Joaquim Sierra Bernat

Redundancy and load balancing at IP layer in access and aggregation networks

Thesis submitted for examination for the degree of
Master of Science in Technology.

Espoo 26.1.2011

Thesis supervisor:

Prof. Jukka Manner

Thesis instructor:

M.Sc. (Tech.) Pasi Kinnari

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Joaquim Sierra Bernat

Title: Redundancy and load balancing at IP layer in access and aggregation networks

Date: 26.1.2011 Language: English Number of pages:9+79

Department of Communications and Networking

Professorship: Networking Code: S-38

Supervisor: Prof. Jukka Manner

Instructor: M.Sc. (Tech.) Pasi Kinnari

Mobile communications trends are towards the convergence of mobile telephone network and Internet. People usage of mobile telecommunications is evolving to be as on fixed broadband devices.
Thus, mobile operators need to evolve their mobile legacy networks, in order to support new services and offer similar availability and reliability than the rest of Internet. The emergence of all-IP standards, like Long Term Evolution, is pushing this evolution to its final step. The challenging and highly variable access and aggregation networks are the scope of such improvements.
The thesis presents in detail different methods for increasing availability on high-end switches, analyzing its strengths and weaknesses. It finally evaluates the implementation of an enhanced VRRP as a solution for high availability testing then feature on a real network.

Keywords: Redundancy, Aggregation networks, Load Balancing, VRRP, Proxy ARP, HSRP, GLBP, CARP, Mobile operator networks

Preface

This thesis has been carried out at Tellabs Oy, at the Control Plane Development department located in Espoo.

Firstly, I want to express my gratitude to my instructor Pasi, who offered me an invaluable opportunity to work on Tellabs as a trainee. I would like to thank everyone I met in Tellabs for their help when difficulties arrived, specially Ari, Jukka and Jari. I would also like to thank Ville and Juhamatti for their impressive help every time I needed.

I wish to thank my supervisor Jukka Manner for his feedback when fixing and improving the thesis.

I have to thank many people who have been important during my thesis work: in despite of the long distance, I often felt like in Barcelona, Cusco, Mataro or Terrassa.

Finally, I would like to thank someone very special for the support and friendship he offered me during this period of my life: Kiitos paljon Teemu!

Otaniemi, 31.12.2010

Joaquim Sierra Bernat

Contents

Symbols and abbreviations

Abbreviations

10GE	10 Gigabit Ethernet
24/7	24 hours per day and seven days per week
3G	Third generation of mobile comunications standard
4G	Fourth generation of mobile comunications standard
ARP	Address Resolution Protocol
APS	Automatic protection switching
ATM	Asynchronous Transfer Mode
AVG	Active Virtual Gateway
AVF	Active Virtual Forwarder
BFD	Bidirectional Forwarding Detection
bps	bits per second
BSC	Base Station Controller
BSD	Berkeley Software Distribution
BGP	Border Gateway Protocol
CAPEX	Capital expenditure
CARP	Common Address Redundancy Protocol
CLI	Command-line interface
DiffServ	Differentiated services
ECMP	Equal Cost Multipath
EDGE	Enhanced Data Rates for GSM Evolution
FE	Fast Ethernet
Gbps	Gigabit per second
GE	Gigabit Ethernet
GGSN	Gateway GPRS Support Node
GLBP	Gateway Load Balancing Protocol
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HMAC	Hash-based Message Authentication Code
HSDPA	High-Speed Downlink Packet Access
HSRP	Hot Standby Router Protocol
ICMP	Internet Control Message Protocol
ID	Identifier
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
IP/MPLS	Internet Protocol/Multiprotocol Label Switching
IRB	Integrated Routing and Bridging
IRC	Internet Relay Chat

IRDP	ICMP Router Discovery Protocol
ISIS	Intermediate System To Intermediate System
ITU-T	Telecommunication Standardization Sector
kbps	kilobits per second
LAN	Local Area Network
LDP	Label Distribution Protocol
LMP	Link Management Protocol
LSP	Label Switched Path
LTE	Long Term Evolution
MAC	Media Access Control
Mbps	Megabits per second
MGW	Media Gateway
MPLS	Multiprotocol Label Switching
MPLS FRR	MPLS Fast Reroute
ms	millisecond
MSP	Multiplex Section Protection
MTBF	Mean Time Between Failure
NAT	Network Address Translation
nm	nanometer
OPEX	operational expenditure
OSI	Open Systems Interconnection
OSPF	Open Shortest Path First
QoS	Quality of Service
PDH	Plesiochronous Digital Hierarchy
PWE3	MPLS Pseudowire
RIP	Routing Information Protocol
RFC	Request for Comments
RNC	Radio Network Controller
RSVP-TE	Resource Reservation Protocol - Traffic Engineering
SAToP	Structure-Agnostic TDM over Packet
SDH	Synchronous Digital Hierarchy
SFP	Small form-factor pluggable transceiver
SGSN	Serving GPRS Support Node
SHA1	Secure Hash Algorithm
SMS	Short Message Service
SNMP	Simple Network Management Protocol
SONET	Synchronous optical networking
SSH	Secure Shell
TDM	Time Division Multiplexing
TDMoIP	TDM over IP
TTL	Time To Live
UMTS	Universal Mobile Telecommunications System
UTP	Unshielded twisted pair
VF	Virtual Forwarder

VLAN	Virtual Local Area Network
VPN	Virtual Private Network
VRF	Virtual Router Forwarding
VRID	Virtual Router Identifier
VRRP	Virtual Router Redundancy Protocol
WCDMA	Wideband Code Division Multiple Access

1 Introduction

The evolution of mobile networks towards the mobile broadband everywhere is creating innovative applications and services. New business models may be created thanks to mobile Internet. Moreover, the capabilities of mobile phones are continuously increasing and their cost decreasing. Such trends results in a proliferation of mobile internet usage worldwide. Consequently, mobile operator networks, which are the result of convergence between the cellular telephone networks and the Internet, keep evolving to support new services at higher bitrates.

Every day, more mobile applications and devices are increasingly consuming more bandwidth. Tellabs, vendor of networking devices focused on mobile services, announced that they expect mobile data traffic to grow from 30 to 50 percent a year during the beginning of the current decade. Almost all the important worldwide mobile operators know that next generation of networks and devices will be a reality in very short term. LTE (Long Term Evolution) and LTE Advanced, commercially known as 4G, will allow an important increase of data rate produced by user. The fourth generation of mobile telecommunications standard is completely based on Internet Protocol. That clearly shows the intention of LTE: extending Internet to the mobile world. Therefore, the next all-IP (Internet Protocol) [17] mobile Internet should be able to carry a very large amount of data, with high bitrates, low latency and high availability for both voice and data services. Such increase in the quantity of transmitted data will stress the current mobile operator networks, which should keep their availability as high as it is today. Consequently, a global network improvement needs to be done.

Such improvement has to be done in every possible layer of mobile networking systems. Mobile operator nodes need optimized hardware and software, but also the network itself need to have a smart design and a reliable architecture. Hardware improvements are evident: third layer switches and routers are rapidly evolving with increased interface speeds and packet forwarding rates. Additionally, the software complexity of these systems is increasing, with more protocols and services offered in a more reliable way. From the network architecture point of view, high availability may be achieved by using protection techniques like redundancy and load balancing. Protection in every part of the network should increase popularity. But legacy devices and expensive deployment of new generation nodes are the constraints to achieve the desired smart and reliable network. Different generations of technologies are already coexisting in the same network, increasing then its management complexity. For that reason, the improvements should keep the complexity of the network management as simple as possible.

Mobile operators and mobile networks are facing important challenges with this rapid increase of the amount of traffic they need to manage. These companies must concentrate more efforts on their own business processes, instead of focusing on first hop redundancy techniques or balancing access layer

traffic. Mobile networking systems should make life easier to mobile operators, not more complicated. Thus, in order to gain competitive advantage from their networked systems, the network should ideally handle first hop outages and load balancing tasks automatically.

In order to increase the availability of the network, downtime must be minimized. Redundancy has to be added in every part of the network, mainly at the network edge, where a failure would have less protection mechanisms. Indeed, the network edge is the scope of the widely called First Hop Redundancy protocols. Such protocols offer IP virtual redundancy, meaning the usage of an already existing device as backup gateway on IP networks. These protocols share the advantage of working in simple devices without routing capabilities. They all have an automatic behavior once configured, which is one of the desired features. Additionally, those protocols allow the load sharing between the actual gateway and the backup device, thus reducing the working stress of many devices while taking advantage of underused ones.

Many mobile networking devices, like Tellabs 8600 Managed Access System, need the implementation of a First Hop Redundancy protocol providing virtual redundancy and load balancing capabilities. After analizing the characteristics and peculiarities of mobile operator networks, a list of possible protocols is produced. The strengths and weaknesses of such candidate protocols are compared and discussed. Thus, this thesis is about the analysis of different protocols which main purpose is the virtual redundancy at IP layer. Among all available choices, the Virtual Router Redundacy Protocol (VRRP)[5] appears as the most suitable in that case. Then, once VRRP is elected, the protocol is implemented in the Tellabs 8600 Managed Access System following the RFC 3768 specification, which is the version 2 of VRRP. Additionally, some enhanced features are also implemented to improve the behavior of standard VRRP. Finally, the enhanced protocol is tested on a network composed by Tellabs 8600 switches and the results are analyzed.

This thesis is structured in different chapters, each of them focusing a concrete topic. Chapter 2 discusses about mobile operator networks, the scope of the problem treated in this thesis. Chapter 3 focuses on the problem of high availability on networking. It presents different methods to achieve it, showing First Hop Redundancy protocols as the possible solution for the problem treated in this thesis. Then, in Chapter 4, a detailed description and analysis of several protocols is done, choosing the VRRP as the most suitable option. Chapter 5 describes the Tellabs devices where the chosen solution is implemented. The following chapter, Chapter 6, explains the implementation of VRRP on Tellabs 8600 switches, describing the additional features which enhances the selected solution. Chapter 7 explains the tests done on real devices once the solution is implemented. It also shows the results and compares them with the theoretical and standard ones. Finally, Chapter 8, as conclusion, describes the benefits and downsides of the implemented solution when solving the problem of high availability in mobile networks.

2 Networking scope

In this chapter, the current trends on network usages are commented, focusing mainly on mobile telecommunications operators and their deployed infrastructures. The most popular protocols on such operator networks, i.e. IP and MPLS (Multiprotocol Label Switching) [16], are described from mobile operator point of view. The different mobile telecommunication generations are briefly commented. At the end of the chapter, the discussion is about the aggregation networks, the part of operator infrastructure that aggregates traffic from different access networks to their core part. Aggregation networks are accurately described as they are the scope where the implementation done in this thesis will be installed.

2.1 Mobile Internet

With the explosion of the Mobile Internet, smartphones and 3G connected laptops have become common devices that continuously send and receive data from Internet. The importance of being connected to the network is increasing every day. Mobile devices using internet are not a minority anymore. Today, the mobile internet is spreading everywhere, resulting in an early outdated mobile phone network. Such networks need to evolve on reliability and supported services should be comparable to the regular internet, as bandwitdh and hardware capabilities of mobile devices are ready for most of internet services.

Many businesses depend on the mobile network to conduct their day-to-day operations, to offer their services, and to sell their products. Further, as a result of the worldwide reach of Internet, businesses need to be open 24/7 to serve one part of the world while the other part sleeps. Global Internet crosses language, geographic and time barriers. Consequently, it has become critical for mobile networks to be working and available 24 hours a day, seven days a week, to serve a multinational customer base. Indeed, as the mobile access and aggregation networks are part of the global internet, identical considerations about high availability, redundancy, traffic sharing, and disaster recovery need to be taken into account on global internet networks and mobile internet networks as well.

2.2 IP/MPLS on radio aggregation networks

2.2.1 Internet Protocol

Using IP technology in mobile aggregation networks was not that common during last few years. IP is placed in the OSI (Open Systems Interconnection) layer 3. However, the other transport technologies that have been used since long time ago, such as SDH (Synchronous Digital Hierarchy) [53] or ATM (Asynchronous Transfer Mode) [52], are located in layer 2 of the OSI

model. That makes compulsory for IP the use of available layer 2 technologies. As the networks are evolving more and more and they are increasing the use of packet based technologies, it is inevitable to question whether the access transport networks should be switched or routed. Mobile operators are currently transforming their access, aggregation and transport networks to IP routed networks. This change provides lots of business benefits to operators, as these radio access networks are much easier to manage and are fully compatible with the rest of Internet services. In the short-term future, the next generation of mobile communications, known as LTE or 4G as well, will produce for the first time the all-IP radio access networks. Depending on the willingness of the leading mobile operators and vendors to make investments, LTE might be commercially available in only a couple of years or less. The all-IP mobile network will be then a reality.

2.2.2 MultiProtocol Label Switching

MPLS has been developed as flexible, low overhead and cost effective network architecture for modern packet switched networks that require large bandwidth management. The main advantages of the MPLS networks are a capability to encapsulate many different transport protocols and the traffic engineering extension which makes possible an intelligent management of network resources. It also includes the necessary signalling protocols to discover, configure and manage connectivity in the network.

The usage of MPLS in the radio access networks allows the tunnelling of some other layer 2 technologies on inexpensive IP MPLS networks. Indeed, the MPLS encapsulated connections are based on the recommendation RFC 3985, which allows the emulation of services like ATM, Frame Relay [31] or TDM (Time Division Multiplexing) on a commonly called pseudowire edge-to-edge emulation (PWE3) [14]. As the RFC details, a pseudowire is established by two unidirectional LSP (Label Switched Path), permitting a simple and useful technique for delivering legacy technologies over IP MPLS networks. It separates the transport protocols from the transmission media, thus providing a more flexible transition from old standards. Pseudowires have become a very common mechanism for backhauling mobile traffic from cell sites to the RNC (Radio Network Controller) or BSC (Base Station Controller). Those pseudowires are usually provisioned from some centralized network management system, which is always the case on mobile operator networks.

