An Evaluation of the Virtual Router Redundancy Protocol Extension with Load Balancing

Jen-Hao Kuo, Siong-Ui Te, Pang-Ting Liao, Chun-Ying Huang, Pan-Lung Tsai,
Chin-Laung Lei, Sy-Yen Kuo, Yennun Huang, Zsehong Tsai
cliff@fractal.ee.ntu.edu.tw, koan@fractal.ee.ntu.edu.tw, b88901138@ntu.edu.tw,
huangant@fractal.ee.ntu.edu.tw, charles@fractal.ee.ntu.edu.tw, lei@cc.ee.ntu.edu.tw,
sykuo@cc.ee.ntu.edu.tw, yen@research.att.com, ztsai@cc.ee.ntu.edu.tw

Abstract

Virtual Router Redundancy Protocol (VRRP) is designed to eliminate the single point of failure in the static default routing environment in LAN. The original VRRP protocol does not support load balancing for both incoming and outgoing traffic. This paper describes EVRRP, i.e. Enhanced VRRP. EVRRP supports an efficient multiple-node cluster and symmetric load balancing among routers. Each router periodically exchanges information to determine the status of the master and backups. The master router distributes and redirects the traffic to one of the backup routers by ICMP redirect message. Backup routers accept the traffic from the master and one of the backup routers takes over the master traffic using a gratuitous ARP message when the master fails. The improved election protocol speeds up the original VRRP election protocol and shortens the failover time by adding a new state in the previous VRRP state diagram and a new protocol type. An extensive evaluation of the EVRRP protocol is described in the paper.

1. Introduction

Virtual Router Redundancy Protocol (VRRP) [3][4][5] is designed to eliminate the single point of failure in the static default routing environment. VRRP became an IETF (RFC2338) standard in 1998. Since then, it has been widely used in a LAN environment to tolerate router/gateway failures. However, most implementations of VRRP today have been limited to a primary-backup configuration where no load balancing of traffic between the primary router and backup routers is supported. The master router in VRRP provides the routing function and sends heartbeat packets to the backup router. The backup router will start to route packets only when the master router fails. Since the backup router will be idle when there is no failure, the resource in the backup router is wasted most of the time.

The EVRRP (Enhanced VRRP) work is inspired from shortcomings of the previous RFC2338 VRRP. The major difference between EVRRP and VRRP is that EVRRP provides an efficient mechanism to do load balancing among routers without the need of running multiple VRRP daemons on each router. Furthermore, by modifying the VRRP state diagram and adding the election protocol to support multiple-router cluster architecture, EVRRP further improves the scalability of the original VRRP protocol. EVRRP is backward compatible and supports all the original VRRP features such as preemption, virtual MAC, etc.

2. Router Clustering

VRRP implementation uses VRRP_TIMER_SKEW to support multiple backup routers in a single virtual router. Since two backup routers may have the same priority or two successive backup routers may try to become the master in 4ms interval, the router cluster could become unstable when the master router is down. EVRRP changes the state diagram and use ELECTION to prevent this from happening and provides a robust method to support a large cluster with many backup routers.

2.1. Election

Election is invoked only when one of the backup routers discovers a failure of the master. In election, all backup routers will exchange election messages to determine which backup router should become the new master. While receiving an ELECTION message, any router in the backup state will compare its priority

setting with the election message to see if it should become the master router. If a backup router receives an ELECTION message which has a higher priority than its own priority, the backup router will remain in the backup state. If the received packet priority is lower than its own priority, the backup router will keep on broadcasting ELECTION messages to check if any other router has a higher priority. After three rounds of sending election messages, the router who is surviving the election will become the master router.

In the current VRRP protocol, there is no checking of the TYPE field in the VRRP control packet. The ELECTION packet will be received and treated as an advertisement packet. Therefore, the EVRRP election messages will be ignored in current VRRP to support the backward compatibility.

3. Load Balancing

There are at least two routers (primary and backup) in use at the same time in VRRP. It's a waste of resource if the backup router just listens to VRRP heartbeat messages without doing anything.

In general, ICMP redirection [7] is used by a router to inform a client that there is a better path than sending packets to itself. The router sends an ICMP redirection packet to the client to point to another router. The EVRRP uses the ICMP redirection messages to redirect traffic to backup routers for load balancing. The EVRRP load balancing diagram is shown in Figure 1.

