLarge\_clientx.xml |
| 20 | Data Collaboration - AdHoc | DataCollabAdHoc\_clientx.xml |
| 21 | Data Collaboration - Conferencing | DataCollab\_clientx.xml |
| 22 | Data Collaboration - External | DataCollabExternal\_clientx.xml |
| 23 | Reach2013 - Instant Messaging | ReachIM\_clientx.xml |
| 24 | Reach2013 - Data Collaboration | ReachDataCollab\_clientx.xml |
| 25 | Reach2013 - Application Sharing | ReachAS\_clientx.xml |
| 26 | Reach2013 - Audio Conferencing | ReachAV\_clientx.xml |
| 27 | Instant Messaging - Large Conference | ImLarge\_clientx.xml |
| 28 | Instant Messaging - Conferencing | ImConferencing\_clientx.xml |
| 29 | Instant Messaging - External | ImExternal\_clientx.xml |
| 30 | Instant Messaging - Peer to Peer | IM\_clientx.xml |
| 31 | Mobility - Presence | MobilityIMP\_clientx.xml |
| 32 | PSTN Gateway Simulator | PSTN*areacode*\_*mediationpoolname*.xml |

LPT load configuration files to simulate large conferences are placed only on a single stress test machine, and the other stress machines are designated as PSTN gateway simulators. To properly simulate external access scenarios, place the external LPT load configuration files on stress test machines designated as external stress test machines.

In the Contoso example stress test exercise, the placement of the LPT load configuration files will be as depicted in the following table.

LPT Load Configuration Files Placement



There are known bugs with the Load Configuration Tool that require you to fix the LPT load configuration files to be able to conduct a successful Lync Server 2013 Stress Test exercise.

* **SipAuthMethod** The value for all external LPT load configuration files is set to Kerberos. You must change this to NTLM.
* **UcwaTargetServerUrl** and **UcwaTargetServerPort** The values for Mobility LPT load configuration file (*MobilityIMP\_clientx.xml*) are left blank. Use the Lync Pool Web Services External URL as value for **UcwaTargetServerUrl** and use 443 as the value for **UcwaTargetServerPort**.
* **UserNumberBeginning** The value in the Audio Conferencing LPT load configuration file (*AV\_clientx.xml*) refers to the same value in one client and the other—in the Contoso example scenario, *AV\_client0.xml* and *AV\_client1.xml* both use the value of **203** for **UserNumberBeginning**. Each client file should refer to a different set of users. For example, if *AV\_client0.xml* uses the value **203** for **UserNumberBeginning**, and **UserNumber** value is **17**, then *AV\_client1.xml* should use the value of **220** for **UserNumberBeginning**.
* All Reach2013 LPT load configuration files (*ReachIM\_clientx.xml*, *ReachDataCollab\_clientx.xml*, *ReachAS\_clientx.xml*, and *ReachAV\_clientx.xml*) are incorrectly configured to enable the **UcwaCommonConfiguration** module. Disable this by changing the value of **Enabled** under **UcwaCommonConfiguration** module to **false**.
* **UserNumberBeginning** Value in Instant Messaging - External LPT load configuration file (*IMExternal\_clientx.xml*) refers to the same value in one client and the other—in the Contoso example scenario, *IMExternal\_client0.xml* and *IMExternal\_client1.xml* both use the value of **2458** for **UserNumberBeginning**. Each client file should refer to a different set of users and not colliding with the previous load configuration file(s). In the Contoso scenario, the last index of user from Instant Messaging Conferencing LPT load configuration file (*ImConferencing\_client1.xml*) is **2458**. Therefore, the **UserNumberBeginning** value of *IMExternal\_client0.xml* must start from **2459**. When the **UserNumberBeginning** value is changed, then **InviteUserStartIndex** must be modified to match **UserNumberBeginning**.
* **InviteUserEndIndex** The value of Instant Messaging - External LPT load configuration file refers to the expected last index of users that are part of Instant Messaging - External LPT user loads—which in the Contoso scenario is **3029**. The combination of **UserNumberBeginning** and **UserNumber** in both *IMExternal\_client0.xml* and *IMExternal\_client1.xml* LPT load configuration files must result in LSS users coverage with index from **2459** to **3029**. To achieve this, reduce the **UserNumber** value in *IMExternal\_client0.xml* to **285**. Therefore, *IMExternal\_client0.xml* will cover user index **2459** to **2743**, and *IMExternal\_client1.xml* will cover user index **2744** to **3029**.
* Instant Messaging - Peer to Peer (*IM\_clientx.xml*) and Mobility - Presence (*MobilityIMP\_clientx.xml*) LPT load configuration files cover the same user index, and UserNumberBeginning value for both Instant Messaging - Peer to Peer (*IM\_clientx.xml*) and Mobility - Presence (*MobilityIMP\_clientx.xml*) LPT load configuration files must be configured to continue from the last index of Instant Messaging - External LPT load configuration. In the Contoso example, **UserNumberBeginning** for *IM\_client0.xml* and *MobilityIMP\_client0.xml* must be set to **3030**. Ensure that the combination of **UserNumberBeginning** and **UserNumber** values of the last LPT load configuration files covers the last index of provisioned in Active Directory.
* When changing **UserNumberBeginning** and **UserNumber** values in Instant Messaging - Peer to Peer (*IM\_clientx.xml*) and Mobility - Presence (*MobilityIMP\_clientx.xml*) LPT load configuration files, make sure you adjust **InviteUserStartIndex** and **InviteUserEndIndex**, and also **UcwaUserStartIndex** and **UcwaUserEndIndex** so that the values are within the coverage of LSS users for these particular user load stress tests.