2.3 Legacy mobile networks

Mobile networks have been continuously evolving during the last decades. New standards, new services, new business possibilities appear pretty often to promise a new era on mobile communications. Even though they might seem a revolution to the end user, from the data network point of view few changes

have strongly affected the operator networks. Those important changes coincide with different generations of mobile telecommunications standards:

- 1st generation

  During the 80's, the beginnings of mobile telecom industry, the information carried by the mobile network infrastructure was exclusively analog voice. Additionally, some signalling data was transported by the network with management and administration purposes.

- 2nd generation

  During the 90's, GSM (Global System for Mobile Communications) became a reality: the first generation of digital mobile communications was deployed, popularizing mobile communications worldwide. For the first time, an important amount of digital data was going through mobile operator networks with SMS (Short Message Service) messages at the beginning. A significant evolution of GSM standard permitted data traffic generated by mobile devices at a higher data rate: GPRS (General Packet Radio Service) and EDGE (Enhanced Data Rates for GSM Evolution) packets appeared later on, they were commercially known as 2.5G.

- 3rd generation

  Since 2000, the UMTS (Universal Mobile Telecommunications System), also known as WCDMA (Wideband Code Division Multiple Access) standard is pushing mobile communications to another step. The UMTS is the first mobile communications system which is mainly designed to transport data, not only packetized voice. Such network is currently the most deployed on developed countries and it has an increasing importance on developing areas. More recent technologies like HSDPA (High-Speed Downlink Packet Access) are offering throughputs around 14 Mbps per single customer. As a result of such high user data rates, massive multimedia usage and intensive internet navigation are stressing the existing operator networks. Both capacities and services of networks should be improved to guarantee a minimum service level to the end customer. For example, the use of redundancy on data communications must be considered in order to increase the network resiliency.

- 4th generation

  LTE is the last standard on mobile communications, being finished and standarized a couple of years ago. During 2010, real environment tests of LTE communications have been done at a small scale. LTE is meant to dominate the mobile industry during the next decade, as it allows very high data rates for a single device. For the first time, all the archicteture of the standard takes into account the network homogeneity at layer 3 of OSI model in almost every operator network. As the majority of network

infrastructure uses IP because its multiple advantages, LTE is designed to communicate and offer its services in an all-IP environment. Although LTE is widely considered as the 4G standard, it does not match the requirements for such name. The future LTE Advanced, which should be ready during 2011, will be considered as the real 4G standard.

As it may be seen in Figure 1, mobile operator networks are currently composed by many different elements using the old and recent technologies previously explained. Nodes are, quite often, very specialized and only perform functionalities. However, several network interconnections can be done using third layer switches like Tellabs 8600, network devices that are compatible with different generations of mobile network technologies. Such devices may support services from different mobile standards at the same time. Figure 1 shows different Tellabs switches interconnecting devices working on GSM, UMTS and LTE services in both access and aggregation networks.

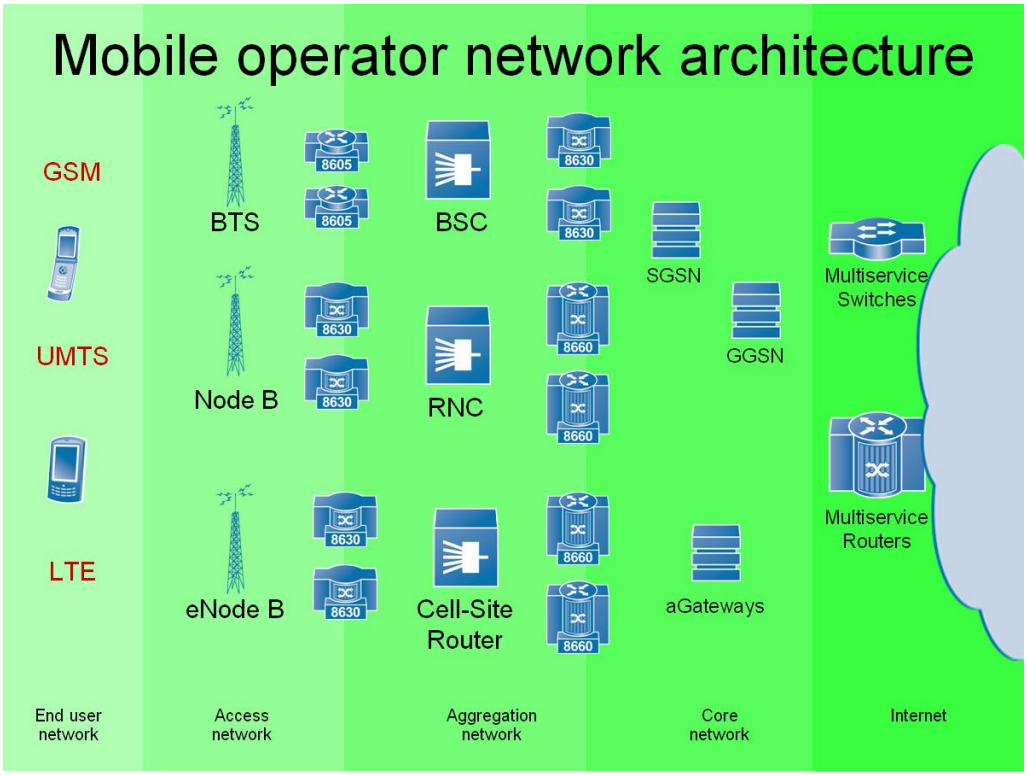

Figure 1: Mobile operator network architecture

To support and follow network evolution at the aggregation level, operators face often the dilemma of either progressively evolving their networks, incurring only incremental CAPEX (Capital expenditure) and OPEX (Operational expenditure), or revolutionizing their networks, leading to potentially huge investment and sharp CAPEX and OPEX increases.

As every mobile standard generation has a profitable period longer than a decade, different generations of mobile infrastructure coexist and must coexist during a certain period of time. No operator can afford the cost of disabling the old infrastructure and creating a new one from scratch every time a new generation of mobile standards appears.

From the economic point of view, common sense and cost saving policies prevent operators of dismantling any part of the network if it is still profitable. For that reason, there are several technologies being used simultaneously. The presence of legacy devices in the mobile network infrastructure is a challenge when evolving services and improving capacities. Today, the presence of GSM networks cohabiting with GPRS and UMTS nodes increase the complexity of the global mobile infrastructure.

2.4 Aggregation Networks

The protocol implemented in this thesis will be working on Tellabs 8600 switches. Such devices have been focused on mobile operator networks. Then, it is important to describe in a more detailed way the type of network where the implementation will operate. More specifically, such devices and their services are meant to be placed in the networks that interconnect the access mobile network and the core internet. Such networks are result of the aggregation of large amount of widely varied flows. The variety of these flows is due to their diverse origin: mobile operator customers use very different kind of radio access networks, thus mixing together in the same network infrastructure GSM calls and WCDMA data connections, for example. This network infrastructure with such variety of aggregated flows is commonly called aggregation network. As other parts of operator networks willing to offer uninterrupted service, aggregation networks need to apply some protection and redundancy to avoid or minimize blackout periods in case of a failing device. Here, VRRP appears as the best possible option offering redundancy and protection against failure.

Most of data networks placed between the access networks and core networks may be considered aggregation networks as they aggregate and concentrate traffic data destined to the core network. As explained, different type of nodes may compose the aggregation networks. For example, a device working as a RNC (Radio Network Controller), a BSC (Base Station Controller) or even a SGSN (Serving GPRS Support Node) might pertain to such type of network. Those nodes can use different technologies at link level as they come from different periods and different vendors, but they may be able to share VRRP feature as they probably use IP as network protocol.

RNC, BSC, MGW (Media Gateway), SGSN and GGSN (Gateway GPRS Support Node) are typical nodes on aggregation networks. Briefly, RNC and BSC are the control and management of base stations and antennas. GGSN is basically the responsible for intercommunication between GPRS network and the global internet. SGSN is a multifunction node with tunnelled communi-

cations with radio network controller and it also builds IP packets for sending to GGSN. Finally, MGW is charged with the translation of multimedia flows into different standards.

Such mobile network nodes do not have the same network capabilities as switches, routers or hosts which are normally present on large data networks and global internet. Those devices have been designed focusing on mobile services and they are specialized on offering concrete mobile services. Such mobile services are today converging to regular data services, but it was not the case during recent past years, however. Thus, most of the components of any aggregation network are not able to run routing protocols or any expensive protection mechanism. They cannot be considered as regular data network nodes as they do not support the most generic standards. For example, as routing protocols are not supported, IP static routes are typically used. For this reason, the implementation of VRRP on Tellabs 8600 switches is useful to their owners, mainly mobile network operators. They can maintain their current static configuration in the networks when adding redundancy and even load sharing at IP layer. For example, RNC are commonly configured with multiple default routes per service. The use of some redundancy protocol like VRRP can provide load balancing through the different connections of this device.

2.5 Summary

Mobile communications are evolving so fast that many different network technologies are still running at the same time. There are different types of nodes, each of them with some specialized functionality inside a complex and challenging network. As the bandwidth and usage of such infrastructure is increasing, network stress may result in an increase of outages. Then, a failover system should be developed on such networks, where dynamic routing and expensive protection mechanism cannot be used. Indeed, a virtual redundancy protection would be an optimized system for failing routers at networks edge, where static routes are used.

3 High availability

This chapter describes different ways of achieving high availability on networks. Those manners may be combined to increase the availability until high-end operator standards, typically around five nines: 99.999%. Among these different possibilities there are protocols, equipment redundancy, load sharing, link redundancy, which are described as follows.

3.1 Problem

Network failures result in losses of several millions of dollars to businesses each week. These network failures can stem from a range of causes: from the failure of a single network device to a disaster striking the area where the key network devices are housed. Network outages may cost a business up to 200.000 dollars per minute.[1] The market stakes may even be higher than the monetary damage. As follows, a good example to relate the impact of unavailability: a user may be satisfied with 99.9 percent of time availability of his or her network. But 0.1 percent of time unavailability can be very damaging in other contexts, especially when real time communication is a must. Thus, Internet has to preserve its all time working status for millions of businesses to avoid those important losts.

For this reason, every part of the internet needs to be robust, and every network attached to the internet must be reachable at all times. As a result, high availability, redundancy, traffic sharing, and disaster recovery are high priority concerns for network manufacturers determined to build accessible and robust networks.

3.2 Alternative uses

Not only unexpected failures provoke the unavailability of a router. There are many other reasons for using a redundancy system on large networks. For example, when there is a backup router, routine maintenance becomes possible without disrupting traffic. Besides, there are some access and core switches, such as Tellabs 8660 and Tellabs 8630, with redundant control cards. This means the node has duplicated power supplies, duplicated hardware and duplicated software containing duplicated protocols status and routing tables. Then, if any of those cards fails, the redundant card takes over all the routing duties without any traffic disruption. With these types of nodes, software maintenance without outage is possible although there is no redundancy protocol running in the network segment.

Another way to avoid losses of traffic during software maintenance is the use of graceful restart routing protocols. A specific version of typical routing

[1] Commented in the preface of *VRRP: Increasing Reliability and Failover with the Virtual Router Redundancy Protocol* [2]

protocols, such as BGP (Border Gateway Protocol) [21], OSPF (Open Shortest Path First) [26] or ISIS (Intermediate System To Intermediate System) [32], allows traffic forwarding during a certain period of time even though a protocol in the neighbor is shutdown. The graceful restart protocol freezes the own routing table during some time, and it extends the validity of the routes longer than a standard protocol does. This mechanism uses the fact that a change of topology is a very rare and uncommon event. For example, during the software update of some device, links are still up and routes are valid despite the node needs to be restarted. In other words, it is very unlikely that a neighbor will change its address or location during the restarting process. Thus, the routes are valid longer time than standard protocols indicate.

3.3 Network equipment redundancy

Today, all manufacturers offer both redundant power supplies and redundant cooling systems for their top-level routers. These components are the most trivial to protect. However, they are also the most prone to errors as they contain mechanical or moving parts and they use high voltages.

Routers intended to core networks often offer redundant forwarding plane. These devices with redundant forwarding plane are capable of keep forwarding, even if some parts of the switching hardware break down. Normally, these components are also hot swappable, thus without any need of shutting down the whole device when changing a forwarding component of the router.

The most high-end routers are built with redundant and hot swappable control cards. Hot swappable control card enables fast replacement of this key component in case of failure, with no need of turning off the device. Using standard routing protocols, when the active control card is removed or failing, routing protocol adjacencies with neighbors are disrupted, even though the backup control card takes the active role immediately. For example, if a control card crashes on a router and the card needs to restart, the neighborhood will detect the restarting process and will react following the instructions of standard protocols. Due to the control card switchover of the router, all sessions between the failing control card and the neighbors will be disrupted, and then closed. When the backup control card becomes active, it reestablishes the sessions. But during the meantime, the neighbors have advertised to their neighborhood a change on the topology: the failing router is not valid anymore as a next hop to any further destination. Consequently, an important part of the network is not aware of the fast control card switchover in the failing router, and many nodes are still searching new routes for reaching the failed router. Such behavior wastes network resources and does not take full advantage from the redundant control card.

3.4 Graceful restart protocols

A possible improvement that networks should implement is the combination of redundant control cards with graceful restart protocol extensions. Such protocol extensions permit that the actual traffic forwarding can still continue without any disruption and without any wasting of network resources. The graceful restart extensions add a grace period to the standard protocols, affecting the way how failures advertisement is done in the network.

In case of control card failure, the neighbors of the failing router do not immediately report to their own neighbors that the failing router is no longer available. Instead, they wait a certain amount of time equal to the grace period before sending any advertisement. Generally, this value may be modified manually by the user or negotiated and agreed between the nodes, typically around several seconds. If the control plane of failing router comes back up and reestablishes its sessions before the grace period expires, as would be the case during an instantaneous control card switchover, the temporarily broken sessions are not visible to the network beyond the neighbors. An improvement in the use of network resources becomes possible with the graceful restart protocols.

An assumption done during the grace period is that the failing node is still forwarding traffic. The router preserves its forwarding state because it should be a control plane failure which is involved in the event. Then, forwarding capabilities of the router are assumed to be intact. But might not be the case if there is a link failure. In fact, such graceful restart extensions permits the creation of *blackhole traffic* if forwarding plane failure is not correctly detected. In other words, the restarting nonstop forwarding node might potentially send traffic to a destination that is not correct anymore, thus network is dropping silently traffic which should be valid.

