Configuring the z/OS Load Balancing Advisor: Making External IP Load Balancers Sysplex Aware

Enterprise Networking and Transformation Solutions, Raleigh

Agenda

- Exploiting and Configuring the z/OS LBA
  - Basic Configuration
  - High Availability Features
  - Operations
  - Comparing IP load balancing solutions

Steps for Configuring and Running the Advisor and Agent

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).
- **lb_connection_v4** and/or **lb_connection_v6**
  - Specifies the local IPv4 or IPv6 address and port the Advisor should listen on for connections from Load Balancers. Use both to listen for either type of connection. Default port value is 3860.
- **lb_id_list**
  - Specifies remote addresses from which Load Balancers are allowed to connect to the Advisor

lb_id_list
If the Load Balancer has multiple source IP addresses it can use, make sure lb_id_list contains the address the Load Balancer will use as a source IP address when connecting as a SASP client. If your Load Balancer-to-Advisor connection is failing, examine the Advisor log for a message with the text, 'Unauthorized connection attempt from '. If this message is present, is the address the load balancer is using as a source IP address for connecting to the Advisor. Insert this address into the lb_id_list statement, restart the Advisor and reconnect from the Load Balancer.
For CISCO CSM, this is the client VLAN interface IP address, not the server VLAN IP address.

General Configuration Hints:
Any statement that contains the phrase "id" specifies a remote endpoint.

Any statement that contains the phrase "connection" specifies a local IP address and possibly port.

Advisor Configuration Statements (Cont)

- **port_list** (optional)
  - Specifies a list of ports and the type of WLM server recommendation that should be used for each. Overrides value from the *wlm* statement on a port basis. Only valid for V1R7.
- **update_interval** (optional)
  - Specifies how often Agents update the Advisor with new information. May also determine how often the Advisor updates the load balancer if the load balancer supports the 'push" flag. Default 60 seconds.
  -
- **wlm** (optional)
  - Specifies the default type of WLM recommendations that will be attempted Choices are *serverwlm* and *basewlm*. Default is *basewlm.* Only valid for V1R7.

Advisor Configuration Example

```
debug_level                                15                    # Error, Warning, Event, Info
update_interval                            120             # Agent updates every 2 minutes
lb_connection_v4       10.67.5.1..3860     # DVIPA
lb_id_list
{
   10.67.1.11
            # SDBAV4
}
agent_connection_port  8100
agent_id_list
{
   10.67.1.1..8000
  # SD1AV4
   10.67.1.2..8000
  # SD2AV4
   10.67.30.22..8000                                                          # SD2A2V4
   10.67.1.10..8000
  # SDAAV4
}

wlm serverwlm                                      # Request server-specific WLM weights
port_list
{
   21       wlm basewlm                            # Use system WLM weights for FTP
}
```

Function Externals
Agent Configuration Statement

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

- **advisor_id**
  - Specifies the remote IP address and port of the Advisor this agent will communicate with
- **host_connection**
  - Specifies the local IP address and port the Agent will bind to for communications with the Advisor.
- **debug_level** (optional)
  - Specifies the level of debug information that will be logged. Default 7 (Error, Warning, Event).

Function Externals
Agent Configuration Examples

Agent #1:

```
debug_level          15              # Error, Warning, Event, Info
advisor_id           10.67.5.1..8100 # DVIPA
host_connection      10.67.1.2..8000 # SD2AV4
```

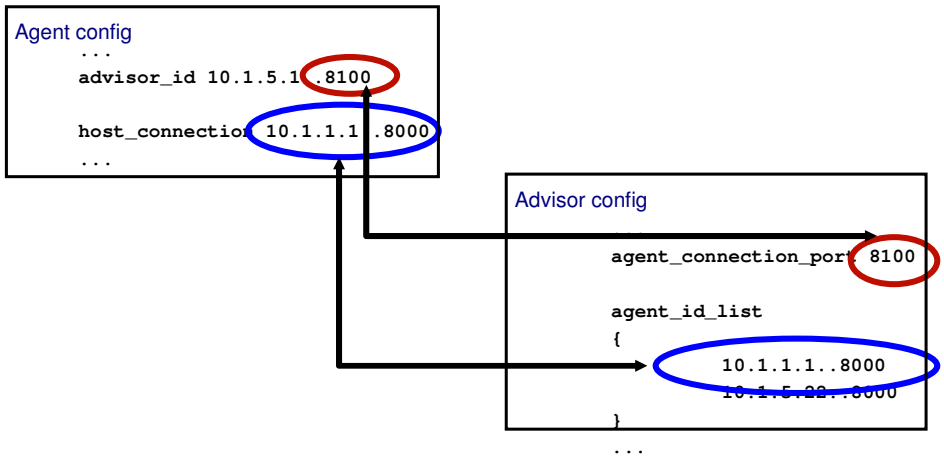
==========

Agent #2: same as above, except:

```
host_connection      10.67.30.22..8000
```

Configuration Hints

- Configuration relationships among the Advisor and Agent
  - See connected arrows
  - IP address in *advisor_id* statement can be any IP address belonging to the TCP/IP stack the Advisor is running on, however, it is recommended this be a unique application-instance DVIPA.



Notes - Advisor Started Procedure Sample (in SEZAINST, alias EZBLBADV)

NOTES

```
//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)
//CEESNAP  DD SYSOUT=*,DCB=(RECFM=FB,LRECL=132,BLKSIZE=132)
//SYSMDUMP DD DISP=SHR,DSN=your.data.set.name
```

Notes - Agent Started Procedure Sample (in SEZAINST, alias EZBLBAGE)


```
//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)
//CEESNAP  DD SYSOUT=*,DCB=(RECFM=FB,LRECL=132,BLKSIZE=132)
//SYSMDUMP DD DISP=SHR,DSN=your.data.set.name
```

Function Externals
Security and Control Considerations

- Define OMVS segments if not already defined
- Configure RACF (or equivalent External Security Manager Product)
  - Define User IDs to RACF, associate with OMVS segment, and assign UIDs
  - Add user IDs to the STARTED class
  - Optionally, restrict which users can start the applications
    - Only one instance of the Advisor is permitted per sysplex
    - Only one instance of the Agent is permitted per MVS system
  - Permit access to the BPX.WLMSERVER FACILITY class profile
    - For RACF, explicit access not required if the resource profile is not defined
    - For other security products, consult the product documentation as to whether explicit access is required for your installation
  - Sample RACF definitions can be found in hlq.SEZAINST(EZARACF)
    - look for "LBADV" and "LBAGENT"

RACF Configuration Notes

NOTES

- Sample RACF definitions can be found in hlq.SEZAINST(EZARACF)
  - look for "LBADV" and "LBAGENT"
- Both require OMVS segments (UID may be zero or non-zero)
  - `ADDUSER LBADV DFLTGRP(OMVSGRP) OMVS(UID(nn) HOME('/'))`
  - `ADDUSER LBAGENT DFLTGRP(OMVSGRP) OMVS(UID(nn) HOME('/'))`
- Add the default users to the started class
  - `RDEFINE STARTED LBADV.* STDATA(USER(LBADV))`
  - `RDEFINE STARTED LBAGENT.* STDATA(USER(LBAGENT))`
- If you would like to control which users can start the applications
  - Ensure that the OPERCMDS class is active and RACLISTed and RACLIST processing is enabled.
    - `SETROPTS CLASSACT(OPERCMDS)`
    - `SETROPTS RACLIST (OPERCMDS)`
  - Define the following OPERCMDS class profiles using a security product like RACF
    - `RDEFINE OPERCMDS (MVS.SERVMGR.LBADV) UACC(NONE)`
    - `RDEFINE OPERCMDS (MVS.SERVMGR.LBAGENT) UACC(NONE)`
  - Grant the Advisor access to the OPERCMDS class.
    - `PERMIT MVS.SERVMGR.LBADV CLASS(OPERCMDS) ACCESS(CONTROL) - ID(userid)`
    - `PERMIT MVS.SERVMGR.LBAGENT CLASS(OPERCMDS) ACCESS(CONTROL) - ID(userid)`
  - Refresh the OPERCMDS class.
    - `SETROPTS RACLIST(OPERCMDS) REFRESH.`

