Administrators Guide

Fault tolerant Hot Standby

Version: 7.x

SUPPORT@TELISCA.COM
TEL. +331 4645 0512

HELP
Open a ticket with your logs on http://support.telisca.com for a prompt and efficient response!
Server: MENU>Support>Zip Logs

Summary

1 Overview

The telisca framework can synchronise its configuration settings with a back-up server out of the box. An IP redirection can be put in place with a load balancer to point to the backup server when the primary fails. However when the applications require CTI functions, **the Hot Standby module provides a mechanism to control which of the telisca CTI service (on the primary or backup server) will be active**.

This document explains which applications can be made fault tolerant with the module Hot Standby, how to configure it and how to create Load Balancing using the Windows Load Balancer or External Load Balancer and some guidance on how to configure an existing setup to work with Fault Tolerance.

1.1.1.1 Redundancy architecture

An optional module allows IPS Framework & Administration and telisca CTI Server to work in Hot Standby mode. On normal process, primary server is running and backup server is idle. An automatic mechanism starts the backup server when the primary server fails.

In this mode, an IP Balancer must be used to redirect the http requests from the IP Phones. Normally, the IP Phones are not configured to use DNS and the http service is called using an IP address. This address must by default be directed to the primary telisca server.

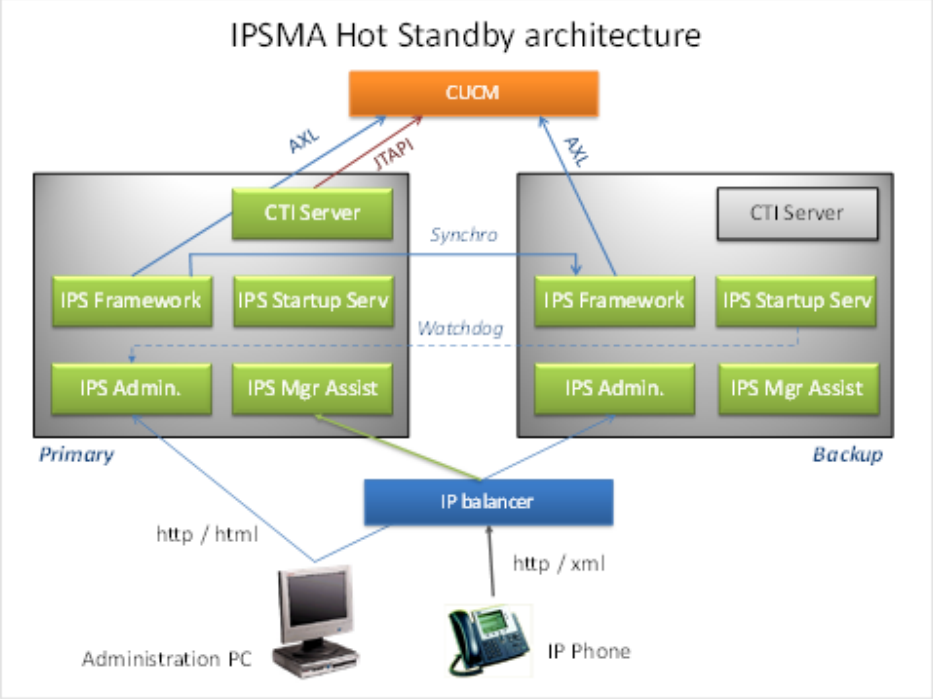
IPS Administration synchronizes the system configuration on the two servers natively, this module is only needed for CTI based applications.

This mode if only supported if a specific license (IPSFT_lic.xml) have been uploaded from IPS Administration. This license is independent of the number of phones/users.

This option is supported with the following applications:

- TAnnounce
- Attendant Console
- Busy Alerter
- Wake Up Call
- Voice Callback
- Voice Alert
- Conference Center
- TSSO
- IPS Popup
- IPS Manager Assistant
- IPS Pager
- IPS Phone Config
- IPS Alarm Callback
- Desktop Popup
- Missed Calls Email Alerter
- Morning Check
- Silent Monitoring

1.1.1.2 Load Balancer architecture



Fault tolerance, architecture for IPSMA (same principle as other apps)

The IP Balancer will check periodically that the primary server is available (ping) and eventually answer to a monitor http URL without http error. After several consecutive errors the IP Balancer must redirect all http requests to the backup server.

In the meantime, the IPS Startup Service on the backup server periodically checks the primary server monitor URL as well. After the same number of consecutive errors he forces the primary server to switch to idle mode and the backup server to run mode. This starts the CTI Server which connects to the CUCM CTI Manager.

The IP Load Balancer and the backup server must switch in approximately the same time frame so that the application is already up and running on the backup server when the IP Load Balancer redirects the http request from the IP Phones to the backup server.

Returning to the normal configuration is a manual process done from the administration, unless the primary server has the option 'First server auto-start Primary/Active'.

When a server restarts and finds that the other server is already running, it synchronizes its configuration from the first one and run-in idle mode. When both servers are restarting at the same time (auto restart during night) a mechanism forces the primary the server to run and the backup to idle.

Important! When using fault tolerance, the two servers must be configured with the same NTP time server. This is because the synchronisation of configuration files is based on timestamps.

2 Pre-requisites, installation

Supported Cisco CUCM:

- CUCM version 10.5, 11.5, 12, 12.5, 14

Windows servers supported:

- Windows Server 2012 R2 Essentials or Standard
- Windows Server 2016 Essentials or Standard
- Windows Server 2019 Essentials or Standard
- Windows Server 2022 Standard

- Minimum configuration: 1 vCPU, 4GB RAM, 70GB disk
- Virtual Machine VMware vSphere, Hyper-V or Cisco UCS, Cisco UCS-E
- Cloud ready

2.1.1.1 Network prerequisites

You have to make sure that the following ports and protocols are allowed on the network for Fault Tolerance, please refer to the telisca framework guide for the base installation requirements.

Source	Destination	Protocols/ports	Delay max RTT
telisca server	telisca server (IPS Startup)	TCP 2011	1000ms
telisca server	telisca server (IPS Framework)	http 80 or https 443	1000ms

Ports and Protocols used by the Fault Tolerance mechanism (all port numbers are configurable)

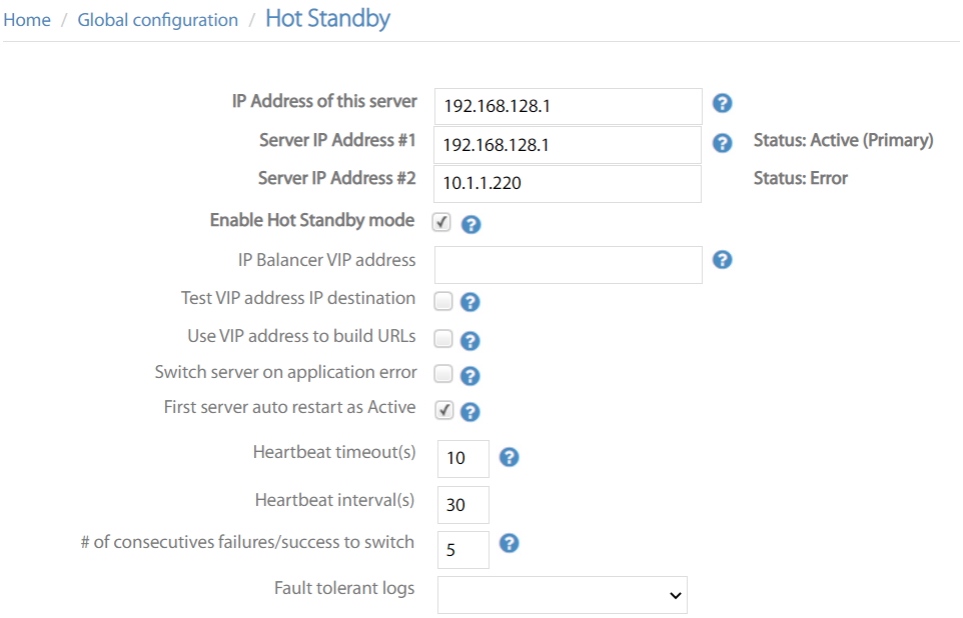
Example of transaction data transfer:

Operation description	Source-destination	Data transfer per operation
Fault tolerance keep-alive	telisca server -> telisca server	180 kB / hour

3 Administration

The fault tolerant configuration is defined in Global / Parameters screen.

It is available only if Fault Tolerant license IPSFT_lic.xml license file has been loaded.



The Hot Standby configuration screen

3.1.1 Steps to configure the hot standby

See here if using Windows NLB, the features required to be setup on both servers.

The two servers must have the same telisca applications installed.

On each server:

- Enter the local IP address

On server one:

- Enter Server IP address #1
- Enter Server IP address #2
- Check 'Enable Hot Standby mode'.
- Optionally, enter the IP address of the IP Load Balancer. It is used to build URL pushed to IP Phone. Do not define it if using Windows Load Balancer.
- Validate

The configuration will be replicated on the secondary server (if not done automatically, validate one of the configuration screens on primary to force the synchronisation, for IPS Global Directory for instance).

3.1.2 Administration tab fields

3.1.2.1 Test VIP address destination

You can optionally enter the IP address of the IP Load Balancer that is redirecting the IP Phone http request to the running server. If this option is enabled, the Hot Standby Module will check the destination of the VIP address by sending http request to the VIP address. The Hot Standby module will switch the server which is the destination of the VIP address, after the number of failures defined, to Primary mode.