The following table shows the **UserNumberBeginning** and **UserNumber** values in all LPT load configuration files after the fixup.

UserNumberBeginning and UserNumber Values Post Fixup



* 1. Run Client Scripts Customization

Prior to executing a stress testing exercise using LSS, ensure that your Lync Server 2013 implementation is operational and has all settings configured to support the specific user load testing scenarios that you're planning to run against the environment.

While the Load Configuration Tool generates the batch files (RunClientx.bat) to run the LPT against load configuration files created by the tool, you must make additional customizations to ensure the test will run properly.

In essence, the batch files consist of the following tasks:

* Register LPT counters (regsvr32 /i /n /s LyncPerfToolPerf.dll)
* Run LPT using configuration files created previously with three variances:
  + **Normal** stress client machine (all LPT load configuration files except the Large conference scenario configuration files and the PSTN Gateway simulator configuration file)
  + **Large** conference stress client machine (all LPT load configuration files and LPT load configuration files for Large conference scenarios, except the PSTN Gateway simulator configuration file)
  + **PSTN** Gateway Simulator stress client machine (just the PSTN Gateway simulator configuration file)

As you want to separate stress client machines simulating internal user loads and external user loads, the batch files will need to be modified so that **Normal** and **Large** stress client machines are only running internal LPT load configuration files. You can then copy the original batch files to the external stress client machines and then modify the batch files so they will only run external LPT load configuration files.

Another issue that needs to be considered is that some of the LPT user load scenarios share the same resource files. As the batch files are originally generated in a way that multiple LPT load configuration files are loaded simultaneously in parallel, there is a risk of resource access contention. To address this, you must modify the batch files so there is a waiting period between loading the next LPT load configuration file.

Note

There are multiple ways to approach this, and one option to consider is to rewrite the batch files using PowerShell script and leverage the **start-sleep** cmdlet to add a waiting time between launching one LPT load configuration file and the next.

Based on the information above, in the Contoso stress test example, five machines will be prepared to simulate internal user loads, external user loads, and a PSTN gateway. All machines will have the Lync Server 2013 Stress and Performance Tool installed, with all of the required LPT load configuration files copied locally, and they will run client scripts created to launch the appropriate LPT load configuration files based on the individual stress machine roles.

* 1. Performance Monitoring Preparation

To determine whether the target infrastructure can support the planned user loads, you should implement performance monitoring on both the Lync Server 2013 server roles and the LSS stress client machines.