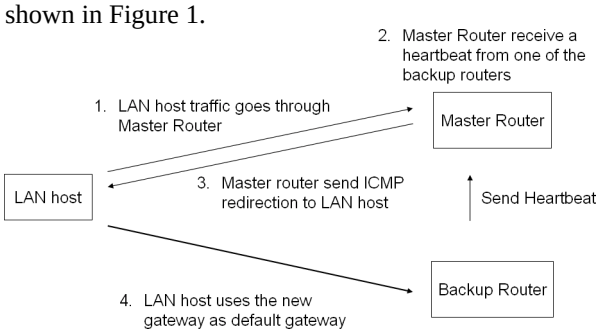


Figure 1. Load Balancing in EVRRP

The load balancing protocol in EVRRP is very straightforward: each backup router periodically sends EVRRP advertisement packets to the master router and the master router keeps a list of living backup routers. If the master router does not receive an EVRRP advertisement packet from a backup router for some time, the backup router is considered failed and is removed from the load-balancing router list. The master router checks all outgoing packets from hosts in LAN and determines what traffic should be redirected to backup routers. Besides using source and destination IPs as the redirection rule in prototype implementation, the redirection rule of EVRRP can be easily enhanced using destination IP, router load, traffic load, etc.

3.1. Advertisement

Master router uses the advertisement message to send heartbeat packets to all backup routers. In EVRRP, a backup router also uses the advertisement control messages to inform the master router of its existence so that the master router can identify where the backup router is and redirect some of the traffic to the backup router

4. Router Redundancy

The redirection algorithm, although simple, creates a new problem: what happens if a backup router fails while a host is sending packets through this failed backup router? As described earlier, if the master router does not receive a VRRP advertisement message from a backup router for some time, the backup router is considered failed and the master router will send a gratuitous ARP [8], which links the IP address of the backup router to the MAC address of the master router. Therefore, the master router could take over the job of forwarding packets for the failed backup router. As a result, without any change of configuration, a host can still send packets to WAN through the IP address of the failed backup router although the MAC address of this failed backup router IP address is now the MAC address of the master router.

The usage of the gratuitous ARP in our protocol is to eliminate the router failure situation. We use ARP Poison to seamlessly move the traffic from a failed router to other working routers. There are 3 scenarios which will invoke the ARP poison:

1. Backup Router Failure: Since there may be traffic dispatched to backup routers, if a backup router fails, the master router sends a gratuitous ARP to notify all hosts in LAN that the IP address of the failed backup router is now mapped to the MAC address of the master router.

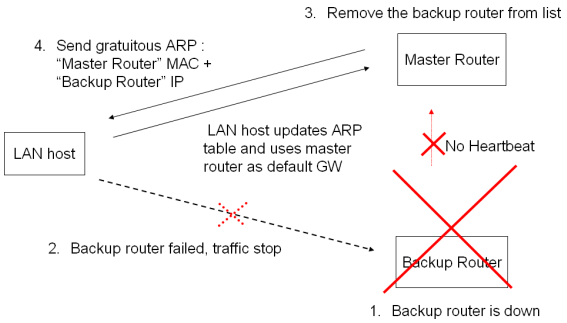


Figure 2. Backup Router Failed in EVRRP

2. Join of New Backup Router: If there is a new backup router joining the router farm, the master router needs to enlist the new router and distribute part of the traffic to the new router. The new joined router needs to broadcast its heartbeats to inform the master router of its existence. Besides, backup routers must send gratuitous ARPs periodically because the IP address of the backup router could have been mapped to the master router earlier.
3. Join of New Master Router: If a router becomes the master, it sends gratuitous ARP packets, using the Virtual IP address of the gateway and the virtual VRRP MAC address. The original master will be demoted to a backup. The demoted router needs to send a gratuitous ARP using its real IP address and MAC address to make sure that other hosts in LAN do not lose their WAN connections while the master router changes.

5. EVRRP State Diagram

EVRRP state diagram is modified from that of the original VRRP. By adding the Election state and modifying the backup/master states, EVRRP can support dynamic changes of configuration and load balancing in a cluster.

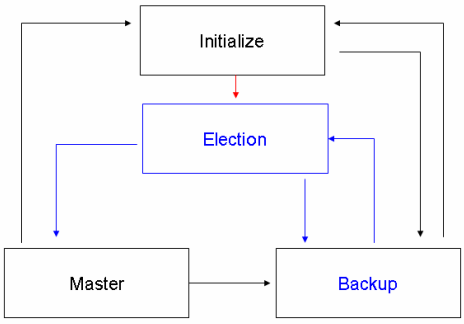


Figure 3. EVRRP State Diagram

5.1. Major differences in State Diagram

In Master State, the master router needs to route the packet and broadcast heartbeats. In EVRRP, the master router must also maintain a redirection list of all living backup routers by listening to heartbeats. Master router needs to distribute traffic to backup routers. All these tasks will increase the CPU usage of the master router a little.

