

---

**Communications Server z/OS V1R5 and V1R6 Technical Update**

---

# **Sysplex Load Balancing for z/OS**

© Copyright International Business Machines Corporation 2004. All rights reserved.

---



© Copyright International Business Machines Corporation 2004. All rights reserved.



## Sysplex Load Balancing:

- ✓ Load balancing into a z/OS Sysplex using Cisco technologies
- ✓ Sysplex Distributor combined with Cisco MNLB forwarding agents
- ✓ Content Switching Module (CSM) in a Cisco Catalyst switch
- ✓ Content Switching Services (CSS)
- ✓ New z/OS external load balancer advisor technology

---

# Load Balancing into a z/OS Sysplex using IBM and Cisco technologies

Copyright International Business Machines Corporation 2004. All rights reserved.

---



### **Best Practices for IP Workload Distribution in an IBM zSeries Server Environment**

#### **An IBM and Cisco Interoperability Study of z/OS Sysplex Distributor with Cisco Multi Node Load Balancing (MNLB), Cisco Content Services Switch (CSS), and Cisco Content Switching Module (CSM)**

➤ The scope of the interoperability test and white paper is:

- Describe the design principles and best practices implemented to achieve high availability in a zSeries server environment with Cisco networking equipment, without compromising future scalability paths.
- Demonstrate the ability to distribute a variety of workloads to various server programs residing on zSeries, using z/OS Sysplex Distributor, Cisco Content Services Switch (CSS), and Cisco Content Switching Modules (CSM) in a Catalyst 6500 environment.
- Describe a variety of failure scenarios and how the cluster and network design reacts to these failures

## Test scenarios

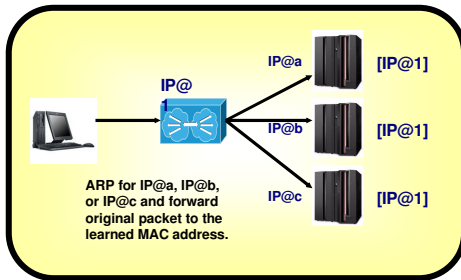


- Sysplex Distributor with MNLB forwarding agents
  
- Catalyst 6509 with CSM blade
  - ƒ Server NAT with Policy Based Routing
  - ƒ Server NAT and Client NAT
  
- Catalyst 6509 and CSS switch
  - ƒ Server NAT with Policy Based Routing
  - ƒ Server NAT and Client NAT
  
- Tests also included SSL modules in both CSM and CSS

This test and white paper focus on the connectivity between the Catalyst 6500, CSS, CSM, and the zSeries server using OSA-Express components.

# Sysplex Distributor/MNLB using dispatch mode forwarding - basics

e-business



➤ The server cluster IP address (IP@1) is from a networking point of view owned by the load-balancing node

➤ The load-balancer forwards inbound IP packets by selecting an appropriate next-hop IP address to route the packet to - there will be one configured next-hop IP address per server instance in the cluster

┆ If servers are directly connected that means ARPing for the next-hop address and forwarding the original packet (including its original destination IP address) towards that discovered MAC address

Sysplex Distributor is a dispatch mode technology

┆ A distributed VIPA is transparently replicated to all LPARs in the Sysplex, but only the distributing LPAR advertizes it to the network

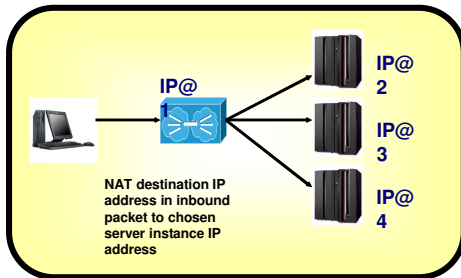
Dispatch mode is also known as MAC-level forwarding.

➤ Only layer-4 load-balancing supported (TCP connections)

➤ If servers are more than one-hop away from the load-balancer, then some encapsulation technology is needed to forward the packets (such as Generic Routing Encapsulation - GRE)

➤ No restrictions on outbound path from server instances (no NATing of either inbound or outbound traffic)

## CSM and CSS using Directed Mode - basics

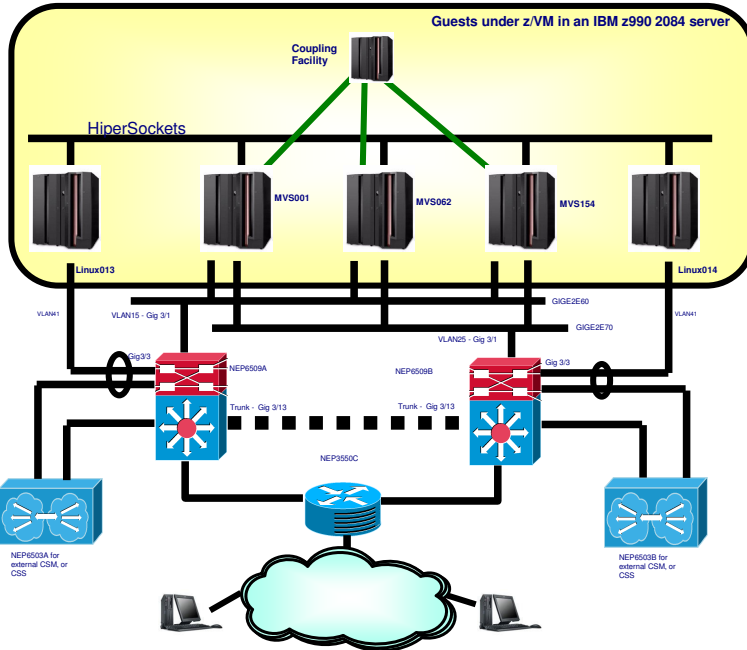


CSM and CSS are almost always configured to operate in directed mode

- They can be configured to operate in dispatch mode, but normally do not operate in that mode with a z/OS Sysplex due to issues with shared OSAs:
  - None of them are able to use GRE, which is required if OSA ports are shared
  - If OSA ports are not shared, then both CSM and CSS can be configured to operate correctly using dispatch mode

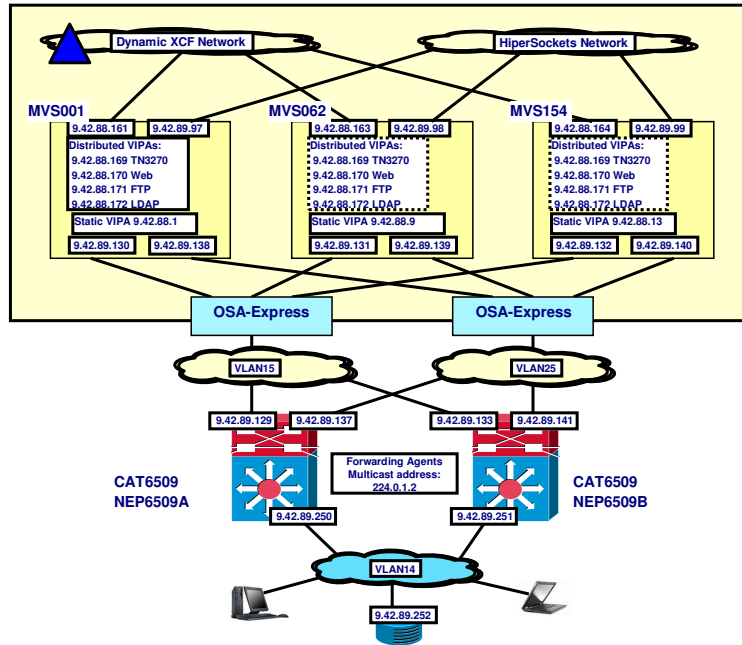
- The server cluster IP address (IP@1) is owned by the load-balancing node
- The load-balancer forwards inbound IP packets by selecting an appropriate server instance specific IP address and changes the destination IP address in the inbound packets to that of the chosen server instance (performs server NAT)
- Outbound packets must be routed back via the load balancer for it to change the server IP address to the cluster IP address
- Layer-4 to layer-7 load-balancing is supported
- No restrictions on network topology between the load balancer and the server nodes

