Faster recovery in Network Virtualized Environment During Failover/Failback

[1]S.Annie Christilla
[1]Associate Professor
Department of Computer Science, St. Francis De Sales College, Bangalore, Karnatka, India

ABSTRACT

On virtualized environment, when there is dual VIOS (Virtual IO Server) present, if any network failure happens it takes very long to recover from it. Dual VIOS environment is basically setup to have failover configuration to achieve relaiablity. One of the VIOS will be primary and will take care of communication to hosts outside. In case if there are any failures, other VIOS will take over and take care of the communication. In this paper, method to recover faster is being proposed in this draft.

Keywords: VIOS(Virtual IO Server), SEA(Shared Ethernet Adapter), NIC (Network Interface Card), LPAR (Logical PARtition)

1. INTRODUCTION

Dual VIOS setup is that it promotes redundancy, accessibility, and serviceability. It also offers load balancing capabilities for Multipath I/O (MPIO) and multiple shared Ethernet adapter configurations. When compared to a single VIOS setup, a dual VIOS setup has the following additional components: An additional VIOS partition. Each VIOS partition consists of an additionalvirtual Ethernet adapter, which is used as the control channel between the two shared Ethernet adapters. Setting the trunk priority
on the virtual Ethernet adapters that are used for bridging two physical adapters in a shared Ethernet adapter configuration.

Both VIOS partitions will be up and running. One of them will be primary and the other will be secondary. Both these partitions will be exchanging heart beats to know the availability of other VIOS. When there is no response for 3 retries, the other VIOS will know something happened to the other VIOS and takeover of primary.
The VIOS which is primary will take care of network communication from local partitions to the outside world and vice versa.Generally both VIOS will be connected to different network switches so that when there is problem with network infrastructures that also can be taken care. Connecting to single switch again might cause network failures.

2. PROBLEM STATEMENT

On virtualized environment, Dual VIOS is configured to achieve reliability and accessibility. Dual VIOS can be configured as backup mode. In this case both VIOS will be up and running. But only one VIOS will take care of network communication. The VIOS which takes care of communication is called primary VIOS. When there is failure detected, the secondary VIOS will become primary and take cares of the communication. The problem with this method is that, switch connected to first VIOS(that was primary earlier) will have MAC address and IP address mapping for all the LPARs. As a result communication from outside hosts will flow in from both the VIOSs(both switches). This will cause poor network throughput/loops. In this paper a new method will be discussed to avoid this problem.

3. EXISTING METHOD

When Dual VIOS is configured, it can be configured in backup mode. In this case both VIOS will be up and running. But only one VIOS will take care of network communication. The VIOS which takes care of communication is called primary VIOS. The other VIOS will be standby. The other configuration is both VIOS will be up and running. One VIOS will be serving for some VLANS(set 1) and backup for other set of VLANS(set 2). The other VIOS will be serving for some VLANS(set 2) and backup for other set of VLANS(set 1). Switch connected to VIOS will keep learning about the LPARs behind the VIOS through ARP/Data packets flowing from LPAR to outside. Switch will have forwarding table populated. Lets consider a example, where there is 2 LPARS present on the system and IPs for those system are 10.0.0.1 and 12.0.0.1. Now the switch connected to VIOS 1 will know for the destinations 10.0.0.1 and 12.0.0.1 traffic should be forwarded to the port where VIOS 1 is attached. As the other VIOS standby, switch attached to VIOS 2 will not have any information related to hosts 10.0.0.1 and 12.0.0.1. Communication will be happening using VIOS1. Both the VIOS will be exchanging Heart Beats through control channel. If there is no response for consecutive 3 Heart Beat requests, the other VIOS(VIOS 2) will take over as primary and VIOS 1 will become secondary. However in this case still switch connected to VIOS 1 will have information about MAC and IP addresses of the 2 LPARs. As a result, when there is data from outside it will be sent through the first switch. As a result data will come through VIOS 1. When LPAR sends data out, data will go through VIOS 2(as this is primary VIOS now) and switch connected to VIOS 2 will learn the IP address and MAC address. Now switch 2 will also have information about IP 10.0.0.1. Similarly it could happen for IP 12.0.0.1. When it happens both switches will try to forward the data and it might result in network loop. In cases where both switches are not connected still data might come through one switch and might go through other switch. This will result in poor network performance.