WLM Considerations for the Advisor/Agents

- Make Advisor and Agent non-swappable
  - V1R7
    - Both will run non-swappable by default. No action required.
  - V1R4, V1R5, and V1R6
    - Configuration of the Program Properties Table (PPT) in the SCHEDxx member is required in order to run non-swappable. See the following "Notes" page for the actual PPT entries.
- Ensure the Advisor and Agents receive the proper dispatching priority
  - Verify Advisor and Agents are assigned to the WLM SYSSTC service class
  - See MVS Planning: Workload Management for more information

Notes - PPT Entries to make Non-swappable

NOTES

```
PPT PGMNAME(EZBLBAGE)          /* z/OS LOAD BALANCING AGENT   */
    KEY(8)                     /* PROTECTION KEY              */
    NOSWAP                     /* NONSWAPPABLE                */
    PRIV                       /* PRIVILEGED                  */
    SYST                       /* SYSTEM TASK , NOT TIMED     */
    AFF(NONE)                  /* NO CPU AFFINITY             */
PPT PGMNAME(EZBLBADV)          /* Z/OS LOAD BALANCING ADVISOR */
    KEY(8)                     /* PROTECTION KEY              */
    NOSWAP                     /* NONSWAPPABLE                */
    PRIV                       /* PRIVILEGED                  */
    SYST                       /* SYSTEM TASK , NOT TIMED     */
    AFF(NONE)                  /* NO CPU AFFINITY             */
```

High Availability Considerations
Automatic Restart of Advisor/Agent

- Ensure that automation is in place to restart
  - the Advisor
    - on the same/other system in the sysplex in cases of failures (of the Advisor or the System)
      - ARM (Automatic Restart Manager) policy or other automation can be used
      - TCP/IP AUTOLOG processing can be used to cover scenario where TCP/IP stack on local system fails
  - the Agent
    - in the same system when the agent terminates abnormally
    - Using ARM or other automation
      - 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
  - *Note:* Internal checks will prevent the starting of multiple Advisors (within the sysplex) or multiple Agents within the same system.

- Special recovery considerations for the Advisor on a multi-stack system (CINET)
  - Refer to "CINET Considerations" in the appendix