In Backup State, the procedure is somewhat different from original VRRP. First, a backup router will enter Election State when it receives an election packet whose priority is lower than itself, or it suspects the Master router is failed because of missing master heartbeats. The Backup router also needs to broadcast gratuitous ARP packets to keep the binding of it IP and MAC address. All living backup routers need to broadcast heartbeats to inform the Master router of its existence so that the Master router will enlist the backup routers in its redirection list.

The Election State is part of the original VRRP Backup State with an extension. In Election State, a router acts as an ordinary backup router and it responds to election packets as soon as it receives lower priority Election/Advertisement packets. It will return to Backup State if it receives a higher priority packet. After the election period is over or election times out without receiving any Advertisement or Election packet, the router will promote itself to the Master State.

6. Compatibility with Original VRRP

By inserting an original VRRP router into the EVRRP router farm to check the EVRRP backward compatible with VRRP, we can generalize them into three conditions, a VRRP router acts as a Master, Slave Router, or how does a VRRP work while receiving Election packets.

1. VRRP Router Acts as Backup Router: When VRRP router acts as a backup router in EVRRP router farm, it can be functional okay but lack of load-balancing capability due to it does not send any heartbeats and master router cannot be aware of its existence. The VRRP router will ignore any other lower priority heartbeats which send by other EVRRP backup routers by default. And as long as there is a higher priority router sending heartbeat, the VRRP backup router will stay in Backup State.

2. VRRP Router Acts as Master Router: After a VRRP router becomes the master router, the whole router farm will make no different between ordinary ones. The VRRP master router will route all traffic through itself since it has no load balancing capability. It will ignore all other Advertisement/Election packets because all the packets have lower priority bit.
3. VRRP Router Receives Election Packets: The implementation of VRRP protocol supports only one type of packet, the ADVERTISEMENT packet. The EVRRP creates a new type of packet, ELECTION packet, which is almost identical with ADVERTISEMENT packet, is used when there is a new master router need to be elected. The current implementation of VRRP on Linux ignores the check of the type of VRRP packet since it assume there is only one type of packet and needless to check it. And also, the ELECTION packets received by original VRRP daemon can be viewed as useless packets and dropped without any error. Because the original VRRP state diagram does not support Election State, it must use the default ms_down_timer to make sure the Master router is down and then transit itself to the Master router.

7. Evaluation

The major benefits from EVRRP are the very low overhead in CPU consumption of its load-balancing and fail-over capabilities and the backward compatibility to the VRRP standard. In this section we perform the following tests to measure EVRRP load-balancing capability, CPU usage comparison, VRRP fail-over compatibility, traffic overhead incurred by the EVRRP protocol with extra ICMP packets, fail-over timeout, and packets forwarding and rebalancing between the backup routers and the master router.

7.1. Testing Environment Setup

In order to simulate real world traffic pattern, we setup a 3-router environment for EVRRPd testing. Routers 13, 14 and 15 use Gentoo Linux [11] w/ EVRRPd installed. There are 200 clients and 12 servers using HTTP as testing tool. At server side we use Spirent's Reflector 220™ [12] as http server, which has a max throughput of 200Mbps[1] while each router has the maximum throughput of 100Mbps. Router 15 has the highest priority and router 13 has the lowest priority to be the primary. In another word, if Router 15 fails, Router 14 will take over as master router, and so on. The topology of the testing is shown in Figure 4.

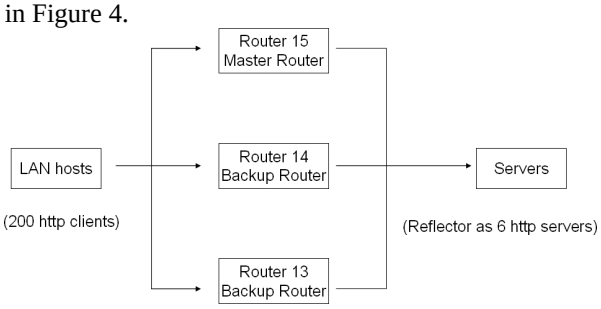


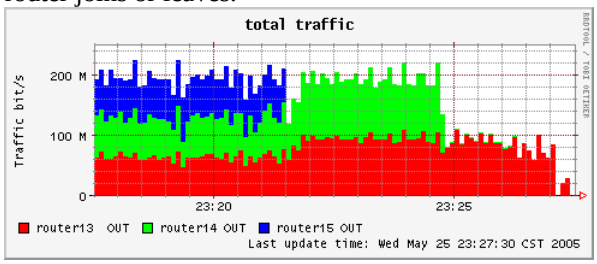
Figure 4. Setting of EVRRP Testing Environment