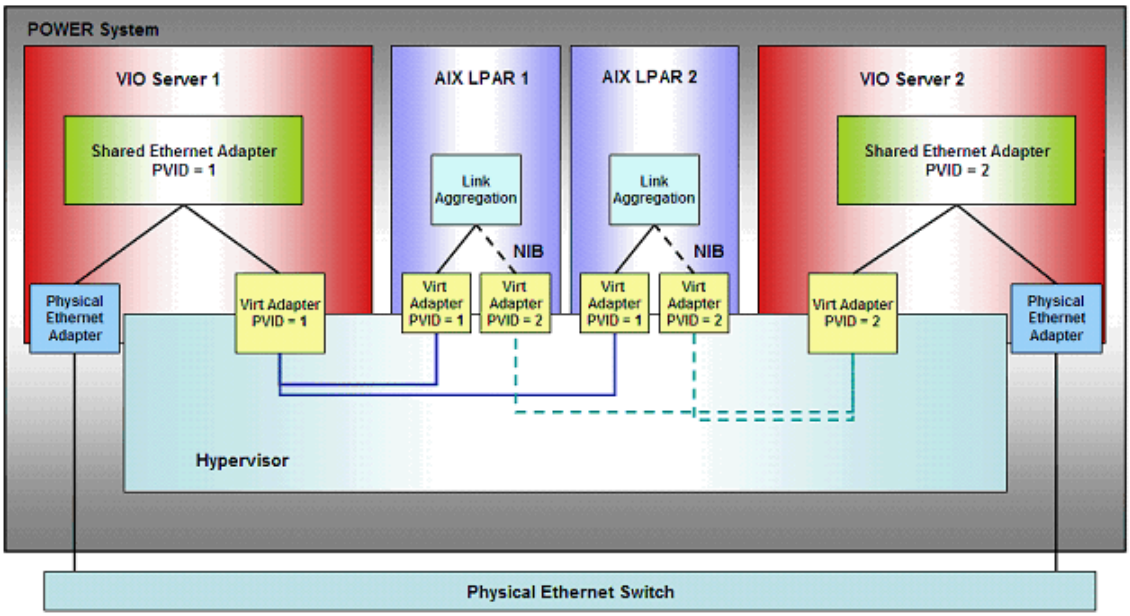

Figure 1. Dual VIOS configuration using single switch

4. PROPOSED METHOD

In this paper, an algorithm to achieve better network recovery is proposed. On VIOS at SEA layer information (Source IP, Source MAC) will be maintained for each LPAR it is serving. When failover happens, In other words when VIOS becomes primary(VIOS 2), using this table, RARP will be send to switch for each address. In our case we have 2 address (10.0.0.1 and 12.0.0.1), for both we will be sending RARP packet to switch. As a result switch will immediately learn those MAC addresses and update its forwarding table. As a result now, switch will be having these 2 MAC address in its forwarding table. When any traffic comes to these MAC addresses the port VIOS 2 is connected should be used.

RARP message is used for communication between SEA and switch. It is used to get IP address from MAC address. SEA will fill both IP and MAC address information and send it to switch. Upoing receiving RARP packet, switch will take MAC and update the forwarding table. Refer Figure 2 for the RARP format.

Algorithm:

1) When LPAR is configured, SEA will learn about new LPAR and add source IP address in its table.
2) When communication happens from the LPAR, SEA will learn source MAC address. (In our case for both LPARs these 2 steps will be learnt and updated in the table maintained by SEA).
3) Both VIOS will be sending heartbeat on the control channel.
4) Both VIOS should respond to the Heart beat and send the acknowledgement.
5) If VIOS doesn't get HB ack from other VIOS it should retry for 3times.If all the 3 times HB Ack is not received, the second VIOS will take over as primary VIOS.
6) While becoming primary, it should send RARP packets to switch for each LPAR it is serving. (In our case it should send 2 RARP packets, one for 10.0.0.1 and the other one for 12.0.0.1).
7) Upon receiving these RARP packets, switch will learn about all these MAC addresses and update its forwarding table. (Traffic to these MAC addresses should be forwarded to the port VIOS is attached).

01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Hardware type																				Protocol type										
Hardware address length								Protocol address length												Opcode										
Source hardware address :::																														
Source protocol address :::																														
Destination hardware address :::																														
Destination protocol address :::																														

Figure 2. RARP Protocol format

5. CONCLUSION

The objective of the proposed method is to faster recovery when failover happens. The proposed method updates the switch using RARP protocol and updates the forwarding table. As a result both switches are having correct information instantly, causing only one switch to take care of the communication. This results in avoiding duplicate data delivery and network loops. Hence the network recovery is improved using this method.

REFERENCES

1. **RFC 903Reverse Address Resolution Protocol -** https://tools.ietf.org/html/rfc903
2. Hai Lin, Lucio Correia, Mel Cordero, Rodrigo Xavier, Scott Vetter, and Vamshikrishna Thatikonda - IBM PowerVM Virtualization
3. Gary R. Wright(Author), W. Richard Stevens - TCP/IP Illustrated, Vol. 2: The Implementation
4. Kumar Reddy, "Network Virtualization".P
5. Network Adapter Specification LSO feature – Intel