On Lync Server 2013 server roles, especially Front End Server, high CPU utilization and memory utilization can easily indicate that the server is not going to be able to handle the end state user loads. Normally, you want to keep CPU utilization to below 80 percent CPU utilization to allow for unplanned server failure in a pool and also to allow room for unexpected user growth.

On LSS stress client machines, high CPU utilization and memory utilization can cause the stress clients to not sufficiently generate the expected user loads to fully stress the Lync servers—thus giving you a wrong expectation around the ability of your Lync Server 2013 server roles to handle the end state user loads.

Based on this information, you need to run the Performance Monitoring tool and capture all of the pertinent counters throughout the stress testing exercise for analysis after completion of a test run.

Obviously, there are more counters to watch for aside from CPU utilization and memory utilization. Therefore, refer to the list below regarding the list of counters that you can consider to include when creating a Perfmon data collector set.

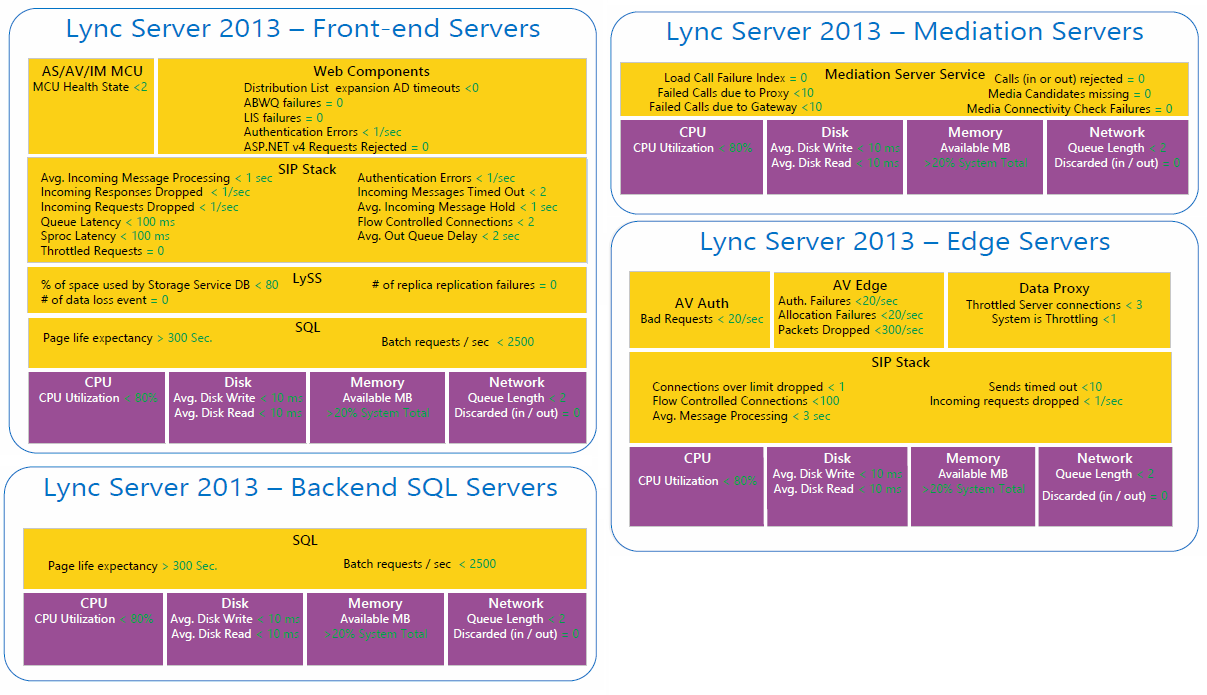


Figure 22 Lync Server 2013 Key Health Indicator (KHI) counters

On the stress client machines, the following counters can be considered to be included as part of the LPT Perfmon data collector set.