- Advisor may be (re)started while Agents already running or vice versa

High Availability Considerations
Dynamic VIPA for Load Balancing Advisor

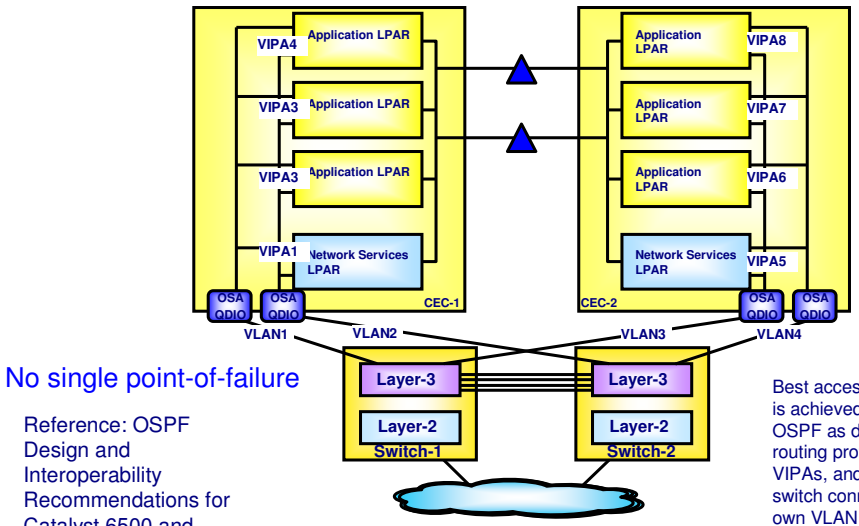
```
IPCONFIG
  DATAGRAMFWD                       ; Enable datagram forwarding
  DYNAMICXCF 10.67.30.1 255.255.255.0 1 ; Enable dynamic XCF
VIPADYNAMIC
  VIPARANGE  DEFINE 255.255.255.0 10.67.5.1
ENDVIPADYNAMIC

DEVICE VIPA41   VIRTUAL  0          ; Static VIPA
 LINK  LVIPA41  VIRTUAL  0   VIPA41
 HOME  10.67.1.1              LVIPA41

PORT
  3860  TCP LBADV          ; SASP Workload Advisor (LB connections)
  8100  TCP LBADV          ; SASP Workload Advisor (Agent connections)
  8000  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

High Availability downstream network connectivity to the Sysplex: OSA-Express with QDIO



No single point-of-failure

Reference: OSPF Design and Interoperability Recommendations for Catalyst 6500 and OSA-Express Environments.

Best access availability is achieved by using OSPF as dynamic routing protocol, using VIPAs, and having each switch connection be its own VLAN and use the switch's layer-3 functions to interconnect them.

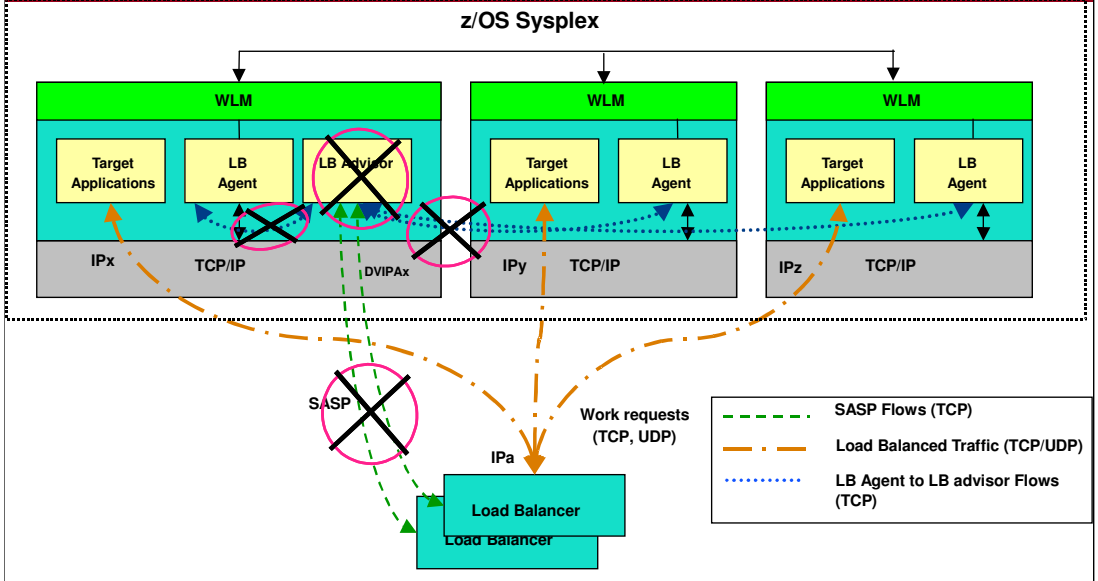
Two OSA adapters per LPAR, each connected to a different Layer 3 switch.

A VLAN per switch port towards OSA adapters are recommended to minimize the effects of spanning tree algorithm when end-stations are disconnected.

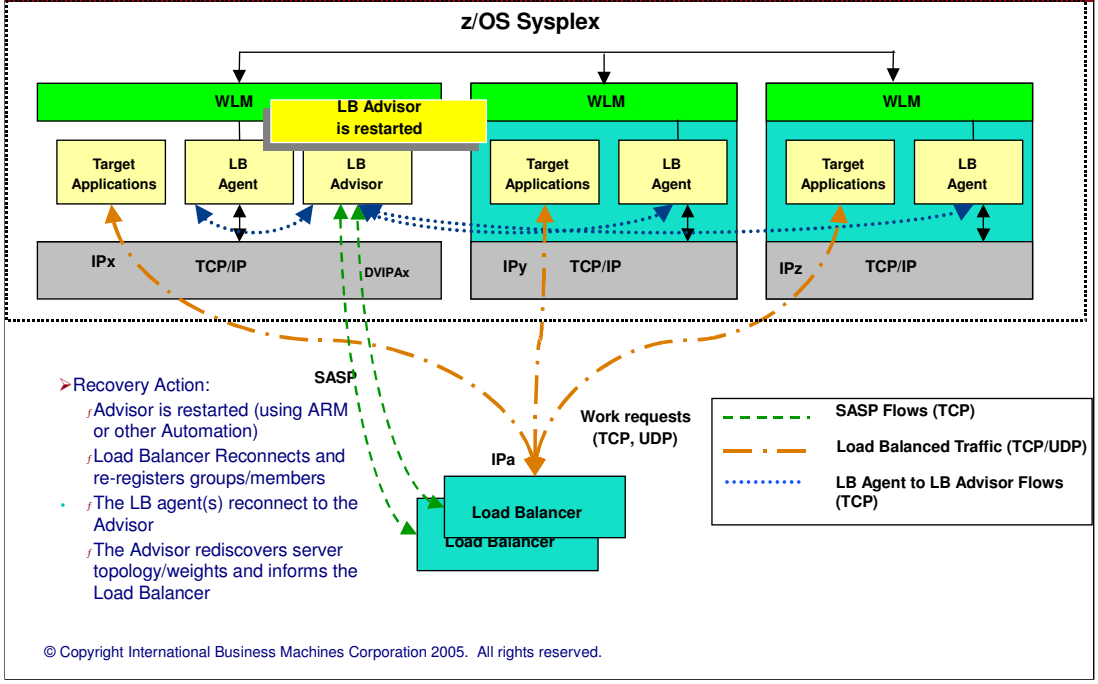
This design allows OSPF to kick in and recalculate routes immediately when a cable is disconnected.

High Availability and Recovery Scenarios

Recovery Scenario A: Advisor Fails

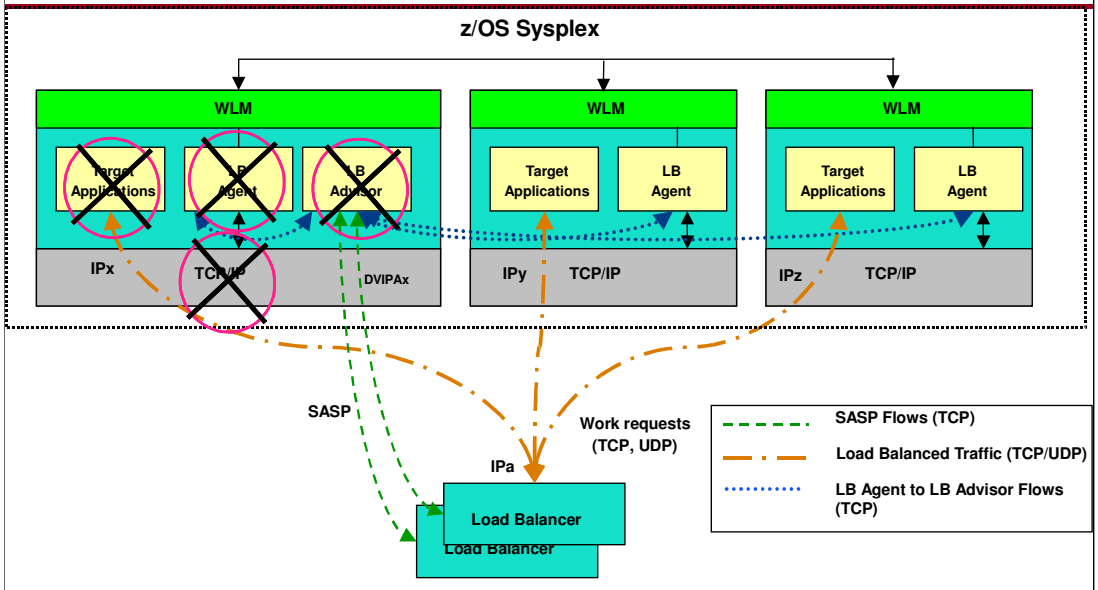


Recovery Scenario A (Cont.): Advisor Fails

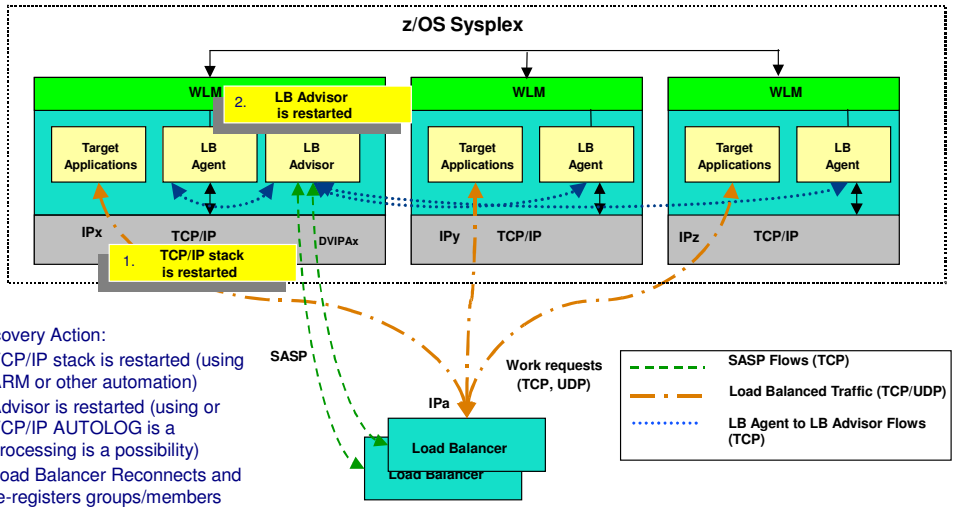


- Recovery Action:
  - Advisor is restarted (using ARM or other Automation)
  - Load Balancer Reconnects and re-registers groups/members
  - The LB agent(s) reconnect to the Advisor
  - The Advisor rediscovers server topology/weights and informs the Load Balancer

Recovery Scenario B:
TCP/IP Stack on Advisor System Fails

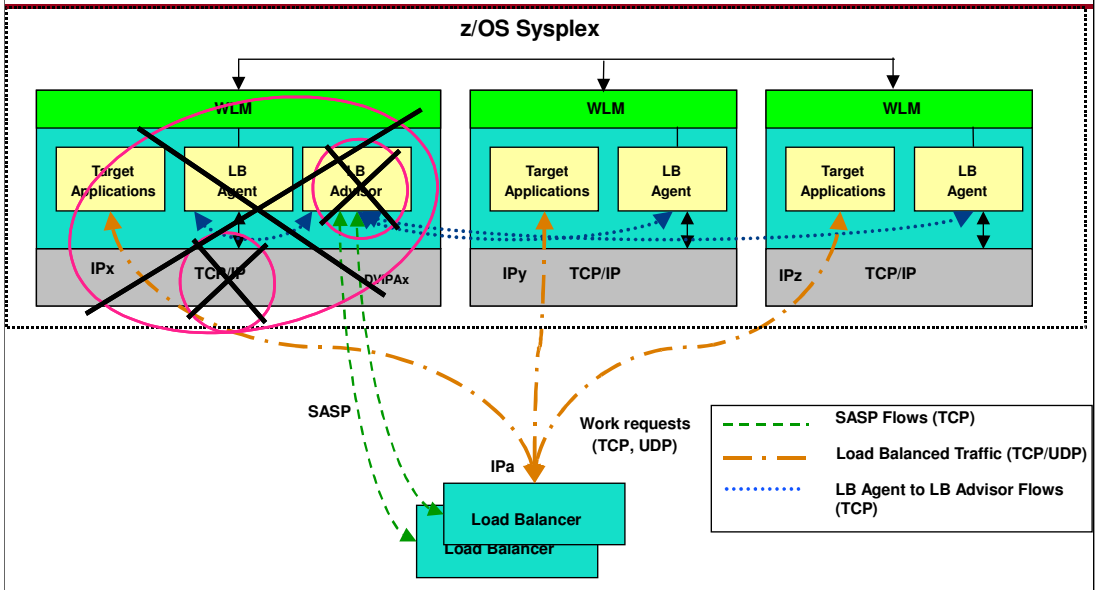


Recovery Scenario B (Cont.): TCP/IP Stack on Advisor System Fails

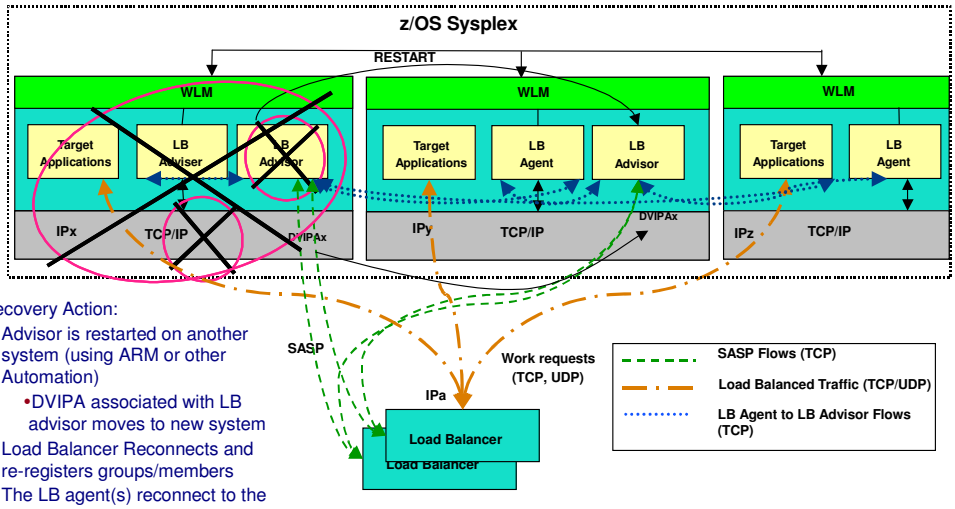


- Recovery Action:
  - TCP/IP stack is restarted (using ARM or other automation)
  - Advisor is restarted (using or TCP/IP AUTOLOG is a processing is a possibility)
  - Load Balancer Reconnects and re-registers groups/members
  - The LB agent(s) reconnect to the Advisor
  - The Advisor rediscovers server topology/weights and informs the Load Balancer

Recovery Scenario C: Advisor System Fails