Even though the nonstop forwarding ability and the optimized network resources allocation, graceful restart protocols have some weaknesses. As commented before, valid traffic could be silently dropped during the grace period if its destination is no longer available. Additionally, all nodes of a network are required to be running the graceful restart extensions. Such fact is particularly bothersome in operator networks, which includes many different vendors and different environments. Another reason for not betting on graceful restart protocols is because of its behavior during the grace period: the network topology is *topologically frozen*. Therefore, any change on network topology during this period will not be taken into account.

In brief, redundant architecture is definitely an obligatory component for any network willing to offer high availability services, like mobile operator networks. Redundancy allows the network to keep working even though an unexpected failure or a scheduled update occurs.

3.5 Usefulness of additional redundancy

In a network redundancy configuration, is there any real benefit of using more than two routers instead of just two? In most cases, there is no difference. Statistically, the probability of failure of one router is relatively small, but not zero. A typical value of MTBF (Mean Time Between Failure) in commercial routers and switches is around 15 years. Considering that it normally takes around 24 hours to repair a broken device, it should be expected to need a backup router about 0.018 percent of the time (1 day every 15 years). Objectively, it is a small percentage of time. However, taking into account the large amount of routers in a mobile data network, the accumulative probability of having a critical failure in some point of the network might become rather large.

Using a redundancy mechanism to automatically and transparently take over all routing functions for a segment it will prevent outages except the case of both routers failing simultaneously. As accumulative probability of independent events indicate, such an event will happen with a probability equal to the square of only one router failing. Therefore, the MTBF of the redundant configuration is about 80 thousand years. Comparing this value with the MTBF of just one device, it is obvious that the advantage of using a backup router is very remarkable.

Then, considering the case of an additional backup router that is added to the same network segment, the network would have three routers. Such number of redundant elements might be considered as additional redundancy, or in other words, the redundant router may be considered as protected by another router: it has its own protection. On such overredundant scenario, the probability of outage becomes very small as it is the product of the three individual probabilities. The resulting MTBF is about several million of years. That sort of reliability is not needed by practically any data network.

In fact, all these statistical arguments assume that the failure of one router is completely unrelated with the failure of the backup. This is very unrealistic assumption: real networks are built with many interdependencies, such as power supply source, location of nodes, link paths, etc. Therefore, a network may frequently suffer from related simultaneous failures. Indeed, such probability needs to be taken in consideration. Adding redundancy might not help the robustness of network at all. However, if properly configured, additional redundancy might be used on a very specific and really critical network segment, i.e. the use of two backup routers instead of only one in some bank network.

3.6 Protection

Protection is a term that is used in telecommunications field to define a high availability architecture where one or more additional components can be used almost instantly to take over the functionality of a failed component.

Protection is commonly used in telecommunication networks to provide high availability telecommunications services, a must in the core networks. Most of protocols present in the core networks, like PDH (Plesiochronous Digital Hierarchy), SDH and ATM follow some protection guidelines. For example, SDH specifies a maximum duration of network failure of 50 ms, meaning that any traffic rerouting must be done during the first 50 ms after the failure. To achieve this instantaneous switchover, SDH uses different states (Loss of Signal, Loss of Pointer) to indicate wether losses of traffic frames are occurring or not.

In case of protection at IP network level, things are different than in SDH because IP is a stateless protocol. Therefore, as long as routing tables remain synchronized, forwarding can easily be replicated, therefore redundant: the only things that affects where a packet is forwarded are forwarding table and the packet itself. Then, protection can be achieved using IP protocols, such as routing protocols or First Hop Redundancy protocols if node capabilities are limited.

Protection may be divided into equipment protection and line protection. Equipment protection refers to the duplication of the critical hardware components on nodes, such as line cards and control cards. Line protection, on the other hand, means that several paths are reserved for the traffic to single destination.

3.6.1 Line protection

The 1+1 protection means that the traffic is sent redundantly through two physically different paths and the receiving node selects the better of the two signals. The 1+1 protection offers extremely fast failover times.

The 1:1 protection means that there are two alternative paths for the traffic. Only one is used at a time, the active path. When the connection is disrupted, the traffic is directed to the alternative path. This protection method allows the secondary path to be used for low priority traffic, which is dropped when failover occurs. The reliance on signaling traffic for the failover decision might make 1:1 protection relatively slow, but is much cheaper to use than 1+1, as redundant capacity can be utilized even during non-failure behavior. Other techniques based on local repair mechanisms are as fast as 1+1 protection. This is the case of MPLS FRR (MPLS Fast Reroute) [30] protection system.

The n:m protection refers to a situation where for n active paths there are m alternative paths. Because telecommunication industry bases failure scenarios in probabilities, the n:m is the preferred protection method, as it allows an optimised trade-off between protection performance and cost-effectiveness.

Using protection is the most efficient way to build high availability networks, since it offers failover times from almost zero, offered by 1+1 protection, to a few milliseconds offered by lesser protection technologies.

3.6.2 Protection on mobile networking scope

On mobile access and aggregation networks, the main reason for using protection is to guarantee that real time traffic is not lost in case of an undesired event, such as equipment or line failure.

Today, phone calls are still the main source of incomings for mobile operators. Even though the economical benefits from data connections are increasing rapidly, real time communications like phone calls are still the *cash cow* for operators. Thus, the use of protection on mobile network infrastructure keeps the main business on track in case of network failure, without causing any delay or disruption.

The other typical uses of protection are software updates and other maintenance work. As mobile networks are evolving fast and mobile operators are willing to improve their coverage and increase their capacity, hardware and software updates are very common events. Similarly, any maintenance work may be considered as an improvement of the network. Protection methods also enable making changes to lines and equipment without interruption.

It is important to notice that the incentive for mobile operators to ensure high reliability comes from both business and legislation reasons. As telecommunication is a cornerstone of the information society, many countries have legislations that require certain level of reliability for the telecommunication and mobile infrastructure.

3.7 Load Balancing

When designing and building a network, whatever is the size or purpose, an estimation of traffic that the network must be able to manage has to be done. Traffic estimation includes peak traffic and average traffic, among other values. Then, the network capacity that should be really built is a trade-off between these two quantities: peak traffic and average traffic. Besides those values, the network is usually built slightly oversized in order to avoid a too early update. This is because the *always increasing* pattern that follows data traffic: by default, year after year, more users and more traffic per user is generated and managed by networks.

Such fluctuations and variations on data traffic make impossible to perfectly match the traffic needs and network capacities. To compensate these imbalances between real time needs and fixed capacity, different load balancing techniques are in use to allow a better allocation of network resources. Among these methods, load sharing between default and redundant devices during regular service allows taking advantage of redundant capacity, unused otherwise.

Although data traffic pattern follows a daily and weekly routine, some large events or unexpected behavior of people may stress data networks, achieving very high peak of data traffic at some location during some time. As network resources are scarce and expensive, this peak usually exceeds the

maximum capacity of the network. On such cases and in many other different cases as well, automatic load balancing methods are compulsory to assure optimized use of network resources. Those methods can spread the traffic load through the network following certain rules, thus permitting the use of network resources that are underutilized otherwise.

Moreover, load balancing provides a basic failover system, as there are always different alternative paths to route data traffic from its origin to its destination. Then, a load balanced network is able to absorb failures without disrupting always-on services.

Finally, considering the mobile access and aggregation networks, some load balance method should be implemented in every network composing the global internet, as the ones already mentioned. The reason is clear from the point of view of the mobile operator: it is still too expensive to underutilize network resources. Taking into account that the amount of mobile data traffic is rapidly increasing up to critical levels, about 30-50% yearly for the next years, and the network resources are still expensive, a cheap optimisation of resources allocation must be done. The downside is that it makes more complex the network management and increases the management overhead. Different options appear when looking for such optimisation of network resources. Indeed, all of them are different implementations of network load splitting techniques. Two protocols seem to match the requirements of load sharing in mobile operator networks: Equal Cost MultiPath (ECMP) [11] and Virtual Router Redundancy (VRRP).

3.7.1 Equal Cost MultiPath (ECMP)

Equal Cost Multipath Routing is a technique described on RFC 2991 which combined with the most common routing protocols (OSPF, ISIS) allows the splitting of a traffic flow in several different paths. Each of these paths has the same cost from the routing protocol point of view, thus, there is no higher or lower priority path when splitting the traffic between different routes. All paths are indifferently used as they cost the same. As result of this splitting, the traffic flow is not affected by any additional delay, as there is no latency associated. Because the traffic flow is spread through the network, the available bandwidth per link is higher. Fast protection against link failure is assured by the already existing alternative paths. On the other side, it becomes more complex for the operator to keep control of the traffic and the physical route is following.

3.7.2 Virtual Router Redundancy (VRRP)

Virtual Router Redundancy Protocol is an IP standard published by RFC 3768 in its version 2, RFC 5798 [6] for the version 3. The main goal of the protocol is to keep the IP connectivity alive on networks working without any routing protocol, even if the default gateway fails. Despite being designed to backup

an IP address [2] , some VRRP configurations are able to do a basic load sharing in the network. Indeed, such configuration takes advantage of the redundant gateways: several VRRP groups are created permitting to use all the routers during steady situation.

Even though such load sharing is scalable to more routers and hosts, this splitting of traffic is too simple and static to be used in very dynamic and challenging networks. For example, if some hosts are temporarily overloading its gateway, the system is not able to autobalance the load for better use of network resources. Similarly, VRRP load sharing configuration cannot balance the traffic automatically when some part of network topology changes: VRRP should be manually reconfigured if needed.

As follows, in Figure 2, a network example of load sharing using two different VRRP groups at the same time on a single network. As explained, the load sharing configuration using VRRP is static. Indeed, hosts have been assigned different VRIDs (Virtual Router Identifier). They are just divided between those using the Router A (VRID 11) as a gateway and the other hosts using the Router B (VRID 21). VRRP is explained with details later on, see Section 4.8 for a detailed understanding of the example.

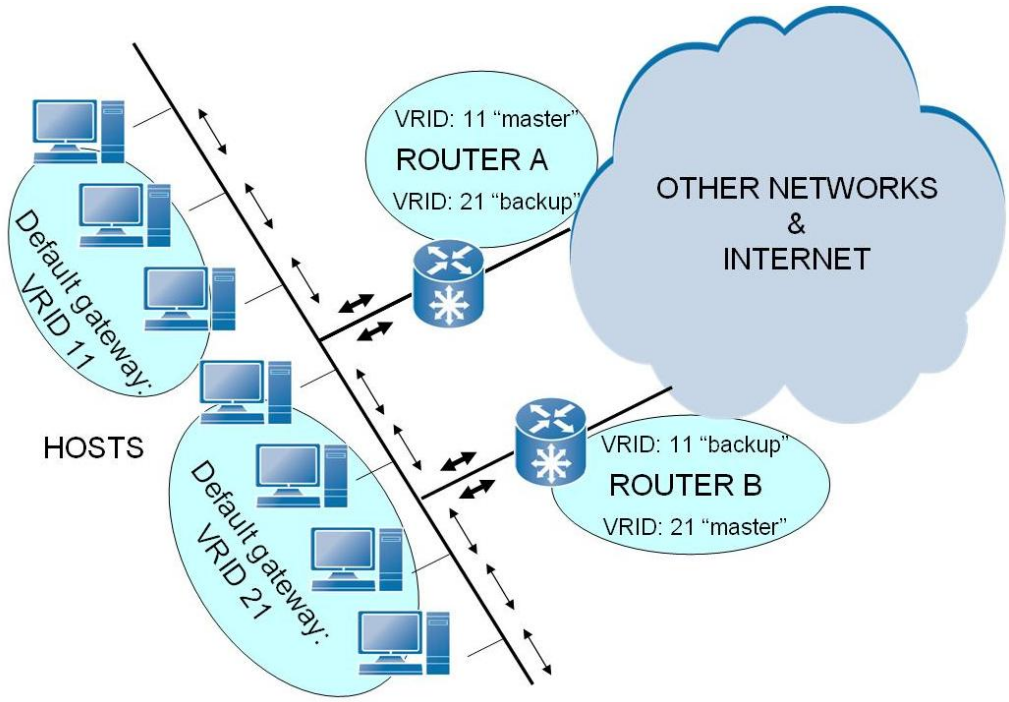

Figure 2: VRRP load sharing between two virtual routers: VRID 11 and VRID 21

[2] See Section 4.8 of this thesis

3.8 Reliability on networks

3.8.1 IP networks

IP networks require that nodes utilize a default gateway to exit the local network and achieve remote networks. Figure 3 shows a representation of a standard IP network. In the picture, all hosts use the same gateway to be able to connect with further networks. On the other side, routing protocols, such as OSPF, RIP (Routing Information Protocol) [27], ISIS or BGP, allow achieving optimized routes for every remote network. These protocols are widely used in the core network. However, when approaching the edge of the Internet, the use of dynamic router discovery mechanism on every host may not be feasible. It may be totally inadvisable for many reasons. For example, administrative overhead, processing overhead, security issues, or lack of a protocol implementation for some platforms too. Consequently, the hosts or endpoints are often statically configured with the IP address of the default gateway. Therefore, in such legacy network cases, there is a long blackout period if the default gateway fails. This outage period is as long as the default gateway recovers or the IP configuration of the hosts is manually changed. Today, no reliable corporative network would allow such a case, neither mobile networks nor Internet.