3.1.2.2 Switch server on application error

If the choice 'Switching on application error' is checked, the backup server will analyze the response of the URL IPSCFG/admin/Monitor.aspx to detect errors (and not only http errors). This option can be selected only if the IP Load Balancer is able to analyze the http response as well and detect that it do not contain an 'OK' answer. Do not use if using Windows Load Balancer.

Note: we do not recommend using this feature as sometimes, minor errors can trigger too many unnecessary switches.

3.1.2.3 First server auto-restart as active/primary

The server #1, will actively try to regain active status when it is restarted or healthy again. If the monitor's URL answers fine for the 'number of failures' defined.

3.1.2.4 Heartbeat timeout

Defines how long IPS Startup will wait for a reply when requesting monitor.aspx URL. By default 10 seconds. The timeout must take into account AXL timeout because when monitor URL is queried an AXL read can be executed. Heart beat timeout shou_ld be equal to AXL timeout (default 7s) + 3s. AXL timeout is defined in menu Global Config, Config CUCM folders, advanced parameters.

3.1.2.5 Heartbeat interval

Defines the period (in seconds) IPS Startup will check the monitor.aspx URL on the alternate server or the local server in some case. By default every 30 seconds.

3.1.2.6 Number of failures to switch

The backup server will become primary after the number of consecutive failures of the primary server. If option 'First server auto restart primary' is enabled the first server will become primary again after the number of consecutives success.The minimum value - 1 x heartbeat interval should be greater than IPS Framework & Administration restart time, which is generally around 60s. Default value is 5 x 30s.

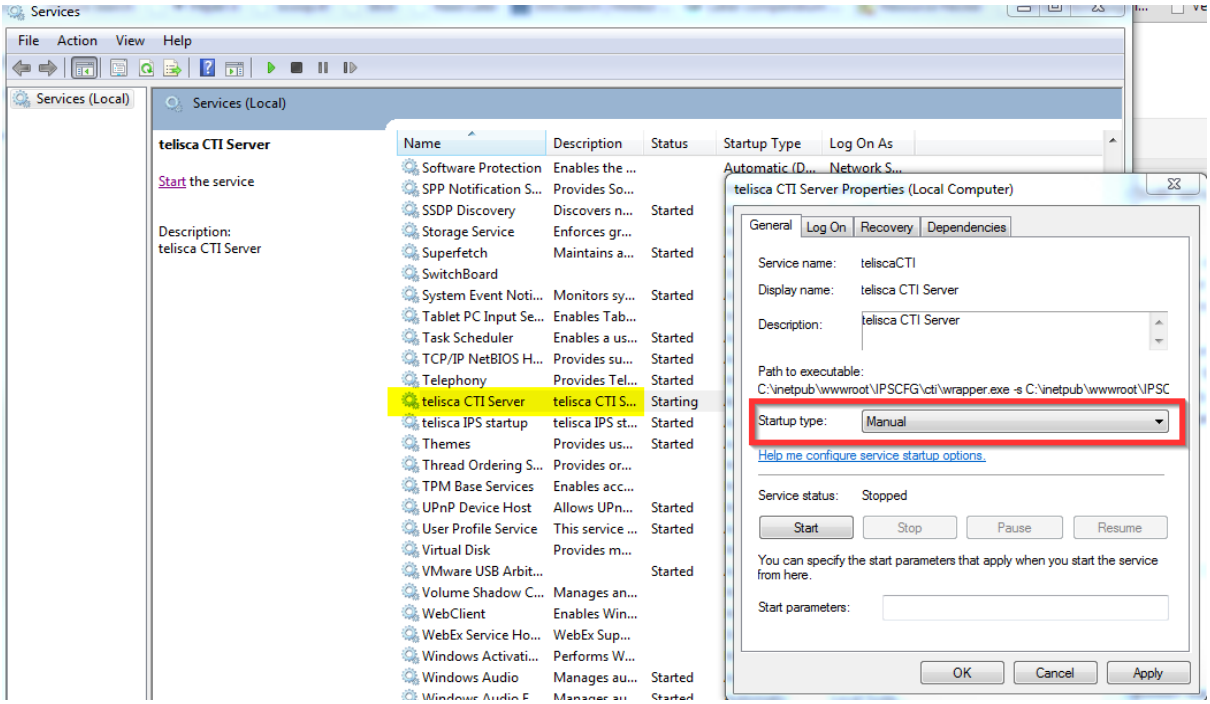
Important: the IP Load Balancer switching parameters must match the number of failures and the heartbeat interval to switch times the interval between heartbeats.

The IP Phones query the servers through the IP load balancer. You must then make sure that the IP Phone Service URL is based on the IP load balancer virtual IP address.

The load balancer should work in failover mode and redirect the queries in priority to the primary server. It will redirect the queries on the backup server after server when several consecutives http queries fail on the primary server.

3.1.3 Change CTI service settings on backup server

Since May 13th 2013, the CTI Server 2.6.0 must stay on Automatic start and be running, so no need to follow the steps below.



3.2 Deploying the Fault Tolerance

The Phone services definitions for the telisca applications must use the VIP address and the subscriptions must be updated.

3.3 Considerations

The Hot Standby module mainly ensures correct switchover of the telisca CTI server. Depending on which application uses this module some things need to be remembered:

- Special User accounts need to be created/configured identically on both servers
- Same for special folder permissions
- Replicated directories of IPS Global Directory are copied to each telisca server independently according to their import schedule, when testing, manual import of the directories should be ran if they haven't been automatically imported yet

4 Enabling Windows Network Load Balancer

Here are the instructions to set up Windows Network Load Balancing in a basic configuration. The Network Load Balancing (NLB) feature is installed on two servers, we use the NLB Manager on the primary to configure it, the settings are replicated on the secondary automatically. Then the active telisca server can be accessed using this new virtual IP from the phones and the web interfaces. The Balancing is done when the Primary server returns an error to the Windows Load Balancer.

4.1 Network prerequisites

1. Switch (layers 2 and 3): must accept the multicast packets.
2. Router: ARP association between IPv4 virtual unicast address and physical multicast mac address.
3. Both servers must be on the same subnet with the same vlan configuration.

4.2 Network cards considerations

NLB does not require more than one network card per host. However, there are several scenarios in which a user may prefer to add another network card:

- **Inter-host communication in unicast mode**
  In unicast mode, each host in the cluster has the same IP Address and the same MAC Address making them look identical from a networking perspective. So, unicast mode has the side effect of disabling communication among the hosts of the cluster.
- **Separating the front-end traffic from the back-end traffic**
  The network adapter that has NLB bound to it can be used to handle incoming connections and connections to a bank-end database, for example, can be made from a separate back-end network adapter.

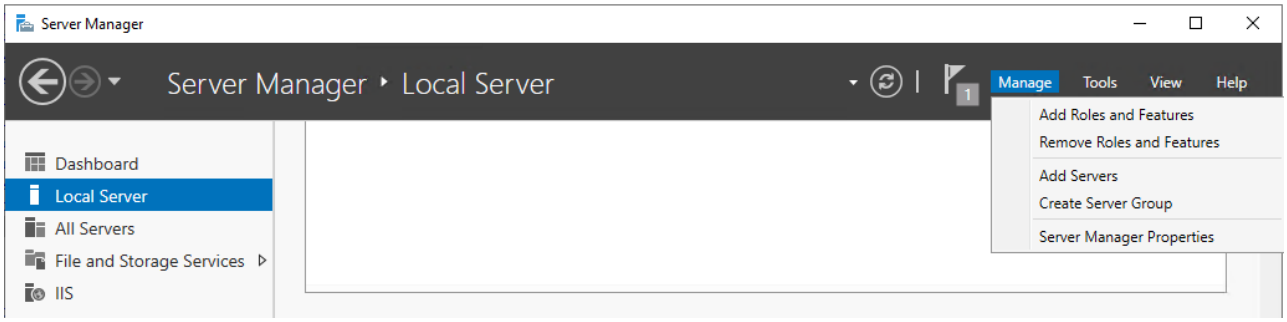
Further information can be found here: http://technet.microsoft.com/en-us/library/cc783135(v=ws.10).aspx

The steps below are made with a unique network card.