# Equipment overview

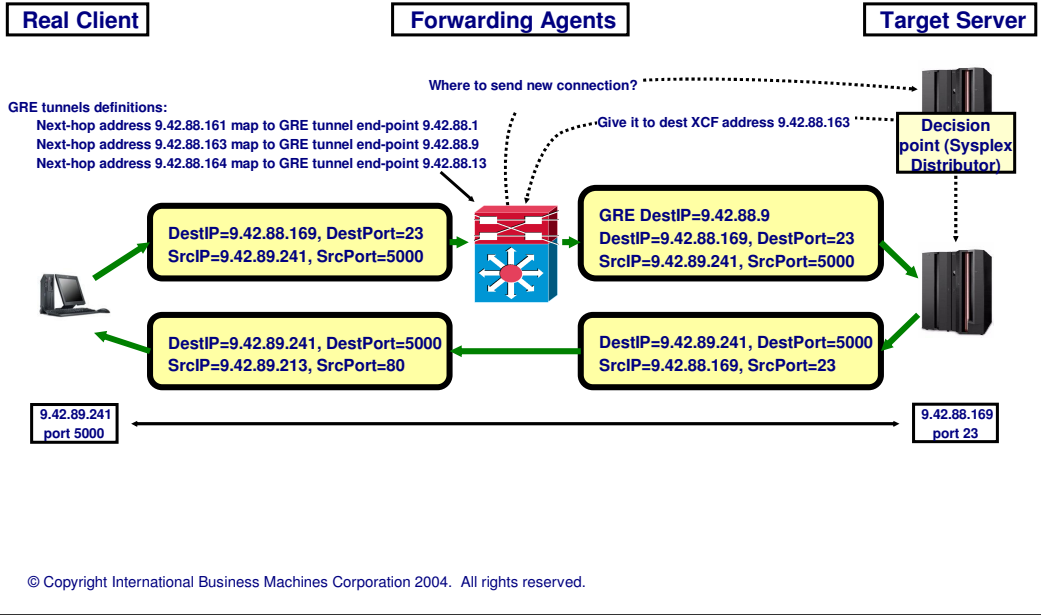




# Sysplex Distributor / MNLB - Dispatch mode Test environment



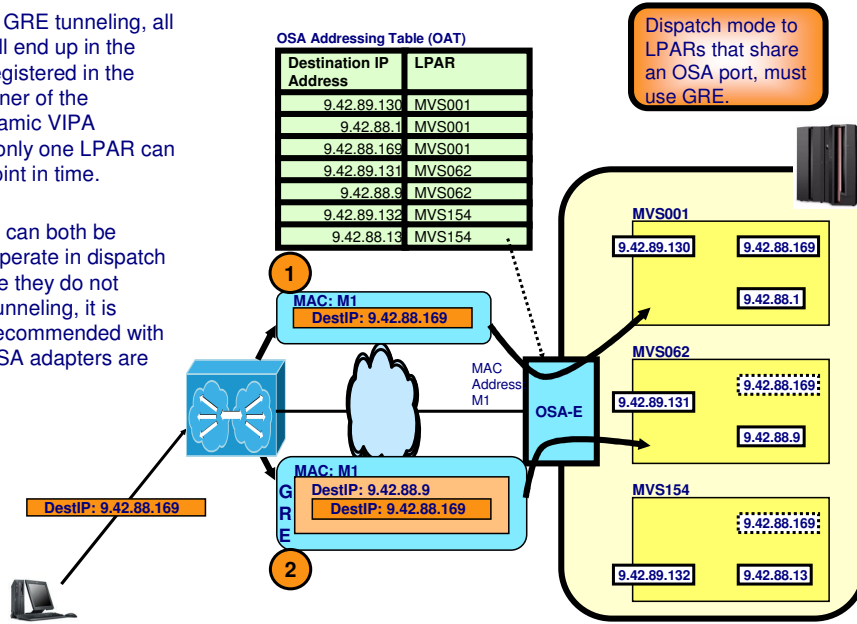
# Sysplex Distributor / MNLB - Dispatch Mode forwarding - flow basics



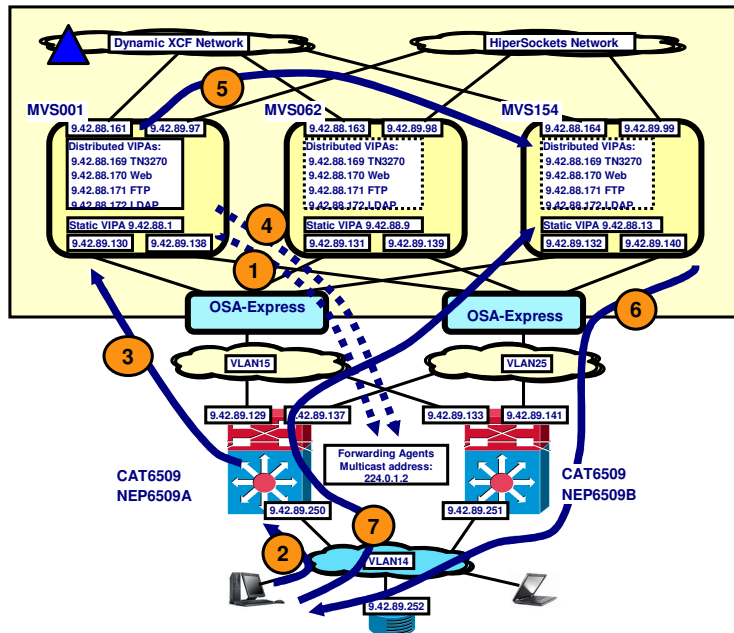
# Sysplex Distributor / MNLB Dispatch Mode use of GRE tunneling

➤ Without use of GRE tunneling, all connections will end up in the LPAR that is registered in the OAT as the owner of the distributed dynamic VIPA address - and only one LPAR can be so at any point in time.

➤ CSM and CSS can both be configured to operate in dispatch mode, but since they do not support GRE tunneling, it is generally not recommended with z/OS unless OSA adapters are non-shared.



# Sysplex Distributor / MNLB flows



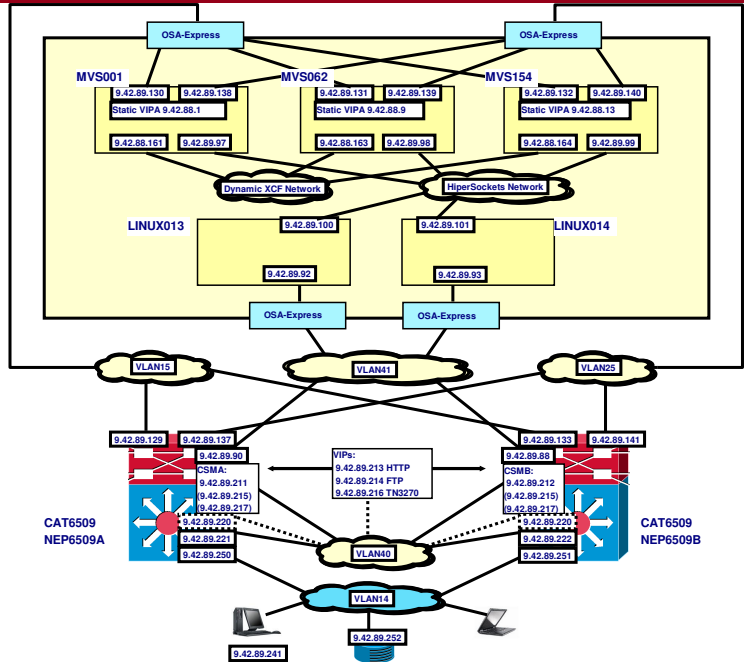
## Sysplex Distributor / MNLB Dispatch Mode forwarding - key points

e-business



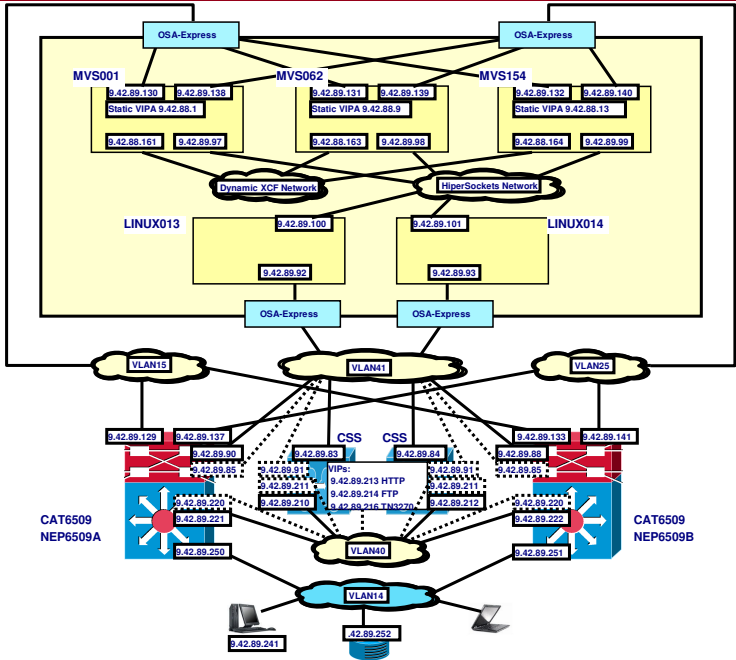
- Is based on Dispatch mode forwarding – also known as MAC-level forwarding
- Load balancing decision point is inside the z/OS Sysplex and can take real-time server availability and LPAR capacity into consideration
- First SYN packet of a new inbound connection is routed via the Sysplex Distributor node and XCF to the target node. Succeeding packets per connection are routed directly from the 6509s to the target nodes
- Server IP address and client IP address are never NATed - there is no requirement for outbound packets to be routed via any specific path
- Dispatch mode forwarding must be combined with use of Generic Routing Encapsulation to overcome one-hop away and shared OSA limitations

