Leveraging DataPower XI50z and Sysplex

Section 9249
August 10, 2011
STG Lab Services
Thomas Cosenza, CISSP
tcosenza@us.ibm.com
What is a Sysplex?

- Sysplex is a clustering technology utilized historically by z/OS
  - You can spread a workload over several LPARs
  - In case of an LPAR failure the workload can be moved without the end users knowledge (in most cases)
- Sysplex uses metrics to balance out workload (not just a spray)
  - CPU Usage
  - Current Capacity
  - Current Load
- Allows applications on z/OS to be dependent on the physical connections
The z/OS High Availability Architecture

Application Layer

Infrastructure
- DB2
- WAS
- CICS
- MQ
- RRS
- ARM

Systems
- WLM
- SYSPLEX
- Coupling Facilities

Networking
- Workload Distribution Techniques
- Dynamic VIPA
- IP Routing
- Network Interfaces
- SFM

Redundancy
What is the XI50z DataPower Device

- Part of zEnterprise
- Secure Appliance
- High Speed Cryptography
- Can secure just parts of data
  - Credit Card PAN
  - SSN numbers
- Built in ESB Device used in SOA
DataPower and zEnterprise

- zEnterprise’s main strength
  - Secure Environment
  - Hybrid Work Load Support
- DataPower XI50z
  - Allows zEnterprise components to communicate without application changes
    - Legacy applications
    - Web Services
    - XML
    - HTTP
Sysplex Enhancement

- Sysplex design originally was only for z/OS systems
- SOA and Cloud are Game Changers
  - Hybrid-Workload support
- In V1R11 the Sysplex could now target IPv4 devices outside of System z
  - IPv6 Devices came in V1R12
- One of the first targets for the Sysplex was DataPower
Open 24/7

- DataPower XI50z does not have a routing daemon
  - Default Gateway
  - Static Routes
- While this works in the distributed world in z we must have almost 100% up time
- If you do not want to connect your IEDN to an external router what do you do? Target OSX on z/OS as default gateway or static route hop?
  - LPAR Failure
  - Stack Abend
  - OSX failure
Going back on what was a best practice

• In the past it was considered a best practice to put Dynamic Vipas on different subnets then your OSA card.
  • Still is for the OSD and OSE cards
• However for IEDN devices it is advisable to actually do the opposite
  • Layer 2 routing
  • Allows us to use a DVIPA as the Default Gateway to the XI50z (or any other zBX type blade)
• I have nicknamed these “IEDN VIPAs”
  • Not an official IBM name just my nickname
  • Same as other type of VIPA however
Way 1. Infrastructure DVIPA

- DataPower does not support any routing protocols
  - Static Routes
  - Default Gateway Routing
- Would you really want to?
  - Want more CPU dedicated for Transactions
- So use a DVIPA with the same IP Address Subnet as the IEDN subnet
  - Using LAYER 2 Routing
What is Layer 2 Routing

- The LAN infrastructure transports “Frames” between Network Interface Cards (NICs)
- Each NIC has a physical hardware address –called Media Access Control (MAC)
- Every frame comes from a MAC and goes to a MAC
- A frame carries a payload of a specified protocol type, such as ARP, IPv4, IPv6, SNA LLC2, etc.
- Uses a Protocol Called ARP in order to discover other MAC address and their corresponding IPv4 addresses
So let's look at this process

1. **Who owns IPv4 address 10.1.1.3??**

2. **I do and my MAC address is MAC3!!**

3. **Here is a unicast packet for 10.1.1.3 at MAC3**

**ARP Cache**

<table>
<thead>
<tr>
<th>IP address</th>
<th>MAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1.1.2</td>
<td>MAC2</td>
</tr>
<tr>
<td>10.1.1.3</td>
<td>MAC3</td>
</tr>
</tbody>
</table>
z/OS VIPA address visibility on the IEDN

- OSX interfaces must be defined with the INTERFACE statement.
- With VMAC and ROUTEALL, only addresses for which OSA has to perform ARP are registered in the OAT.
- In all other cases, all HOME IP addresses will be registered in the OAT and the OAT content will be changed as the HOME lists change due to (dynamic) movement of IP addresses.
- OSX interfaces will do gratuitous ARP for the OSA interface IP address and for VIPA addresses that belong to the same subnet as the OSA interface.
Network connectivity resilience on the IEDN

z/OS TCP/IP supports interface recovery if multiple network interfaces to the same subnet exist. In this example, both OSA PORTA and PORTB are connected to the IEDN (10.1.1.0/24 subnet).

