ISSN 2277-3061
Volume 15 Number 10
International Journal of Computers & Technology

Simulation and Analysis of Distributed Gateway System for First-Hop Redundancy

Haidlir Achmad Naqvi[1], Sofia NaningHertiana[2],Ridha Muldina Negara[3],Rohmatullah[4]
[1,2,3,4] Fakultas Teknik Elektro,Telkom University, Indonesia
[1]haidlir@gmail.com,[2]Sofiananing16@gmail.com, [3]ridhanegara@gmail.com

ABSTRACT

Internet is expected to be available every time, so that the services run on it are able to be used any time. To achieve high availability Internet network, redundancy is used by employing more than one gateway so if one gateway is down, the other gateway will take the job of it. In the access layer, we use first-hop redundancy as the redundancy system. In this research, author design a system that has first-hop redundancy. The system is designed using software defined networking paradigm. The system uses POX as controller and Openflow as Controller Data-path Protocol Interface between switches and controller. The system is simulated on the mininet to examine the capability of the system by measuring some parameters: fail over delay, resource utilization and overhead on the bipartite topology. Of the examination results and analysis, the system is able to run the first-hop redundancy which yields fail-over delay below 140 ms for single flow.

Indexing terms/Keywords

First-Hop Redundancy; Software Defined Networking; Openflow

Academic Discipline And Sub-Disciplines

Computer communication, Internet

SUBJECT CLASSIFICATION

System Design

TYPE (METHOD/APPROACH)

Quasi-Experimental

1. INTRODUCTION

In the design of network, redundant link is mandatory to mitigate the effect when the primary link is down, in other word network redundancy. With network redundancy, the availability of the network is expected to be 99.999% (five nines)[2]. Network redundancy can be implemented either in the core, distribution, or access infrastructure layer.

Ethernet is the most widely used layer two technology recently. Ethernet network redundancy mechanism is different to network redundancy mechanism in Internet Protocol (IP), since there are no hierarchical addressing scheme and routing mechanism in the Ethernet. Actually there are some effort to get gateway redundancy in the Ethernet network or first-hop redundancy[3], such as Proxy Address Resolution Protocol[4], Virtual Router Redundancy Protocol[5], Hot Standby Router Protocol[6], and Common Address Redundancy Protocol[8]. Those protocol works in the conventional network control which is distributed system, every gateway maintain their own state and communicate with other devices periodically, so that the system is very complex. Moreover, the control communication (control channel) uses same infrastructure (link) along with data channel, hence it reduces the available bandwidth for users. As more gateway employed in the first-hop redundancy, the system becomes more complex and hence greater overhead traffic.

In the other hand, there is a new emerging approach in networking world, namely software defined networking (SDN). SDN uses logically centralized controller to control networking infrastructure and devices [9, 10, 11], hence it promises more fine-grained management. The communication between controller and infrastructure uses control channel, which is separated from data channel, hence it will not reduce the available bandwidth for users. Therefore, in this research, we create a system which runs first-hop redundancy function using SDN approach. This research focus on the uplink direction of unicast traffic for first-hop redundancy function design and analysis. The system runs on Ethernet or IEEE 802.3. The system is built on POX Controller and uses Openflow Protocol 1.0 as Control Data Plane Interface (CDPI).

2. RELATED WORK

Recently, there are some efforts to build first-hop redundancy into the network. The most widely used first-hop redundancy protocol is Virtual Router Redundancy Protocol (VRPP) [5] standardized by IETF (Internet Engineering Task Force) (rfc 3768). VRRP has load sharing capability to share load into several gateways. Another way to achieve first-hop redundancy is using Proxy ARP[4]. In order to handle IP-to-MAC mapping, Proxy ARP necessitates host to have large ARP tables as more destination in other network. There are also Hot Standby Router Protocol (HSRP)[6] that developed and patented by vendor, therefore not all devices support those protocols. Those protocols run on the conventional distributed system of network. They use a message exchange mechanism to get information of each other node, to ultimately achieve first-hop redundancy capability. It means that as more node is involved in the first-hop redundancy, overhead packet will be greater. And since the control channel uses same links with data channel, the available bandwidth

for host is reduced. A solution to address this problem is by separating data channel and control channel or using centralized controller, which is used in the software defined networking

3. DESIGN OF SYSTEM

The system uses software defined networking (SDN) approach. It uses single controller that is connected to every switch in the infrastructure layer as depicted figure 1. The switches comprise of two groups of switch: gateway switch which is connected to other network and access switch which is connected to the host. The switches are programmable switch that supports Openflow Protocol 1.0. The system runs on Ethernet network with single at Internet Protocol (IP) address. The gateway switches do not have IP address on its uplink interface, instead, the IP address is saved in the controller as well as the ARP mechanism.