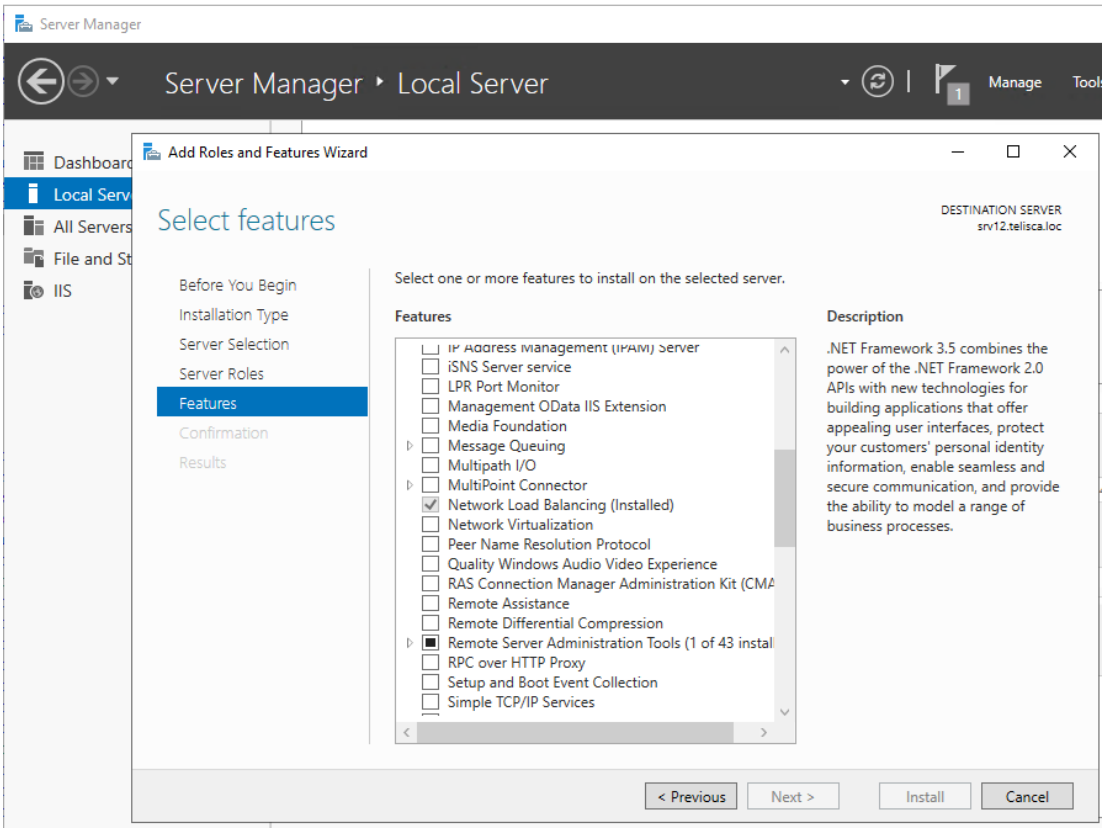
4.3 Activate the Network Load Balancer feature

The feature needs to be enabled on both servers: primary and backup.

Launch 'Server Manager', select Local Server, Manage and add Roles and Features.



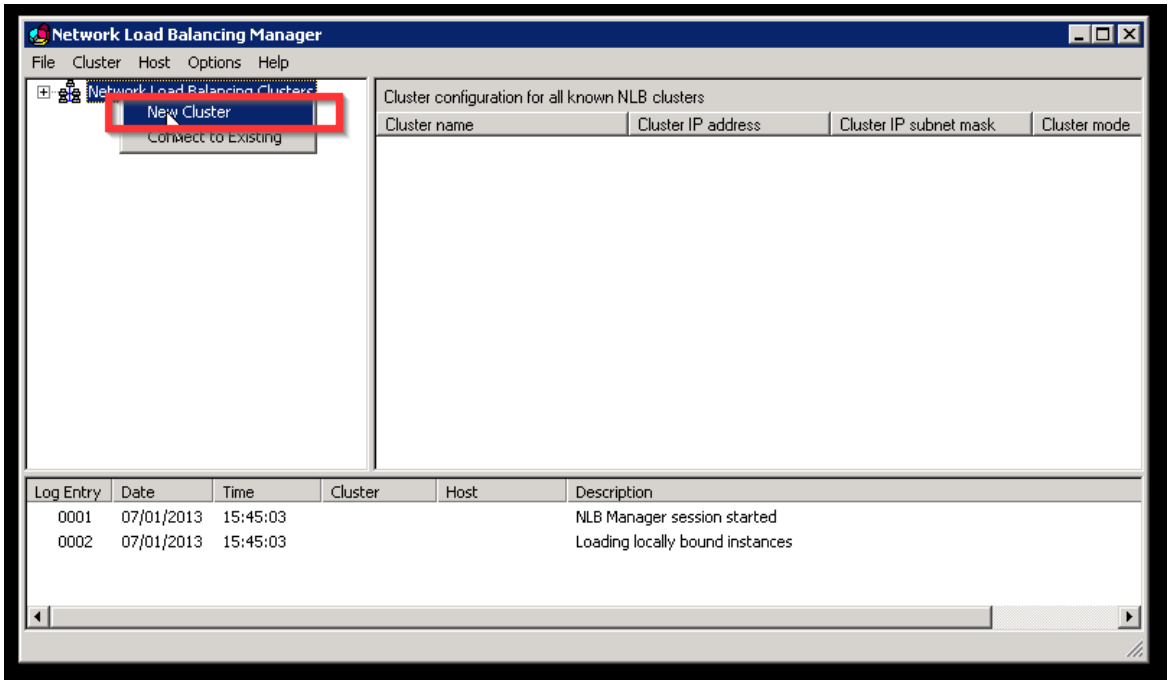
Then click on Features and select Network Load Balancing.



4.4 Cluster creation

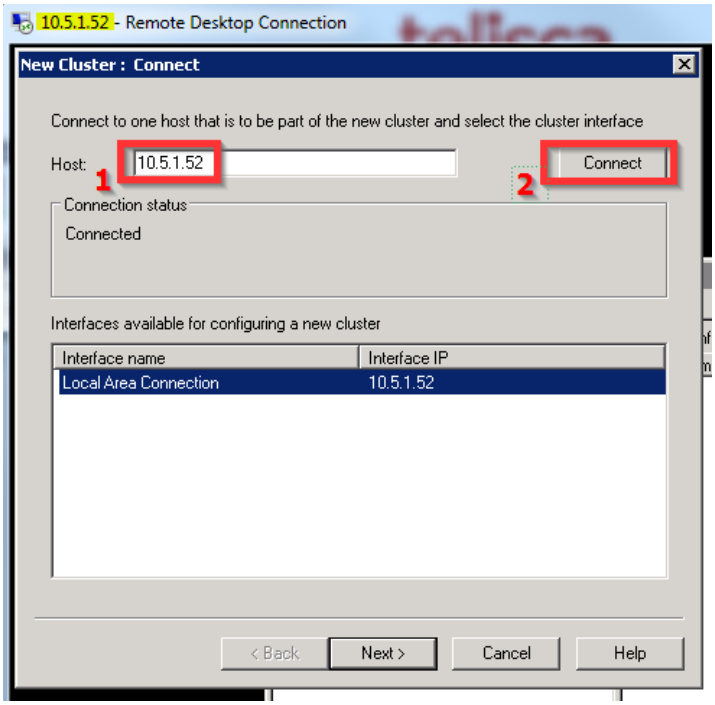
It is recommended to use a local admin account whenever prompted, even if the servers are on a domain (unless you know what you are doing).

Launch nlbmgr from the command prompt. Then Right-click "Create new cluster"

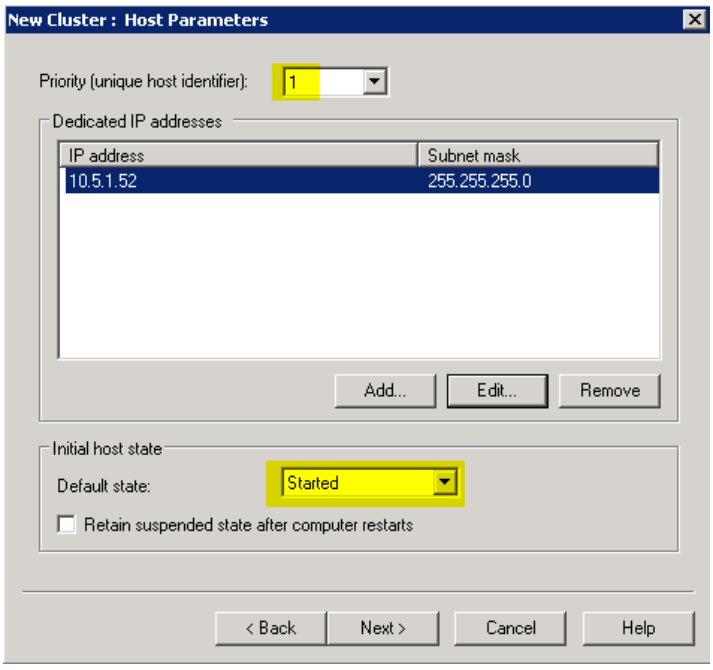


4.4.1 Connect first host

You enter the IP Address of the primary telisca server.



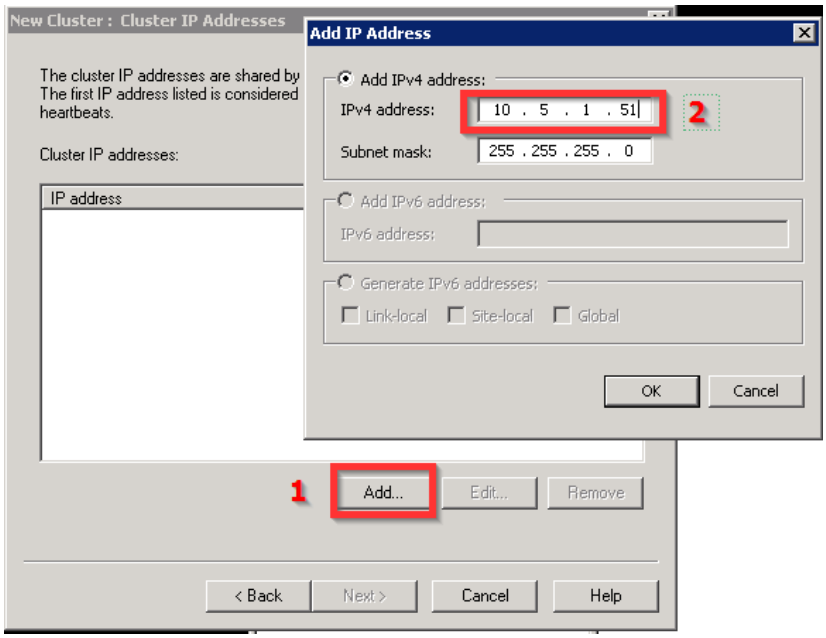
4.4.2 Set the parameters of the primary



4.4.3 Parameters of the virtual IP address.

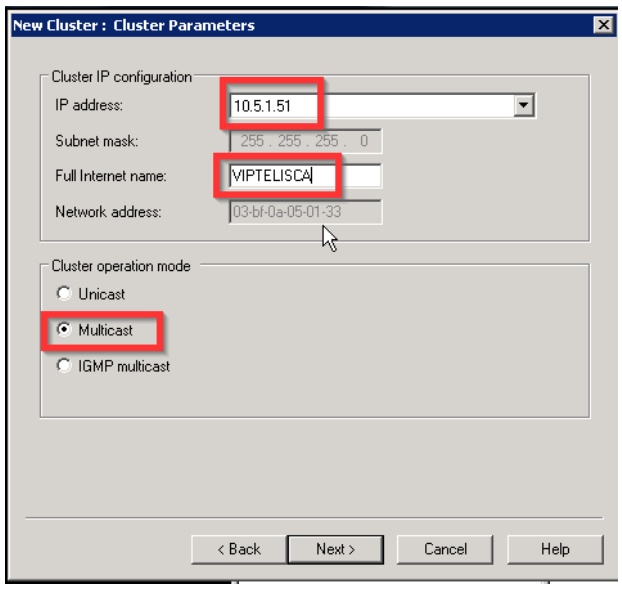
In this example the IPs are:

- Primary 10.5.1.52
- Secondary 10.5.1.53
- VIP 10.5.1.51



4.4.4 Cluster parameters

The Full Internet Name is up to you.
Multicast Cluster Operation Mode must be selected.



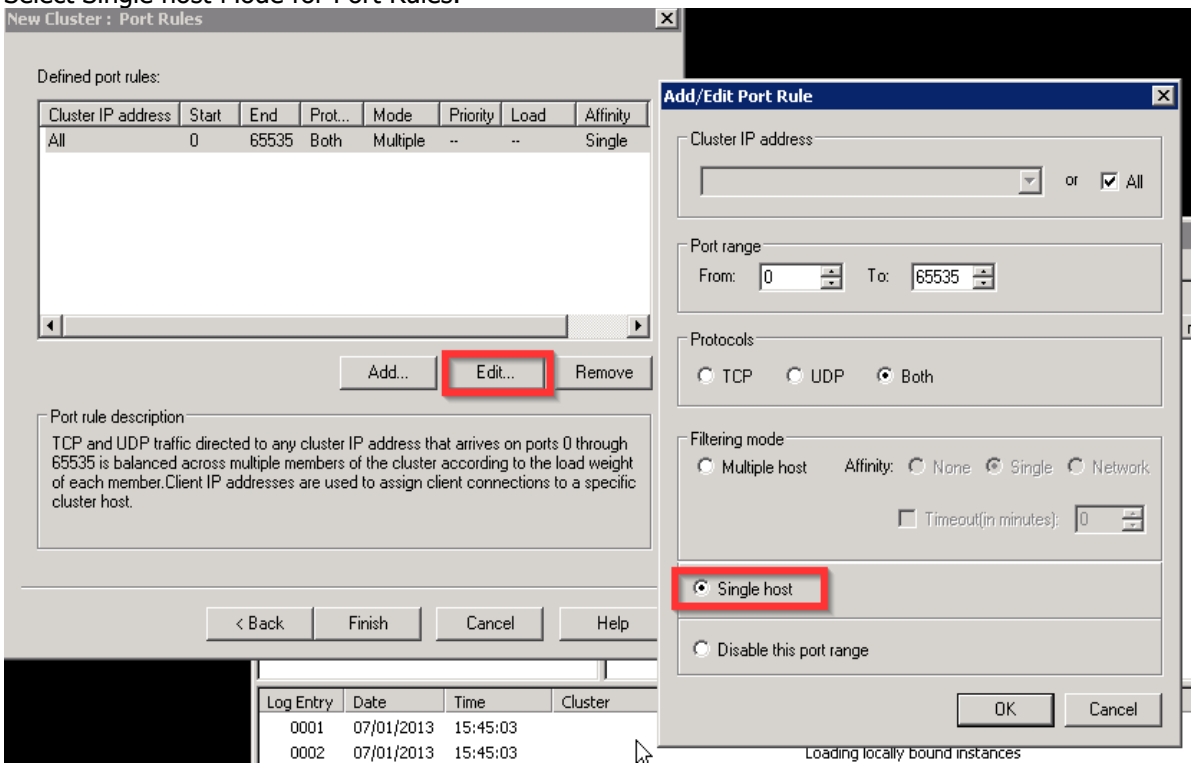
Important: some routers may prevent the use of an IPv4 address for Multicast. A special config must be added to the router.

- In this case, arp entry must be configured on vlan gateway of telisca server, to be able to reach VIP from network other than telisca server subnet.
- This arp entry will match unicast IPv4 VIP address (10.5.1.51) and multicast mac-address (03-bf-0a-05-01-33).

Important: Full Internet name must be provided even if it is not used explicitly anywhere.

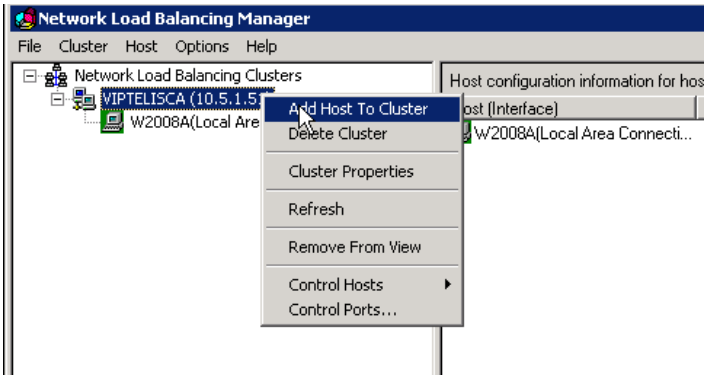
4.4.5 Cluster Port Rules

Select Single host Mode for Port Rules.

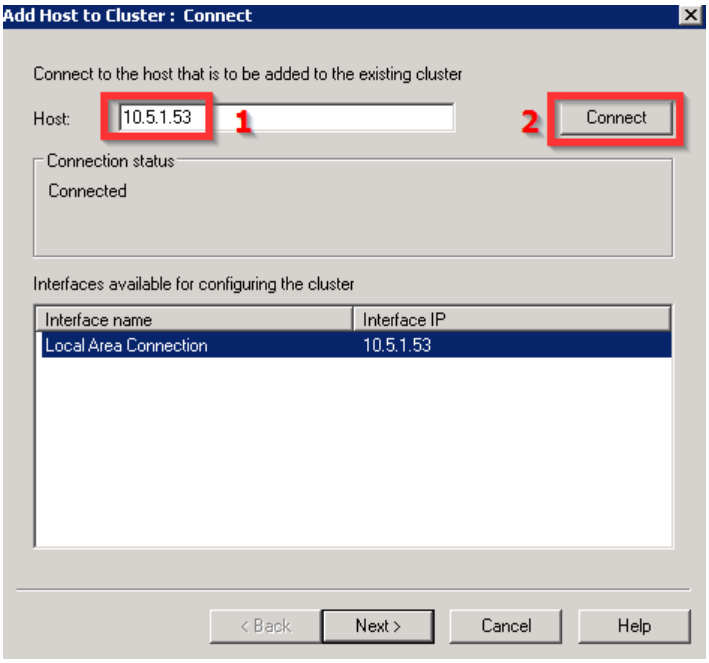


Click Finish.

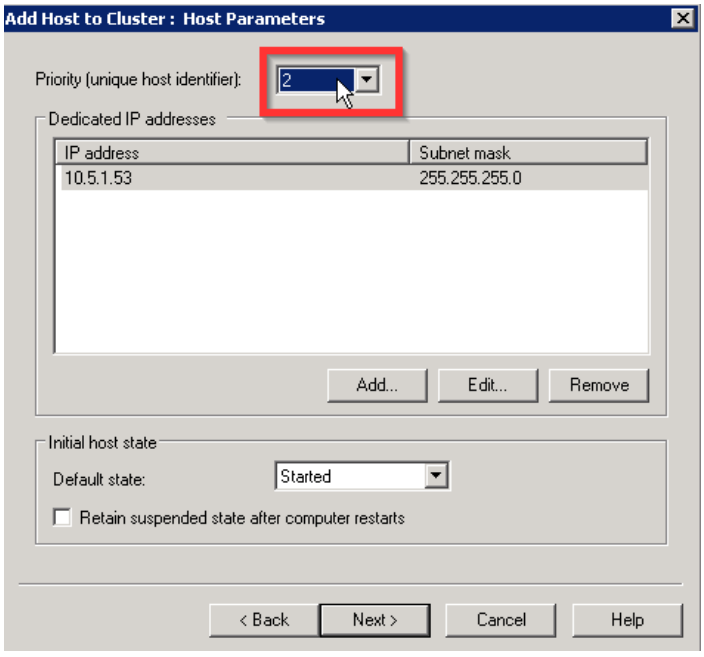
4.5 Adding the secondary Node



4.5.1 Enter secondary server IP



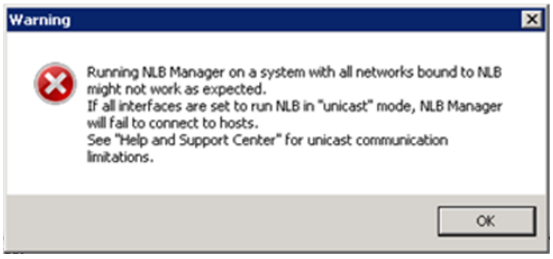
Set priority 2



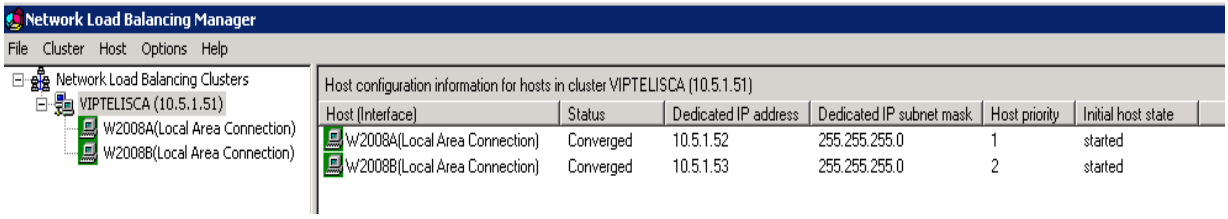
Next and Finish.