# CSM - Directed mode - test environment



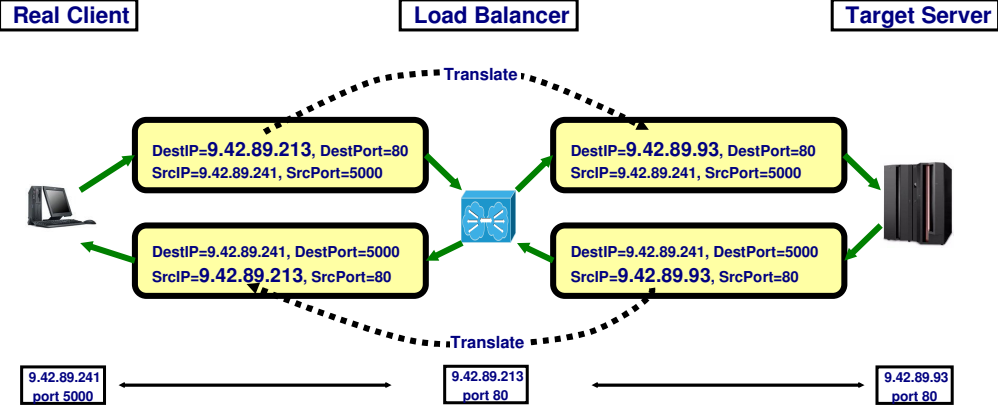
© Copyright International Business Machines Corporation 2004. All rights reserved.

# CSS - Directed mode - test environment



© Copyright International Business Machines Corporation 2004. All rights reserved.

# Cisco CSS/CSM - Directed mode - Server NAT and Policy Based Routing - flow basics





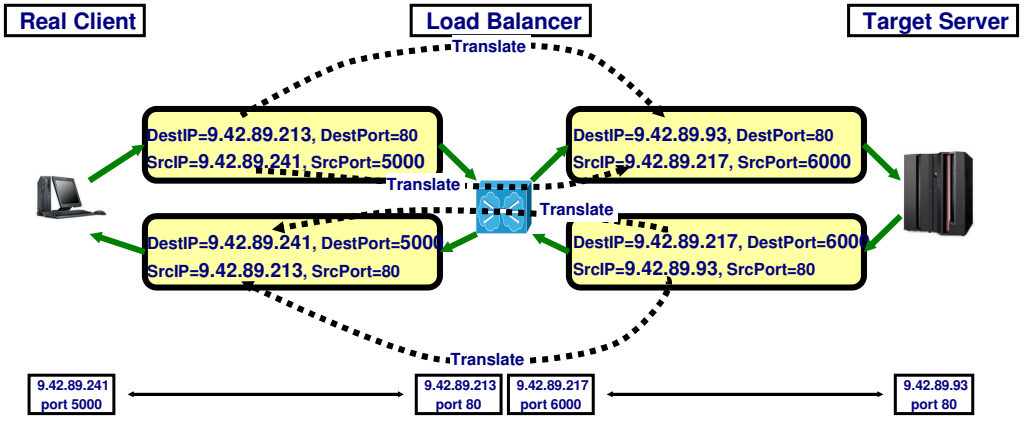
## Cisco CSS/CSM - Directed mode - Server NAT and Policy Based Routing - key points

e-business



- Only the server IP address is NATed (destination IP address on inbound and source IP address on outbound)
- Policy-based routing in routing infrastructure re-directs outbound IP packets from target servers to the load-balancer so it can NAT the source IP address in outbound packets
- Outbound packets that do not need NATing of the source IP address are routed using normal IP routing table processing
- Real client IP address information is available to target servers

# Cisco CSS/CSM - Directed mode - Server NAT and Client NAT - flow basics



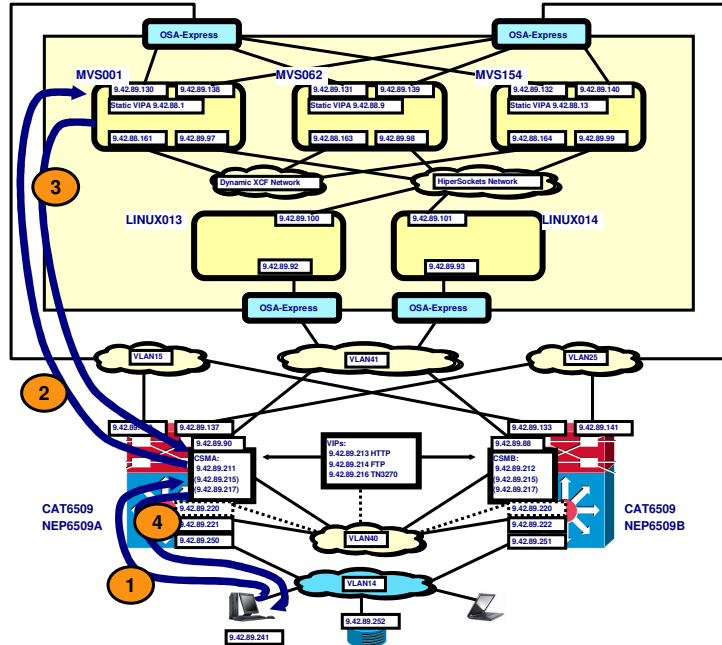
## Cisco CSS/CSM - Directed mode - Server NAT and Client NAT - key points

e-business



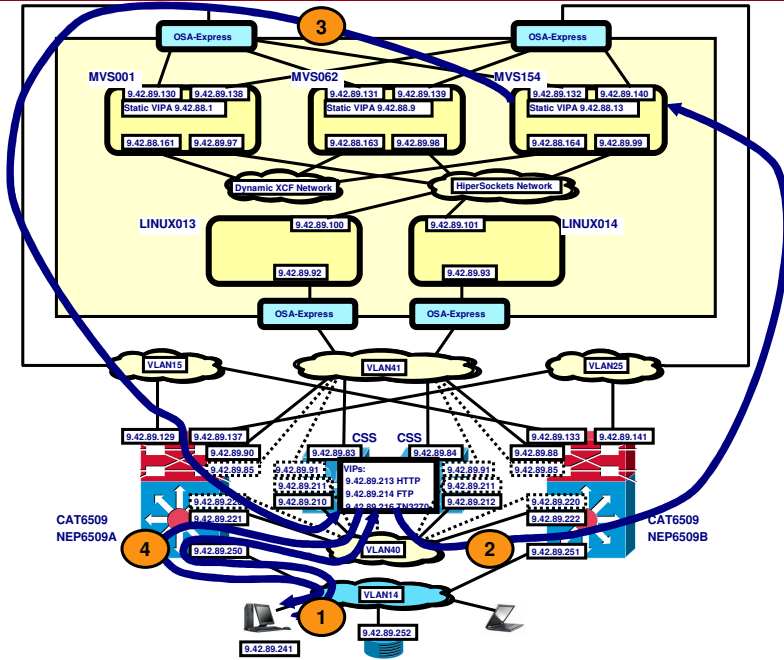
- Both server IP address and client IP address are NATed by the load balancer - no need for use of Policy Based Routing since outbound packets from target servers are destined for a load balancer IP address
- Client IP address seen by target servers is an IP address on the load balancer and not the real client IP address
- Should be used with care if any of the following functions on server nodes are in use:
  - ┆ Networking policy conditions based on client IP address information
  - ┆ NETACCESS rules for access control and/or MLS label assignment
  - ┆ Server configuration options based on source IP address information, such as TN3270 server LU name assignment
- May also complicate diagnosing certain error cases where real client IP address and port number are unknown on the server node

# Cisco CSM flows

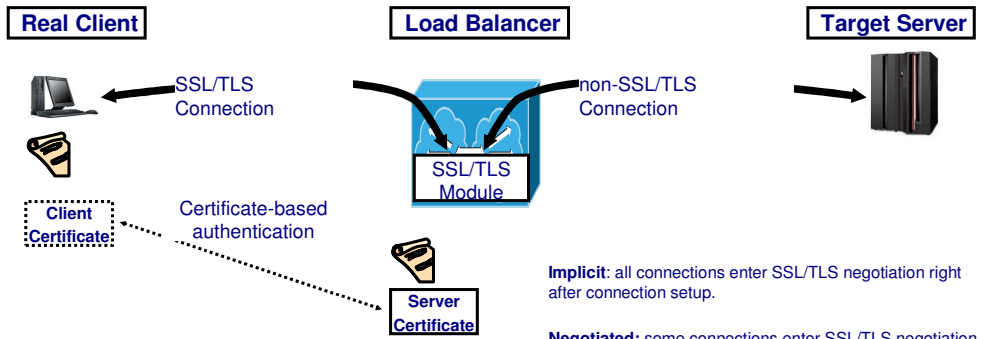


© Copyright International Business Machines Corporation 2004. All rights reserved.