7.2. Data Collection

We use SNMPD on each router to collect CPU load and traffic data in real-time and use RRDTools to graph the numeric data for easy reading. The reason for choosing this way to collect data is that all the routers traffic can be monitored at the same time and the traffic changes can be easily shown.

7.3. Load Balancing Capability Test

First we generate a stable traffic load for Routers 13, 14 and 15. Each router has 67 (=200/3) Mbps http traffic. Then we disconnect Router 15 and 14 sequentially until only Router 13 is alive and acts as the master router to route all traffic. The master router transfers traffic from Router 15 to Router 13. Since the maximum physical bandwidth is 100Mbps for each router, the total traffic drops from 200Mbps to 100Mbps. From Figure 5, we can clearly see the benefits of EVRRP router fail-over and load balancing as clients can receive data at about 200 Mbps as long as more than one routers are alive. Besides, the traffic can be evenly distributed to every router after a new router joins or leaves.



[1] We choose custom an http client instead of the Spirent's Avalanche 220™ due to the unsupported ICMP redirection.

Figure 5. LB Test – Total Traffic Measured at Client Side

Besides, the traffic can be evenly distributed to every router after a new router joins or leaves. The following figures show the traffic how to distribute among the routers.

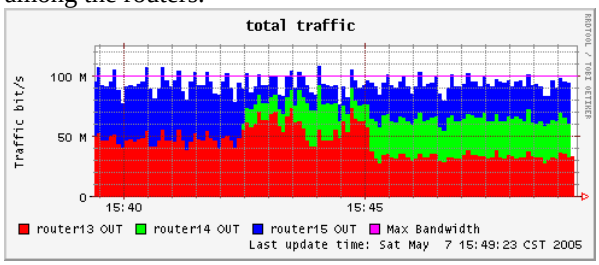


Figure 6. New Backup Router Joins to Share Traffic

Figure 6 shows a new router (Router 14) joins the EVRRP router farm and acts as backup router. The new join router can only share the master router traffic. After the ICMP redirection timeout, the total traffic is redistributed evenly to all routers.

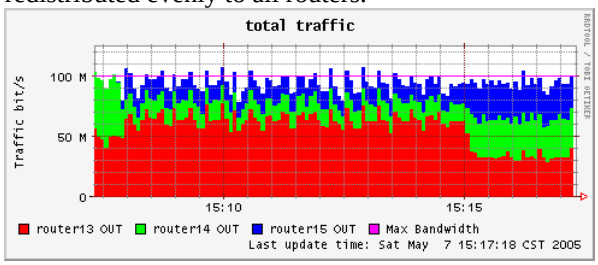


Figure 7. New Master Router Joins to Share Traffic

Figure 7 shows a new master router (Router 15) joins the EVRRP router farm. Router 15 will take over the Router 14's traffic and redistribute the original traffic of Router 14 to Routers 14 and 15. Previously dispatched traffic through Router 13 must wait until an ICMP redirection timeout to redistribute the traffic evenly.

7.4. CPU Usage Comparisons

In these tests, we compare the CPU load among Linux routers – Linux router with VRRPd, and Linux router with EVRRPd. Since the first two cases do not support load-balancing, to be fair we generate only 100Mbps traffic for testing. All the data are the measurements from the master router since backup routers are idle in the first two cases. Most of the CPU load is generated by System process. The Linux router and VRRPd consume about 18% of CPU time. The EVRRPd consumes about 30% CPU time when there is no backup router and drops to 10% when there are two backup routers for load sharing.

7.5. EVRRP Compatibility Test

In this test we substitute EVRRPd Router 14 with a standard VRRPd. After the traffic becomes stable, Router 15 shutdown to promote Router 14 to the master router. We want to make sure that the fail-over of EVRRPd is compatible with the standard VRRPd. Figure 8 shows the traffic monitored by client side.