4.5.2 Warning message

When configuring NLB, you may see this message. It is only a Warning that tells that you need to support multicast to operate NLB, which is the case with the instructions provided above.



4.5.3 Final result



Important: reboot! it is recommended to reboot the whole server after NLB has been installed, you can experience very slow performance on IIS and network access to the backup server if not restarted.

Further information on how to setup a network load balancer on Windows
http://www.techotopia.com/index.php/Building_a_Windows_Server_2008_R2_Network_Load_Balancing_Cluster

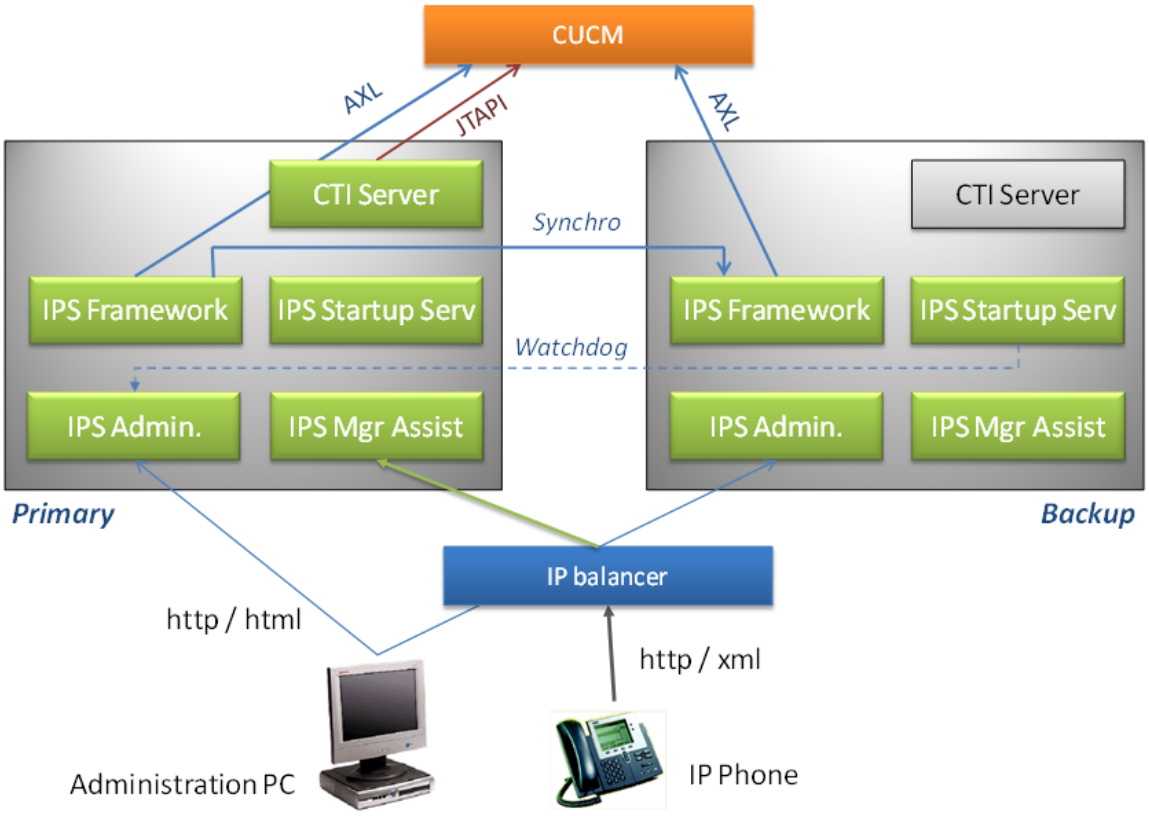
4.6 Verifying the setup

We can now test if we access the server using the VIP. Go to:
`http://[VIP]/IPSCFG/admin`

And check the Hot Standby Tab, it should display the local IP for the active/primary server.

5 Configuring external Load Balancer

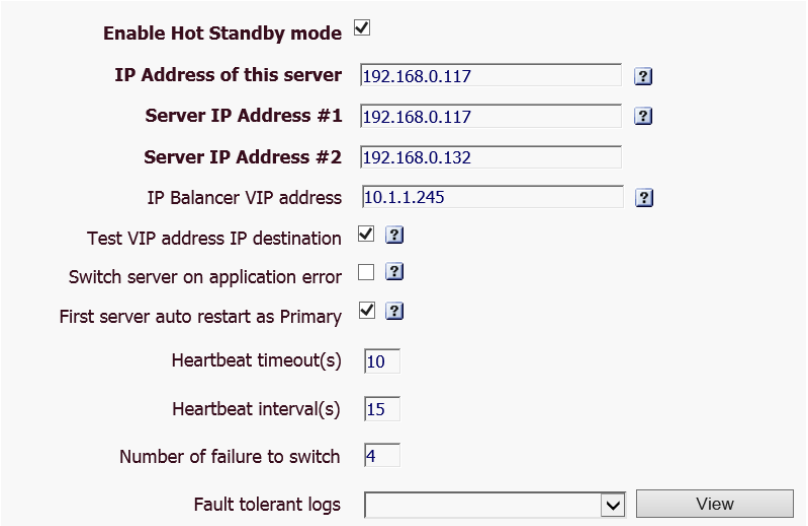
If the company has an external load balancer solution available, it may be used with telisca's applications. It must be configured in Failover mode.



The Load Balancer should check periodically that the primary server is available (ping) and eventually answer to a monitor http URL without http error. After several consecutives errors the Load Balancer must redirect all http requests to the backup server.

In the meantime, the IPS Startup Service on the backup server periodically checks the primary server monitor URL as well. After the same number of consecutive errors he forces the primary server to switch to idle mode and the backup server to run mode.

When a server restarts and finds out that the other server is already running, it synchronizes its configuration from the first one and run in idle mode. When both servers are restarting at the same time (auto restart during night) a mechanism forces the primary the server to run and the backup to idle.



5.1 Test VIP address destination

If a the Virtual IP (VIP) address of the Load Balancer is defined, and the parameter 'Test VIP address IP destination' is checked, the application will use it to check on which server the load balancer is redirecting the http requests. Then IPS Startup will switch the active server to match the one that receives the requests. This option may be useful to accelerate the failover in some case. However, if the load balancer switch to often, it will make telisca application switch to often as well.

5.2 Failover parameters

It is important that the Load Balancer and IPS Startup behave the same on failover, so the different parameters set in IPS Administration should be in balanced with Load Balancer's parameters.

If 'Switch server on application error' is set to false, the load balancer will check the server is running fine by calling the URL : http://host/IPSCFG/admin/monitor.aspx?mod=NONE (replace by https, if SSL is enabled) . It should then test that there is an answer (no timeout) and no http error (HTTP 200).

If 'Switch server on application error' has been enabled in Hot Standby configuration, the load balancer will check the server is running fine by calling the URL:
http://host/IPSCFG/admin/monitor.aspx?mod=IPSWS (replace by https, if SSL is enabled). The Load Balancer should then check that the value returned is IPSWS=OK.

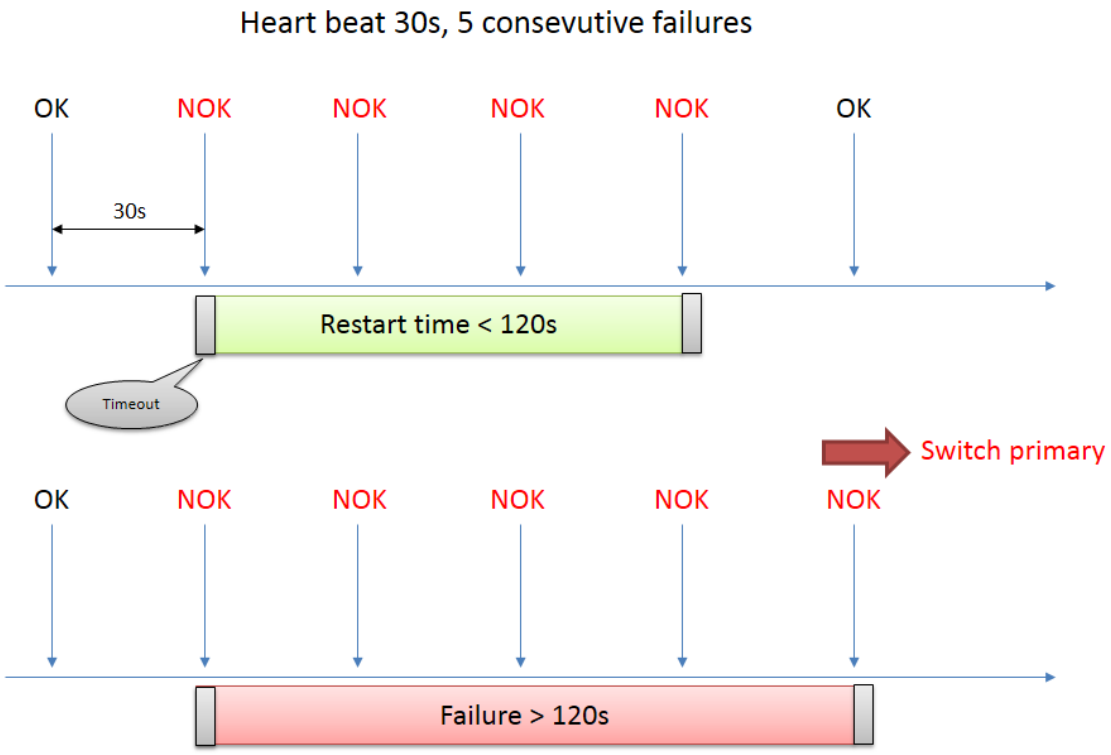
If the load balancer checks which server is ACTIVE, it should call the URL:
http://host/IPSCFG/admin/monitor.aspx?mod=FT (replace by https, if SSL is enabled). It should then test that there is an answer is ACTIVE and not STANDBY.