When PORTA fails, PORTB is given ARP ownership of the addresses PORTA previously had. PORTB sends gratuitous ARPs to enable downstream hosts to update their ARP cache.

When PORTA fails, PORTB is given ARP ownership of the addresses PORTA previously had. PORTB sends gratuitous ARPs to enable downstream hosts to update their ARP cache.

<table>
<thead>
<tr>
<th>IP Address</th>
<th>MAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1.1.1</td>
<td>MAC1</td>
</tr>
<tr>
<td>10.1.1.2</td>
<td>MAC2</td>
</tr>
<tr>
<td>10.1.1.10</td>
<td>MAC1</td>
</tr>
</tbody>
</table>

OSA PORTA’s OAT

<table>
<thead>
<tr>
<th>IP</th>
<th>ARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1.1.1</td>
<td>Yes</td>
</tr>
<tr>
<td>10.1.1.10</td>
<td>Yes</td>
</tr>
<tr>
<td>10.1.1.2</td>
<td>No</td>
</tr>
<tr>
<td>10.1.3.1</td>
<td>No</td>
</tr>
</tbody>
</table>

OSA PORTB’s OAT

<table>
<thead>
<tr>
<th>IP</th>
<th>ARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1.1.1</td>
<td>No</td>
</tr>
<tr>
<td>10.1.1.10</td>
<td>No</td>
</tr>
<tr>
<td>10.1.1.2</td>
<td>Yes</td>
</tr>
<tr>
<td>10.1.3.1</td>
<td>No</td>
</tr>
</tbody>
</table>

XCF IP address: 10.1.3.1

IEDN-Host-1’s ARP cache

<table>
<thead>
<tr>
<th>IP Address</th>
<th>MAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1.1.1</td>
<td>MAC2</td>
</tr>
<tr>
<td>10.1.1.2</td>
<td>MAC2</td>
</tr>
<tr>
<td>10.1.1.10</td>
<td>MAC2</td>
</tr>
</tbody>
</table>

IEDN-Host-1’s ARP cache

<table>
<thead>
<tr>
<th>IP Address</th>
<th>MAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1.1.1</td>
<td>MAC2</td>
</tr>
<tr>
<td>10.1.1.2</td>
<td>MAC2</td>
</tr>
<tr>
<td>10.1.1.10</td>
<td>MAC2</td>
</tr>
</tbody>
</table>

z/OS TCP/IP supports interface recovery if multiple network interfaces to the same subnet exist. In this example, both OSA PORTA and PORTB are connected to the IEDN (10.1.1.0/24 subnet).

TEPIPA

VIPA 10.1.1.10

CIDR: 10.1.1.0/24

<table>
<thead>
<tr>
<th>Port name</th>
<th>PORTA</th>
<th>Port name</th>
<th>PORTB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home IP</td>
<td>10.1.1.1</td>
<td>Home IP</td>
<td>10.1.1.2</td>
</tr>
<tr>
<td>Mac address</td>
<td>MAC1</td>
<td>Mac address</td>
<td>MAC2</td>
</tr>
</tbody>
</table>

OSA PORTA fails

When PORTA fails, PORTB is given ARP ownership of the addresses PORTA previously had. PORTB sends gratuitous ARPs to enable downstream hosts to update their ARP cache.
So let's look at how the infrastructure DVIPA would work.
zBX Backend Application

- This recommendation is much like the last however the focus has changed
- Instead of routing of the blade
- We will now look at Backend Applications
What if DVIPA was on a different subnet