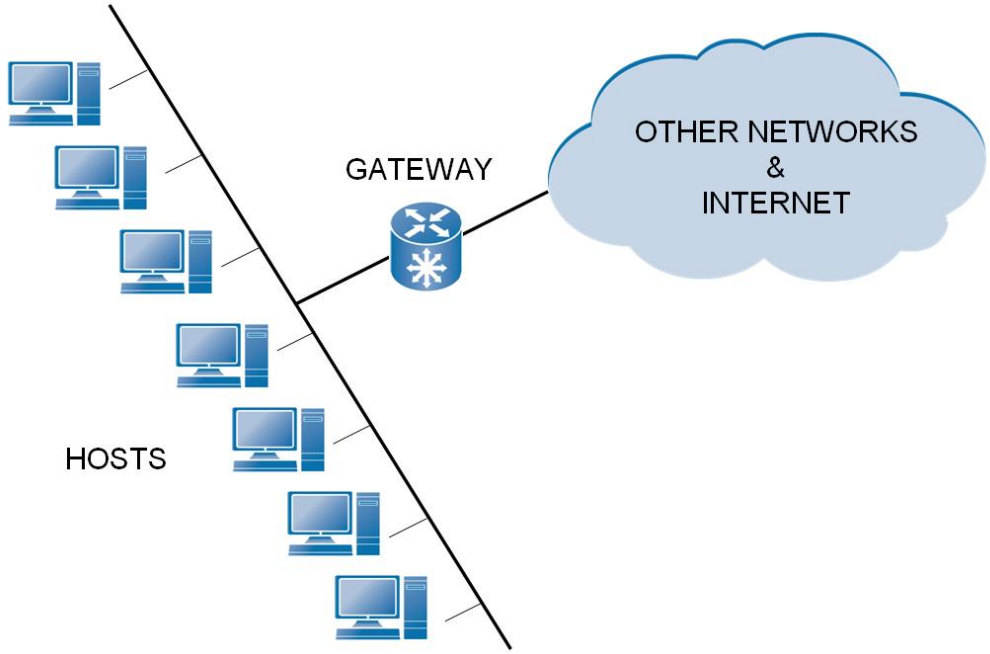

Figure 3: Typical IP network where several hosts are connected to a single gateway

Then, unless Proxy ARP mechanism is used [3] , every endpoint on every

[3] See Section 4.3 of this thesis

access network must know the IP address of the gateway to further networks. Typically, this is the address of the router or 3-layer switch that connects the local network with external networks. Thus, hosts use normally a fixed default gateway address for all traffic with destination beyond the local network.

The hosts in this shared network segment are usually configured with a single default gateway address that points to the router that connects to the rest of the network. The problem is that even if there is a second router in the segment that is also capable of being the default gateway, the end devices do not use it. Therefore, if the first default gateway router fails, the network stops working.

The most common way to direct traffic from access layer endpoints to a particular default gateway is to let the gateway respond to the ARP (Address Resolution Protocol) [19] request with its own unique MAC (Media Access Control) address. Once a particular gateway responds to an ARP request from an endpoint with its MAC address, that endpoint caches the response and will continue to use the discovered gateway for all transmissions destined beyond to the local subnet. In a standard protocol approach, all endpoints in a common subnet will use the same default gateway and uplink path for all transmissions, which leaves a possible second path unused. Thus, the standard IP-ARP behavior drives to an inefficient use of the possible redundant resources, which might be dramatic in case of outage of the gateway.

3.8.2 Mobile operator networks

LTE (Long Term Evolution) is the last step towards convergence of mobile phone and mobile data on a merged IP network infrastructure. The new mobile communications standard includes Internet Protocol in all parts of its architecture, thus becoming much easier to solve all management and administration problems. From network level, IP will imply both simplicity and flexibility on mobile operator networks. However, very specialized devices and backwards compatibility issues are creating a challenge when evolving networks to newer standards. For example, very high availability is a requirement on every mobile operator network, although not easy to achieve.

High availability is especially important in mobile operator networks, as in order to carry voice and real-time traffic, the network infrastructure must deliver the same level of availability as the old public switched telephone network. Today, a downtime on a single company network can be very expensive: it could be customers, business transactions or time-critical communications. If scaling to an operator network, then a downtime might become really expensive if several enterprise networks are affected by a blackout. For such reason, high availability is a key characteristic of any operator network.

When operators design and build network infrastructure, one possible way of achieving high availability on core services is to provide redundancy of critical components. Redundancy is typically achieved by duplicated components running in parallel, thus providing automatic backup in case of failure. Such

components might be the forwarding plane of a switch, the link path itself, the control plane of a node, the power supply of a node, or even the entire node. When properly duplicating routers, switches and links to ensure continuity of service across failures, network availability and resiliency becomes higher. In brief, such resilience in the network ensures that no single point of failure might disrupt any part of the network.

Additionally, as the forecast for the next years announces an increase of the quantity of traffic managed by mobile operator networks, load sharing techniques should be considered seriously. Such methods allow multiple components to run at the same time, with the intelligence to determine which components are available and algorithms that determine how the load is spread in the network.

In many parts of the network, dynamic routing protocols are used to keep the network running even if network problems appear. However, one place where it is difficult to provide redundancy is at the endpoints of the network. Such problem appears by different reasons, like prohibitive cost of multiple network connectivity for endpoints or its inability for running dynamic routing protocols. The impossibility of running routing protocols could be originated by the low capacity of the endpoints or by a network management decision to avoid network overhead and routing complexity. Indeed, some devices present in the operator networks do not support all the protocols, especially routing protocols. For this reason, static addressing and default gateway compose very often the entire routing table of endpoints. However, such network configuration would be invalid in case of failure of the default gateway as endpoints would not be able to reach their default gateway, thus losing its connectivity.

3.8.3 First Hop Redundancy Protocols as solution

This problem can be solved using one of First Hop Redundancy protocols. These protocols protect against the failure of the gateway when hosts are not able to learn the address of the default gateway dynamically.

First Hop Redundancy protocols emulate a cluster of default routers using other routing devices which are present in the same network. When a router member of the cluster fails, another will step in and take care of the tasks that belonged to the failed router as fast as possible. Hosts are not aware of the failure: they just keep working as they were used to, without any change on its network configuration. Thus, First Hop Redundancy protocols offer a solution that preserves original paradigm of non-intelligent hosts while removing the single point of failure.

3.9 Summary

There are several aspects for building a robust and reliable network that can survive a large-scale disaster involving component, link, and device failures.

To achieve such reliability, a good solution is to build networks with redundant physical components. In that case, networks need associated protocols and mechanisms that let the duplicated components appear as a single entity to the rest of the network. For instance, VRRP is such a protocol that permits to different routers to appear as a single node to the rest of the network, although they are physically apart from each other. In case of router or link failure, VRRP keeps alive an IP address of the failing device, thereby avoiding loss of connectivity in an automatic manner. Such automatic behavior avoids the need of any reconfiguration of hosts. Additionally, when all the components are working properly, VRRP is able of taking advantage of redundant components while distributing the network traffic load among them.

4 First Hop Redundancy protocols

The following chapter introduces different protocols globally called First Hop Redundancy Protocols. Those protocols are all pretty simple, thus they might be implemented in devices working on aggregation and mobile access networks. Consequently, they are the candidate protocols to be implemented on mobile network switches in order to increase availability of the network.

4.1 Background

Commonly, IP network hosts do not contain any routing intelligence. Indeed, the information regarding network routing is located in the routers. In practice, this philosophy assumes that most hosts have a default gateway that handles forwarding of packets from the local network to an external one. This network architecture reduces management overhead. However, if the default router fails, hosts within the router network are unable to communicate with the rest of the world.

As said before, a lot of legacy host implementations cannot manage dynamic discovery routines. However, those systems are capable of having a static default gateway configured. Therefore, redundancy and load sharing should be implemented to the legacy networks without adding any complexity to the hosts.

A method for addressing this problem is to use one of the First Hop Redundancy Protocols. These protocols select a router on a LAN (Local Area Network) segment to automatically take over if the default router fails. They were developed to solve a common problem in shared networks such as Ethernet or Token Ring: to provide an alternative gateway if it goes down. For this reason, First Hop Redundancy protocols are a suitable option for increasing redundancy on already deployed networks, such as mobile operator networks, too large to be totally renewed. These protocols try to solve and fix this problem in many different ways, with different strengths and weaknesses, but similar at some point. They usually fake the IP-ARP resolution to permit the switchover without any modification of the configuration at IP level.

4.2 Challenges

Many possible solutions to the gateway failover problem have come and gone over the years. For example, letting the end users to have to reconfigure their own default gateway address in the endpoints is a very bad option for several reasons. In addition to large chance of typographical errors, the work can require a reboot of the endpoint. Moreover, it is unlikely that the end user will think about changing back the address of default gateway when the original node recovers. And it also requires that there is somebody going to the endpoint place to make the change, which is not easily feasible in the case of mobile networks, where endpoints can be separated by few kilometers.

As previously discussed, the use of dynamic routing protocols at endpoints is not a good solution to the problem. Most routing protocols do not converge well when the number of nodes is large, typical case for mobile operator networks. Additionally, it is not a good idea that endpoints affect the global routing tables of the global network. If one of these devices is not configured properly, it could cause serious global routing problems. It is always a good principle of network design to keep network functions on network devices, far away from the endpoints. In other words, end devices should not need to worry about performing routing tasks.

4.3 Proxy ARP

Proxy ARP is a helpful method for enabling machines in a subnet to connect with other remote subnets, without any need to configure routing or identify a default gateway.

Proxy ARP refers to a technique by which one host, normally a router, answers an ARP (Address Resolution Protocol) request originally intended for another node. Proxy ARP router splits an IP network in two separate segments. Hosts on one segment can only reach hosts in the other segment through the proxy ARP router. The device performing the Proxy ARP procedure allows traffic going through itself by faking its identity: it answers the ARP requests coming from one segment and looking for a device placed in the other one. It then accepts the responsibility of routing packets to the machine for which the ARP request is intended for. To properly work, the proxy ARP device, which is located between two parts of an IP network, must know routes to all hosts on both segments.

In a typical Proxy ARP configuration, the end devices are not configured with a default gateway at all. Instead, they discover the path to remote devices in a similar way that they find devices in the local LAN segment, using Address Resolution Protocol (ARP). When routers run Proxy ARP, they respond to ARP requests on behalf of the remote device. After that, the source device simply sends a packet to the remote destination IP address using the MAC address of the Proxy ARP router as the destination MAC address, which is exactly the desired behaviour. Proxy ARP will take care of this packet after receiving it, forwarding it to the real destination device.

4.3.1 Example of Proxy ARP communication

A description of an ARP communication between a Host A in Subnet A and a Host C in Subnet B is detailed as follows:

1. Host A needs to send packets to Host C. Host A believes that it is directly connected to Host C's subnet so the regular ARP would work properly when asking for the MAC address of C.

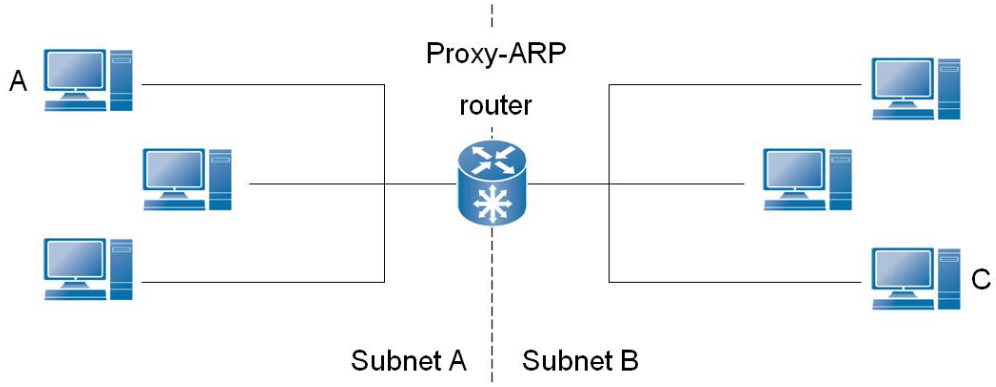

Figure 4: Proxy ARP network

2. For Host A to connect to Host C, Host A should first determine Host C's MAC address. To do this, Host A broadcasts a new ARP request on Subnet A. The ARP request is included in an Ethernet frame with Host A's MAC address as source address. The ARP request reaches all nodes in Subnet A, including the interface of the router. The request, however, does not actually reach Host C.

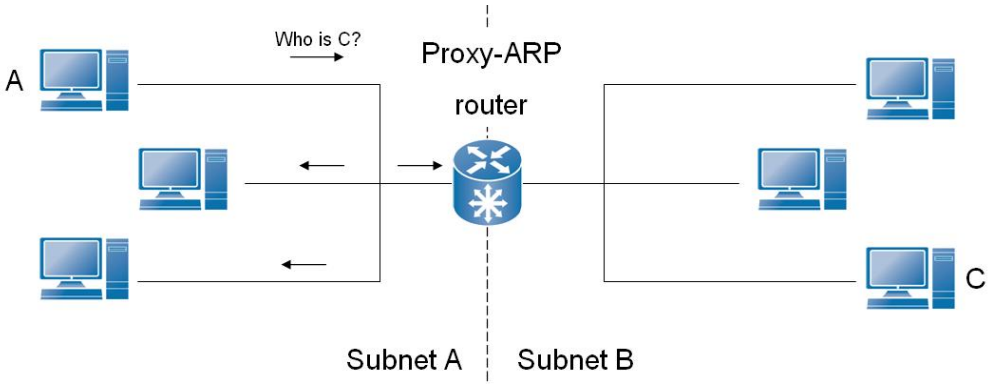

Figure 5: Broadcasting of ARP Request from A

3. The router then replies to Host A with the router's own MAC address. This is called the proxy ARP reply given by the router to Host A.