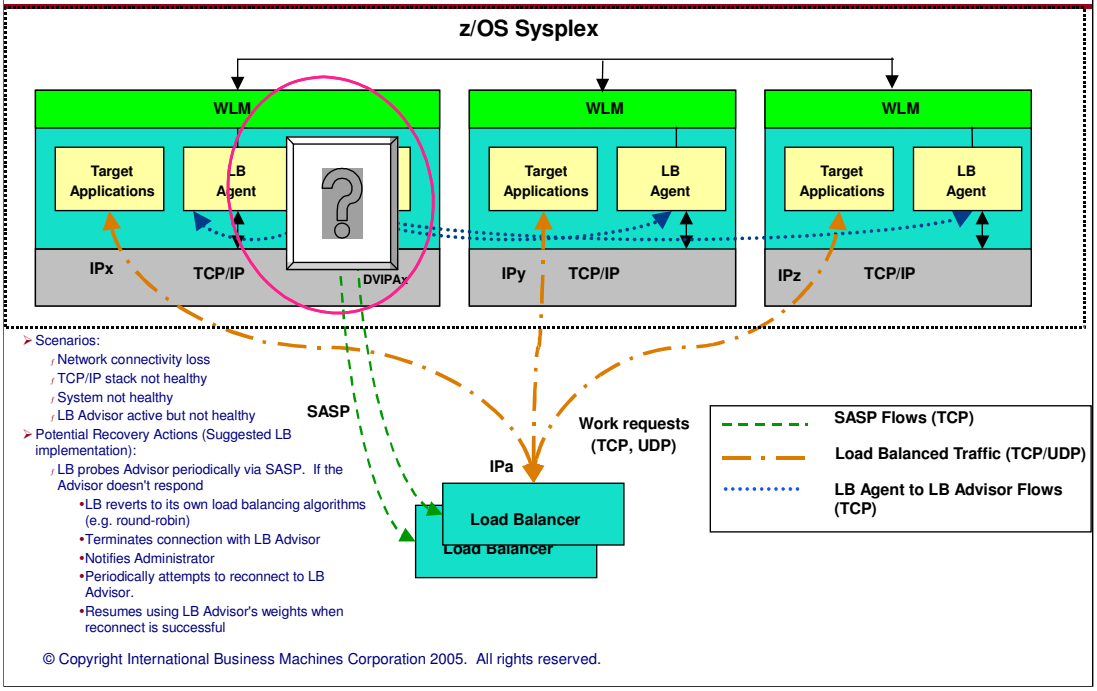


Recovery Scenario C (Cont.): Advisor System Fails



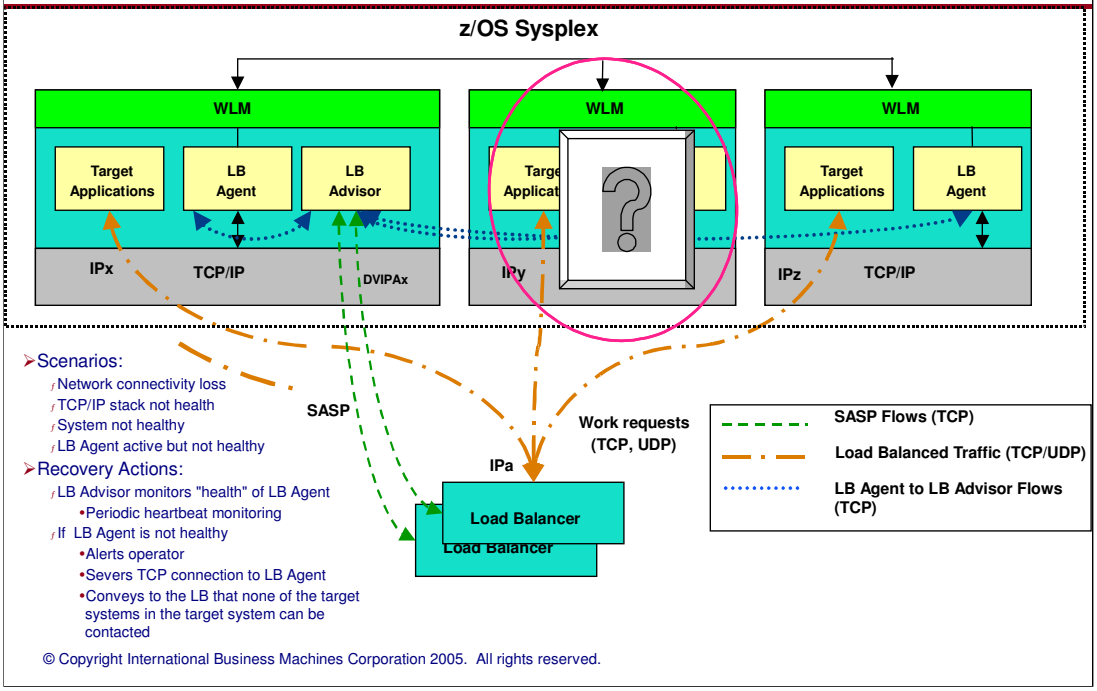
- Recovery Action:
  - Advisor is restarted on another system (using ARM or other Automation)
    - DVIPA associated with LB advisor moves to new system
  - Load Balancer Reconnects and re-registers groups/members
  - The LB agent(s) reconnect to the Advisor
  - The Advisor rediscovers server topology/weights and informs the Load Balancer

Recovery Scenario D: Advisor Not Responding



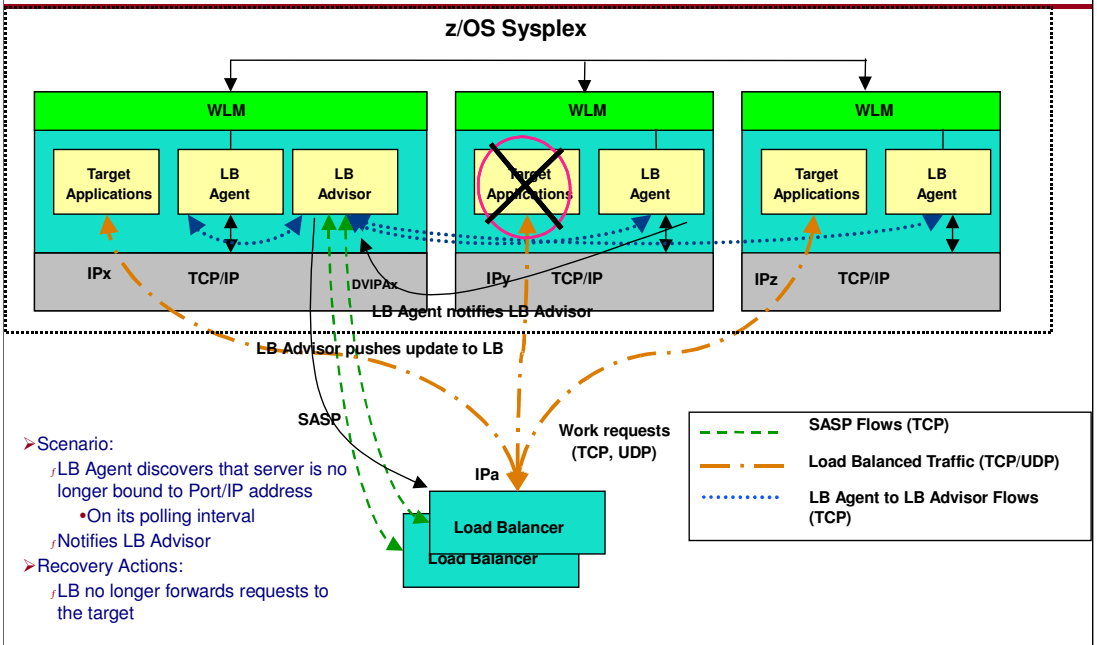
- Scenarios:
  - Network connectivity loss
  - TCP/IP stack not healthy
  - System not healthy
  - LB Advisor active but not healthy
- Potential Recovery Actions (Suggested LB implementation):
  - LB probes Advisor periodically via SASP. If the Advisor doesn't respond
    - LB reverts to its own load balancing algorithms (e.g. round-robin)
    - Terminates connection with LB Advisor
    - Notifies Administrator
    - Periodically attempts to reconnect to LB Advisor.
    - Resumes using LB Advisor's weights when reconnect is successful

Recovery Scenario E: Agent Not Responding



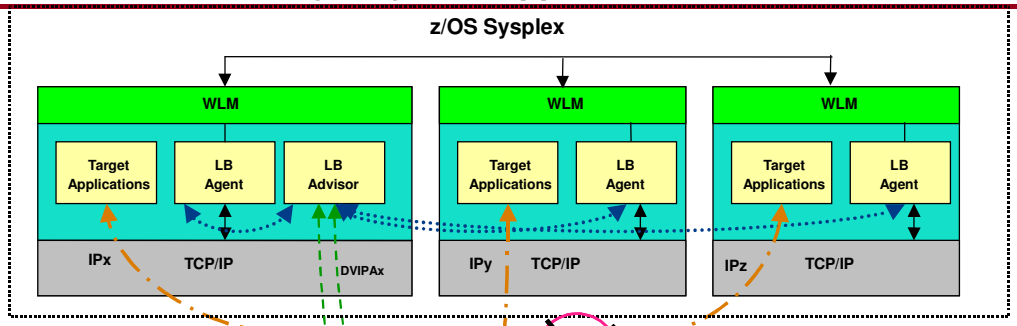
- Scenarios:
  - Network connectivity loss
  - TCP/IP stack not health
  - System not healthy
  - LB Agent active but not healthy
- Recovery Actions:
  - LB Advisor monitors "health" of LB Agent
    - Periodic heartbeat monitoring
  - If LB Agent is not healthy
    - Alerts operator
    - Severs TCP connection to LB Agent
    - Conveys to the LB that none of the target systems in the target system can be contacted

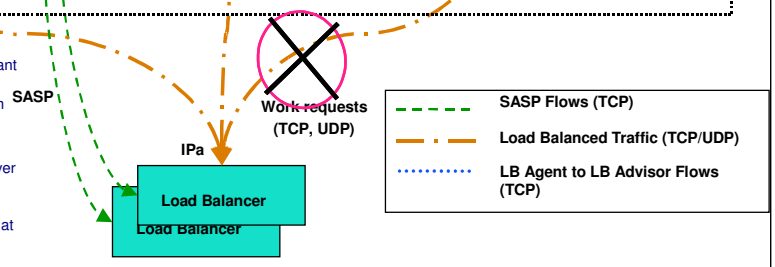
Recovery Scenario F: Target Application Fails



- Scenario:
  - LB Agent discovers that server is no longer bound to Port/IP address
    - On its polling interval
  - Notifies LB Advisor
- Recovery Actions:
  - LB no longer forwards requests to the target

Recovery Scenario G: Loss of Network Connectivity Between LB and Target System/Application





- Scenario:
  - Network connectivity loss and no redundant paths to target system
  - LB Advisor may still believe target system is reachable and healthy
- Potential Recovery Actions:
  - LB should be configured to perform IP layer health probes (to target systems)
  - If target system is not reachable the LB should suspend forwarding requests to that system
  - IP layer health probes should continue to be issued so that distribution to the target system can once again resume if connectivity to that system is restored

Advisor and Agent Console Commands

Advisor Console Command Examples (Displaying Status Information)