External customer network
Any subnet/prefix

DVIPA 155.155.1.1 VIPADEFINED
DVIPA 155.155.1.1 VIPABACKUP

DVIPA 10.1.1.210 VIPADEFINED
DVIPA 10.1.1.210 VIPABACKUP

OSX

OSD

OSX

OSD

OSX

OSD

OSD

10.1.1.1 10.1.1.2 10.1.1.3 10.1.1.4

10.1.1.100

10.1.1.210

192.168.1.1 192.168.1.2

10.1.1.100

Default router: 10.1.1.210

Transaction to 155.155.1.1

Default router: 10.1.1.210

Transaction to 155.155.1.1
What if DVIPA was on a different subnet

External customer network
Any subnet/prefix

DVIPA 10.1.1.200
VIPABACKUP

DVIPA 10.1.1.200
VIPADEFINED

OSD
OSD

OSX
OSX

192.168.1.1
192.168.1.2

10.1.1.1
10.1.1.2
10.1.1.3
10.1.1.4

Default router: 10.1.1.210

Transaction to 155.155.1.1
3) Tier Support for DataPower

- Communication Server has added support to expand targets beyond zOS
- DataPower was an early adapter of this technology
- DataPower can give Sysplex metrics so that Sysplex can make smart decision on what DataPower Device it should target
IEDN Default VLAN ID IPv4 subnet 10.1.1.0/24

10.1.1.1 10.1.1.2 10.1.1.3 10.1.1.4

OSX

OSD

External customer network Any subnet/prefix

DVIPA 155.155.1.1 VIPADEFINE

192.168.1.1 192.168.1.2

OSD

OSD

OSX

OSX

10.1.1.1

10.1.1.2

10.1.1.3

10.1.1.4

SDA

SDA

SDA

SDA

Data Power Tier

VIPADISTRIBUTE TIER1 GROUP1
GRE CONTROLPORT 1702
DISTMETHOD TARGCONTROLLED
155.155.1.1 Port 10000
DESTIP 10.1.1.100 10.1.1.101
10.1.1.102 10.1.1.103

IEDN Default VLAN ID IPv4 subnet 10.1.1.0/24
What if my Sysplex was larger

- What if you deployed an Ensemble within two zEnterprise systems that are physical separated?

![Diagram showing VLAN IDs, IP addresses, and network components.]
Multi-Tier Weighting

- As part of the Sysplex we can now look at the whole transaction.
- Very important when talking about DataPower since it is rarely the end point for any transaction.
- Sysplex Distributor can add up the weights of any defined Tier 1 and Tier 2 (backend) connection.
External customer network
Any subnet/prefix

VLAN ID 100
10.1.1.0/24

10.1.1.1 10.1.1.2 10.2.1.3 10.2.1.4

OSX

OSD

External customer network
Any subnet/prefix

VLAN ID 101
10.2.1.0/24

10.1.1.100
10.1.1.101
10.2.1.100
10.2.1.101

SDA

DVIPA 155.155.1.1
VIPADefined

DVIPA 10.1.1.210

OSX

DB2

80
20
10

SDA

DVIPA 10.2.1.210

OSX

DB2

90
10
10

SDA

Say all DP weights are equal
The Transactions will never leave the zEnterpise on the left since the zEnterpise on the Left has better weights on the DataPower devices.
VLAN ID 100 10.1.1.0/24

External customer network
Any subnet/prefix

VLAN ID 101 10.2.1.0/24

VIPADEFINE TIER1
255.255.255.0 155.155.1.1
VIPADEFINE TIER2 CPCSCOPE
255.255.255.0 10.1.1.0

VIPADEFINE TIER1 GROUP1
GRE CONTROLPORT 1702
DISTMETHOD TARGCONTROLLED
155.155.1.1 Port 10000

DESTIP 10.1.1.100 10.1.1.101 10.2.1.100 10.2.1.101

VIPADEFINE CPCSCOPE
255.255.255.0 10.1.1.0
VIPADEFINE TIER2 CPCSCOPE
255.255.255.0 10.1.1.0
VIPADEFINE TIER2 GROUP1
DISTMETHOD SERVERWLM
10.2.1.210 Port 10002

DESTIP ALL
Questions?

Thomas Cosenza
System z I/T Specialist
IBM STG Lab Services
XI50z Team Lead

IBM
3031 N Rocky Point DR
Tampa, FL 33607-5878
Tel 720-395-7392
Mobile 813-270-9911
Email: tcosenza@us.ibm.com