Intelligent Server Load Balancing in IBM 390 Environments: A Cisco/IBM Alliance Case Study

Session 2700
Agenda

- What Is Server Load Balancing?
- MultiNode Load Balancing Overview, Components, and Protocols
- MultiNode Load Balancing Operation
- MultiNode Load Balancing Implementation
- MultiNode Load Balancing Roadmap
- The Cisco/IBM Test

What Is Server Load Balancing?
**What Is Load Balancing the Data Center?**

- Continuous availability
- Distribution of workload equally across servers
- Simplifying the process of client connections

**Limitations with Today’s Solutions**

- Restricted scalability for unknown growth
- No provision for continuous availability
- Disruptive session failover
- Single point of failure solutions
- Limited server availability knowledge (external)
- Limited data throughput
- No Sysplex support for an IP network in S/390
Why MultiNode Load Balancing?

- MultiNode Load Balancing has no single point of failure and will allow traffic to traverse multiple routers.
- MultiNode Load Balancing is Cisco IOS® based on both routers and switches providing highest performance and scalability.
- MultiNode Load Balancing architecture provides for linear scaling of routers/switches allowing for network growth to be dynamically activated.
- A new protocol provides server feedback to increase the intelligence of load balancing decisions.

MultiNode Load Balancing Overview, Components, and Protocols
MultiNode Load Balancing

- MultiNode Load Balancing is a Cisco hardware and Cisco IOS software solution that distributes IP traffic across server farms and IBM mainframes.
- MultiNode Load Balancing is available in most router and switch-based platforms.
- MultiNode Load Balancing leverages LocalDirector capabilities for load balancing and server control.
- MultiNode Load Balancing uses server feedback for load balancing decisions based on application status.

Components and Protocols

Components
- MultiNode Load Balancing Services Manager (MNLB SM)
- Backup Services Manager (MNLB BSM)
- Forwarding Agent (FA)
- Workload Agent (WLA)

Protocols
- MultiNode Load Balancing architecture
  - Cisco Application Services Architecture (CASA)
  - Dynamic Feedback Protocol (DFP)
Components

- **MultiNode Load Balancing Services Manager (MNLB SM)**
  
  Makes all Load Balancing decisions
  
  Uses server and application feedback to make balancing decisions
  
  Also, may use existing LocalDirector Load Balancing Algorithms
  
  Tells the forwarding agent destination of connections

- **MultiNode Load Balancing Backup Services Manager**
  
  Supports the Services Manager for MNLB SM failures
  
  Provides the same functionality as the MNLB SM

Components (Cont.)

- **Forwarding agent**
  
  Gathers information multicasted from the services manager
  
  Advertises the virtual addresses to the IP network
  
  Caches wildcards and creates affinities
  
  Communicates to the services manager on behalf of a client requesting a server and during session flow
  
  Directs the client to the best server

- **Workload agent/s**
  
  Monitors the resources per server
  
  Supplies the services manager/s with the server application port availability and host workload (WLM) information
Protocols

- Cisco Application Service Architecture (CASA)—part of ContentFlow architecture
  Communication protocol between the SM/s and FA/s
  Transports the wildcards
  Transports the action

- Dynamic Feedback Protocol (DFP)
  Communication protocol between the WLA/s and the SM/s
  Passes WLA load balancing data to the SM for inclusion to the load balancing algorithm

MultiNode Load Balancing Operation
Steps of Load Balancing

- Step 1. Configuration
- Step 2. Client request server
- Step 3. Forwarding agent intercepts request
- Step 4. Request forwarded to server
- Step 5. Server responds to client
- Step 6. Session begins

Network Configuration—Initial

- Services Manager
  - Real Addresses:
    - x.x.x.1
    - x.x.x.2
    - x.x.x.3
    - x.x.x.4
  - Virtual Address: x.x.x.100
- Forwarding Agent
- Wildcard Cache: x.x.x.100
- Source Address: x.x.x.20
- IGMP Multicast
Client Requests Server

- Note: Prior to the connection request;
  Services manager and forwarding agents have been configured
  The FA/s have learned the virtuals from the MNLB SM and will advertise the virtual addresses to the network
  MNLB SM locates WLAs and receives server operating and application information
- Client sends a request to connect to a server
  Clients sends a TCP/IP SYN packet, the SYN packet is a request for connection

Load Balancing Service (LBS) Intercepts Request

- The FA receives the SYN packet, the packet matches the Wildcard list so the FA requests the MNLB SM for the “best” server using CASA
- MNLB SM chooses the best server based on;
  Existing LocalDirector algorithms
  Workload agent information input
  Dynamic Feedback Protocol (DFP)
  Server probe feedback
- The SM informs the FA, using CASA, of the address of the server selected as the “best” server
Request Forwarded to Server

- The forwarding agent forwards the clients' SYN packet to the server using Layer 2 (dispatch mode)
  - Uses ARP to obtain Layer 2 address
  - IP header still contains virtual IP address
- The server will receive the client’s SYN packet, realize the destination address is the loopback/real address of that server and accept the connect request