5.2.1 First server auto restarts as primary

If the load balancer after a failover continues to check the primary server and redirect the http request again as soon as it answers correctly to several consecutive request, the you must check the option 'First server auto restarts as primary' in IPS Administration. You should also make sure that the rollback condition (number of consecutive errors, heart beat) are set the same as for the failover.

5.2.2 Heart beat interval and consecutive failures

The Load Balancer and IPS Startup service must switch in approximately the same delay so that the application is already up and running on the backup server when the Load Balancer redirects the http request to the backup server. The minimum time to switch to backup server is 60s as it takes usually 60s to start IPS Framework & Administration. However it may take more time depending of server or VM sizing and other VM installed on the same server . It also leaves the time to CUCM's CTI Manager to unregister the CTI Route Points and CTI Ports used by the application (around 60s). Depending of the time the check is initiated it is necessary to make sure that the time to make n-1 consecutive checks is longer than the maximum time it takes for the server to restart. Default values are 5 consecutives check every 30s.


Heart beat 30s, 5 consevutive failures
OK
NOK
NOK
NOK
NOK
OK
30s
Restart time < 120s
Timeout
Switch primary
OK
NOK
NOK
NOK
NOK
NOK
Failure > 120s

6 Appendix

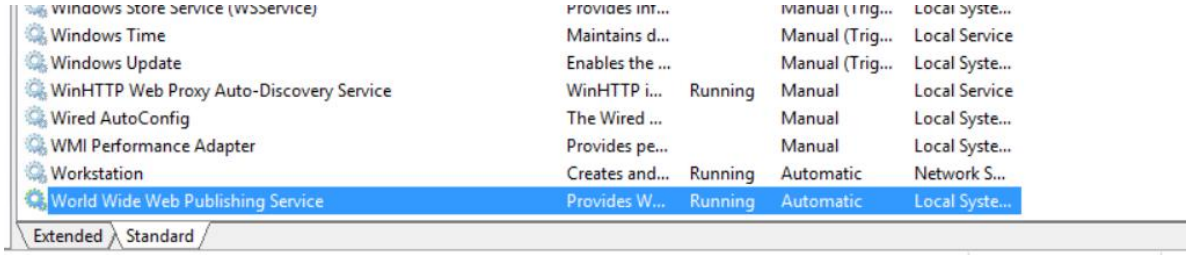
6.1 Test Hot Standby process

This chapter describes a process to test the Hot Standby mechanism. How to simulate the primary server failover, so that the backup server switch from standby mode to active mode.

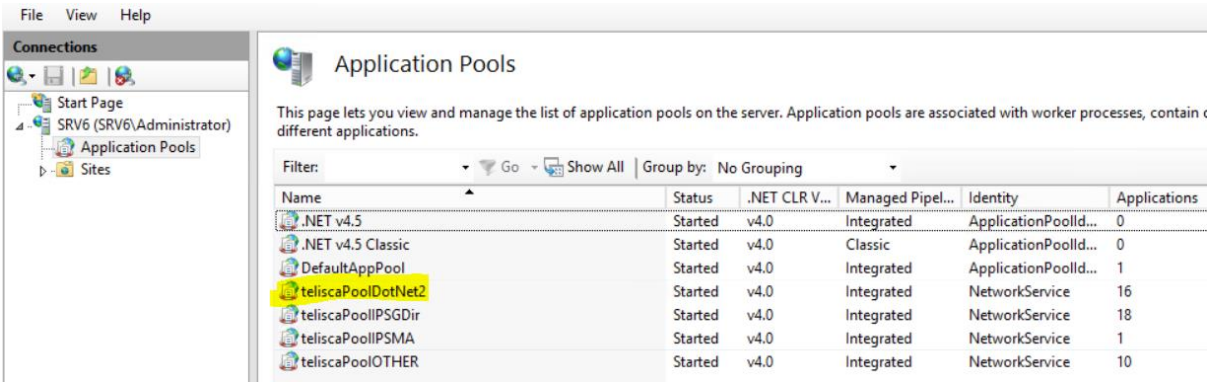
6.1.1 Stopping the primary server

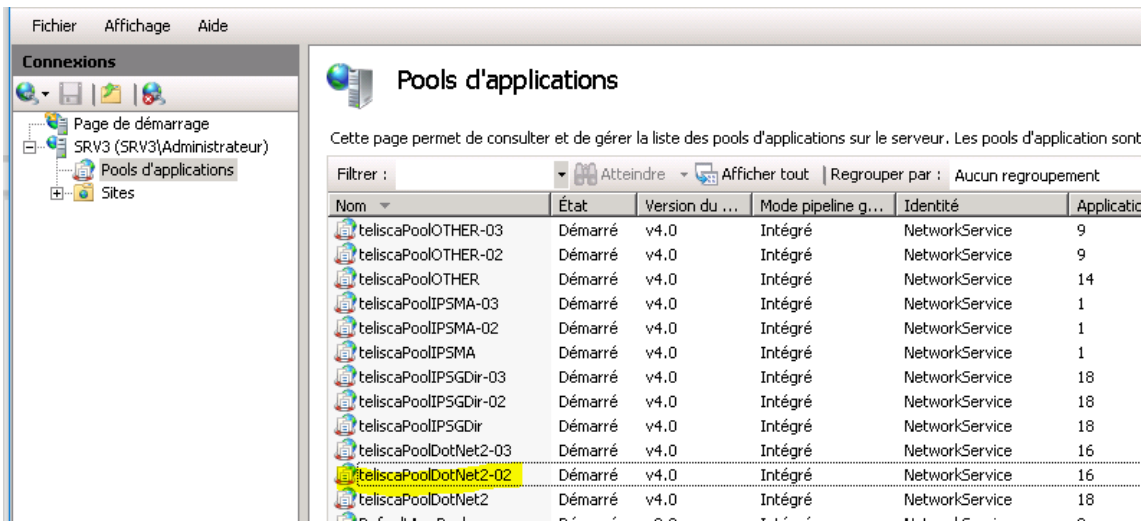
In order to simulate the primary server, you can do different things:

- Stop the Virtual Machine on which the telisca applications is running. This can be done from the Virtual Machine administration, for example from VSphere (for VmWare).
- OR Disconnect the Virtual Machine from the network. This can be done from the Virtual Machine administration, for example from VSphere (for VmWare).
- OR Stop IIS Server. This can be done by stopping the Windows Service 'World Wide Web Publishing Service'.



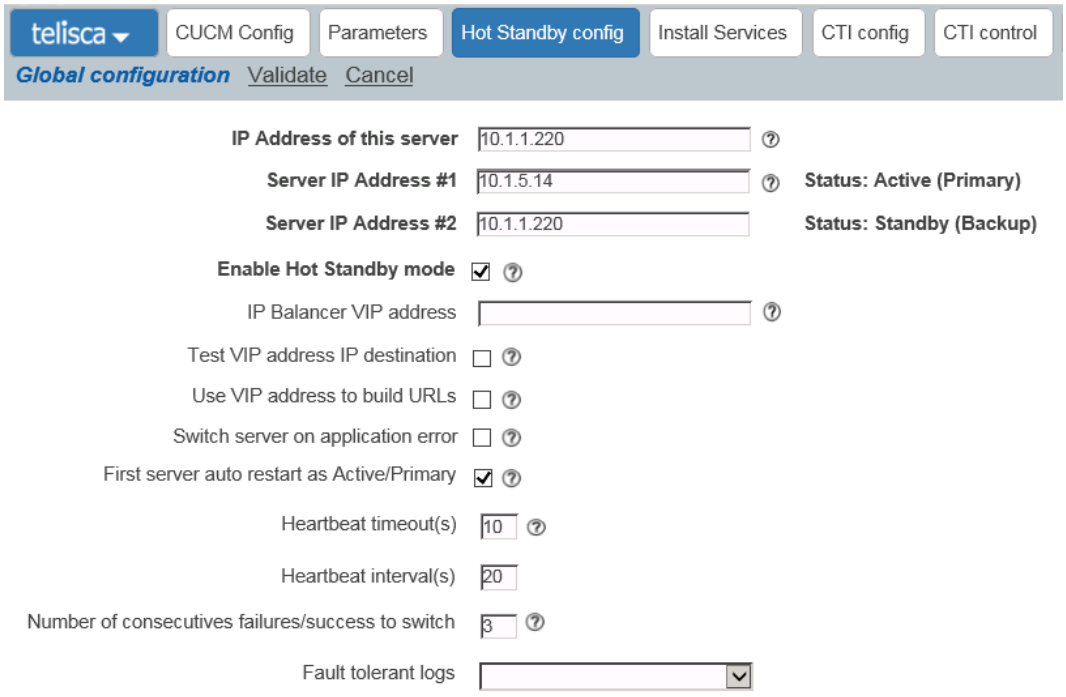
If several telisca instances are installed on the same server (same Virtual Machine), you may want to stop one instance, without stopping the other one. In this case, you should only stop the main telisca IIS Application Pool from IIS Administration. The application pool name is teliscaPoolDotNet2 for the default instance and teliscaPoolDotNet-XXX where XXX is the instance prefix (by default 02) for the additional instances.