# Cisco CSS flows



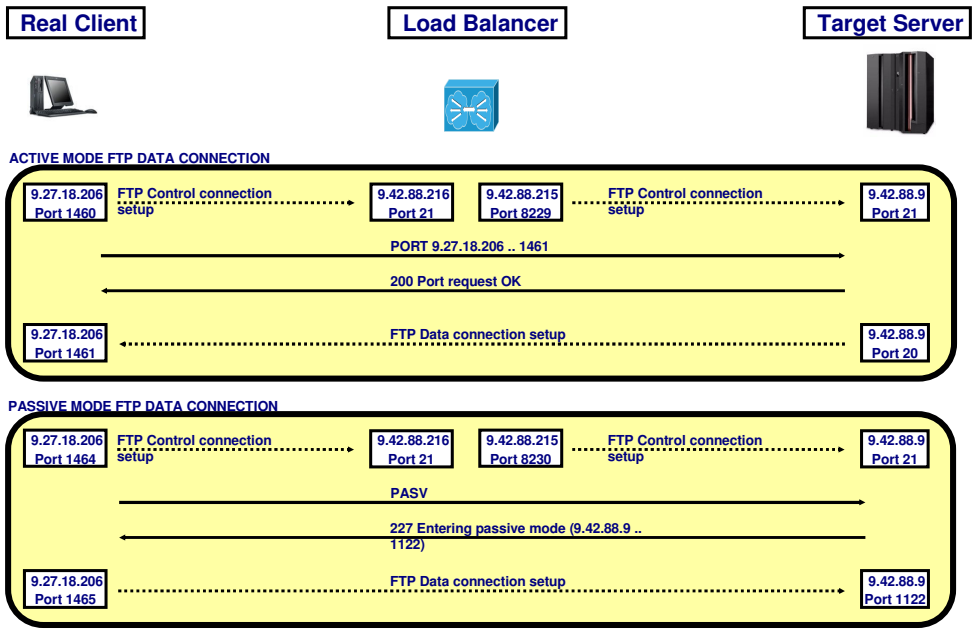
# SSL/TLS offloading to CSS or CSM



**Implicit:** all connections enter SSL/TLS negotiation right after connection setup.  
**Negotiated:** some connections enter SSL/TLS negotiation after an application protocol exchange, others do not.

- Implicit mode SSL/TLS workload can be offloaded (specific SSL/TLS server port number) - negotiated SSL/TLS can in general not be offloaded (shared port number)
  - ⌘ FTP is negotiated
  - ⌘ TN3270 can be either
  - ⌘ HTTPS is implicit
- SSL/TLS processing cannot be offloaded if client authentication is required on target server

# FTP workload CSM/CSS Server NAT and Client NAT



© Copyright International Business Machines Corporation 2004. All rights reserved.

## FTP workload key points

e-business



### ➤ Both CSM and CSS stay out of the loop for FTP data connections

- ‡ Direct data connection between real client IP address and real server IP address
  - The case for both server NAT with PBR and Server NAT with Client NAT
- ‡ SSL/TLS FTP (not the same as SFTP in SSH) with the SSL/TLS end-point on z/OS does work through load balancer
  - The CSM and CSS do not need to investigate the PORT command or the PASV reply
- ‡ The new extended passive mode (EPSV) is not supported when using CSM or CSS in directed mode
  - This could be an issue if SSL/TLS FTP sessions are needed and those sessions traverse a traditional NAT firewall somewhere in the network before the connections hit the CSM/CSS load balancers - such an access path will not work

### ➤ Sysplex Distributor / MNLB works for all combinations of FTP workload

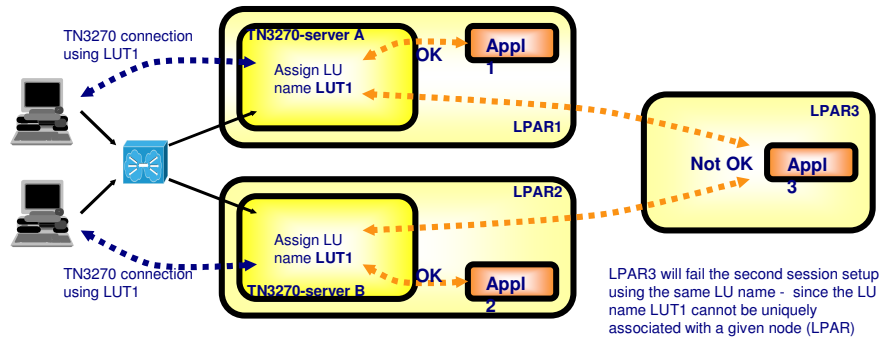
- ‡ non-SSL/TLS active mode, passive mode, and extended passive mode
- ‡ SSL/TLS active mode, passive mode, extended passive mode
  - NB: If SSL/TLS sessions pass through NAT firewalls on the path towards the Sysplex, then extended passive mode must be used
- ‡ Error discovered during testing: all inbound data in passive mode is sent via SD stack – being researched.



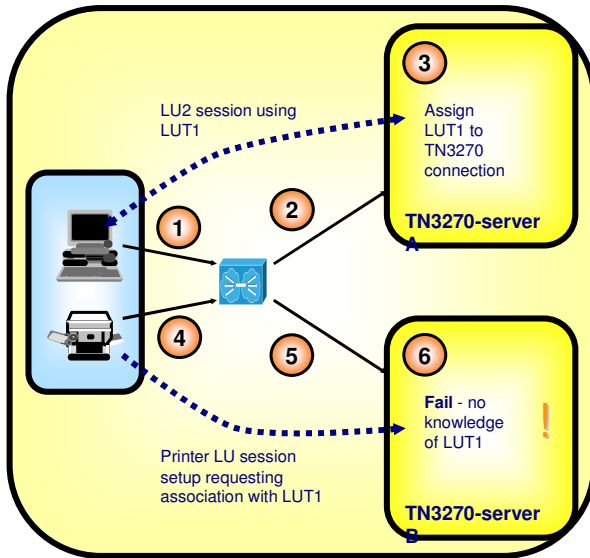
# Assigning LU names when connection balancing TN3270 connections

## Three main ways LU names can be assigned:

- 1 Generic request** - server decides entirely which LU name to assign
  - Each TN3270 server must have its own group of LU names (generic pools)
- 2 Specific request with pool name** - server validates and decides which LU name in the named pool to assign
  - Pool names can be the same in more TN3270 servers, but each server must have its own group of LU names assigned to those pool names
- 3 Specific request with LU name** - server validates and assigns requested LU name
  - Generally TN3270 servers can allow assigning the same LU names since theoretically a client workstation will only request an LU name once



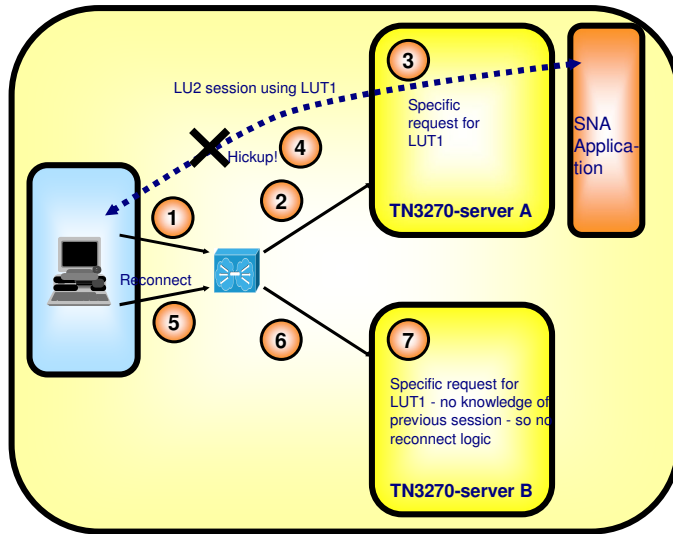
## TN3270 connection balancing and printer association issues