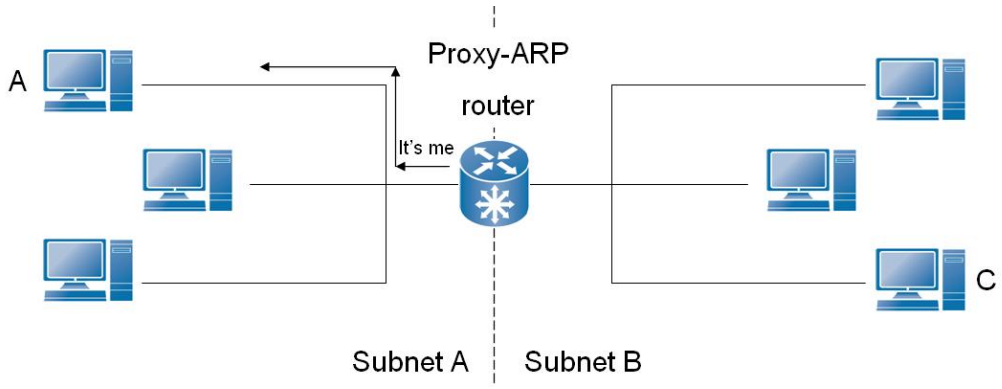

Figure 6: Proxy ARP router replies as it was the host C

4. Host A then updates its ARP table. It sends the traffic directed to Host C to the router.

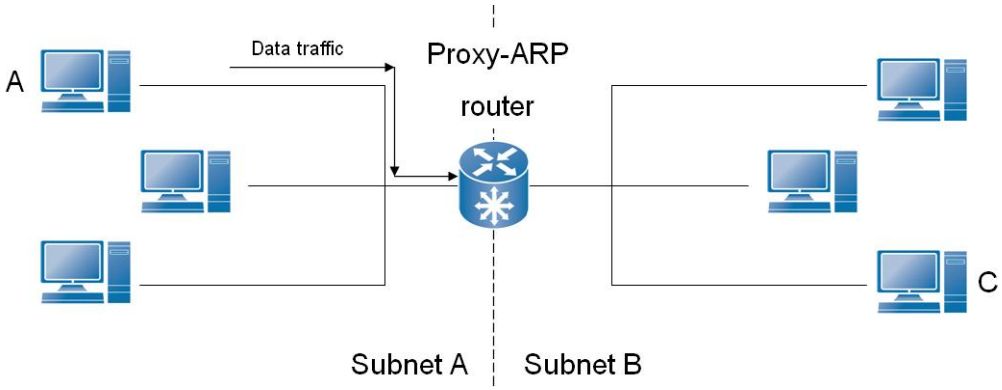

Figure 7: Traffic directed to C is sent to the Proxy ARP router

5. When router receives traffic directed to C, it forwards the packets to the Host C.

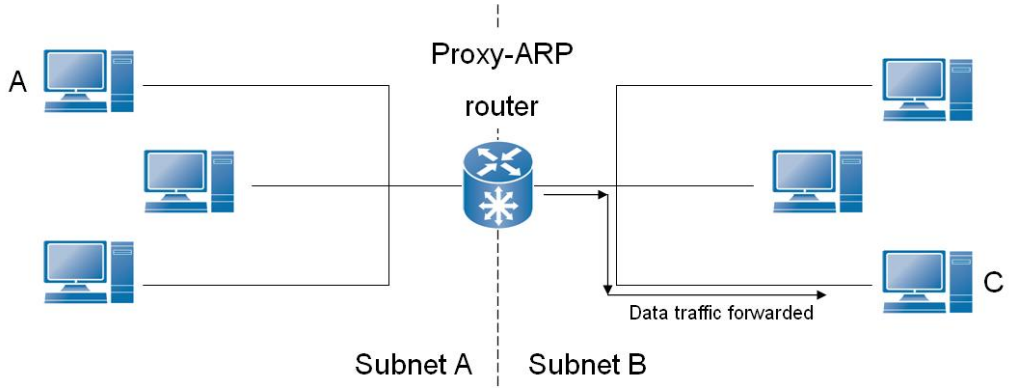

Figure 8: Traffic towards C is correctly forwarded by the Proxy ARP router

4.3.2 Proxy ARP details

From real time applications point of view, Proxy ARP protocol does not switch to a backup router fast enough when the primary router fails. This is because end devices do not change their MAC addresses very often, thus ARP entries keep its validity during pretty long time.

Indeed, ARP cache entries remain valid unless a gratuitous ARP explicitly modifies such address. Anyway, most of devices will remove a stale ARP entry if the device does not answer for around 20 minutes. Such long delay makes not feasible to use the regular ARP procedure as reliable failover system. A possible way to speed up this procedure is to clear the ARP cache of the end device by rebooting or by manual configuration, but it is still not fast enough to keep connectivity alive in case of failure.

4.3.3 Advantages of Proxy ARP

The most important advantages of Proxy ARP are listed as follows:

- It can be added to a single router on a network without disturbing the routing tables of the other routers in the network.
- Proxy ARP has no minimum requirements: it must be used on a network where IP hosts are not configured with a default gateway or do not have any routing intelligence.

4.3.4 Disadvantages of Proxy ARP

Using Proxy ARP as First Hop Redundancy protocol would have some downsides:

- It locally increases the amount of ARP traffic: it may be a problem on a busy segment.
- Hosts need larger ARP tables in order to handle IP-to-MAC address mappings.
- Low security level as *spoofing* is possible.
- It does not work for networks that do not use ARP for address resolution.

4.4 ICMP Router Discovery Protocol

Described in RFC 1256, ICMP Router Discovery Protocol (IRDP) [10] is an extension of ICMP (Internet Control Message Protocol) [18] that allows hosts to actively find a new router when their default gateway fails. As other First Hop Redundancy protocols, routers send periodic multicast hello messages to the LAN segment when IRDP is enabled. End devices listen for these messages and use them to build their internal routing tables. In case of a host not receiving these hello messages for a while, the host assumes that the router must have failed and it starts asking for a new gateway. Indeed, the end device sends a multicast query looking for a new router to take over.

4.4.1 Suitability of IRDP

This protocol is not a common protocol on operator networks because of several reasons. One of the weaknesses of IRDP is that it requires special software to be running on end devices. As end devices and hosts are commonly legacy devices with very limited capabilities, IRDP cannot be running on such old devices. Another disadvantage of that protocol is its low popularity. Due to its small acceptance, few devices support IRDP although is a standard protocol.

4.5 Hot Standby Router Protocol

Hot Standby Router Protocol (HSRP) [12] is a redundancy protocol developed by Cisco looking for solving the problem of network blackout in case of failure, providing dynamic switchover of the forwarding responsibility when an active router becomes unavailable. HSRP is, in summary, a non-disruptive and reliable failover protocol that can work with very simple and basic hosts. HSRP is documented in RFC 2281, although it is a Cisco proprietary standard.

HSRP works allowing two routers to share the same virtual IP and MAC addresses. End devices simply send their packets to further destination through these virtual addresses, as a standard default gateway. One of the routers, called the active router, will receive and forward the packets. The other router, called standby router, is just tracking the state of the active router, it is not

participating on traffic forwarding. Thus either may fail without any disruption on the traffic flow. In brief, one router is always active, and the other acts as a standby, switching to active role in case the first should fail. Additional standby routers may be configured if necessary for critical segments needing very high availability.

In a simple way, the procedure of the protocol is as follows. HSRP routers sharing a virtual IP address communicate each other by sending multicast packets periodically. If the active router stops sending these packets for any reason, one of the standby routers will immediately take the responsibility of the IP and MAC addresses, keeping alive the traffic forwarding.

4.5.1 HSRP operation

When HSRP is in operation on a LAN, two or more routers or layer-3 switches share a virtual IP address and a virtual MAC address to form a HSRP group. In other words, each HSRP group forms a single virtual router. HSRP elects two routers based on pre-configured priority. The protocol assigns the mastership of the virtual router to one of the HSRP routers participating in the HSRP group. This HSRP router controlling the virtual IP address is known as Active router. Another router participating in the HSRP group is chosen as Standby router. After election procedure, only these two routers transmit HELLO packets while the others remain silent. This behaviour saves network bandwidth.

Then, during a steady situation, Active and Standby routers send periodically HSRP messages. In case of failure of the Active router, the Standby router takes over as the Active router. If the Standby router fails or becomes the Active router, then another router is elected as the Standby router.

In a typical HSRP network configuration, hosts in the subnet set their static default gateway to the virtual IP address of the HSRP group. If the standby router of a HSRP group do not receive three consecutive Hello messages, the active router is considered down. Then, the standby router with highest priority assumes the virtual MAC address and begins the forwarding of the traffic directed to the virtual IP address. The time of this switchover procedure may be adjusted modifying the Hello messages interval, which is possible to modify manually.

On a particular LAN, multiple HSRP groups may coexist and overlap. Each group emulates a single virtual router, namely a single virtual address. In every HSRP group, the Active router is the only one which responds to all ARP requests for the virtual IP address and forwards all packets sent to this IP address. Therefore, none of the standby routers forward any traffic sent to the virtual IP address. Consequently, all the uplinks on these standby routers remain idle since there is no load balancing option in HSRP. In brief, HSRP permits to increase the network availability by cancelling the impact of a failure at the default gateway. However, the protocol itself does not allow for an efficient use of network resources when multiple available paths exist.

That is always the case unless additional configuration steps are taken. On the following paragraphs, such HSRP configuration is detailed and explained.

4.5.2 HSRP load sharing

Load sharing across multiple gateways and multiple uplinks represents increased performance and take advantage of the costly, redundant facilities that ensure high availability during times when both the primary and redundant paths are active.

HSRP allows the use of network resources in more efficient way, but it has to be configured in a specific manner. By default, when HSRP is operating on a LAN segment, all the traffic goes through whichever router is currently active. This means that the second router and its links are generally idle. And that might suppose to pay for a connection that is almost always unused. To avoid any unused link, a specific HSRP configuration allows the sharing of the traffic among available routers.

For example, to share the traffic among two routers is enough creating two separate HSRP groups. When everything is working normally, one of the routers is the active router for one of the groups. And another router is the active one for the other HSRP group. A router that is active for one group is standby for the other, and vice versa. Then, if either of these routers fails, the other takes over and becomes the active router for both groups. The hosts or endpoints are also divided in two groups: the ones using the virtual address of one HSRP group as a default gateway, and the other hosts using the virtual address of the other HSRP group.

This method only affects the uplink, which is the outgoing traffic from the hosts to the routers. In order to load balance the downlink (from the remote network towards the LAN), the routing protocol running on external part of the router should be taken into account. Such configuration increases the complexity of the network administration however. Another negative issue is that splitting customers or hosts on a common LAN among multiple default gateways reduces the flexibility of local network, which is an important characteristic.

It is also possible to achieve load sharing network alternating HSRP groups in the hosts and using two or more VLAN (Virtual Local Area Network) or subnets. [35]

4.5.3 HSRP parameters

Each HSRP group is identified with a single MAC address, as well as a virtual IP address. Such virtual IP address must belong to the range of possible IP addresses of the LAN. It is exactly the same case as it was an IP address of a device physically connected to the local network. Therefore, this virtual IP address must differ from the addresses allocated as interface addresses in all

routers and hosts of the LAN, including virtual IP addresses assigned to other HSRP groups. This one is a key difference with VRRP.

As follows, a list of the most important parameters present of HSRP hello messages:

- Hellotime. 8 bits. Interval between hello messages. Default = 3 seconds.
- Holdtime. 8 bits. Validity of hello messages. Default = 10 seconds.
- Priority. 8 bits. Value used for electing active and standby routers. Higher priority wins. In case of two routers with equal priority, tie break algorithm is applied: the router with the higher IP address wins.
- Authentication Data. 8 bytes.A password in clear text.
- Virtual IP Address. 32 bits. The virtual IP address used by this group.

4.5.4 HSRP details

HSRP has some security issues, even following recommendations and precautions. However, in most cases there is not any problem as HSRP is mostly used on trusted LAN segments.

HSRP has two main security-related problems. The first potential security problem is caused by incorrect router configuration resulting in several routers becoming active or none of them being Active. The second security issue is related with a hostile user willing to capture confidential traffic or to cause Denial of Service. If an untrusted user is capable of configuring a device to take over as the HSRP active router, the security threat is evident. However, the combined use of local multicast address 224.0.0.2 and a Time To Live (TTL) of 1, makes very difficult an HSRP attack without being directly connected to this network segment.

HSRP basic authentication helps in preventing misconfigured routers to become active on a LAN. But the security weakness is clear, as the authentication password is sent through the network as multicast message and without any encryption. Consequently, it is relatively easy for any device in the LAN segment to capture the password.

Another problem related with HSRP authentication is the password disagreement. If the passwords of two routers belonging to the same group do not match, there is no possible way of knowing which one is correct. Then, they both assume that the other is wrong. When such situation happens, both routers might become active, which is a potentially dangerous misbehavior. In brief, HSRP authentication is not a good way of preventing a malicious user from taking over control of the gateway.

There is another way of remarking how insecure HSRP is: checking that all HSRP parameters are travelling through the network in clear, without any encryption. Just capturing a single hello message, it is fairly easy to know the entire HSRP configuration, like IDs, priorities, timers and virtual addresses.

There is definitely not enough protection against malicious users who would like to create a false HSRP device to intercept and disrupt traffic.

4.5.5 Advantages of HSRP

HSRP as First Hop Redundancy protocol would have some good characteristics:

- Easy to configure, the protocol does not affect the routing tables or hosts configuration.
- The traffic increase caused by HSRP is minimal.

4.5.6 Disadvantages of HSRP

On the other side, the use of HSRP would imply many weaknesses:

- Three second recovery time is hardly acceptable for real time traffic, such as voice over IP traffic.
- HSRP is a weak protocol from the security point of view (see Section 4.5.4).
- HSRP is a Cisco proprietary protocol, while in a free patent protocol, further development is feasible.

4.6 Gateway Load Balancing Protocol

GLBP (Gateway Load Balancing Protocol) [33] has a similar purpose as other First Hop Redundancy protocols. GLBP provides failure protection as a main feature, but it additionally permits load sharing of traffic coming from hosts within a common subnet through redundant default gateways.

4.6.1 GLBP operation

Similarly to HSRP, GLBP builds a cluster of routers that cooperate with each other to be seen by the hosts of the LAN as a single virtual router. But there is a key difference between GBLP and the rest of protocols: more than a single router is elected from the group to take the responsibility of forwarding the traffic that hosts send to the virtual router. Therefore, instead of only one router forwarding packets as in HSRP or VRRP, the traffic is distributed among several elected routers.

The main difference between GLBP and the other First Hop Redundancy Protocols, such as HSRP and VRRP, is the load sharing ability. GBLP permits the distribution of traffic among multiple gateways in a simultaneous manner, following any of its specific load splitting rules. Indeed, all routers composing a GBLP group are able to share the total load created by hosts

with destination to the default gateway of a LAN. It is important to highlight that distribution of traffic occurs inside every group and between its members, never across different GBLP groups. With a GBLP properly configured, the efficiency of use of network resources is increased. GBLP improves the overall performance when several uplink paths are available. Instead of a single active router which is the only one forwarding traffic, i. e. the master of a group of routers, GLBP elects an Active Virtual Gateway (AVG). The AVG, chosen by election, assigns a virtual MAC address to each of the other GLBP routers of the group. The AVG is also responsible for answering ARP replies for the virtual IP address. Thus, another function of the AVG is the assignment of a virtual MAC to every host. In other words, the AVG relates every host with a default gateway. The load sharing process occurs through the combination of both assignments and the different possibilities they offer. Those routers receiving this virtual MAC address assignment and participating on the traffic forwarding are called Active Virtual Forwarders (AVF).