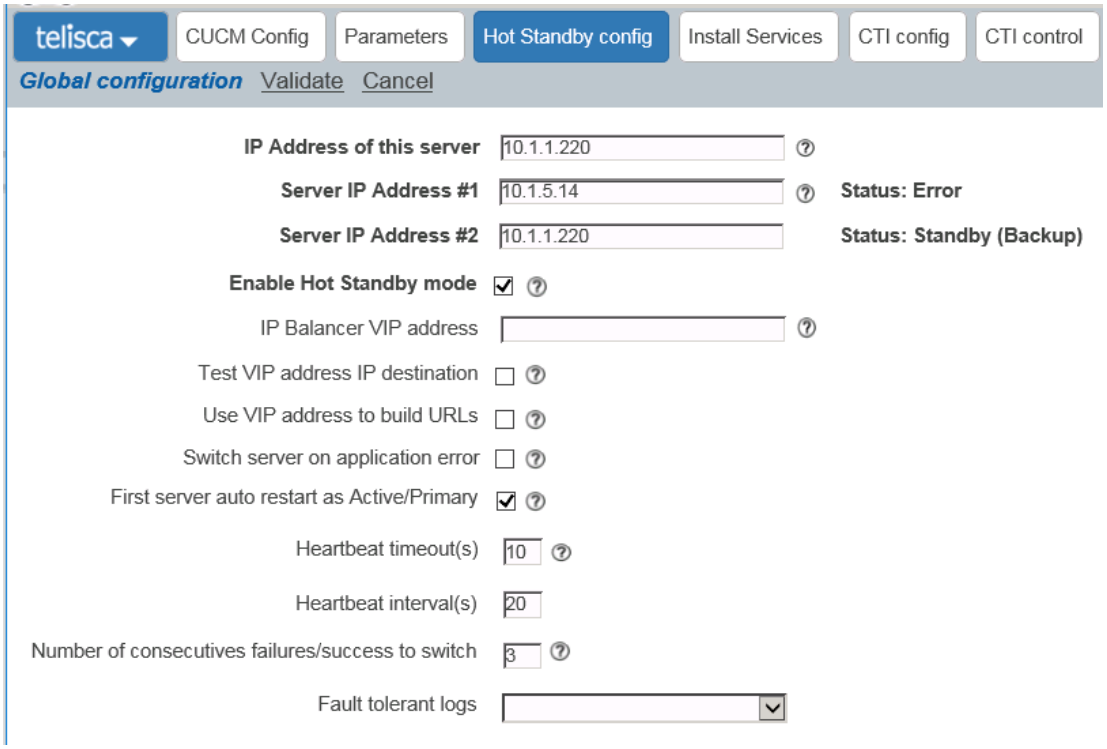


6.1.2 Check the switchover process

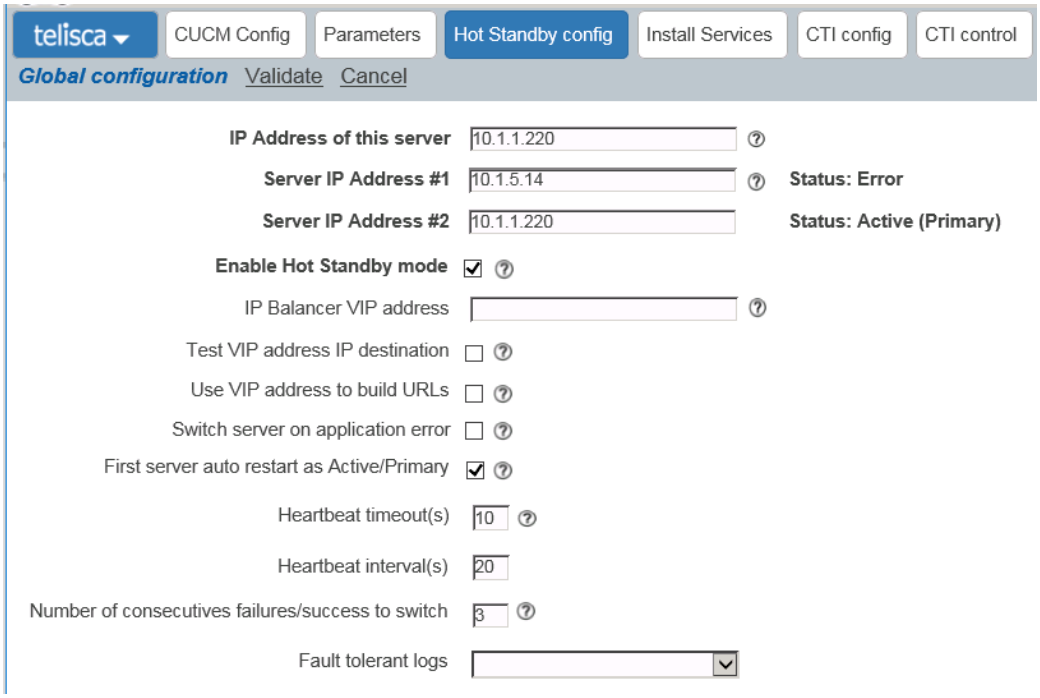
Connect to the backup server administration (http://backup-host/IPSCFG/admin), then select Global Configuration menu and Hot Standby Config tab. In the normal mode, the primary server (IP Address #1) is Active and the backup server (IP Address #2) is Standby.



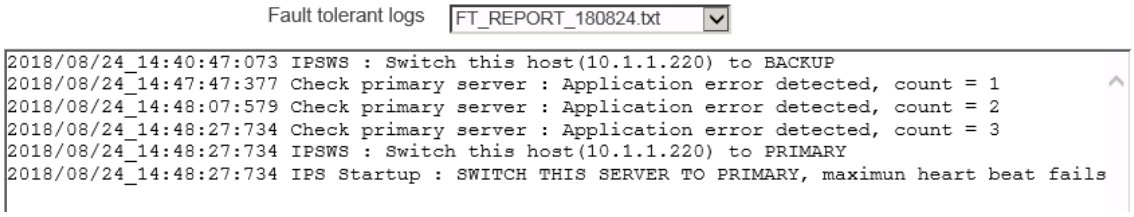
After stopping the primary server, temporarily, the primary server status is error and backup server still standby mode. You need to refresh the page by clicking on Hot Standby config tab.



After 2 or 3 minutes, the backup server switch to Active mode. You need to refresh the page by clicking on Hot Standby config tab.

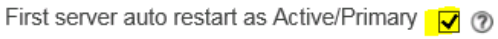


You can check the hot standby process by selecting the last Fault tolerant log

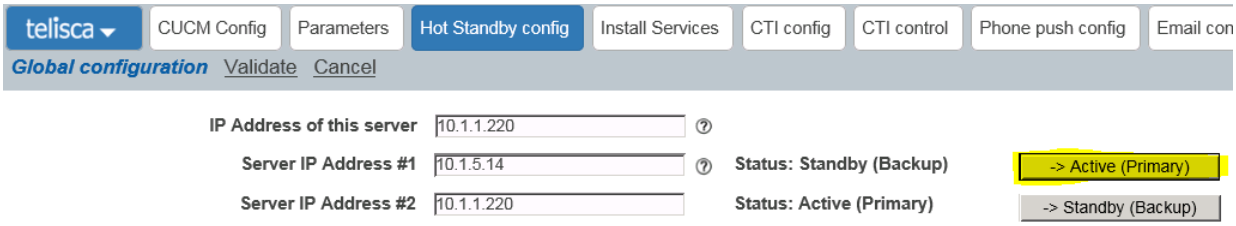


6.1.3 Rollback to primary server

If the parameter 'First server auto restart as Active/Primary' is checked, the rollback will be automatic as soon as the primary server is available again (by undoing the stopping operation chosen).



Otherwise, after restarting the primary server, you can switch back to the primary server manually, by clicking on the '-> Active (Primary)' button in front of Server IP Address #1.



6.2 Expected delay in failover

The expected time for the backup server with CTI to become active depends on these parameters:

The expected delay can be calculated by adding the failover delay (the number of failures x heart beat frequency) + heartbeat timeout + CTI startup delay.

CTI startup delay is the time it takes to the CTI Server to:

- Connect to CTI Manager (normally a few seconds)
- Register and CTI Monitor CTI Ports (200ms/CTI port or more, depending of treatments)
- CTI monitor devices (200ms / device).