- 1 Initiate connection for establishing an LU2 session
- 2 This connection is sent to TN3270-server A
- 3 TN3270-server A selects an LU2 LU name - for the purpose of this example it doesn't matter if it is a generic request or a specific request. As an example an LU named LUT1 is chosen
- 4 The workstation user now starts a printer emulator and a request for a new connection with the TN3270 server is initiated.
- 5 This connection is sent to TN3270-server B
- 6 The workstation emulator now requests a printer LU name that is associated with the LU2 LU name from the first session LUT1 - and TN3270-server B is going to reject that since it doesn't know anything about the LU name that was assigned by TN3270-server A

**Affinity (or stickyness) needed by Load Balancer to handle this case so all connections from the same client go to the same server instance**

# TN3270 connection balancing and reconnects



- 1 Initiate connection for establishing an LU2 session
- 2 This connection is sent to TN3270-server A
- 3 TN3270-server A allows LU name LUT1 and user connects to SNA application
- 4 Network has hick-ups and client believes connection is broken
- 5 Client sends a new connection (reconnect) request
- 6 Reconnect request is sent to TN3270-server B
- 7 If this connection had ended up in server A, reconnect logic would have kicked in and reconnected the user with the SNA application. But Server B doesn't know anything about the previous session and treats it as a new connection and allows LU name LUT1, which now is active on both LPARs

**Affinity (or stickyness) needed by Load Balancer to handle this case so all connections from the same client go to the same server instance**

## TN3270 workload load balancing key points



- Ensure TN3270 server instances do not assign the same LU names
  - ƒ Separate generic LU name pools
  - ƒ Same LU group names, but with different LU names
  - ƒ Stickyness!
  
- If printer association is used, printer connection request must go to same TN3270 server instance as LU2 connection request was processed by
  - ƒ Stickyness!
  
- For TN3270 server reconnect, reconnect request must go to same TN3270 server instance as original connection request
  - ƒ Stickyness!
  
- If client IP address or hostname is used in mapping rules, ensure client NATing is not used by load balancer

## HTTP(S) workload load balancing key points



- Both CSM and CSS allow load-balancing based on content of the HTTP request
  - ┆ Sysplex Distributor/MNLB does not
- HTTPS connections must be terminated on the load balancer if content-based load balancing is required
  - ┆ SSL/TLS client authentication based on client's digital certificate is not possible on the real HTTP servers
- CSM and CSS can parse all HTTP header fields – including affinity cookies
  - ┆ Can direct connections with affinity to specific server instance

---

# **z/OS Sysplex - external load balancing advisor (LBA)**

Copyright International Business Machines Corporation 2004. All rights reserved.

---



# Sysplex internal vs. external workload balancing - which technology is best for me?

e-business



**Internal (SD)**

- Has realtime information available...
  - more timely capacity information
  - QoS from Service Policy Agent
  - application-independent server availability
- No problems with shared OSA adapters; no intermediate routers
- Uses expensive Sysplex resources for inbound traffic
  - Inbound traffic funneled through single point
  - Routing stack uses zSeries MIPs for inbound routing

**External (ND, Cisco CSS or CSM, etc.)**

- Specialized routing hardware may be more cost-effective
- May be configured for no single point of traffic flow
- So far no general Sysplex-capacity feedback technology (some specialized attempts have been made)
- Requires application-specific health probes ("application ping")
- Problems with shared OSA adapters or other intermediate routers

**External/Internal (SD with Cisco MNLB forwarding agents)**

- Combines the advantages of internal decision making with routing efficiency of external control point
- Shared OSA adapters require use of Generic Routing Encapsulation by the switch.
- Preferred technology especially when workload includes large amounts of inbound data flows.

! Denotes decision point

© 2004 International Business Machines Corporation 2004. All rights reserved.

## Function comparison Load Balancing: DNS, SD, Outboard LBs

e-business



	DNS	Sysplex Distributor	Outboard Load Balancers
<b>When is server instance decision made?</b>	Name resolution	Connection setup (in-line SYN segment)	Connection setup (in-line SYN segment / HTTP GET request arrival)
<b>TCP connections or UDP associations</b>	Both TCP and UDP (at time of name resolution) [UDP: EE as an example]	TCP only	Both TCP and UDP (at time of packet arrival)
<b>Real-time application availability information available</b>	DNS/WLM: Yes otherwise: No (at time of resolution)	Yes	Based on a polling interval. (z/OS LBA: feedback)
<b>LPAR WLM information available</b>	DNS/WLM: Yes otherwise: No (at time of resolution)	Yes	No (z/OS LBA: yes)
<b>Network QoS performance available</b>	No	Yes	No
<b>Extra network flows</b>	Each new contact preceded by DNS flows	Inbound via distributing stack. Virtually none if combined with Cisco MNLB forwarding agents.	No (some minimal extra paths inside the switch environment)
<b>Potential issues</b>	DNS and resolver caches. Clients hardcoding IP addresses.	TCP only - no UDP support	No real time Sysplex information available (z/OS LBA: will be). Need to understand NATing implications.

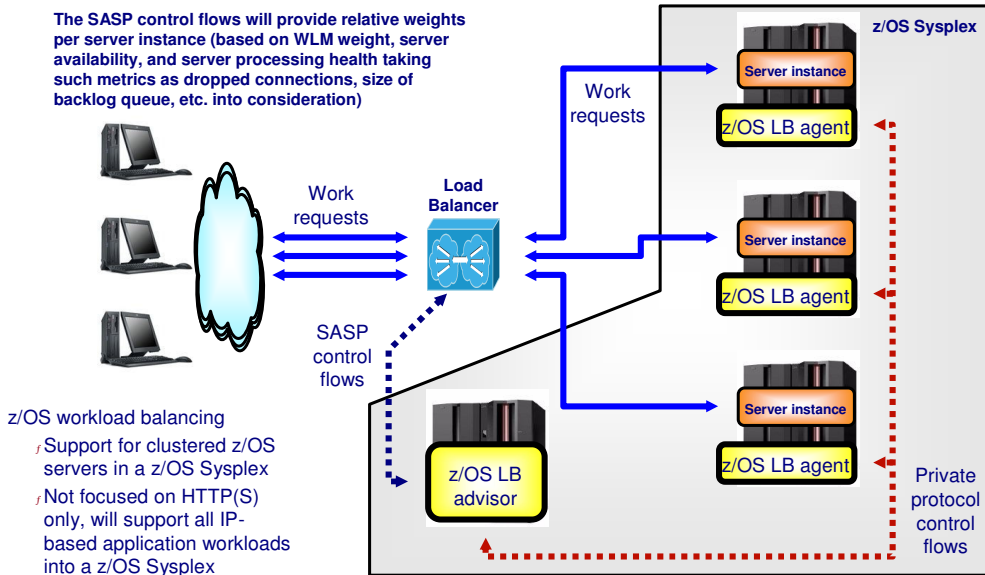
© Copyright International Business Machines Corporation 2004. All rights reserved.



# z/OS Load Balancing Advisor (LBA) for outboard load balancers



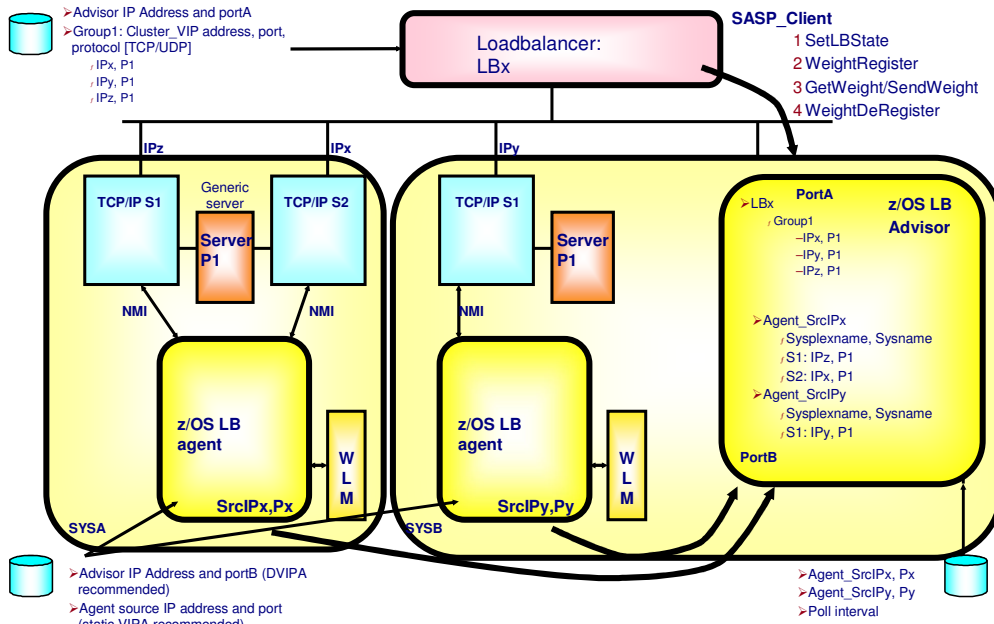
The SASP control flows will provide relative weights per server instance (based on WLM weight, server availability, and server processing health taking such metrics as dropped connections, size of backlog queue, etc. into consideration)