Network Configuration—Contact Server

- Forwarding Agent
- Wildcard Cache: x.x.x.100
- Affinity Cache:
  - Source .20
  - Destination .100
  - Real .1
- Services Manager
- Affinity Cache:
  - Source .20
  - Destination .100
  - Real .1
- Source Address: x.x.x.20
Server Responds to Client

- Server sends a connect acknowledgement back to the client
- As the packet passes through the FA, the type of packet is “listened” for and packets such as SYN, ACK, RST and FIN are noted back to the SM for client connection tracking
- Note: The response back to the client can pass back out any router/switch that has enabled the forwarding agent feature

Session Begins

- Session begins and packets flow until completion or an error
- Fault tolerance
- Listening actions
  As session traffic flows the FA keeps the SM aware of any RST or FINs from the client
Network Configuration—Traffic Flows

MultiNode Load Balancing Implementation
Commands to Define MNLB: Services Manager

casa service-manager multicast-ttl value Configures the Services Manager multicast connection time-to-live value.

virtual virtual_name | virtual_ip [:virtual_port]:[bind-id]:[protocol] [service-state] Defines the virtual servers.

real real_name | real_ip[:[port]:[bind-id]:[protocol]] [service-state] Defines the real servers.

bind virtual_id real_id [real_id...] Associates the virtual server the real server and establishes the load-balancing relationship between the virtual and real servers.

redirection virtual_id dispatched casa wildcard-ttl value fixed-ttl value igmp multicast-group-address port port Specifies the type of load balancing redirection for the virtual server.

predictor virtual_name | virtual_ip [:virtual_port]:[bind-id]:[protocol] [fastest | roundrobin | leastconns | loaded | weighted [roundrobin | none] Sets the load balancing algorithm for a virtual server.

route forwarding_agent_control_address 255.255.255.255 forwarding_agent_lan_address 1

dynamic-feedback ip_address:port [retry retry] [attempts attempts] [timeout timeout]

Commands to Define MNLB: Forwarding Agent

ip cef distributed Enables CEF.

ip igmp join-group multicast-group-address Interface command to join an igmp multicast group.

ip casa control-address igmp-address Specify the IP address and IGMP address of the forwarding agent. The recommended IGMP address is 224.0.1.2.

forwarding-agent num [passwd [timeout]] Specify the port number. The default is port 1637.

show ip casa affinities

show ip casa stats

show ip casa wildcard

show ip casa oper
### MNLB: Putting it Together

#### Forwarding Agent 1
```
ip flow-cache feature-accelerate
ip cef distributed
ip casa 206.10.20.1 224.0.1.2
Forwarding-agent 1637
interface Ethernet0/0
no ip directed-broadcast
ip route-cache flow
ip igmp join-group 224.0.1.2
no ip mroute-cache
```

#### Forwarding Agent 2
```
ip flow-cache feature-accelerate
ip cef distributed
ip casa 206.10.20.2 224.0.1.2
Forwarding-agent 1637
interface Ethernet0/0
no ip directed-broadcast
ip route-cache flow
ip igmp join-group 224.0.1.2
no ip mroute-cache
```

#### Services Manager
```
ip address 172.26.56.30 255.255.255.0
route 206.10.20.2 255.255.255.255 172.26.56.20 1
route 206.10.20.1 255.255.255.255 172.26.56.10 1
casa service-manager port 1638
casa service-manager multicast ttl 60
virtual 172.26.57.100:0:0:tcp is
    redirection 172.26.57.100:0:0:tcp directed basica wildcard t6 60 fixed ttl 60 igmp
    real 172.26.56.1:0:0:tcp is
    real 172.26.56.2:0:0:tcp is
    bind 172.26.57.100:0:0:tcp 172.26.56.1:0:0:tcp
    bind 172.26.57.100:0:0:tcp 172.26.56.2:0:0:tcp
```

### Configuring the WorkLoad Agent (WLA)

```
Dfpfg00 * DFP Agent configuration
* Clusterip is Cluster VIPA address
CLUSTER=1.1.1.201
* WLMpoll is Time interval in seconds for Agent to poll WLM; Default=60
WLMPOOLL=60
* WLM delta weight increment from last reported to mgr; Default=1
WLMINCRT=-1
* SYSPARM Member name containing Application to Port/Protocol mappings
SERVICEMAP=DFPMAPO0
* DFP listening port number for Load Manager TCP connections
* Port numbers below 1024 require OE UID=0
TCPPORT=2025
Dfpmap00 * APPLICATION SERVERS TO PORT MAPPING FILE
* ONE RECORD FOR EACH DEFINITION
* GROUPNAME= APPLICATION GROUP/WLM_LOCATION NAME
* PROTOCOL= INTERNET PROTOCOL TCP OR UDP
* PORT= PORT
GROUPNAME=IMWEBSRV,PROTOCOL=TCP,PORT=80
```

**LocalDirector**
MNLB Forwarding Agent Scaling: Connection Rate

Cisco 7500 Routers with Two 400M FastEther Channels WebStone 1 Byte Gets:

<table>
<thead>
<tr>
<th>Connections Per Second</th>
<th>One FA</th>
<th>Two FA's</th>
<th>Four FA's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3,210.61</td>
<td>6,395.27</td>
<td>10,666.64</td>
</tr>
</tbody>
</table>

What Software Is Required?