4.6.2 GLBP load sharing

As already said, hosts need to solve an IP-ARP resolution to access the MAC address of the default gateway in order to forward traffic to the default gateway through Layer-2 protocols. The default gateway normally replies to the ARP query, allowing the endpoint to know the Layer 2 address of the gateway and forward data traffic there via a link layer transfer. Therefore, if all the redundant gateways in a redundancy group selectively respond to ARP queries in an ordered manner, traffic coming from hosts and going to further networks will be intelligently divided across all redundant gateways.

In case of GLBP, load sharing is provided in addition to resiliency. Resiliency is provided in a similar way as other First Hop Redundancy protocols: if a gateway is failing, a remaining gateway will take over its load in addition to its own. Thus, the local endpoints are not aware of the failing gateway and there is no traffic disruption. Additionally to failover services, load sharing is also provided by GLBP: one of the redundant Layer 3 gateways of a GLBP group, the AVG, handles the assignment algorithm of virtual MAC addresses and responds to ARP requests on behalf of the entire GLBP redundancy group. All intelligence related with load sharing system is managed by the AVG. Then, load balance of traffic is done without affecting the configuration of hosts.

GBLP offers four different algorithms to determine how the assignment of network hosts to each GLBP router is done.

1. None.
   All ARP Replies from the AVG indicate its own Virtual Forwarder (VF) MAC address. Then, all traffic will be directed to the AVG without any load sharing.

2. Weighted.
   GLBP sets a weight on each device which is proportional to the amount of traffic willing to be received by the node. Consequently, some of the routers will receive more traffic than others depending on their weight.

3. Host dependent.
   To decide which Virtual Forwarder MAC address is assigned to each host, the standard MAC address of the end point is used. As the MAC address of each host is unique, all traffic generated by a single user is directed to the same router. It is important to highlight that host dependent load sharing must be used when routers have NAT (Network Address Translation) active.

4. Round robin.
   Round Robin method uses each VF MAC address on a sequentially way when answering ARP request for the virtual IP address.

4.6.3 Advantages of GLBP

GLBP has important strengths, the most important are listed as follows:

- Load sharing ability permits the simultaneous utilization of multiple paths, resulting in a more efficient use of network resources.
- Automatic and personalized load sharing. Traffic may be distributed among available gateways following the rules of the most desirable load-balancing algorithm.
- Simplicity of the access layer design. Instead of additional configuration needed in HSRP or VRRP, GLBP makes possible to achieve more efficient use of network resources without any additional VLANs and subnets in the hosts.

4.6.4 Disadvantages of GLBP

Using GLBP as First Hop Redundancy protocol would have some negative issues compared to other alternatives:

- Cisco proprietary protocol.
- Higher complexity on network management as a result of high number of configurable parameters to take into consideration.

4.7 Common Address Redundancy Protocol

CARP [48] is the Common Address Redundancy Protocol. That protocol is part of the network module of BSD (Berkeley Software Distribution) operative system and its main function is achieving system redundancy on IP networks.

Indeed, CARP groups several physical computers together under a virtual address. Of these systems, just one is the responsible of responding to all packets destined for the group, the other systems being hot spares. In other words, CARP provides a backup if the default gateway fails. In such case, the backup device has the permission to respond instead. Additionally, it allows some degree of configurable load sharing between systems and is able to support both IPv4 and IPv6.

CARP has been developed after VRRP because the RFC 3768 has a possible overlapping with a Cisco patent. Thus, CARP can be considered a secure and free alternative to the Virtual Router Redundancy Protocol and the Hot Standby Router Protocol. To avoid legal conflicts with the previous two protocols, CARP was designed to be somehow different. The inclusion of cryptography is the most remarkable difference. The CARP advertisement is protected by a SHA1 HMAC (Hash-based Message Authentication Code) , providing more advanced security than the other analyzed protocols.

4.7.1 CARP operation

CARP, as many other First Hop Redundancy Protocols, allows the sharing of an IP address among a group of routers in the same network segment. The group of routers sharing the same IP, called redundancy group, has a single IP address assigned which is shared among all the group members. One of the routers composing the group is designated as the master. The rest of members are the backups. The master is the router that actually holds the shared IP address. As the responsible of the shared IP, the master responds to any traffic or any ARP requests directed towards the shared IP. Routers may belong to several redundancy groups at the same time, with fully CARP functionalities each of them.

The master of the redundancy group sends regularly CARP advertisements. With these advertisements, every device connected to the local network and acting as a backup of this redundancy group is aware of the master state. If the backup devices are missing the CARP advertisements during a predefined period of time, one of the backup routers will take the role of master. CARP advertisements, among other data, contain the Virtual Host ID. Such identifier permits to all CARP devices be aware of which is the redundancy group of the advertisement. Thus, devices may get and read the message or just drop it silently depending on the participated groups of the receiving device. So, many different CARP redundancy groups may coexist in the same network segment without disturbing to each other.

CARP also includes a basic authentication method to avoid the spoofing of CARP advertisements from a malicious user. If authentication is used, each redundancy group is configured with a password. Consequently, every CARP packet sent to that CARP group is encrypted, and no device outside this group can understand the message content.

As other First Hop Redundancy protocols, CARP is a multicast protocol.

It uses IP multicast abilities for messaging and protocol operation. The advertisements are sent by the master at configurable intervals, using the IP protocol number 112. If the master fails, the backup devices in the CARP group begin to advertise automatically. The router being able to advertise with higher frequency is the device with the lowest configured advbase and advskew values. This backup router becomes the new master. If the old master comes back up, it becomes backup host by default. However, it is possible to modify such behavior if there is a device deserving to become the master whenever possible.

4.7.2 CARP load sharing

CARP can balance the incoming traffic on a redundancy group in two different ways. These two load balancing mechanisms are called ARP balancing and IP balancing.

ARP balancing has a limited scope. It is only working for clients in the same network segment. Such limited location is produced by the nature of ARP traffic, which never crosses any router or layer-3 switch. Indeed, ARP balancing spreads the load by changing the content of responses of ARP queries sent by the hosts. In detail, when an ARP query is received, CARP protocol uses the source MAC address of the ARP request to assign a router to the client. The ARP request is only answered if the selected router is in master state. If the router is in backup state, the ARP request will be ignored. Such load balancing method is not normally used as ARP load balancing cannot balance traffic that crosses a router to a further destination.

By default, IP balancing is used instead of ARP balancing when CARP load balancing is utilized in IP networks. This mechanism also works for traffic that crosses a router. IP load balancing distributes the incoming traffic to all nodes in the CARP group. More precisely, it uses a multicast MAC address as destination for traffic. Consequently, incoming traffic reaches all CARP nodes. Of course, only one node in the CARP group accepts the traffic, the rest of group just silently drops it. The node accepting the traffic may be different for every packet. It depends on the load balancing algorithm followed by the group. Anyway, the decision about which node will accept the packet depends on the source and destination IP addresses of the packet itself. CARP makes an operation with both addresses and compares the result against the CARP state of the node. This mechanism works in all IP environments compatible with MAC multicast addresses and provides more precise load balancing than ARP balancing.

To perform IP balancing, all CARP devices in the network have to receive the traffic that is destined towards the load balanced IP. Such replication of traffic is already done in networks with hubs, IP balancing thus not affecting. However, the downside of such traffic multiplicity is clear in switched networks: it will imply a higher network load.

As said on manual page of CARP [4], one of the weaknesses of CARP is its inability of assure synchronization between the members of a redundancy group. About ARP load balancing, the text explains: *ARP load balancing can lead to asymmetric routing of incoming and outgoing traffic, and thus combining it with packet filter state table logging is dangerous, because this creates a race condition between balanced routers and the host they are serving. For example, an incoming packet creating state in the first router, being forwarded to its destination, and destination replying faster than the state information is packed and synchronized with the second router. If the reply would be load balanced to second router, it will be dropped due to no state of second router.*

Even though CARP load balancing is able to split the traffic among available resources, it is difficult to achieve an equilibrated load balancing in a multiple machines system. This is because CARP uses an operation based in the source IP address of the packet to determine which device must handle the traffic. In other words, it does not actually depend on the load. Therefore, some of the load balanced machines normally handles higher loads than the others in most cases.

4.7.3 CARP details

Services that require a constant connection, like SSH (Secure Shell) or IRC (Internet Relay Chat) [24] , do not support CARP. In other words, services not permitting any disruption with the server cannot be properly transferred through a switchover in case of failure. However, CARP can help when reducing downtime as short as possible in other type of services. In such case, it is important to take into account that CARP does not synchronize data between applications, meaning that a restoration of session state during a failover has to be done through other ways in parallel to CARP functionalities.

A typical use of CARP is the redundancy of firewalls. If network clients have been configured with the virtual IP address of a redundancy group as the default gateway, all traffic originated by the clients is controlled by firewall acting as a master. If the master firewall fails, the responsibility of the redundancy group is taken by a backup firewall. Thus, firewall services keep working in spite of failure and clients do not notice any service disruption.

4.7.4 Advantages of CARP

CARP would have important strengths, listed as follows:

- CARP is designed to be free patent.
- Offers good load balance features and high security because encrypted advertisements.

[4] Manual page of CARP on OpenBSD [49]

4.7.5 Disadvantages of CARP

Using CARP would imply some negative issues compared to other candidates. They are listed as follows:

- CARP is a single platform protocol, as it is based on BSD systems.
- Synchronisation is not assured in case of failure, applications like SSH would not work.

4.8 Virtual Router Redundancy Protocol

Virtual Router Redundancy Protocol (VRRP) is an IETF (Internet Engineering Task Force) standard, with the version 2 published by the RFC 3768. VRRP is similar to previously detailed HSRP in purpose and in functionality, but not from a legal point of view. Unlike HSRP, which is a Cisco proprietary protocol, VRRP is an open standard, widely used on the telecommunications industry. The version 2 is widely implemented on networking devices from many manufacturers, but it only works with IPv4. Recently, during March 2010, the version 3 of VRRP has been standarized as RFC 5798. This third version of VRRP is fully compatible and it normally works on IPv6 networks, in addition to be compatible with IPv4 networks. Due to the brief period of time since the publication of the official standard, none of the network equipment vendors has implemented yet the version 3. The VRRP description that follows only refers to its version 2, the RFC 3768.

4.8.1 VRRP operation

A group of routers or Layer-3 switches participating in the protection of a single interface on a router is identified with a unique virtual router ID (VRID). All those routers must have interfaces which belong to the same local IP network as the protected one. It is important to remember that no routing protocols are working inside a local IP network, which is the scope of a VRRP session. Indeed, VRRP is normally used as a redundancy and protection protocol when, for some reason, no routing protocol can be used on the network devices.

A single VRRP group is composed by all the routers protecting a concrete IP address: the Virtual IP. Only one router will be the active responsible for that Virtual IP. Such router is chosen among the other candidates because of its VRRP priority. All interfaces participating on a VRRP group have a configured priority value. That priority is an 8 bits integer, thus ranging from 0 to 255. From that value, the protocol is able to calculate which interface should be the responsible for the Virtual IP. The router taking the responsibility of the Virtual IP address is considered the Master of the VRRP group. The rest of devices participating in the same group are considered Backup. As priorities may change or a failure of master can happen, VRRP advertisements are

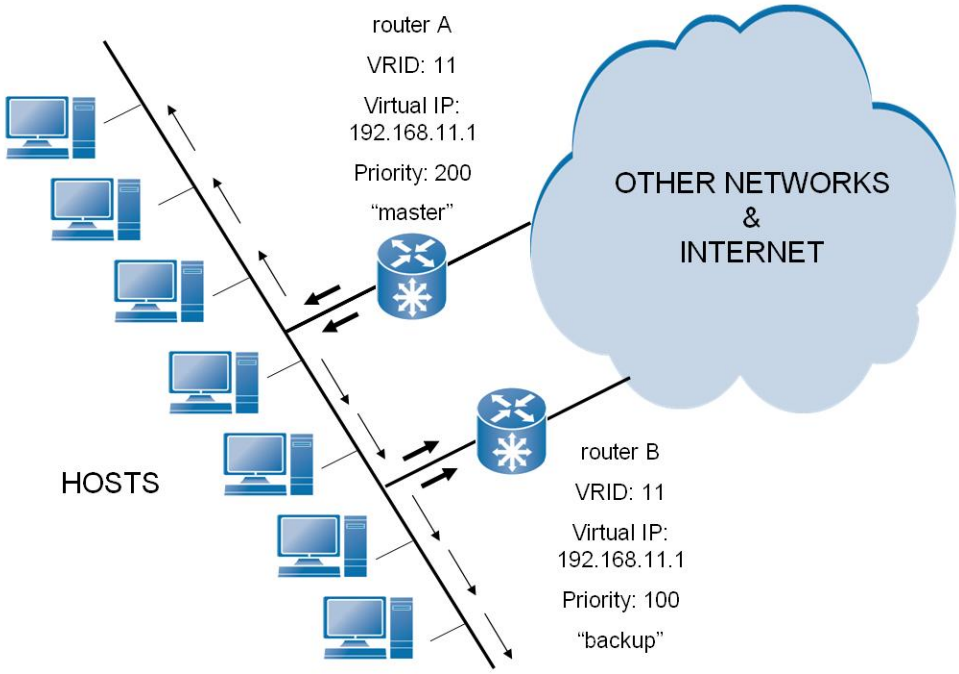

Figure 9: VRRP basic network example

regularly sent from the Master to all the Backup routers to describe the currently active Master. A possible example of VRRP network is shown by Figure 9. The VRRP configuration chooses the router A as master of the VRRP group with VRID 11. For that reason, VRRP advertisements are sent from router A, as the direction of arrows shows. Such advertisements achieve the whole IP network, but they do not cross any router.