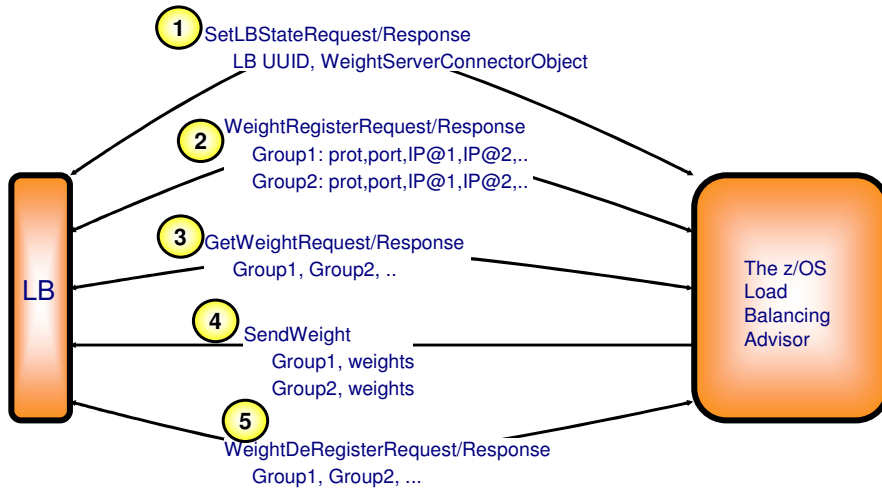
- z/OS workload balancing
- ⌈ Support for clustered z/OS servers in a z/OS Sysplex
  - ⌈ Not focused on HTTP(S) only, will support all IP-based application workloads into a z/OS Sysplex
  - ⌈ Based on Sysplex-wide WLM policy
  - ⌈ Scope is a z/OS Sysplex

The z/OS Load Balancing Advisor technology is a new z/OS Communications Server technology that is planned to be made generally available 4Q2004

# z/OS LB Advisor/Agent structure Overview



# Load Balancer - Load Balancing Advisor interactions - SASP protocol elements



SASP: Server Application State Protocol

## The weights



➤ The weights are composed of two main elements:

┆ WLM weight

- The WLM weight based on displaceable LPAR capacity as we know from other WLM-based load balancing solutions, such as Sysplex Distributor
- ✓ A numeric value between 0 and 64

┆ Communications Server weight

- This weight is calculated based on the availability of the actual server instances (are they up and ready to accept workload) and how well TCP/IP and the individual server instances process the workload that is sent to them.
- ✓ Expressed as a numeric percentage value between 0 and 100
- Purpose of calculations is to:
  - ✓ Prevent stalled server from being sent more work (accepting no new connections and new connections are being dropped due to backlog queue full condition)
  - ✓ Proactively react to server that is getting overloaded (accepting new connections, but size of backlog queue increases over time approaching the max backlog queue size)

➤ The final weight is calculated by combining the WLM and the CS weights into a single metric

┆ Final weight = WLM weight \* CS weight / 100

➤ Due to current external load balancer behavior when a weight of zero is returned, the z/OS LBA currently will never return a zero weight - the lowest weight it will return is a weight of one.

┆ Weights that are returned to the load balancer are normalized to values between 1 and 64

- If all server instances have the same final weight (example 32), then a 1 will be returned for all server instances

## Load balancer registrations



The load balancer may register two types of groups for which it wants weights:

### A A system group

- ⌘ Represented by a list of IP addresses only.
- ⌘ IP addresses are matched to TCP/IP stacks in the Sysplex.
- ⌘ WLM weights for the LPARs are retrieved.
- ⌘ CS weight indicates if IP address is active in the Sysplex or not (0 or 100).
- ⌘ LBA displays will show a protocol value of zero for system group registrations.

### B An application group

- ⌘ Represented by a list of IP address, Protocol (TCP or UDP), and port.
- ⌘ Server address spaces are matched to registrations.
- ⌘ WLM weights for the LPARs are retrieved.
- ⌘ CS weights are calculated factoring in how well the server instances are performing.
- ⌘ LBA displays will show protocol as TCP or UDP with the registered port numbers

When an external load balancer connects to the z/OS load balancing advisor, it instructs the advisor how it wants weights presented:

- ⌘ The load balancer will poll every so often to obtain the current weights
- ⌘ The load balancer requests the advisor to push weights down at certain intervals or when the weights change
- ⌘ This is how a Cisco CSM external load balancer behaves



### ➤ z/OS LB Advisor

#### f New, stand-alone application

- Runs in Unix System Services environment
- Looks like an MVS Started Task
- Can start it from JCL, issue MVS operator commands, etc.

#### f Executes on any system within the sysplex

- Provides Load Balancing advice for any TCP/UDP server applications within the sysplex
- Acts as a TCP server application supporting SASP (port 3860 by default, but can be customized)
- Supports multiple external LBs concurrently
- Does not require Sysplex Distributor to be configured

#### f Communicates with local Load Balancing Agents

- Uses TCP connections, acts as TCP server (on separate port from SASP)
- Obtains server topology information and workload balancing recommendations from each target system and for each target application

#### f Configuration

- Must identify all eligible Load Balancing Agents (by source IP address and source port)
- Must identify all eligible Load Balancers by source IP address
- IP address and port it should listen to (Application Specific Dynamic DVIPA strongly encouraged)
- Other parameters (debug levels, polling interval, etc.)



### ➤ z/OS LB Agent

#### f New, stand-alone application

- Runs in Unix System Services environment
- Looks like an MVS Started Task
- Can start it from JCL, issue MVS operator commands, etc.

#### f Executes on every target system in the sysplex

- Or at least on every system in the sysplex that is a target of a load balanced request
- Provides Load Balancing advice for specified TCP/UDP server applications on local system

#### f Supports multiple TCP/IP stacks and all known server types: stack-affinity, generic, bind-specific, Shareport, etc.

- Computes weights based on WLM, server availability, server health (dropped connections due to backlog queue full or dropped datagrams due to UDP queue limit exceeded)
- When WLM starts supporting server-specific weights, the LB agent will pick those up

#### f Simple Configuration

- Specify IP address and port for Load Balancing Advisor
- Specify Source IP address/port to be used in connecting to Advisor
  - Static VIPA recommended (allows for failures in physical interfaces)
  - The same source port can be used by all Agents (simplifies configuration)

## Advisor Configuration Statements



Sample is in `hlq.SEZAINST(EZBLBADC)` (alias `LBADVCNF`)

➤ **agent\_connection\_port**

┆ Specifies the port the Advisor should listen on for connections from Agents

➤ **agent\_id\_list**

┆ Specifies which agents are allowed to connect to the Advisor

➤ **debug\_level** (optional)

┆ Specifies the level of debug information that will be logged. Default 7 (Error, Warning, Event).

┆ Recommended level is 79 (defaults plus Info and Debug).

➤ **lb\_connection\_v4**

┆ Specifies the IPv4 address and port the Advisor should listen on for IPv4 connections from load balancers. (There is **lb\_connection\_v6** for IPv6 addresses.)

➤ **lb\_id\_list**

┆ Specifies which load balancers are allowed to connect to the Advisor

➤ **update\_interval** (optional)

┆ Specifies how often agents update the Advisor with new information. Default 60 seconds.



## Advisor Configuration Example



```
debug_level          15      # Error, Warning, Events, Info
update_interval      120     # Agent updates every 2 minutes
lb_connection_v4     9.67.5.1..3860 # DVIPA
lb_id_list
{
  9.67.1.11           # SDBAV4
}
agent_connection_port 8100
agent_id_list
{
  9.67.1.1..8000     # SD1AV4
  9.67.1.2..8000     # SD2AV4
  9.67.30.22..8000   # SD2A2V4
  9.67.1.10..8000    # SDAAV4
}
```



### Sample is in hlq.SEZAINST(EZBLBAGC) (alias LBAGECNF)

➤ **advisor\_id**

f Specifies the IP address and port of the Advisor this agent will communicate with

➤ **host\_connection**

f Specifies the source IP address and port the Agent will bind to for communications with the Advisor.