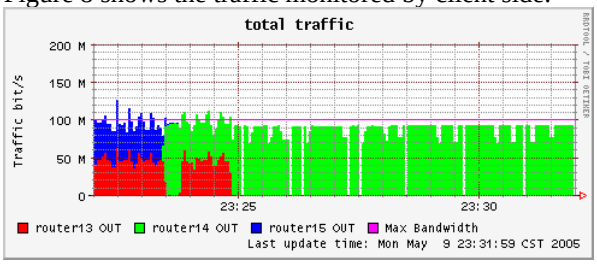


Figure 8. EVRRP Compatibility Test

While the Router 15 is working, it can redirect part of the traffic to Router 13. But there is no heartbeat from Router 14 and thus the master router can't tell if the Router 14 is alive or not. As a result, the master router (Router 15) will not redirect any traffic to Router 14. After Router 15 is failed, Router 14 takes over as the master router. Since Router 14 is running VRRPd, it cannot distribute traffic to Router 13. Therefore, all traffic is taken over by to router 14 as in the VRRP standard. The above tests show that EVRRP routers can work with VRRP routers seamlessly.

7.6. Traffic Consumed by extra ICMP packets

The EVRRPd uses ICMP redirection packets to redirect traffic among routers. So beside the data traffic, ICMP redirection packets also consume part of the total available bandwidth. We use the 3-router test (1 master router, 2 backup routers) to analyze the ICMP traffic load. In the tests, we reduce the ICMP redirection timeout from 900 to 60 seconds on client side to generate more ICMP packets for analysis. In these tests, ICMP packets only represent 0.22% of the total packets. Besides, the percentage of ARP packets which we use for redirecting traffic among routers is only 0.09%. We only monitor the master router for the number of total ICMP packets since the master router is the only source which generates ICMP packets.

The idle master router generates about 0.5kbps traffic for heartbeats. Each backup router generates about 1k bps traffic, which is a little more than the master router due to the extra gratuitous ARP broadcast.

7.7. Fail-over Test Environment

Since a fail-over only happens between one backup router and the master router, we simplify the test to two EVRRP routers to measure the fail-over time.

In this section, we measure the fail-over time of the master and the backup router. Host A sends test data (FTPput) packets to host B through the master router or the backup router during a period of one minute. Our testing and measurement software is the popular NetIQ Chariot [10]. During each testing period, we unplug the Ethernet connection of the master or backup routers at 10 second to simulate router failure and restore the connection at 30 second to simulate router recovery.

7.8. Fail-over Timeout

We compare the original VRRP and EVRRP fail-over. Figure 9 shows the VRRP failover timeout which is about 5 seconds under test. Figure 10 show the EVRRP failover timeout. Even though the implementation of EVRRP using both ELECTION state and the heartbeat mechanism, the timeout of failover in EVRRP shows no different from original VRRP, which is also about 5 seconds.

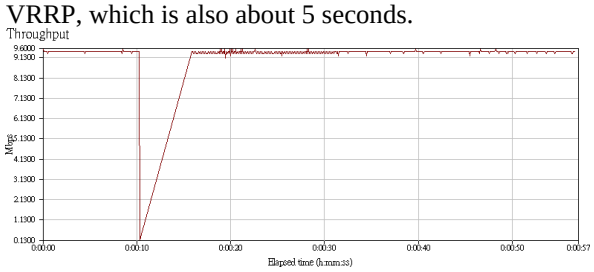


Figure 9. The Fail-over Timeout of VRRP

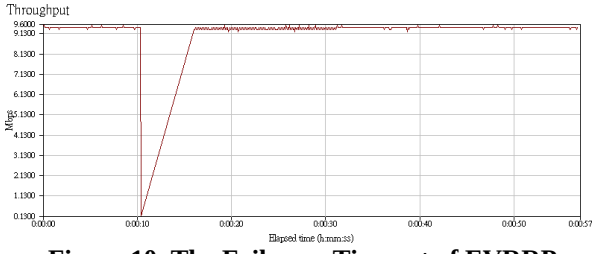


Figure 10. The Fail-over Timeout of EVRRP

7.9. Traffic Redirection and Balancing

The EVRRP will redirect some traffic to idle backup routers for load balancing. The traffic pattern will change when the master router or the backup router fails. Figure 11 shows the effect of the status change of the master router. Test data packets sent through the backup router as the master router fails and then recovers. Note that the gaps in the figure at 10 second and 30 second arise from the change of MAC address of the backup router. Figure 12 shows the condition of the backup router fails. Test data packets sent through the backup router as the backup router fails and then recovers. It shows a large timeout gap because the master router will notice the backup router has failed after not receiving heartbeat from the backup router and then use ARP poison to take over the traffic. The traffic drop in the figure on the left arises from the redirection of data packets from the backup to the master, while the gap on the right in the future is due to the rebalancing of the data packets from the master to the backup.