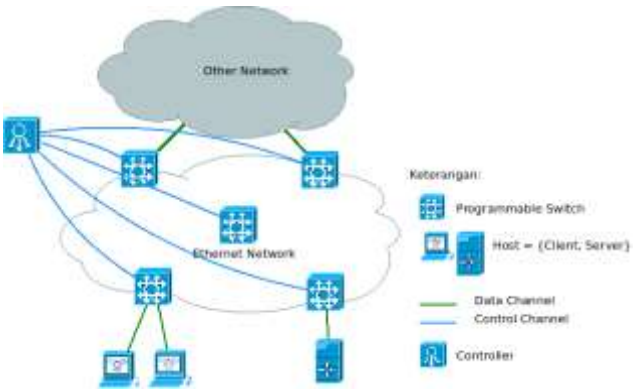

Fig. 1. System Architecture

Host can be either a client computer, server, or other devices. It gets dynamic IP address from dynamic host control protocol (DHCP) service run on controller. Switch acts as a relay between host and controller in the DHCP discover-offer-request-ack process. Only one subnet of IP address in single network broadcast multiple-access, namely Ethernet. The system is made in the form of SDN application resided in the controller. The system rules the forwarding and monitoring process, as well as DHCP service and Proxy ARP service. Particularly, the first-hop redundancy mechanism will be explained later.

3. 1 First-Hop Redundancy

First-hop redundancy is a network redundancy mechanism for uplink direction in the layer 2 technology such as Ethernet network. In the implementation, first-hop redundancy is used when host only configured to only has single gateway address. Since host only has single gateway address, then while the gateway is down, consequently the host will be isolated from the outside of the network. That host is only able to communicate with the internal network.To prevent this accident, additional gateways are added to the network. Since host only configured with single gateway address, then only a gateway address that can be used in a single time. Therefore, in the first-hop redundancy mechanism, all gateways have same IP address and in this research same MAC address too. Normally, if there are more than one devices with same IP address, those device will be intentionally down because of IP address conflict, but in the first-hop redundancy this case will not occur. So while a gateway is down, other gateway will take that gateway's responsibility.

3. 2 System Structure

The system comprises some subsystem to run packet forwarding, monitoring, DHCP, and Proxy ARP services, they are main subsystem, bucket subsystem, forwarding subsystem, monitoring subsystem, supporting subsystem.

3.2.1 Main Subsystem

Main subsystem is responsible for receiving event-trigge such as: (1)switch-controller connection established event, (2)link local discovery protocol (LLDP) event, (3)packet-in event, (4)flow-statistic event, (5)port-status event, (6)switch-controller connection down event. Additionally, the main subsystem also acts as scheduler for some activity such as: (1)path lookup every ten seconds, (2)sending flow-statistic request every second, (3)sending ARP request to the next-hop gateways every ten second.

3.2.2 Bucket Subsystem

Bucket subsystem contains some information such as: (1)matrix adjacency, (2)all available path between node pairs, (3)port information of all switches, (4)ARP table, (5) flow entry of all switches. The system use Depth-First Search (DFS) to find all possible path between all node as explained algorithm 1. Matrix adjacency is used to capture topology $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_n\}$ is switch set and $E = \{e_{1,1}, e_{2,2}, \dots, e_n\}$, is link between switches. Every $E_{i,j}$ saves metric information conform with equation 2. The DFS algorithm yields $T_{k,l}$ paths which stored into bucket Path or T. Every $T_{k,l}$ has metric value conform with equation 3.

$$metric_{i,j} = \frac{10^2}{(capacity_{E_{i,j}} - load_{E_{i,j}})} \quad (2)$$

$$metric(T_{k,l}) = \sum_k^l metric(E_{i,j}) \quad (3)$$

3.2.3 Forwarding Subsystem

Forwarding Subsystem primary responsibility is installing flow entry into the switch to forward flows in the infrastructure layer. When the first packet of the flow come into the switch, the switch will pass the packet to the controller. The controller will find the path to deliver that packet to the destination based in the header information of that packet. Furthermore, this subsystem is also responsible to sort packets that come into the controller, so ARP packet can be passed into the ARP function, and DHCP packet can be passed into DHCP function, and LLDP packet can be passed into openflow.discovery.