➤ **debug\_level** (optional)

f Specifies the level of debug information that will be logged. Default 7 (Error, Warning, Event).

f Recommended level is 79 (defaults plus Info and Debug).



**Agent #1:**

```
debug_level          15      # Error, Warning, Events, Info
advisor_id           9.67.5.1..8100 # DVIPA
host_connection      9.67.1.2..8000 # SD2AV4
```

=====

**Agent #2: same as above, except:**

```
host_connection      9.67.30.22..8000
```

# Advisor Started Procedure Sample (in SEZAINST)



```
//LBADV      PROC
/**      IBM Communications Server for z/OS
/**      SMP/E distribution name: EZBLBADV
/**      Licensed Materials - Property of IBM
/**      (C) Copyright IBM Corp. 2004
/**      Function: Sample procedure for running the
/**              z/OS Load Balancing Advisor
/**
//LBADV EXEC PGM=EZBLBADV,REGION=0K,TIME=NOLIMIT,
//      PARM='POSIX(ON) ALL31(ON)'/
/**** Notes:
/** - The system link list concatenation must contain the TCP/IP
/**   runtime libraries and the C runtime libraries. If they are
/**   not in the link list concatenation, this procedure will need
/**   to be changed to STEPLIB to them.
/**   If you add them to STEPLIB, they must be APF authorized.
/** - The z/OS Load Balancing Advisor requires a configuration file
/**   which can be a member of an MVS PDS(E), an MVS sequential file,
/**   or an HFS file.
/**
//CONFIG DD DSN=TCPIP.TCPPARMS(LBADVCNF),DISP=SHR
/**CONFIG DD DSN=TCPIP.CONFIG.LBADV,DISP=SHR
/**CONFIG DD PATH='/etc/lbadv.conf',PATHOPTS=(ORDONLY)
//STDENV DD DUMMY
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)
//SYSIN DD DUMMY
//SYSERR DD SYSOUT=*
//SYSOUT DD SYSOUT=*,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)
//CEEDUMP DD SYSOUT=*,DCB=(RECFM=FB,LRECL=132,BLKSIZE=132)
//SYSMDUMP DD DISP=SHR,DSN=your.data.set.name
```

© Copyright International Business Machines Corporation 2004. All rights reserved.

# Agent Started Procedure Sample (in SEZAINST)



```
//LBAGENT PROC
/** IBM Communications Server for z/OS
/** SMP/E distribution name: EZBLBAGE
/** Licensed Materials - Property of IBM
/** (C) Copyright IBM Corp. 2004
/** Function: Sample procedure for running the
/** z/OS Load Balancing Agent
/**
//LBAGENT EXEC PGM=EZBLBAGE,REGION=0K,TIME=NOLIMIT,
// PARM='POSIX(ON) ALL31(ON)'/
/** Notes:
/** - The system link list concatenation must contain the TCP/IP
/** runtime libraries and the C runtime libraries. If they are
/** not in the link list concatenation, this procedure will need
/** to be changed to STEPLIB to them.
/** If you add them to STEPLIB, they must be APF authorized.
/** - The z/OS Load Balancing Agent requires a configuration file
/** which can be a member of an MVS PDS(E), an MVS sequential file,
/** or an HFS file.
/**
//CONFIG DD DSN=TCPIP.TCPPARMS(LBAGECNF),DISP=SHR
/**CONFIG DD DSN=TCPIP.CONFIG.LBAGENT,DISP=SHR
/**CONFIG DD PATH='/etc/lbagent.conf',PATHOPTS=(ORDONLY)
//STDENV DD DUMMY
//SYSPRINT DD SYSOUT=*,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)
//SYSIN DD DUMMY
//SYSERR DD SYSOUT=*
//SYSOUT DD SYSOUT=*,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)
//CEEDUMP DD SYSOUT=*,DCB=(RECFM=FB,LRECL=132,BLKSIZE=132)
//SYSDUMP DD DISP=SHR,DSN=your.data.set.name
```

© Copyright International Business Machines Corporation 2004. All rights reserved.

## TCP/IP Profile Example



```
IPCONFIG
  DATAGRAMFWD                ; Enable datagram forwarding
  DYNAMICXCF 9.67.30.1 255.255.255.0 1 ; Enable dynamic XCF
VIPADYNAMIC
  VIPARANGE DEFINE DISRUPTIVE 255.255.255.0 9.67.5.1
ENDVIPADYNAMIC

DEVICE VIPA41 VIRTUAL 0      ; Static VIPA
LINK LVIPA41 VIRTUAL 0 VIPA41
HOME 9.67.1.1                IVIPA41

PORT
  3860 TCP LBADV              ; SASP Workload Advisor (LB connections)
  8100 TCP LBADV              ; SASP Workload Advisor (Agent connections)
  8888 TCP LBAGENT            ; SASP Workload Agent (Advisor connection)
```

- These statements can be duplicated in all TCP/IP stack profiles in the sysplex
  - ⌘ Simplifies port reservation for this function
  - ⌘ Allows Advisor to be restarted on any system in the sysplex

## Security Considerations



- Sample RACF definitions can be found in hlq.SEZAINST(EZARACF)

- f look for "LBADV" and "LBAGENT"

- Assign a userid for each Started Task (LB Advisor and Agent)

- f Using RACF STARTED Class profiles

```
RDEFINE  STARTED  LBADV.*      STDATA (USER (LBADV) )
RDEFINE  STARTED  LBAGENT.*    STDATA (USER (LBAGENT) )
```

- f Or by updating RACF started procedures table (ICHRIN03)

- f Both started tasks can share the same userid if desired

- Define the userid for each started task

- f Both require OMVS segment (no requirement for UID 0)

```
ADDUSER  LBADV      DFLTGRP (OMVSGRP)  OMVS (UID (nn1)  HOME ('/'))
ADDUSER  LBAGENT   DFLTGRP (OMVSGRP)  OMVS (UID (nn2)  HOME ('/'))
```

- Optionally make the Advisor and Agent non-swappable (highly recommended)

- f Add a PPT entry

- The only way to do so without requiring UID 0

## Security Considerations....



### ➤ Prevent unauthorized start of the Advisor and Agent

⌘ via OPERCMDS Class profile

```
SETROPTS CLASSACT (OPERCMD5)
SETROPTS RACLIST (OPERCMD5)
RDEFINE OPERCMDS (MVS.SERVGR.LBADV) UACC (NONE)
RDEFINE OPERCMDS (MVS.SERVGR.LBAGENT) UACC (NONE)
PERMIT MVS.SERVGR.LBADV CLASS (OPERCMD5) ACCESS (CONTROL) ID (LBADV)
PERMIT MVS.SERVGR.LBAGENT CLASS (OPERCMD5) ACCESS (CONTROL) ID (LBAGENT)
```

⌘ The userids associated with these started procedure tasks must be authorized (READ access) to these profiles

- Prevents unauthorized starting of these servers from USS shell or batch environments (which is not supported)

### ➤ Remember to refresh RACF cache prior to starting the servers

```
SETROPTS RACLIST (FACILITY) REFRESH
SETROPTS RACLIST (STARTED) REFRESH
SETROPTS GENERIC (STARTED) REFRESH
SETROPTS RACLIST (OPERCMD5) REFRESH
```



## Automatic Restart of Advisor/Agent



- Must ensure that automation is in place to restart
  - f the Advisor
    - on the same/other system in the sysplex in cases of failures (of the Advisor or the System)
  - f the Agent
    - in the same system when the agents terminates abnormally
  
- Can be accomplished with ARM (Automatic Restart Manager) policy or other automation
  - f Note that while AUTOLOG can be used to start the Agent, it can NOT be used to monitor the availability of the Agent after initial startup.
  
- Only 1 Advisor can be active per sysplex and only 1 agent per system
  - f Note: Internal checks will prevent the starting of multiple Advisors (within the sysplex) or multiple Agents within the same system.