```
 1 F LBADV,DISP,LB
 2 EZD1242I LOAD BALANCER SUMMARY
 3 LB INDEX      : 00        UUID      : 637FFF175C
 4  IPADDR..PORT : 10.42.154.105..50005
 5  HEALTH       : 20        FLAGS     : NOCHANGE PUSH TRUST
 6 LB INDEX      : 01        UUID      : 207FFF175C
 7  IPADDR..PORT : 10.42.154.160..50006
 8  HEALTH       : 7F        FLAGS     : PUSH TRUST
 9 2 OF 2 RECORDS DISPLAYED

10
11 F LBADV,DISP,LB,I=0
12 EZD1243I LOAD BALANCER DETAILS
13 LB INDEX      : 00        UUID      : 637FFF175C
14  IPADDR..PORT : 10.42.154.105..50005
15  HEALTH       : 20        FLAGS     : NOCHANGE PUSH TRUST
16  GROUP NAME   : SYSTEMFARM
17   GROUP FLAGS : BASEWLM
18   IPADDR..PORT: 10.42.105.154..0
19    SYSTEM NAME: MVS209     PROTOCOL  : 000  AVAIL     : YES
20    WLM WEIGHT : 00021      CS WEIGHT : 100 NET WEIGHT: 00001
21    FLAGS      :
22   IPADDR..PORT: 10.42.105.60..0
23    SYSTEM NAME: VIC007     PROTOCOL  : 000  AVAIL     : YES
24    WLM WEIGHT : 00045      CS WEIGHT : 100  NET WEIGHT: 00002
25    FLAGS      :
... continued on following page ...
```

Two Load Balancers registered

Members in group SYSTEMFARM

Load Balancing Weights

Quiescing/Enabling Workload Distribution

- Quiesce/Enable (Modify *procname*,ENABLE|QUIESCE)
  - Load Balancing Agent command (Issued on the system to be quiesced/enabled)
  - Three levels of control:
    - System
    - TCP/IP stack
    - Specific Application
- "Quiesce" removes member from future workload distribution eligibility
  - Existing Connections are not affected
- "Enable" restores member's eligibility for workload balancing
- Only affects workload arriving through external Load Balancer
  - Uses:
    - Planned outage of an MVS system, TCP/IP stack, application, or homogeneous group of applications
    - Misbehaving application
  - Two points of Quiesce/Enable control
    - MVS operator
    - Load Balancer administrator
      - –Availability may be implementation dependent

Comparing IP Load Balancing Solutions

Comparing IP load balancing technologies

	Sysplex Distributor	External Load Balancers	External Load Balancers with SASP
When is the server instance decision made?	Connection Setup (in line Syn segment)	Connection Setup (in line Syn segment)	Connection Setup (in line Syn segment)
Support for TCP and UDP applications	TCP only	Depends on the load balancer implementation	Depends on the load balancer implementation (SASP supports both TCP and UDP)
Support for affinities between TCP connection requests based on data content	No, support does however exist for timer based affinities	Depends on implementation, some support affinities for HTTP/HTTPS requests by inspecting data content (correlating cookies, jsessionid)	Depends on implementation, some support affinities for HTTP/HTTPS requests by inspecting data content (correlating cookies, jsessionid)
Extra Network Flows	Yes for inbound traffic. Inbound traffic must traverse the Sysplex Distributor node. If Sysplex Distributor is configured as Service Manager for CISCO routers then the inbound traffic can flow directly to the target application. No for outbound traffic	Depends on the load balancer implementation (can be avoided if the load balancer is implemented as part of a router/switch)	Depends on the load balancer implementation (can be avoided if the load balancer is implemented as part of router/switch)

Comparing IP load balancing technologies...

	Sysplex Distributor	External Load Balancers	External Load Balancers with SASP
Network Address Translation	Not needed (client and server IP addresses are not modified)	May be required by some implementations (client and/or server IP addresses may be translated)	May be required by some implementations (client and/or server IP addresses may be translated)
Support for IPv6	Yes	Depends on the load balancer implementation	Depends on the load balancer implementation (SASP supports both IPv4 and IPv6)
z/OS WLM recommendations	Yes (System level and Server level WLM recommendations are available)	Depends on the load balancer implementation	Yes (System level and Server level WLM recommendations)
z/OS Network QoS recommendations	Yes (based on z/OS QoS policy)	No	No

Comparing IP load balancing technologies...

	Sysplex Distributor	External Load Balancers	External Load Balancers with SASP
z/OS TCP/IP server health information	Yes	No	Yes
Detection of target application and/or target system state changes (active or inactive).	Yes, application and system state changes are detected in near real-time fashion.	Depends on the load balancer implementation	Yes, the z/OS load balancing Advisor and Agents detect application and system state changes within a configurable time period (60 seconds by default).
High availability solution (load balancing continues even if the primary load balancing component becomes unavailable)	Yes, one or more backups can be configured to allow for dynamic take over in cases where the TCP/IP stack or system that is acting as the distributor fails.	For failures to the load balancer, it depends on the load balancer implementation. Some solutions provide for backup load balancers that can dynamically take over load balancing responsibilities in cases of failures.	For failures to the load balancer, it depends on the load balancer implementation. The z/OS Load Balancing Advisor and Agents can be configured for high availability to minimize the impact of an Advisor, Agent or system failure).

Comparing IP load balancing technologies...

	Sysplex Distributor	External Load Balancers	External Load Balancers with SASP
How is the solution administered/configured?	Initial setup may require some interactions with the network (dynamic routing protocols, DNS updates for Dynamic VIPAs, etc.). Ongoing administration (adding/removing target server applications and or systems) typically confined within z/OS systems.	Initial setup/configuration on load balancer, some configuration on z/OS may be required. Ongoing administration should be mostly confined to the load balancer (although z/OS configuration may be necessary when adding new target systems, etc.)	Initial setup/configuration on load balancer and on z/OS. Ongoing administration may need to be performed on both the load balancer and the z/OS systems.

Summary
Deciding on an IP Load Balancing Solution

- Decide on type of solution (Internal vs External Load Balancing solution)
  - Compare your requirements with the attributes of each solution being considered
- Internal Load Balancing (Sysplex Distributor)
  - Well suited to non-web traffic
  - Built on dynamic VIPA support for simple implementation and maximum availability
  - No new equipment required, maintains administrative controls within z/OS
- External Load Balancing Solutions
  - Typically better suited for workload distribution of Web traffic
    - Allows distribution to outboard caches as well as target z/OS systems
  - Can handle non-web traffic and non z/OS platforms
  - May or may not require additional equipment
  - Administrative functions typically performed by network administrator
  - Selecting a solution that supports SASP can greatly improve the quality of the load balancing decision
    - Using the z/OS Load Balancing Advisor significantly improves visibility into the Sysplex environment and the current workload conditions
    - SASP will be an open architecture so it is expected that there will be a wider range of load balancer choices in the future

Additional Information
New IP Load Balancing white paper

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
- You can download the white paper as a PDF file from:
  - http://www-306.ibm.com/software/network/commserver/os390/support/ - select White Papers
- A future update of the white paper may include test scenarios using SASP and the z/OS Load Balancing Advisor

For More Information....

URL	Content
http://www.ibm.com/servers/eserver/zseries	IBM eServer zSeries Mainframe Servers
http://www.ibm.com/servers/eserver/zseries/networking	Networking: IBM zSeries Servers
http://www.ibm.com/servers/eserver/zseries/networking/technology.html	IBM Enterprise Servers: Networking Technologies
http://www.ibm.com/software/network/commserver	Communications Server product overview
http://www.ibm.com/software/network/commserver/zos/	z/OS Communications Server
http://www.ibm.com/software/network/commserver/z_lin/	Communications Server for Linux on zSeries
http://www.ibm.com/software/network/ccl	Communication Controller for Linux on zSeries
http://www.ibm.com/software/network/commserver/library	Communications Server products - white papers, product documentation, etc.
http://www.redbooks.ibm.com	ITSO redbooks
http://www.ibm.com/software/network/commserver/support	Communications Server technical Support
http://www.ibm.com/support/techdocs/	Technical support documentation (techdocs, flashes, presentations, white papers, etc.)
http://www.rfc-editor.org/rfcsearch.html	Request For Comments (RFC)

Appendix- Additional Information

CINET Considerations

- Advisor
  - AUTOLOG on multiple TCP/IP stacks may cause the Advisor to start on an unpredictable stack since only one instance may be active in the sysplex.
  - Defining listening sockets/ports
    - If using a unique application-instance DVIPA (recommended), all TCP/IP stacks on that system must code the VIPARANGE statement for that DVIPA
      - Using optional (but not recommended) stack affinity to an application-instance DVIPA instead may be done but can be administratively difficult
    - Both the IPv4 and IPv6 Load Balancer listening sockets should belong to the same TCP/IP stack to keep administration simple
- Agent
  - Address in *host_connection* statements should be DVIPAs
  - If using unique application-instance DVIPAs, same recommendations as Advisor
  - Optional (but not recommended) stack affinity will restrict workload balancing to that stack. Remainder of stacks on that system will not participate in workload balancing.

Operator Commands

Function Externals
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 10.67.1.10
EZD1263I LBADV LOAD BALANCER CONNECTED FROM 10.67.1.11

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 10.67.1.2
EZD1259I LBAGENT CONNECTED TO ADVISOR 10.67.5.1
```