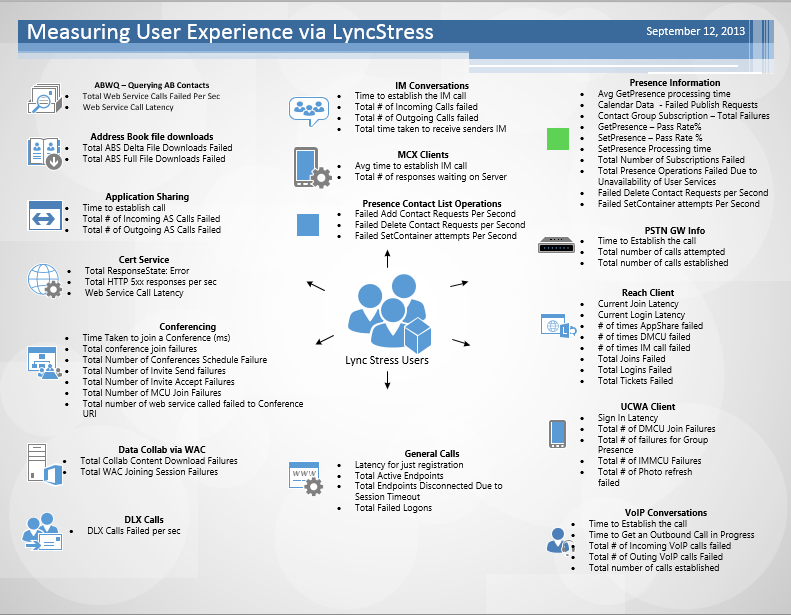


Figure 23 - Lync Performance Tool counters

When you run a stress test exercise against a Lync Server 2013 implementation running on a server virtualization platform such as Hyper-V, capturing performance counters on host servers can help you correlate virtualized Lync Server performance against host server performance throughout the stress test exercise. You can use this to understand the impact of operating Lync Server 2013 server roles on your server virtualization platform. In the case of Hyper-V, you can use the Perfmon data collector set template for Hyper-V from PAL (Performance Analysis of Logs) Tool that can be downloaded from <http://pal.codeplex.com/> as a baseline of important counters to capture on the host servers when running a Lync Server 2013 stress test exercise.

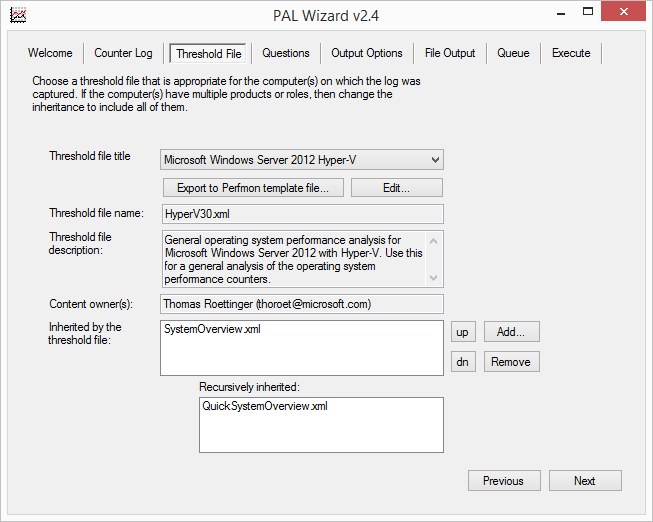


Figure 24 - Hyper-V Perfmon Data Collector Set template from PAL

1. Executing a Lync Server Stress Test Exercise

Running a Lync Server stress test exercise includes:

* **Initialization** Use this step to to initialize the LSS users that were provisioned using **User Provisioning Tool** and prepare all of the control client machines.
* **Test Run 0** Use this step to validate Lync Performance Tool (LPT) load configurations and Lync server configuration are correctly configured. This must be repeated until all LPT load configurations have been validated to run properly.
* **Test Run 1** Use this step to establish the performance baseline of the Lync Server 2013 implementation under load.
* **Test Run 2** Use this step to validate the performance baseline and data source for analysis and reporting of a stress test exercise.

Note

Minor deviation between Test Run 1 and Test Run 2 is expected. However, if you find major deviation between the results from Test Run 1 and Test Run 2, you need to rerun Test Run 1 and Test Run 2 to ensure the validity of the stress test exercise.

* 1. Initialization

To initialize the LSS users that were provisioned using the **User Provisioning Tool**, you can use the Instant Messaging - Peer to Peer LPT load configuration files as a template to create LSS users initialization LPT load configuration files.