It is possible to configure how often the VRRP advertisements are sent. By default, the Master sends a VRRP advertisement every second. But the RFC standard accepts any integer value between 1 and 10 seconds as advertisement interval. With the advertisement, the Master router informs the rest of the VRRP group members about its own current priority, its own advertisement interval and many other parameters of that VRRP session. In a steady state behavior, all backup routers receive VRRP advertisements in time announcing a Master priority that is higher than their own priority. In case that a backup router is missing to receive 3 consecutive advertisements it would declare itself as a Master. Identically, if three consecutive packets advertise a Master priority lower than its own backup priority, the Backup router would declare itself as a new Master for that group.

In Figure 10, the behavior of VRRP in case of failure is detailed:

1. Traffic generated by hosts flows to other networks through the VRRP master router, the router A.

2. When a failure reaches the master router, traffic to further networks is

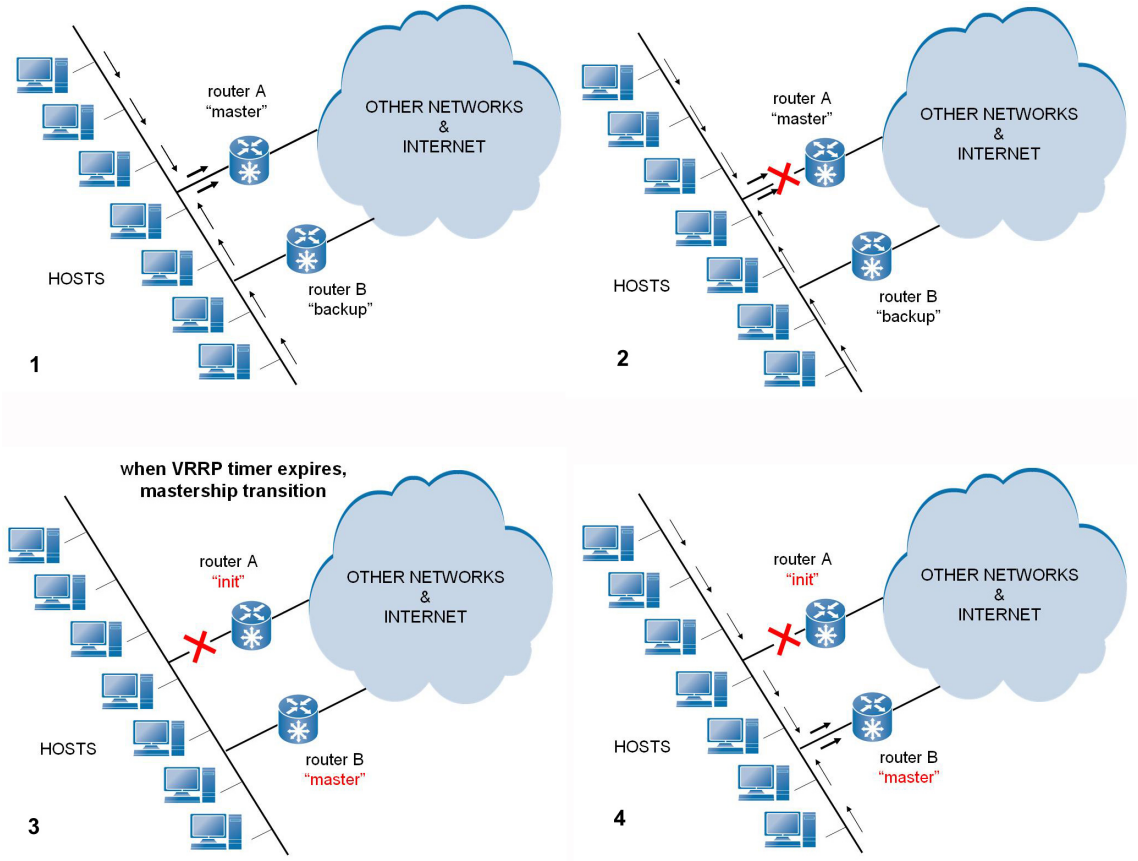

Figure 10: VRRP master transition in case of failure

disrupted.

3. When the backup router is missing 3 VRRP advertisements, it takes the mastership of the VRRP group.

4. Immediately, traffic coming from hosts is forwarded to the current master, the router B.

A special case is the priority 0 of a VRRP advertisement. In such a case, the VRRP advertisement is not specifying the current priority of the Master. Instead, it is indicating that the current master is leaving the master role immediately. Therefore, the Backup router with the highest priority will take the Master responsibility without waiting the mandatory three advertisements intervals.

As backup routers never send any advertisement, it is impossible for them to know the priorities of the other backup routers in the network. To solve that, every VRRP interface implements a skew timer. The skew timer is a short time that every Backup router willing to be Master must wait before switching to Master state. Thus, this timer is only fired when the backup router knows that the current Master is giving up its responsibility. The

value of this timer is inversely proportional to priority. Consequently, the skew timer of the highest priority Backup router is the first one to expire, allowing this router to be the first switching to Master state. As it immediately sends its first advertisement, the rest of Backup routers would receive it and react in consequence. That means that they will cancel their own skew timer and starting again their standard behavior as Backup routers, keeping their Backup state for the time being. This method solves in an elegant way the possibility of some other router with lower priority to become Master unexpectedly.

4.8.2 VRRP load sharing

As one router can participate in several protection groups, specific configuration enables several routers to protect each other splitting the traffic load between them. Such arrangement permits load balance the outgoing traffic of the LAN, but it is not the case for the incoming traffic. Such traffic should be load balanced by a protocol running out of the LAN, thus, out of the scope of VRRP.

For example, in case of a network manager willing to have three routers as gateway of a LAN, it is possible to implement redundancy and load sharing with a simple configuration of VRRP on those routers. The creation of three VRRP groups is a must, with their own VRID and Virtual IP. Each of the routers should be the default Master for one group and Backup for the other two. Finally, dividing somehow all hosts of the network among the three routers allows load sharing of traffic created by the hosts. A typical way to assign the hosts to the routers it is like follows: to configure the default gateway of one third of the routers as the Virtual IP of one of the VRRP groups; another third of the hosts would be configured with another Virtual IP, and the last third would be configured with the last available Virtual IP as default gateway. Then, in a scenario without failure, every router would be used by one third of the hosts: the total load is shared between the redundant routers. A concrete VRRP configuration with load sharing between three router is shown by Figure 11.

4.8.3 VRRP details

VRRP had some legal issue because its similarity with HSRP. Indeed, both protocols share some functionalities and are similar in many aspects. However, there are some differences with VRRP, related to the Virtual IP address. VRRP may use a real IP address of an interface as virtual address. In other words, the VRRP Virtual IP address can be an IP address that is installed in one of the interfaces participating on such VRRP group. However, HSRP can only use a unique virtual IP address which is not used as real IP address on any interface. This difference has a few practical consequences, but it shows a significant change in design principles. Another difference between VRRP

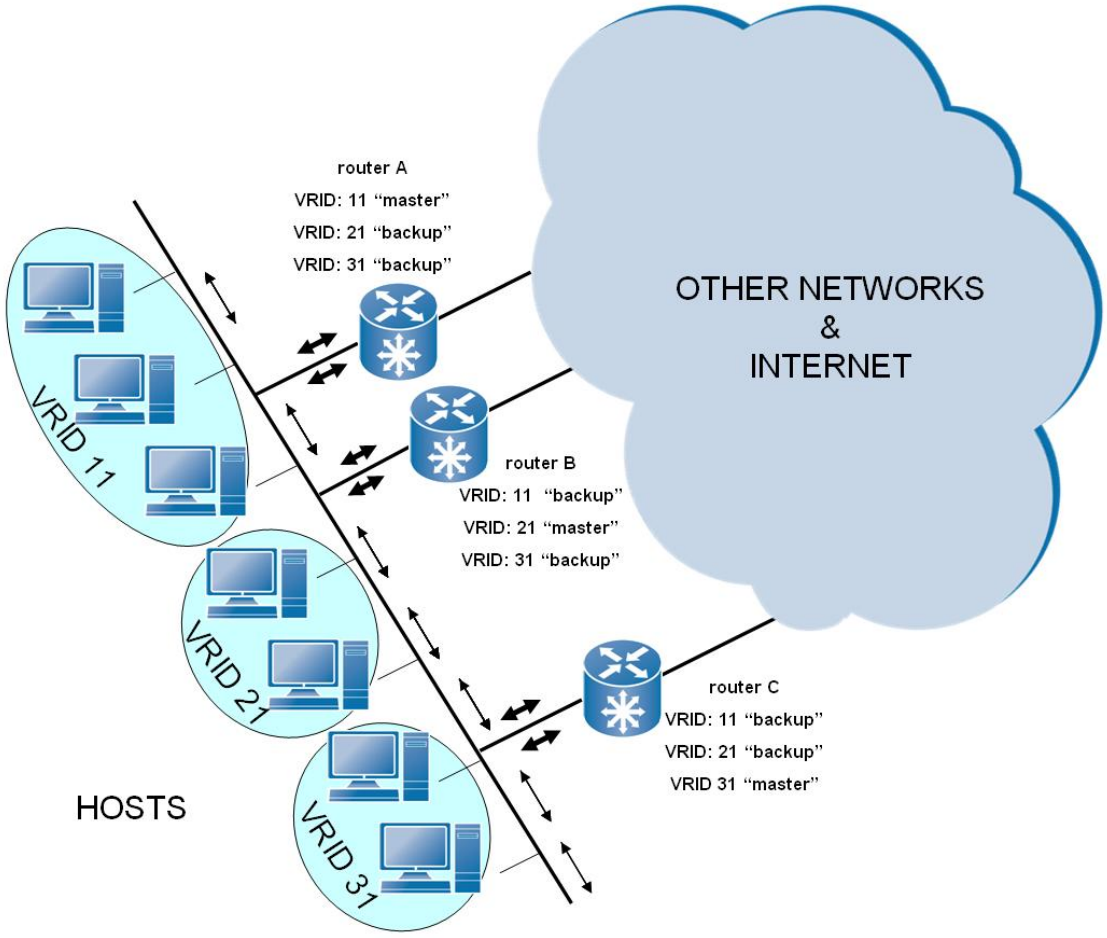

Figure 11: VRRP load sharing with 3 redundant routers

and HSRP is the number and functionality of backup routers. In HSRP, only one router is elected as Backup, called Standby if following HSRP nomenclature. Moreover, this HSRP backup router also sends advertisements in a regular way, contrarily to the VRRP Backup routers, which remain passive from the network activity point of view. This difference with VRRP has to be taken into consideration when calculating the traffic load caused by those protocols. Finally, another interesting difference is that VRRP uses raw IP datagrams, instead of UDP datagrams used by HSRP.

It is important to highlight that each VRRP group has a virtual MAC address, as well as a Virtual IP address. The virtual MAC address is dependent of the Virtual Router Identifier: as described in the RFC 3768, the Virtual MAC address is the source MAC address of VRRP advertisements. All Virtual MAC address follows the structure 00-00-5E-00-01-(VRID). About the Virtual IP address, it has to belong to the network range address used at the LAN. It can be one of the addresses allocated as interface address on some device, although might totally virtual with no interface using it.

4.8.4 VRRP parameters

VRRP advertisement contains many different parameters, which are described as follows:

- Version. 4 bits. It specifies the VRRP version. As commented before, the RFC 3768 specifies the version 2, followed by this thesis.
- Type. 4 bits. Type of VRRP packet. On VRRP version 2, the only type of packet is Advertisement.
- Virtual Router Identifier (VRID). 8 bits. Number between 1 and 255 used as identifier of the VRRP group (Virtual Router using the standarized name found in the RFC).
- Priority. 8 bits. This field is used to elect the master router. Possible priority values are ranging from 1 to 255. Higher value means higher priority. If the device owns the IP address that matches the Virtual IP, the priority must be 255. This value is reserved and compulsory for Virtual IP owners, not being a valid priority if the device is not the owner of the related Virtual IP. As a result, the owner of the Virtual IP is the master router by default. In case of several routers having equal priority, tie break algorithm is applied: the router with the higher IP address wins. Default priority value for a non owner device is 100.
- Count IP address. 8 bits. Specifies the number of IP addresses contained in the VRRP advertisement. Even though the protocol allows multiple IP addresses per VRRP group, only one is used in Tellabs implementation, the same as many other vendors. Thus, the only possible value for this field is 1.
- Authentication Type. 8 bits. This field is maintained to keep backward compatibility with older VRRP versions. Identically as many other vendors, Tellabs equipment will not use any authentication. Then, the fix value for this field is *No Authentication*.
- Advertisement interval. 8 bits. It specifies, in seconds, the time interval between VRRP advertisements. The default and minimum value is 1 second. All members of a VRRP group must have the same advertisement interval configured to avoid conflict. For example, if the advertisement interval declared on a received advertisement does not match the interval locally configured for the same VRID, the packet is discarded. This situation, repetead for three consecutive advertisements, might drive to an incorrect mastership transition.
- Checksum. 16 bits. Used to detect data corruption in the VRRP fields.
- Virtual IP Address. 32 bits. IP address that the master is backing up. There is not default value.

– Authentication Data. 64 bits. As authentication is not used, this field is empty and useless in the VRRP implementation for Tellabs devices.

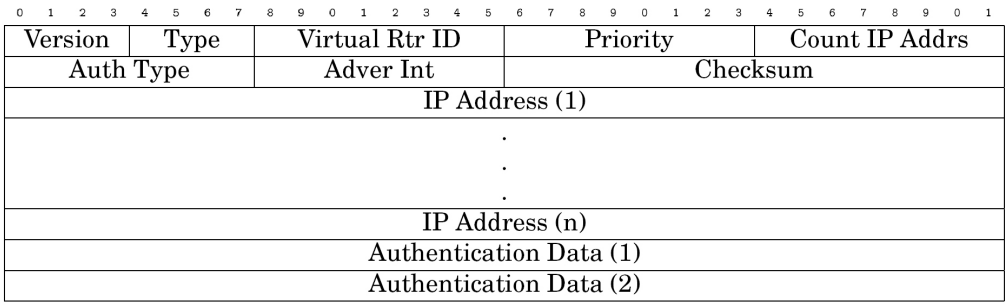

Figure 12: Structure of VRRP advertisement

The VRRP advertisement is the only type of packet existing on the second version of VRRP. The scheme of Figure 12 shows the standarized format of a VRRP advertisement on its version 2. Because VRRP is an IP protocol, this advertisement is always sent as part of an IP packet. Thus, layer-3 protocol of every VRRP advertisement must be IP. The VRRP specific data showed in Figure 12 is placed just after the IP header in the data packet.