The Advisor and Agent can be started in any order.
LBADV is an alias of the sample start procedure, EZBLBADV

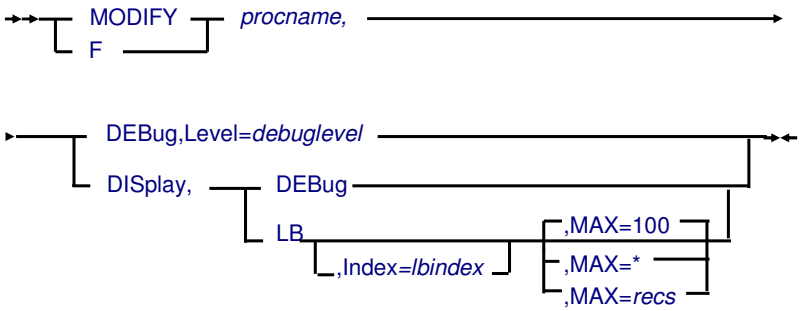
LBAGENT is an alias of the sample start procedure, EZBLBAGE
This example shows...
One LB connecting to the Advisor from 10.67.1.11

Two Agents connecting to the Advisor

One remotely from 10.67.1.10

One locally from 10.67.1.2
Both the Agent and Advisor issue messages when they successfully connect to each other
Netstat displays can be used to determine which Agents have active connections to the Advisor (not shown)