- **Forwarding Agent**: Cisco IOS software release 12.05(T) or above
  - Cisco 75xx, 72xx, 45xx, 36xx, RSM, MSFC (12.0.7-XE)
- **Services Manager**: LocalDirector 3.1.4 or above
  - Model 430 or 416
Network Load Balancing Design
How to Maximize Use of Multiple Paths

- Host IP stack
  IPCONFIG MULTIPATH
- Balance by packet
- Balance by destination
- Switching paths

<table>
<thead>
<tr>
<th>Switching Path</th>
<th>Cisco Express Forwarding</th>
<th>NetFlow Switching</th>
<th>Fast Switching</th>
<th>Process Switching</th>
</tr>
</thead>
<tbody>
<tr>
<td># Routes (11.3–12.0)</td>
<td>6 (12.0 only)</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td># Routes (11.2)</td>
<td>N/A</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Algorithm Used</td>
<td>Round Robin Per Packet or Per Destination</td>
<td>Pick Route Based on Source and Destination Address</td>
<td>Pick Route for Each Destination IP Address</td>
<td>Round Robin Per Packet</td>
</tr>
</tbody>
</table>

Designing for CIP TN3270 Server

- Move the LocalDirector out of the traffic path—**hub or catalyst?**
- Design to not use the existing CIP routers as forwarding agents—**optional**
- Continue to use the existing real/virtual definitions from the LocalDirector
- Continue to use the existing load balancing algorithms
- Install a backup MNLB SM to remove the SPOF—**only required for stage 1**
Designing for Host Availability

- Move the LocalDirector out of the traffic path—hub or catalyst?
- MNLB SM balances based on IBM Work Load Manager (WLM) information, WLM reports server capacity and application availability via a Cisco workload agent
- Provides generic resources equivalence for TCP/IP in a parallel sysplex, install the same VIPA on multiple stacks

Designing for Multiple Platforms

- MultiNode load balancing was not designed for only the IBM environment
- Any environment that requires load balancing can work simultaneously, each cluster of servers will have a unique virtual address and port
- Still use distributed director for geographic distribution
MNLB Roadmap

Stage 1

- MultiNode load balancing components

  Cisco IOS-based forwarding agents in Cisco 75xx, 72xx, 4x00, 3600, Catalyst 5K RSM and Catalyst 6000 MSFC

  MNLB SM runs on LocalDirector chassis

  LD hot-standby for phase one backup manager

  WorkLoad agents for IBM OS/390, NT and UNIX
MNLB Roadmap

- MNLB SM ported to Cisco IOS
- LocalDirector hardware is optional
- MNLB SM backup between routers
- MNLB SM may share primary and backup function
- MNLB SM runs on router blade

SLB Roadmap

- Continued appliance development
- Time to market features

Cisco IOS

Server load balancing
FMA incorporated into Cisco IOS
Cross platform support
Cisco/IBM Case Study

High-Availability Services for Large Enterprises

- S/390 workload manager plus Cisco MultiNode Load Balancing
  - Continuous application availability
  - Optimal resource utilization
  - Scalable Web solution
Test Inputs and Outputs

- **Inputs**
  - WWW-based workload called AKSTRESS
  - IBM Websphere WWW server

- **Outputs**
  - RMF data on S/390 service units for the Websphere server

What Is a MIP Burner?

Represented by this Symbol in Subsequent Slides
Also Known as a Looper Internally, Sometimes Referred to as a Chalupa

S/390 CPU % Looper Constraint Quantification
The Configuration

- IBM OS/390 workload manager
- Cisco S/390 workload agent
- Cisco IOS software with Cisco 7500
- Cisco MultiNode load balancing feature of LocalDirector

Distributing Connections
TCP/IP Stack Failure

Server VIPA 1.1.2.201

Cluster VIPA 1.1.1.201

Server VIPA 1.1.3.201

Cisco 6500

Cisco 7507

LocalDirector Failure

Server VIPA 1.1.2.201

Cluster VIPA 1.1.1.201

Server VIPA 1.1.3.201

Cisco 6500

Cisco 7507

MNLB CASA 192.168.0.0 VLAN1

LocalDirector Failover Cable

LD 430’s
CIP Router Failure

Combined OSA/2 FE and CIP Test Results Summary

Combined OSA/CIP MNLB Results
Test Objectives Met

- Intelligent workload-based load balancing
- Workload availability under failure
- Workload availability under stress

MNLB Load Distribution

% of Web Server Service Units

System 1
System 2

No Constraint
Med. Constraint System
Router Failure
Gateway Failure

Are You
Ready?
Intelligent Server Load Balancing in IBM 390 Environments: A Cisco/IBM Alliance Case Study

Course 2700

Please Complete Your Evaluation Form

Course 2700