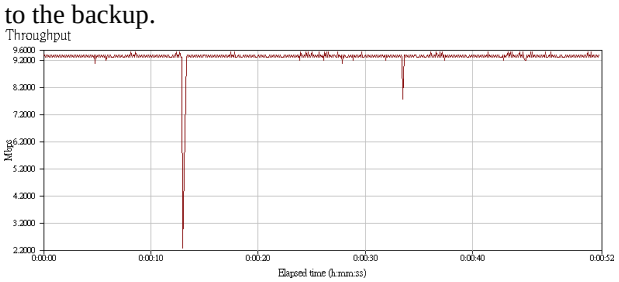


Figure 11. Backup Router Traffic while Master Router Changes

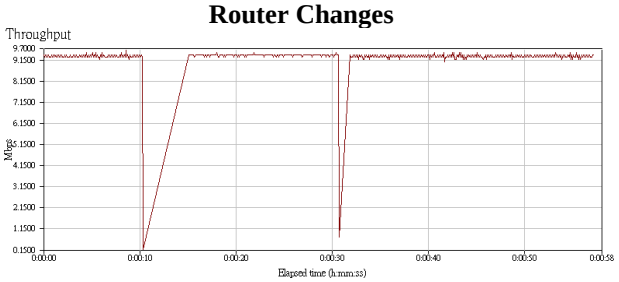


Figure 12. Backup Router Traffic while it Changes

8. Concluding Remarks

The VRRP protocol is an efficient fault tolerant networking solution and is widely used in LAN. Its simplicity and short fail-over time outperform other dynamic routing protocols such as RIP and deploying the protocol does not require any modification of network settings for hosts in LAN. However, VRRP does not support load-balancing and its scalability is limited. In our EVRRP effort, we show that the VRRP protocol can be extended easily and efficiently to support load balancing and high scalability. We believe that EVRRP protocol will be important for small to medium enterprise or campus networks as an economical solution to achieve high dependability in LAN. The EVRRP protocol has been extensively tested and used in our lab for almost one year. We are very confident in its correctness and robustness due to its simplicity, backward compatibility and the extensive testing of the protocol. In the near future, we intend to work with router manufacturers in Taiwan

and submit the EVRRP work to IETF as an enhancement for RFC 2338 and RFC 3768.

9. References

[1] J. Etienne, “VRRPd: overview, implementation and usage,” Ottawa Linux Symposium 2001, July 2001.

[2] J. Ranta, “Router Redundancy and Scalability Using Clustering,” Seminar on Internetworking, Spring 2004, eds. A. Ylä-Jääski, N. Kasinskaja, [Online] Available: http://www.tml.hut.fi/Studies/T-110.551/2004/papers/Ranta.pdf, June 2004.

[3] R. Hinden, D. Mitzel, P. Hunt, P. Higginson, M. Shand, A. Lindem, S. Knight, D. Weaver and D. Whipple, “Virtual Router Redundancy Protocol,” Internet Draft, draft-ietf-vrrpspec-v2-06.txt, Feburary 2002.

[4] [VRRP] R. Hinden, Ed., “Virtual Router Redundancy Protocol,” RFC 3768, April 2004.

[5] [VRRP] S. Knight, D. Weaver, D.Whipple, R.Hinden, D. Mitzel, P. Hunt, P. Higginson, M. Shand, and A. Lindem, “Virtual Router Redundancy Protocol,” RFC 2338, April 1998.

[6] [HSRP] T. Li, B. Cole, P. Morton and D. Li, “Cisco Hot Standby Router Protocol (HSRP),” RFC 2281, March 1998.

[7] [ICMP] J. Postel, “Internet Control Message Protocol,” RFC 792, September 1981.

[8] [ARP] D. Plummer, “An Ethernet Address Resolution Protocol,” RFC 826, November 1982.

[9] VRRPd Linux Implementation, ImageStream Internet Solution, Inc., http://www.imagestream.com/VRRP.html

[10] Chariot, NetIQ Corporation, http://www.netiq.com/products/chr/default.asp

[11] Gentoo Linux, Gentoo Foundation, Inc., http://www.gentoo.org

[12] Avalanche 220™ and Reflector 220™, Spirent Communications, http://www.spirentcom.com/