Function Externals
Advisor Console Command Syntax



Advisor Console Command Examples (Displaying Status Information)

```
 1 F LBADV,DISP,LB
 2 EZD1242I LOAD BALANCER SUMMARY
 3 LB INDEX      : 00        UUID      : 637FFF175C
 4  IPADDR..PORT : 10.42.154.105..50005
 5  HEALTH       : 20        FLAGS     : NOCHANGE PUSH TRUST
 6 LB INDEX      : 01        UUID      : 207FFF175C
 7  IPADDR..PORT : 10.42.154.160..50006
 8  HEALTH       : 7F        FLAGS     : PUSH TRUST
 9 2 OF 2 RECORDS DISPLAYED

10
11 F LBADV,DISP,LB,I=0
12 EZD1243I LOAD BALANCER DETAILS
13 LB INDEX      : 00        UUID      : 637FFF175C
14  IPADDR..PORT : 10.42.154.105..50005
15  HEALTH       : 20        FLAGS     : NOCHANGE PUSH TRUST
16  GROUP NAME   : SYSTEMFARM
17   GROUP FLAGS : BASEWLM
18   IPADDR..PORT: 10.42.105.154..0
19    SYSTEM NAME: MVS209    PROTOCOL  : 000  AVAIL     : YES
20    WLM WEIGHT : 00021     CS WEIGHT : 100  NET WEIGHT: 00001
21    FLAGS      :
22   IPADDR..PORT: 10.42.105.60..0
23    SYSTEM NAME: VIC007    PROTOCOL  : 000  AVAIL     : YES
24    WLM WEIGHT : 00045     CS WEIGHT : 100  NET WEIGHT: 00002
25    FLAGS      :
... continued on following page ...
```

Function Externals
Advisor Console Command Examples

```
... continued from previous page ...

 IPADDR..PORT: 10.42.105.22..0
  SYSTEM NAME: N/A       PROTOCOL  : 000  AVAIL     : NO
  WLM WEIGHT : 00000     CS WEIGHT : 000  NET WEIGHT: 00000
  FLAGS      : NOTARGETSYS
 IPADDR..PORT: 10:1::4:5..0
  SYSTEM NAME: MVS209    PROTOCOL  : 000  AVAIL     : NO
  WLM WEIGHT : 00021     CS WEIGHT : 000  NET WEIGHT: 00000
  FLAGS      : NOTARGETIP
GROUP NAME   : UDP_SERVER_FARM
 GROUP FLAGS : SERVERWLM
 IPADDR..PORT: 10.42.105.154..7777
  SYSTEM NAME: MVS209    PROTOCOL  : UDP  AVAIL     : YES
  WLM WEIGHT : 00021     CS WEIGHT : 100  NET WEIGHT: 00001
  FLAGS      :
 IPADDR..PORT: 2001:DB8::10:5:6:2..7777
  SYSTEM NAME: MVS209    PROTOCOL  : UDP  AVAIL     : YES
  WLM WEIGHT : 00021     CS WEIGHT : 100  NET WEIGHT: 00001
  FLAGS      :
 IPADDR..PORT: 10.42.105.60..7777
  SYSTEM NAME: VIC007    PROTOCOL  : UDP  AVAIL     : YES
  WLM WEIGHT : 00045     CS WEIGHT : 100  NET WEIGHT: 00002
  FLAGS      :
7 OF 7 RECORDS DISPLAYED
```

Notes - Advisor Console Command Examples

NOTES

- Line 1: Command issued on the Advisor system to display all LBs connected to the Advisor
  - Two LBs displayed, one starting on line 3 and one on line 6
  - Line 3: LB with UID=637FFF175C was assigned an LB Index of '00' by the Advisor
  - Line 4: The LB connected from 10.42.154.105, port 50005
  - Line 5: Shows information set by the LB
    - 'Health' is assigned by the LB and not used by the z/OS Load Balancing Advisor
    - 'Flags' are set by the LB
      - 'NOCHANGE' indicates the LB is requesting that only new and changed data be sent to it rather than a complete set of data
      - 'PUSH' indicates that the LB requests the Advisor to periodically send it data rather than to have to request it
      - 'TRUST' is not used by the z/OS Load Balancing Advisor
  - Line 9: Indicates there are two LBs connected presently

- Line 11: Command issued on the Advisor system to display details about the LB which is assigned the LB Index of '00'
  - Line 16: Shows a group by the name of 'SYSTEMFARM' which this LB registered. This group is a system group as opposed to an application group
    - Line 17: Flags for this group. BASEWLM indicates group is using System WLM recommendations rather than Server-specific WLM recommendations
    - Line 18: Shows the first member of the group which is at 10.42.105.154, port 0
    - Line 19: Details on this member
      - SYSTEM NAME is the MVS name of the system where the member resides
      - PROTOCOL is 0, requirement of a system member
      - AVAIL = YES indicates the member is available for load balancing
    - Line 20: More details on the member
      - WLM returned a system weight of 21 on a scale of 0-64 for system, MVS209. Higher weights mean higher capacity.

Notes - Advisor Console Command Example (Continued)

NOTES

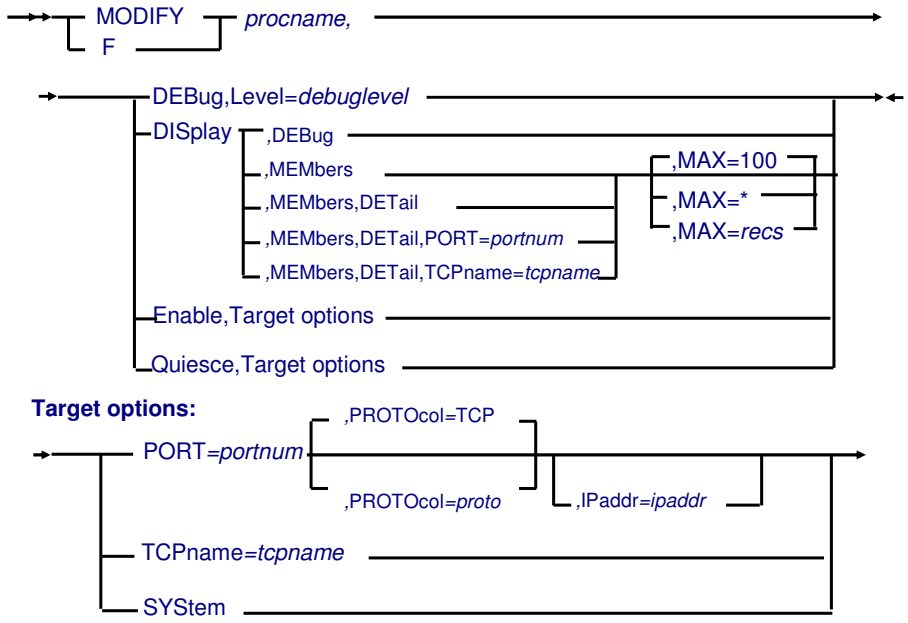
- Line 11: Continued, display details of LB with Index of 00
  - Line 16: Continued, Group, SYSTEMFARM
    - Line 20: Continued, Member IP =10.42.105.154, Port 0, Protocol 0
      - The Agent returned a Comm Server Weight (CS Weight) of 100. For system members, this is always 100.
      - The Net Weight which will be sent to the load balancer is 1. This is the relative capacity of the member when compared to the other Net Weights in the group.
      - Line 21: Flags for this member. Flags are only displayed if they are on. No flags are on for this member. Possible flags are...
        - LBQ - meaning the load balancer administrator quiesced the member
        - OPQ - meaning the Agent operator quiesced the member
        - NOTARGETSYS - meaning no Agents have reported any information on this member
        - NOTARGETIP - meaning this is an invalid system member
        - NOTARGETAPP - meaning this is an invalid application member
        - NODATA - meaning not enough time has elapsed for the Agent to calculate a CS Weight for the member
    - Line 26: Shows the start of the member with IP address 10.42.105.22
      - Line 27: AVAIL = NO indicates the member is unavailable
      - Line 28: Unavailable members will have a Net Weight of 0
      - Line 29: Flag of NOTARGETSYS indicates this member has not been reported by any Agent
    - Line 33: Flag of NOTARGETIP indicates this member is not a valid system member. See bullet, "Should NOT contain the following types of addresses" very early in the presentation for types of IP addresses that can cause this

Notes - Advisor Console Command Example (Continued)