In the Contoso example, copy *IM\_client0.xml* and *IM\_client1.xml* to new files *IM\_client0\_initialization.xml* and *IM\_client1\_initialization.xml*.

For *IM\_client0\_initialization.xml*, modify the **UserNumberBeginning** value to **0** and **UserNumber** to **2250**. For *IM\_client1\_initialization.xml*, modify the **UserNumberBeginning** value to **2250** and **UserNumber** to **2250**. On both files, change **MPOPEndPointPercentage** to **0**.

Run LPT using the modified configuration files on both internal stress client machines manually from the command line: **LyncPerfTool.exe /ConfigurationFile:IM\_clientx\_initialization.xml**.

When you use Perfmon on the stress client machines, watch the number of endpoints registered on each stress client machine. With the default rate of sign in per second of 1, in the Contoso example, within ~38 minutes, you should see that each stress client will register 2,250 endpoints.

After each stress client machine successfully registers the expected number of endpoints, let it run for at least 4 hours, and then you can safely close the LPT instance on the stress client machine to stop the initialization process.

Note

The initialization using Instant Messaging - Peer to Peer LPT load configuration as a template simulates a very simple test, and therefore if the stress client machine is showing instances of disconnection, this might indicate that the Front End server(s) are not able to handle a simple IM and Presence scenario—let alone the rest of the types of user loads.

* + 1. Control Client Users Initialization

To initialize the control client users, you must sign in manually from each control client machine using Lync 2013 client.

* 1. Test Runs

At a minimum, you should perform three test runs as part of a stress testing exercise.

The first test run—let’s call it **Test Run 0**—is meant to ensure that all load configurations are correctly configured. Extensive performance monitoring might not be required in **Test Run 0**. However, you need to use CLS (Centralized Logging Service) or OCSLogger to troubleshoot any issues with the LPT load configuration files.

Example

Failing to change the authentication type to NTLM for external load configuration files generates a lot of authentication failures on the Edge servers. This is one example where CLS or OCSLogger can help you identify an issue with the LPT load configuration files.

**Test Run 1** helps you to establish a performance baseline. This test run should be at least 6 hours. You must run performance monitoring on all Lync Server roles using a Perfmon data collector set that captures all Lync Server 2013 KHIs. On the stress client machines, you must run performance monitoring of all LPT instances using a Perfmon data collector set that captures important LPT counters.

When you test on a server virtualization platform, ensure that the hypervisor performance counters are recorded throughout the stress testing exercise.

To validate the end user experience of the test environment under load, you must run user acceptance test (UAT) test scripts using the control client machines.

Note

A UAT test scripts document is assumed to be available as part of functional testing phase that was performed prior to the stress testing exercise. The test would normally cover activities that users will do on a daily basis as Lync users—such as signing in, conducting peer-to-peer IMs, voice-only conferencing, video conferencing, web conferencing, etc.

With the control client, you must manually capture your observation of the user’s experience, such as the time it takes to sign in, the time it takes to start a conference, voice and video quality, etc.

**Test Run 2** is used to validate performance baseline. If the result of Test Run 2 deviates greatly from Test Run 1, you need to analyze all data points, make adjustment to the environment, and start over.

Note

Between test runs, it is recommended for you to perform clean-up or maintenance, such as running a SQL backup on the backend database servers to trim the transaction logs and to clean up IIS logs on the Front End servers. This is to ensure that you are not encountering disk space issues during a test run that can influence the test results.

1. Analyzing Stress Test Results

As data collection is performed using Perfmon, analysis of stress test results involves the comparison of captured information against KHI thresholds. You can do this using Perfmon or other analysis tools.