Algorithm 1 ~~Find~~ all possible path from graph G Adapted ~~Fro~~m DFS algorithm

Input: network topology $G = (V, E)$ represented in dictionary Adj, where V, E represents DPID and link between DPID

Output: a routing table for source-destination DPID pairs represented in dictionary T

```
T ← {}
procedure PER_SOURCE_DFS(source, origin = None, path ← [])
    if origin ≡ None then
        origin ← source
    end if
    for i ∈ M[source] do
        if i ∈ path ∨ i ≡ origin then
            continue
        else
            if i ∉ T[origin] then
                T[origin][i] ← [path + [i])]
            else
                append [path + [i])] to T[origin][i]
            end if
            PER_SOURCE_DFS(i, origin, path + [i])
        end if
    end for
end procedure
for i ∈ M do
    path[i] ← {}
    PER_SOURCE_DFS(i)
end for
```

3.2.4 Monitoring Subsystem

Monitoring subsystem maintain the state of every switch, every port in every switch and every flow in every port. It runs after main subsystem receives FlowStatsReceiver, PortStatus, and ConnectionDown event. Monitoring subsystem stores data into the bucket.matrix-adjacency, bucket.port-information, and bucket.flow-entry**.**

3.2.5 Supporting Subsystem

Supporting subsystem acts as Proxy ARP and DHCP Server. Proxy ARP is responsible to answer every ARP Request and issue ARP Reply packet. Whereas DHCP Server is responsible to allocate IP address for every host, and maintain the DHCP discover-offer-request-ack process.

4. EVALUATION AND RESULT

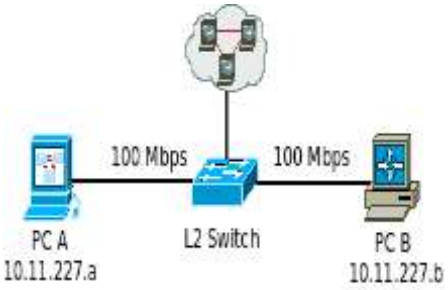

Fig. 2. Simulation Setup

This research usesMininet [13] to simulate the infrastructure of the network. Mininet is installed in the PC A and Controller is lied in the PC B as depicted in figure 2. PC A is powered with 2.90 Ghz x 4 core processor and 3.8 GB of RAM, and uses Ubuntu 13.03 LTS 64 as operating system. Distributed Internet Traffic Generator (D-ITG) and Iperf are also installed in the PC A to generate traffic for measurement. PC B is powered with 2.30 Ghz x 4 core processor and 3.7 GB of RAM, and uses Ubuntu 14.04 LTS 64 bit. Pox Controller and its complementary application is installed in PC B. The average delay between PC A and PC B is about 0.352 ms RTT, measured using ping command.

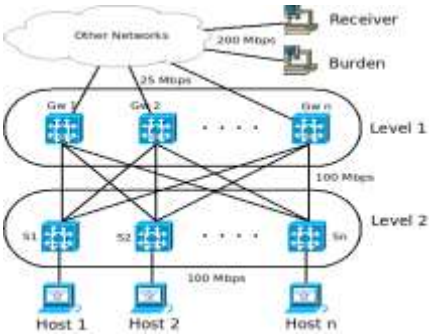

Fig. 3. Bipartite Topology

Table 1. Number of Switch and Host along traffic type for every measurement

Measurement	L1	L2	Host	Traffic
Fail-Over Delay	2 - 4	1 - 8	1 - 8	Single ICMP Ping
Overhead Size	2	1	1	25 - 150 UDP Flows
Memory Consumption: Switch	1 - 4	1 - 8	0	-
Memory Consumption: Host	1	1	0 - 200	-

The measurement uses bipartite topology which comprises two sets of switch as depicted in the figure 3. L1 is the switches that connected to the other network, whereas L2 is the switches that connected to the end-host. This research examine four parameters to analys the system performance, they are: (1)fail-over delay, (2)Overhead Size, (3)Memory Consumption: in the switch and host, as summarized in the table 1.