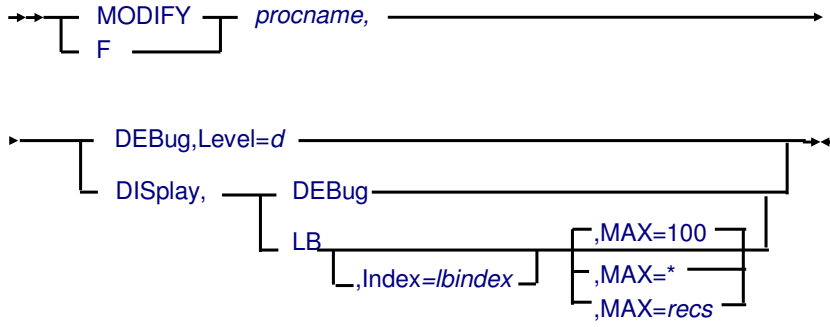
## Starting the Advisor and Agent



```
S LBADV
IEF695I START LBADV WITH JOBNAME LBADV IS ASSIGNED TO USER LBADV
, GROUP SYS1
EZD1231I LBADV STARTING
EZD1232I LBADV INITIALIZATION COMPLETE
EZD1261I LBADV AGENT CONNECTED FROM 9.67.1.10
EZD1263I LBADV LOAD BALANCER CONNECTED FROM 9.67.1.11

S LBAGENT
IEF695I START LBAGENT WITH JOBNAME LBAGENT IS ASSIGNED TO USER LBAGENT
, GROUP SYS1
EZD1231I LBAGENT STARTING
EZD1232I LBAGENT INITIALIZATION COMPLETE
EZD1261I LBADV AGENT CONNECTED FROM 9.67.1.2
EZD1259I LBAGENT CONNECTED TO ADVISOR 9.67.5.1
```

# Function Externals Advisor Console Command Syntax



## Advisor Console Command Examples



```
1 F LBADV,DISP,LB
2 EZD1242I LOAD BALANCER SUMMARY
3 LB INDEX      : 00      UUID      : 85EB355C
4 IPADDR..PORT : 9.67.1.11..50000
5 HEALTH       : 20      FLAGS     : NOCHANGE
6 1 OF 1 RECORDS DISPLAYED
7
8 F LBADV,DISP,LB,I=0
9 EZD1243I LOAD BALANCER DETAILS
10 LB INDEX     : 00      UUID      : 85EB355C
11 IPADDR..PORT : 9.67.5.1..50000
12 HEALTH      : 20      FLAGS     : NOCHANGE
13 GROUP NAME   : MISCAPPSFARM
14 IPADDR..PORT: 9.67.1.10..1023
15   SYSTEM NAME: VIC007   PROTOCOL  : UDP   AVAIL    : YES
16   WLM WEIGHT  : 00032   CS WEIGHT : 100   NET WEIGHT: 00001
17   FLAGS      :
18 IPADDR..PORT: 9.67.1.10..10007
19   SYSTEM NAME: VIC007   PROTOCOL  : TCP   AVAIL    : YES
20   WLM WEIGHT  : 00032   CS WEIGHT : 100   NET WEIGHT: 00001
21   FLAGS      :
22 IPADDR..PORT: 9.67.1.2..623
23   SYSTEM NAME: N/A     PROTOCOL  : UDP   AVAIL    : NO
24   WLM WEIGHT  : 00000   CS WEIGHT : 000   NET WEIGHT: 00000
25   FLAGS      : NOCON
```

## Advisor Console Command Examples



```
... continued from previous page ...
26 IPADDR..PORT: 9.67.1.2..514
27 SYSTEM NAME: MVS209 PROTOCOL : UDP AVAIL : YES
28 WLM WEIGHT : 00030 CS WEIGHT : 100 NET WEIGHT: 00001
29 FLAGS :
30 IPADDR..PORT: 9.67.1.2..513
31 SYSTEM NAME: MVS209 PROTOCOL : TCP AVAIL : YES
32 WLM WEIGHT : 00030 CS WEIGHT : 100 NET WEIGHT: 00001
33 FLAGS :
34 GROUP NAME : ONEMEMBERFARM
35 IPADDR..PORT: 9.67.1.2..0
36 SYSTEM NAME: MVS209 PROTOCOL : 000 AVAIL : YES
37 WLM WEIGHT : 00030 CS WEIGHT : 100 NET WEIGHT: 00030
38 FLAGS :
39 GROUP NAME : SYSTEMFARM
40 IPADDR..PORT: 9.67.1.10..0
41 SYSTEM NAME: VIC007 PROTOCOL : 000 AVAIL : YES
42 WLM WEIGHT : 00032 CS WEIGHT : 100 NET WEIGHT: 00001
43 FLAGS :
44 IPADDR..PORT: 9.67.1.2..0
45 SYSTEM NAME: MVS209 PROTOCOL : 000 AVAIL : YES
46 WLM WEIGHT : 00030 CS WEIGHT : 100 NET WEIGHT: 00001
47 FLAGS :
48 8 OF 8 RECORDS DISPLAYED
```

# Agent Console Command Syntax



→ MODIFY *procname*, →  
F

→ DEBUg,Level=*d* →  
DISplay { ,DEBUg →  
          ,MEMbers → { ,MAX=100 →  
                          ,MAX=\* →  
                          ,MAX=recs →  
          ,MEMbers,DETail →  
          ,MEMbers,DETail,PORT=*nnnnn* →  
          ,MEMbers,DETail,TCPname=*tcpname* →  
Enable,Target options →  
Quiesce,Target options →

### Target options:

→ PORT=*nnnnn* { ,PROTOcol=TCP →  
                          ,PROTOcol=*proto* → } ,IPaddr=*ipaddr* →  
TCPname=*tcpname* →  
SYStem →

## Function External Agent Console Command Examples



```
1 F LBAGENT, DISP, MEM, DET
2 EZD1245I MEMBER DETAILS 348
3 LB INDEX      : 00          UUID      : 85EB355C
4 GROUP NAME   : SYSTEMFARM
5 IPADDR..PORT: 9.67.1.2..0
6 TCPNAME     : TCPCS      MATCHES   : 001  PROTOCOL : 000
7 FLAGS       :
8 JOBNAME     : N/A        ASID       : N/A   RESOURCE : N/A
9 GROUP NAME   : ONEMEMBERFARM
10 IPADDR..PORT: 9.67.1.2..0
11 TCPNAME     : TCPCS      MATCHES   : 001  PROTOCOL : 000
12 FLAGS       :
13 JOBNAME     : N/A        ASID       : N/A   RESOURCE : N/A
14 GROUP NAME   : MISCAPPSFARM
15 IPADDR..PORT: 9.67.1.2..513
16 TCPNAME     : TCPCS      MATCHES   : 001  PROTOCOL : TCP
17 FLAGS       : ANY
18 JOBNAME     : INETD1    ASID       : 002A  RESOURCE : 00000052
19 IPADDR..PORT: 9.67.1.2..514
20 TCPNAME     : TCPCS      MATCHES   : 001  PROTOCOL : UDP
21 FLAGS       : ANY
22 JOBNAME     : SYSLOGD1  ASID       : 002B  RESOURCE : 0000001F
23 4 OF 4 RECORDS DISPLAYED
```

## Some strategies for workload request balancing into a z/OS Sysplex

e-business



### **A DNS/WLM as workload balancing should not be used any longer**

### **B Where HTTP workload is to be balanced based on content of HTTP requests, an outboard load balancer that supports contents inspection must be deployed**

,If HTTPS workload is to be included, the load balancing node must be accompanied by an SSL/TLS offload technology

,Can be combined with a cache appliance for improved performance

### **C UDP workload balancing must be deployed using an outboard load balancer - SD does not support UDP balancing**

### **D Remaining TCP connection balancing can be deployed using either SD or an outboard load balancer:**

,SD has more real-time information available than outboard load balancers - even with outboard load balancers using the SASP protocol

,Who is to apply management control over the workload balancing function will be a major factor in deciding which solution to use

,If installation cannot combine use of SD with Cisco MNLB forwarding agent support, an outboard LB may from a performance point of view be preferable to SD

© Copyright International Business Machines Corporation 2004. All rights reserved.



# Trademarks, Copyrights, and Disclaimers e-business

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:

IBM	CICS	IMS	MQSeries	Tivoli
IBM (logo)	Cloudscape	Informix	OS/390	WebSphere
e (logo) business	DB2	iSeries	OS/400	xSeries
AIX	DB2 Universal Database	Lotus	pSeries	zSeries

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds.

Other company, product and service names may be trademarks or service marks of others.

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or program(s) described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectual property rights, may be used instead.

Information is provided "AS IS" without warranty of any kind. THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. IBM makes no representations or warranties, express or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

© Copyright International Business Machines Corporation 2004. All rights reserved.

Note to U.S. Government Users - Documentation related to restricted rights-Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract and IBM Corp.

© Copyright International Business Machines Corporation 2004. All rights reserved.