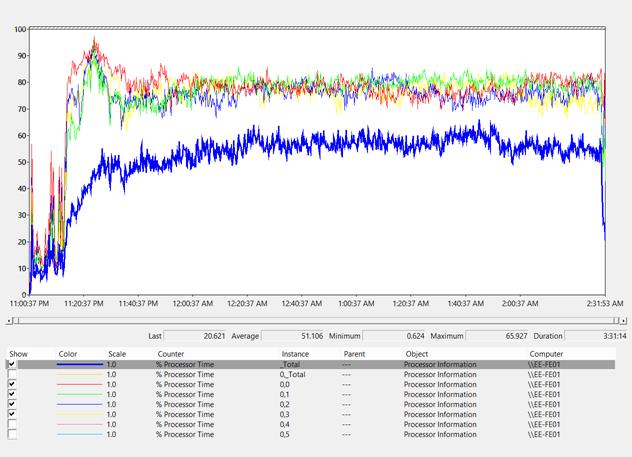
There are two key areas that can be analyzed, **system performance** and **role performance**. Comparison of stress test results compared to system-related KHI thresholds will determine whether the test environment meets the requirements for a **system pass**. After a **system pass** is achieved, further analysis of stress test results can be done to determine whether the stress test results meet the requirements for a **role pass**.



The example in the following section illustrates two key areas, CPU utilization and User Services queue and sproc latencies, that can be used to represent **system performance** and **role performance** to analyze a stress test result and declare whether the test environment can keep up with intended user loads.

* 1. System Performance: CPU Utilization

The following shows CPU utilization data of a single Front End Server from a stress test involving an Enterprise Edition pool with three Front End Servers.



This diagram shows that as we are loading LPT load configuration files gradually with one minute wait time between each, the CPU utilization gradually increases within 30 minutes. As LPT instances start to simulate workloads, there are spikes of CPU utilizations that can be considered as normal given the ramp up of user loads.

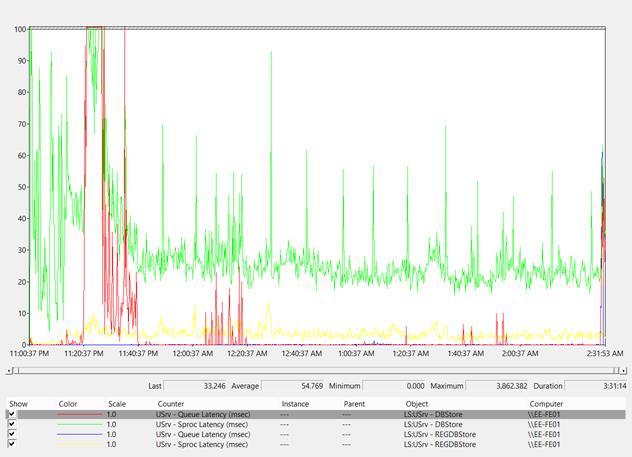
There are four processors that remained highly utilized over time, and upon deeper analysis during the test run, it was found that those processors are associated with RTCLOCAL SQL Server Express Edition instance. As SQL Server Express Edition can only utilize four logical processors, then this result indicates that the Enterprise Edition pool with three Front End servers is already at full capacity and therefore to accommodate more users you will need to add additional Front End servers.

Note

Should you decided to add additional Front End servers, you will need to repeat **Test Run 1** and **Test Run 2** and re-analyze the result of the additional test runs.

* 1. Role Performance: User Services Queue and Sproc Latency

In relation to the CPU utilization diagram above, below is the Queue and Sproc Latency of the user services databases.



The diagram above shows that as we load LPT load configuration files gradually, the queue and sproc latencies are randomly spiking above the threshold (100 ms) within the first 45 minutes. As LPT instances start to simulate workloads, these random spikes can be considered as normal given the ramp up of user loads.

The result is showing that the environment is still capable of handling the user loads, indicated by queue and sproc latencies that remain below the threshold over time.

* 1. Verdict

Since a **system pass** is not achieved here, even though the **role pass** is achieved, this stress test exercise is showing that the design is already at full capacity and not in a state where it can sustain a single Front End server for a long period of time without impacting the user experience.

1. Summary

Using The Lync Server 2013 Stress and Performance Tool (LSS) to perform a stress testing exercise of a Lync Server 2013 implementation can help you to validate the design in terms of capacity planning.

Adjustment to the design, if required, can be implemented in the testing phase and prevent you from placing an undersized design into production.