4.1 Fail-Over Delay Measurement

Fail-over delay is described as time taken while a gateway is down until there is another gateway takes over that dead gateway's job. The measurement unit is millisecond. The measurement is done in the infrastructure layer using ping diagnostic tool and controller using time-stamp. Bipartite topology is used with varied number L1 and L2 switch refers to the table 1. Of the measurement result, the performance of the first-hop redundancy is considered stable since the fail-over delay in the infrastructure layer is below 140 ms and in the infrastructure layer is below 1000 ns for all scenarios, as depicted in figure4 and 5. The 140 ms of the first-hop redundancy's fail-over delay is better than VRRP's 145 ms with 220 backup priority [3], so the performance is considered good.

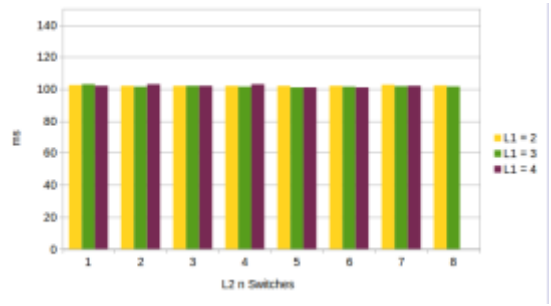

Fig. 4. Transition Delay Measurement in Infrastructure

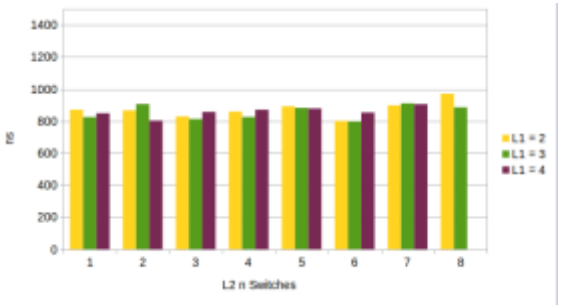

Fig. 5. Transition Delay Measurement in Controler

However, the system is still unable to handle the L1 = 4 and L2 = 8 topology combination. It caused by insufficient resource required to run the DFS routing algorithm. It is the drawback of the system, and can be addressed by changing the recursive DFS algorithm to another routing algorithm.

4.2 Overhead Measurement

Overhead is amount of control channel needed in the fail-over process. The unit of the measurement is byte. The measurement is done by generating varied flows from host to receiver. Then uplink of a gateway is turned into down-state intentionally. Capturing packet of overhead is conducted in the controller using wireshark. The amount of overhead is counted from the link is down until modified flow packet (flow-mod) is sent to the last selected switch.

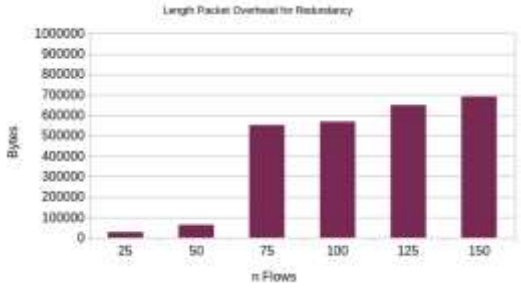

Fig. 6. Fail-Over Process Overhead

Of the measurement result, as more flow involved in the fail-over process, the overhead packet is greater as depicted in figure 6. The overhead packet only involved port-status and flow-mod message from zero to 50 flows. But for flow greater than 50, barrier request and barrier reply is also involved in the overhead packet, it causes big inflation of overhead packet between 50 flow and 75 flow.

4.3 Memory Consumption Measurement

This measurement is conducted to know how much memory consumption for every switch and host used in the network. The measurement unit is Megabyte (MB). As more switch is used in the network, the memory consumption is higher as depicted in figure 7. The greatest in ation of memory consumption occurs when L1 = 4, event the system is unable to handle L1/L2 = 4/8 topology combination. It is caused by as more switch used in network, more circuit exists. As more circuit exists more resource is needed to calculate the routing algorithm and store the result. To improve the system, better and more efficient routing algorithm is maybe the alternative.