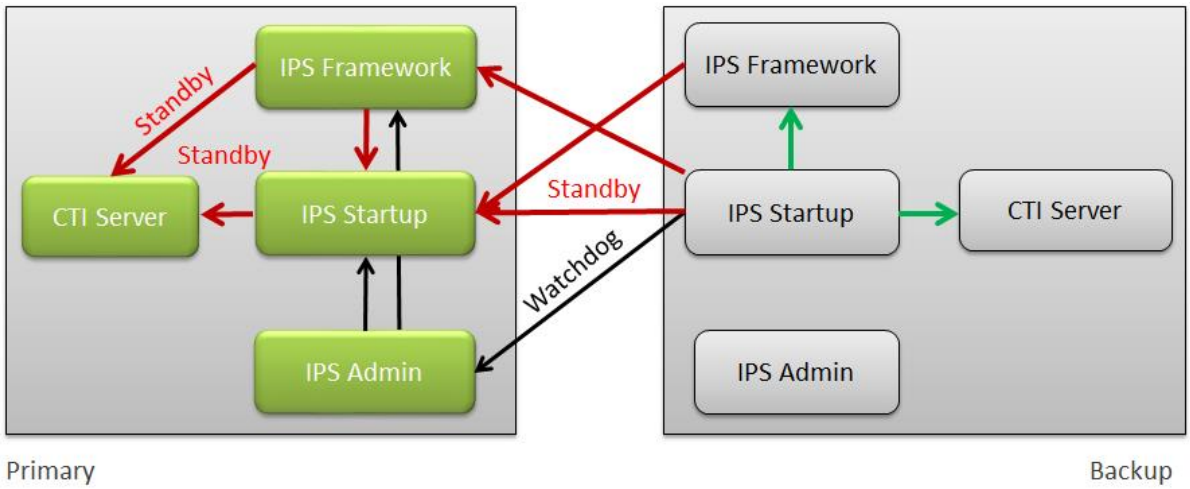
During failover delay, the applications may not be available or work in a specific safe mode. See applications' Administrators Guide for more information.

6.3 Detailed failover/rollback algorithm

The following chapters describe the algorithm used to decide which server is active and which server is standby. It depends of the different settings defined in Hot Standby configuration screen.

6.3.1 Failover initiated by alternate server on active server failure

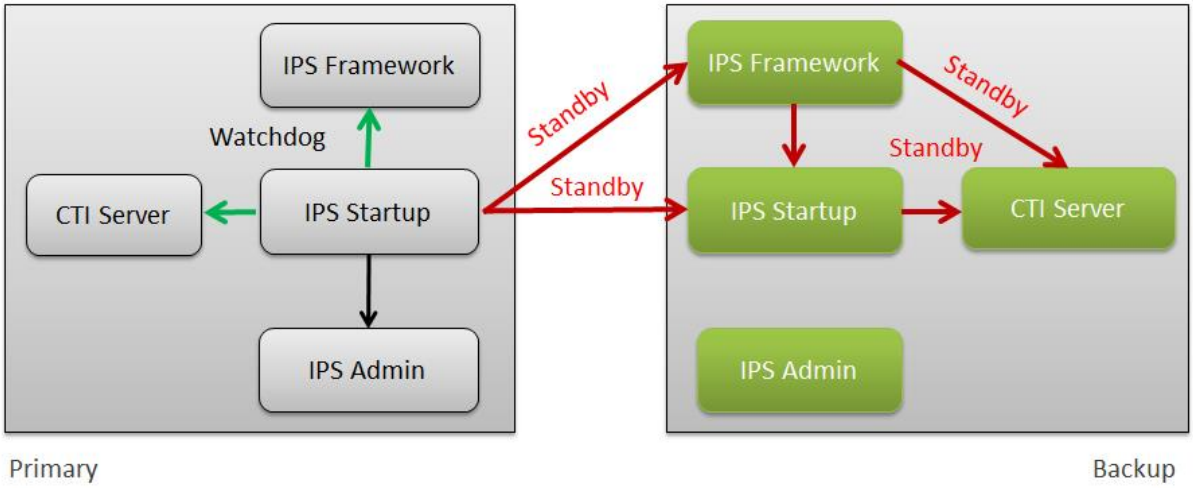
This failover is initiated by IPS Startup Service from the standby server which send periodically http or https requests to http://host/IPSCFG/admin/monitor.aspx?mod=IPSWS to test the health of the active server. In normal mode standby server is the backup server, but can it be the primary server after having switched to standby.



If the server is in Standby mode (not active)
If the server is connected to the LAN
If the defined consecutive failures[1] has been reached
Switch local IPS Startup to Active
Switch local IPS Framework to Active
Switch local CTI Server to Active
Switch remote IPS Framework to Standby
Switch remote IPS Startup to Standby
Switch remote CTI Server to Standby

6.3.2 Rollback initiated by primary server

Rollback is initiated by IPS Startup Service from primary server which send periodically http or https requests to http://host/IPSCFG/admin/monitor.aspx?mod=IPSWS to test the status of the backup.

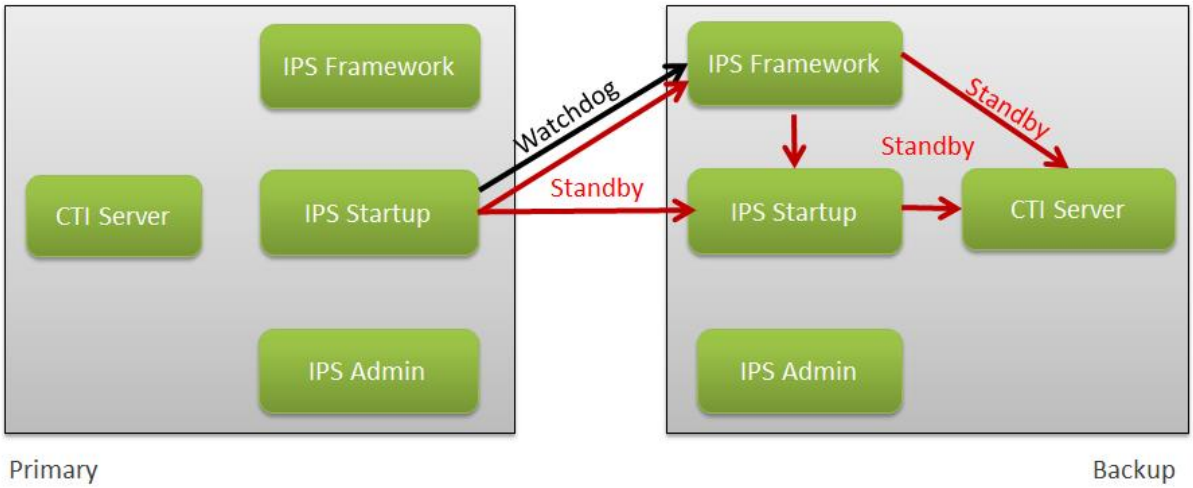


If First Server auto start in Active mode' option is set
If primary server is in standby mode
If local server answers OK to monitor.aspx URL, the consecutive number of times defined or IPS Framework has successfully started
If VIP address send request to this IP address
or if option test VIP destination is not set
Switch local IPS Startup to Active
Switch local IPS Framework to Active
Switch local CTI Server to Active
Switch remote IPS Framework to Standby
Switch remote IPS Startup to Standby
Switch remote CTI Server to Standby

6.3.3 Avoid two servers are in Active mode

It may happen after the LAN between the two servers has been disconnected that the two servers are active at the same time. The algorithm will decide which server should switch to standby taking into account if it is primary server or backup and to which server the VIP address send the requests.

[1] If Switch Server on application error is enabled, then it check the answer returned is IPSWS=OK otherwise it just check an http response is returned.



If the local server is in Active mode
If the local server is connected to the LAN
If other server is Active the consecutive number of times defined
If the local server is backup server
Switch local IPS Startup to Standby
Switch local IPS Framework to Standby
Switch local CTI Server to Standby
else
If VIP address send request to this IP address
or option test VIP destination not set
Switch remote IPS Startup to Standby
Switch remote IPS Framework to Standby
Switch remote CTI Server to Standby
else
If the other server is Active two times the number of consecutive failure/success defined
Switch local IPS Startup to Standby
Switch local IPS Framework to Standby
Switch local CTI Server to Standby

6.3.4 Active server disconnected from the LAN

If the Active Server is disconnected from the LAN the number of consecutive failure times defined
Switch local IPS Startup to Standby
Switch local IPS Framework to Standby
Switch local CTI Server to Standby

6.3.4.1 CTI Server watchdog

IPS Startup is also checking that the local CTI Server is up and running. If it stopped, it restarts the CTI Server service.

IPS Startup also stop and restarts the CTI Server at 'CTI Server Restart time' (visible in Global Config menu, Parameters folder) and calculated automatically a few minutes after IIS Application Pool has shut down and IPS Framework and Administration has been restarted by IPS Startup.

6.3.4.2 Load Balancer / Hot Standby synchro

The goal is to have the Load Balancer failover and the Hot Standby failover almost at the same time to reduce the time when requests are sent to the server in standby or in error mode. So the number of consecutive ticks with failure, the ticks period and the ticks timeout should be the set the same on Load Balancer and in Hot Standby configuration. However as the tics of the Load Balancer and the hot standby are not synchronized they

may be out of synchro for the duration of a tick period. It may also happen that the requests from the Load Balancer and Hot Standby does not answer the same. There is two mechanism to minimize this situation:
- The server may send an http request to the VIP address and wait for the request to hit the server, which means that the load balancer send the requests to this server. This will accelerate the switch of the hot standby module.
- There is special feature for IPS Manager Assistant. It IPSMA user interface receives an http request when the server is in standby mode, the requests are redirected to the active server.