NOTES

- Line 11: Continued, display details of LB with Index of 00
  - Line 34: Begin display of the group, UDP_SERVER_FARM
    - Line 35: Group Flags = serverwlm indicates server-specific WLM recommendations are used for this group instead of system WLM weights
    - Line 36: Port is non-zero (7777) and protocol is non-zero (line 37), so these are application members and not system members
    - Lines 36, 40, and 44: Each represents a member. So this group contains three members.
    - Lines 37 and 41: Both members are on the same MVS system, MVS209, using the same port, but different addresses. These members represent two applications bound to different addresses.
    - Lines 38, 42, 46: Net Weights of 1, 1, and 2, respectively indicate the last member will get 50% of the connections, while the other two will get 25% each.
  - Line 48: Indicates 7 members were displayed

Function Externals
Agent Console Command Syntax



Target options:

Function Externals
Agent Console Command Examples

```
 1  F LBAGENT,DISP,MEM,DET
 2  EZD1245I MEMBER DETAILS
 3  LB INDEX      : 00        UUID      : 637FFF175C
 4   GROUP NAME   : SYSTEMFARM
 5    IPADDR..PORT: 10.42.105.154..0
 6     TCPNAME    : TCPCS     MATCHES   : 001  PROTOCOL  : 000
 7     FLAGS      :
 8     JOBNAME    : N/A       ASID      : N/A  RESOURCE  : N/A
 9    IPADDR..PORT: 10:1::4:5..0
10     TCPNAME    : TCPCS5    MATCHES   : 001  PROTOCOL  : 000
11     FLAGS      :
12     JOBNAME    : N/A       ASID      : N/A  RESOURCE  : N/A
13   GROUP NAME   : UDP_SERVER_FARM
14    IPADDR..PORT: 10.42.105.154..7777
15     TCPNAME    : TCPCS     MATCHES   : 001  PROTOCOL  : UDP
16     FLAGS      : ANY
17     JOBNAME    : TESTD1    ASID      : 0035 RESOURCE  : 000000A3
18    IPADDR..PORT: 2001:DB8::10:5:6:2..7777
19     TCPNAME    : TCPCS2    MATCHES   : 001  PROTOCOL  : UDP
20     FLAGS      : ANY V6
21     JOBNAME    : TESTD2    ASID      : 002A RESOURCE  : 00000031
22  4 OF 4 RECORDS DISPLAYED
```

Notes - Agent Console Command Example

NOTES

- Line 1: Command issued on an Agent system to display details about all of the members owned by this Agent
  - Line 3: This Agent owns at least one member registered by the LB with an Index of 00.
    - Line 4: This Agent owns at least one member in the group, SYSTEMFARM
      - Line 5: This Agent owns the member described on this line. A port of 0 indicates this is a system member.
      - Lines 6, 7, and 8: Contain detailed member information
        - TCPNAME indicates the TCP/IP stack named TCPCS owns the IP address for this member
        - MATCHES is a count of how many applications are listening on this port. Will be greater than 0 only when the member represents a port sharing group
        - PROTOCOL is zero because this is a system member
        - FLAGS is empty because none of the flags are on. Possible values are...
        - -- ANY indicates the target is bound to the unspecified address (0.0.0.0 or ::)
        - -- V6 indicates the target application was bound using the IPV6_V6ONLY socket option
        - -- APPQ indicates the target application is quiesced
        - -- TCPQ indicates the target TCP/IP stack is quiesced
        - -- SYSQ indicates the target system is quiesced
        - JOBNAME, ASID, and RESOURCE are N/A because this is a system member
    - Line 13: This agent owns at least one member in the group, UDP_SERVER_FARM
      - Line 14: This Agent owns the member described on this line. A non-zero port indicates this is an application member.
        - Line 16: FLAGS = ANY indicates this application is bound to the unspecified address
        - Line 17:
        - -- JOBNAME = TESTD1 is the jobname the application is running under
        - -- ASID = 0035 is the address space ID the application is running in
        - -- RESOURCE = 000000A3 is the resource ID assigned to the application by the TCP/IP stack
  - Line 22: Indicates 4 members are owned by this Agent

Quiescing/Enabling Workload Distribution (Notes)

- Temporal issues of Quiesce (inheritance)
  - Quiesce at SYSTEM and TCPNAME (stack) levels apply to existing members and any newly registered members that fall within that scope
  - Quiesce at the PORT level will apply to newly registered members only if a member matching the Quiesce criteria currently exists at the time of the command

- Rejected Quiesce/Enable commands
  - Whole command is rejected if it fails for a single member

- MVS Quiesce/Enable Interactions
  - Quiesce/Enable are hierarchical
    - SYSTEM = highest
      - TCPNAME (stack) = middle
        - PORT (member) = lowest
  - Enable must be done at the same level of the most recent Quiesce which affected the member
  - Already quiesced member may only be quiesced at a higher level
    - Corollary: Quiesce will be rejected if any affected member is already quiesced at a higher level
    - Quiesce level is "promoted" to the higher level if accepted

Quiescing/Enabling Workload Distribution (Notes)...

- Quiesce/Enable sequences allowed according to scope
  - The following table shows which Quiesce and Enable commands are valid if a prior Quiesce command had affected one of the same members
  - A "dot" at the intersection of a row and column indicates the combination is valid
    - Example: A 'Quiesce,Tcpname' command is issued which affects member 'A' (middle column). If a subsequent Quiesce or Enable command were issued which also affected member 'A', only the 'Quiesce,System' command (first row) and the 'Enable,Tcpname' command (5th row) would be allowed.
  - This table summarizes the rules under "MVS Quiesce/Enable Interactions" bullet on the previous page.

		Prior Command		
		Quiesce,System	Quiesce,Tcpname	Quiesce,Port
Current Command	Quiesce,System		.	.
	Quiesce,Tcpname			.
	Quiesce,Port			
	Enable,System	.		
	Enable,Tcpname		.	
	Enable,Port			.

Diagnosis

Diagnosis - Data to Collect for Problems

- syslogd files
  - Essential for debugging problems.
  - Examples for /etc/syslog.conf

```
*.lbadv.daemon.* /tmp/lbadv.log
*.lbagent.daemon.* /tmp/lbagent.log
```

- Console log

- Dumps (data set names are specified in started procedure)
  - CEEDUMP
  - CEESNAP
  - SYSMDUMP

- For configuration problems, send the Advisor and Agent configuration files and the related statements in the TCP/IP profile.

- For connectivity and other problems between the Load Balancer and Advisor, send the TCP/IP CTRACE (with the TC option) and Packet trace.

Dump Dataset Notes

- CEEDUMP and CEESNAP datasets
  - Used for certain unexpected error conditions when the program itself decides to abnormally terminate
- SYSMDUMP dataset
  - Used when the program is terminated abnormally by the operating system (e.g. 0Cx abend)
  - May use CEEDUMP and CEESNAP instead if TERMTHDACT Language Environment runtime option is overridden by installation or start procedure
    - Overriding TERMTHDACT is not recommended

Function Externals
Logging

- Advisor and Agent both log to syslogd, daemon facility
- syslogd must be configured and started before starting the Advisor or Agent
- See "Diagnosis" section for suggested syslogd configuration
- Default debug level (7) should normally be used unless problem documentation needs to be gathered
  - Increased levels of debug may result in voluminous amounts of information

Trademarks and notices

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

  - AIX®
  - AnyNet®
  - AS/400®
  - Candle®
  - CICS®
  - CICSPlex®
  - CICS/ESA®
  - DB2®
  - DB2 Connect™
  - DPI®
  - DRDA®
  - e business(logo)®
  - ESCON®
  - eServer™
  - ECKD™
  - FFST™
  - GDDM®
  - GDPS®
  - HiperSockets™
  - IBM®
  - Infoprint®
  - IMS™
  - IP PrintWay™
  - iSeries™
  - Language Environment®
  - MQSeries®
  - MVS™
  - MVS/ESA™
  - NetView®
  - OS/2®
  - OS/390®
  - Parallel Sysplex®
  - PrintWay™
  - PR/SM™
  - pSeries®
  - RACF®
  - Redbooks™
  - Redbooks (logo)™
  - S/390®
  - System/390®
  - ThinkPad®
  - Tivoli®
  - Tivoli (logo)®
  - VM/ESA®
  - VSE/ESA™
  - VTAM®
  - WebSphere®
  - xSeries®
  - z/Architecture™
  - z/OS®
  - z/VM®
  - zSeries®

- Cisco, Cisco Systems, the Cisco Systems logo, Catalyst, and Cisco IOS are registered trademarks or trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.