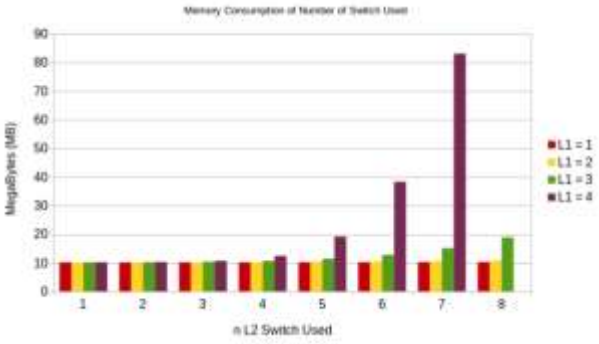

Fig. 7. Memory Consumption of Switch Used in Controller

As more host is used in network, more memory is need to store the infor-mation corresponding to the host in the controller as depicted in figure 8. However, the growth of memory is only 4 KB per-host. In other word, the system can handle 5000 hosts only with 30 MB of memory.

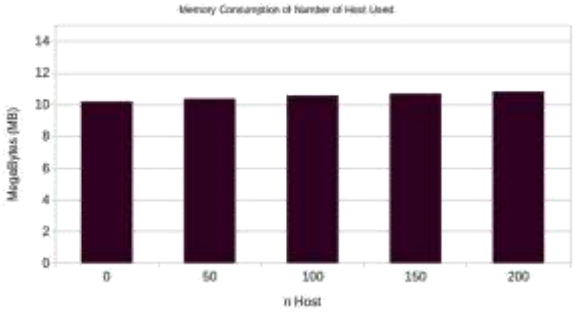

Fig. 8. Memory Consumption of Host Used in Controller

5. CONCLUSION

The system which implements software defined networking approach for first-hop redundancy function can be simulated in a network using bipartite topology with varied switch and host. The first-hop redundancy can address the link failure occurred in the uplink of the gateway. The first-hop redundancy performance is stable in term of fail-over delay in the infrastructure layer is below 140 ms, less than VRRPs with 145 ms (on backup priority 220 con guration). However, the system can improve the quality of service as more gateway employed. Furthermore, the memory consumption of switch is increasing rapidly as more circuit exists in the topology, hence this system is recommended to handle topology with small number of circuit.

6. REFERENCES

I. Chui, Michael, et al. 2012. The Social Economy: Unlocking Value and Productivity Through Social Technologies. Report. McKinsey Global Institute.

II. Piedad, Floyd and Michael Hawkins. 2001. High Availability: Design, Techniques, and Processes. London: Prentice Hall.

III. Bernat, Joaquim S. 2011. Redundancy and Load Balancing at IP layer in Access and Aggregation Networks. Master Thesis. Aalto University.

IV. Cal-Mitchell, Smoot and John S. Quarterman. 1987. Using ARP to Implement Transparent Subnet Gateways. RFC 1027. Internet Engineering Task Force.

V. Hinden, R. 2004. Virtual Router Redundancy Protocol (VRRP). RFC 3768. Internet Engineering Task Force.

VI. T, Li, et al. 1998. Cisco Hot Standby Router Protocol (HSRP). RFC 2281. Internet Engineering Task Force.

VII. FreeBSD Manual Common Address Redundancy Protocol. Last Reviewed (10/06/2015). http://www.freebsd.org/cgi/man.cgi?query=carp&sektion=4

VIII. Open Network Foundation.201). Software-Defined Networking: The New Norm for Networks.

IX. Open Network Foundation.2013. SDN in the Campus Environment.

X. Open Network Foundation.2009. OpenFlow Switch Specification version 1.0.0 (Wire Protocol 0x01).

XI. Ejaz, Syed K. 2011. Analysis of the trade-off between performance and energy consumption of existing load balancing algorithms. GroerBeleg. TechnischeUniversitt Dresden.

XII. Lantz, Bob, Brandon Heller dan Nick McKeown. (2010). A Network in a Laptop: Rapid Prototyping for Software-defined Networks. Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks. New York: ACM.

XIII. A, Sefano, Donato Emma, Antonio Pescapedan Giorgio Ventre. (2004). A Practical Demonstration of Network Traffic Generation. Proceedings of the 8th IASTED International Conference. Hawaii.