4.8.5 Advantages of VRRP

Here are detailed the main strengths of VRRP as First Hop Redundancy protocol:

– Simple configuration, VRRP does not need to modify any routing tables or hosts configuration.

– Free patent protocol, although it exist a complaint because its similarity with HSRP.

4.8.6 Disadvantages of VRRP

Although it is a very adequate option, VRRP has a negative characteristic:

– No security is used, as the offered authentication method is weak.

4.9 Summary

Chapter 4 shows a detailed description of every eligible protocol. The text includes an explanation of the most remarkable parameters. At the end of each protocol analysis, the text shows their specific particularities, divided between strengths and downsides for each protocol. Once all possible options are analyzed, a summarizing table showing the most important advantages

Table 1: Comparison of First Hop Redundancy protocols

Protocol	**Strengths**	**Downsides**
Proxy ARP	Widely used	Reduced scalability
IRDP	-	Not used by other vendors
HSRP	Simple to configure	Cisco propietary
GLBP	Configurable load balance	Cisco propietary
CARP	High security	Single platform: BSD
VRRP	Free patent and simple	No security

and disadvantages of every First Hop Redundancy protocol is shown (Table 1). Such table makes easier a fast comparison and analysis of all the protocols. As previously said in the thesis, the most suitable candidate for the Tellabs implementation is VRRP.

5 Tellabs 8600 Managed Edge System

This chapter introduces the hardware and software family of Tellabs 8600 Managed Edge System where the actual implementation has been done. Details about hardware architecture, software platform used, offered services and device capabilities are shown as follows.

5.1 Main Applications of the System

The Tellabs 8600 Managed Edge System is a next-generation and scalable packet-based platform that is suitable for access networks in mobile transport and converged networks. The Tellabs 8600 has been designed focusing on mobile service providers: it supplies QoS (Quality of Service) and service management features and provides a way to migrate their current radio access networks based on ATM into a manageable IP infrastructure. Additionally, the Tellabs 8600 system supports multiple legacy and recent technologies. It is mainly targeted to be responsible for the transport part of the mobile access networks from the base station sites to the RNC/BSC sites.

The main applications of the Tellabs 8600 Managed Edge System are:

- 2G transport with TDM pseudo wires.
- 3G transport with ATM pseudo wires or over IP/MPLS network.
- 4G - LTE transport over IP. The first all-IP standard on mobile communications is totally supported.
- Managed voice and data leased line.
- Managed LAN interconnection services.
- Managed IP VPNs (Virtual Private Networks).
- Broadband service aggregation.

5.2 Service management

The Tellabs 8600 system is fully supported by the Network Management System (NMS) and the Tellabs 8000 Network Manager, providing altogether an easy and efficient way of end-to-end service management. All the different Tellabs 8600 devices are able to use such network management system as they all share part of the software platform. There are only few differences on the available services depending on the model and the software version running in the device.

5.3 Hardware architecture

As any other similar platform, the high-end models of Tellabs 8600 series have three main network components: backplane, control cards and line cards. Those models, called Tellabs 8660 and Tellabs 8630, have been designed to provide the level of reliability and redundancy commonly associated with high-end telecommunications equipment. They contain a redundant control card and their line cards may be protected by some load balancing mechanism or a link level protection. Briefly, all network parts except the backplane are redundant. Figure 13 shows the hardware architecture of Tellabs 8630 Access Switch, including 2 control cards and 4 line cards connected through a full mesh backplane.

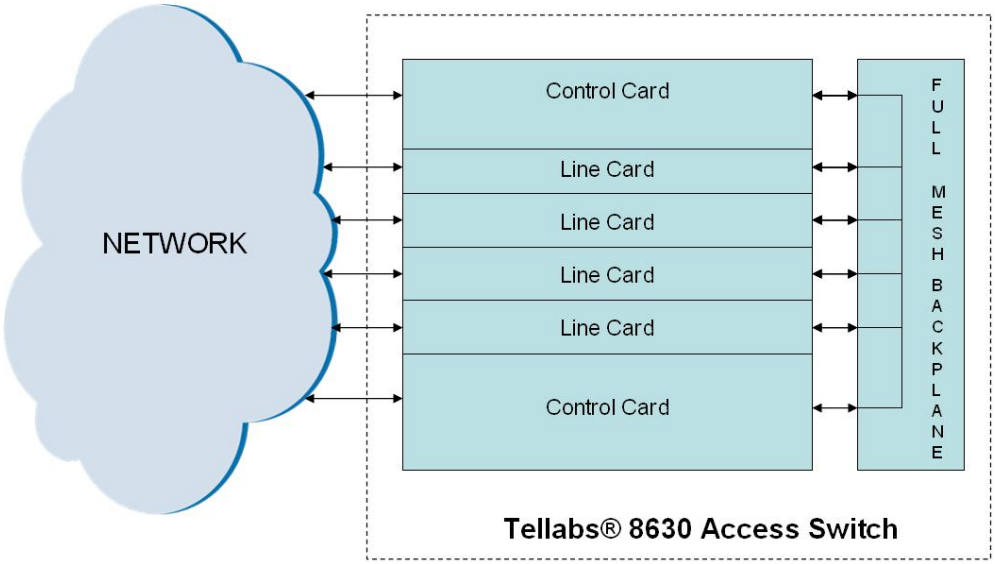

Figure 13: Architecture of Tellabs 8630 Access Switch

5.3.1 Backplane

All cards are connected to the backplane. The backplane is passive and it is responsible for providing the high bitrate level low delay point-to-point links between all possible card positions. Then, the backplane interconnects all card positions between each other in a full mesh topology. Other functions of the backplane are to provide electricity and communication channels to support facilities, such as power units and cooling equipment. The backplane contains hot swap slots for two control cards and for several line cards, twelve slots on 8660 switch and four slots in 8630 switch. Any card can be attached or removed on the fly without disturbing the current process of the device.

5.3.2 Control card

Control cards do not forward any customer traffic. Their duties are mainly related with the execution of protocols. For example, Tellabs 8600 control card keeps control of the different routing protocols active in the node. It also runs the instances of provisioning and signalling protocols like LDP (Label Distribution Protocol) [25] and RSVP-TE (Resource Reservation Protocol - Traffic Engineering) [28]. Other responsibilities are maintenance of the Network Management System connections, key part of the management system. Both active and passive control cards do not share any processor or clock source, being every control card totally independent and self-sufficient. If the active control card fails, the passive control card takes all the responsibilities from the active control card and becomes active. For example, a device provided with redundant control card is the Tellabs 8660 Edge Switch. Its hardware structure is shown in Figure 14.

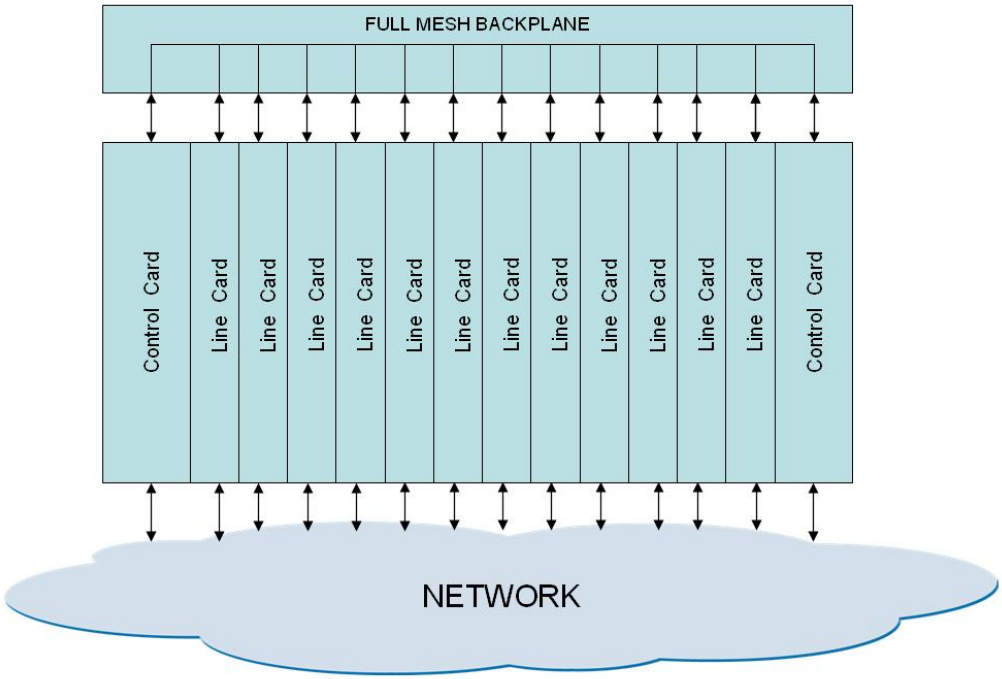

Figure 14: Architecture of Tellabs 8660 Edge Switch

5.3.3 Line card

Line cards are the network components responsible of the traffic forwarding. They also collect information and statistics from traffic, which are sent to the control card for solving many routing and management issues. Line cards do not have any additional functionality, their design being focused on forwarding performance. They are hierarchically dependent on the control card for

routing operations and overall system management. Anyway, line cards keep working without the supervision of the control card as they are autonomous devices with a high degree of independence. Indeed, forwarding and switching is performed by dedicated hardware.

5.4 Software architecture

The software running on every Tellabs 8600 node follows a subsystem structure. Composed by modules, every subsystem has different responsibilities in the software, going from low level routines, common in any generic software, to network and link protocols, specific for high-end network equipment. Every subsystem is very isolated from the others, as they provide different services to the software. In brief, the global software is composed by several subsystems that work independently as separate software pieces, just communicating with each other when needed through some specific internal messaging.

5.4.1 Tellabs 8600 routing subsystem

Originally from a commercial vendor but intensely modified to match Tellabs needs, the routing subsystem is a key part of the control plane software. Tellabs 8600 routing subsystem is a scalable and robust carrier-class switching and routing software. It permits a lot of flexibility to easily add networking capabilities to the new or existing telecommunications products. Indeed, a main characteristic of this routing software is its high reliability and full interoperability with most of the current vendors. The subsystem also includes a High Availability (HA) module which provides enough control plane redundancy to meet the demanding 99.999% or 99.99999% uptime requirements of operator networks. As VRRP cancels the single point of failure in static networks, VRRP module might be considered part of this High Availability solution. It supports IPv4, IPv6, MPLS, Mobile IP, DiffServ (Differentiated services) [22] extensions, DiffServ-TE, Multicast, and Layer 2/Layer 3 based protocols.

5.5 Tellabs 8600 models

A basic description of the different models of Tellabs 8600 product family is provided as follows:

- Tellabs 8605

  Tellabs 8605 is the least capable of the Tellabs 8600 series, this device has fixed interface modules, no change is possible. The total forwarding capacity of the node is 300 Mbps. It is a suitable equipment to be running next to the cell site, at the first step of the wired part of the radio acces network.

– Tellabs 8607

  Tellabs 8607 is a node which is physically pretty similar to the 8605, achieving 500 Mbps of forwarding capability. It contains three slots for reduced sized interface modules, which have lower number of ports and are less powerful than the standard ones.

– Tellabs 8620

  Tellabs 8620 node may contain up to 2 full-sized interface modules, but they do not have the hot swappability characteristic of other interface modules because of its internal installation. The total forwarding capacity of the node is 3.5 Gbps.

– Tellabs 8630

  Tellabs 8630 is the smallest node with standard interface card slots. Four of those slots are for interface cards, allowing a maximum of 8 interface modules in the node. It also contains two slots for control cards, meaning that the control card may be redundant. The total forwarding ability is 14 Gbps.

– Tellabs 8660

  Tellabs 8660: the biggest node of the Tellabs 8600 platform series is able to hold up to 12 interface cards, meaning a total of 24 interface modules. Exactly as the previous device, it contains two slots for control cards, making possible the 1+1 protection, which implies a fully-protected control card. The Tellabs 8660 switch has a maximum forwarding capacity of 42 Gbps.

Figure 15: Family of Tellabs 8600 switches

6 VRRP implementation

In this chapter, the actual implementation of the VRRP protocol on Tellabs 8600 systems is commented. The additional features implemented on the protocol are described with more detail than the standarized features, available in the RFC 3768. After that, all CLI (Command Line Interface) commands permitting to configure and check VRRP parameters are explained. The configuration examples are plotted in the chapter exactly as they would be printed on the screen if using CLI interface.

6.1 Background

As previously commented, VRRP feature is a module inside Tellabs 8600 routing subsystem, containing the protocol data and logic, including all the VRRP specific algorithms. Additionally, some parts of the implemented code are related to internal communications between subsystems, then out of the scope of this thesis.

Moreover, part of implementation must take into account interactions with other protocols and services available on Tellabs software. As part of a commercial product, VRRP feature must work in many different environments, hardware devices and software versions as well. The protocol, as it is RFC standard 3768 must correctly interact with VRRP implementations of other vendors and manufacturers, and any possible misbehavior should be avoided.

6.2 Additional features of VRRP

In addition to the standarised VRRP protocol, advanced features must be created to increase the usefulness of the protocol inside High Availability module of Tellabs software. These additional features are object tracking, accept ping request, fast preemption and configurable preemption delay.

6.2.1 Object tracking

This feature has been implemented simultaneously as VRRP, but cannot be considered as an exclusive part of VRRP module. This is because object tracking can be used by other protocols and other applications. The current version of object tracking feature allows the tracking of three different objects:

- Interface
- IP route
- BFD (Bidirectional Forwarding Detection) [20] session established with a neighbor

If there is not object tracking option, VRRP might be useless for the uplink traffic when external part of master router fails and switchover should happen. On that case, VRRP keeps working as there was no failure inside the IP network where VRRP is running. Thus, VRRP is still pointing the failing device as the responsible of the traffic coming from hosts, even though an external interface is down.

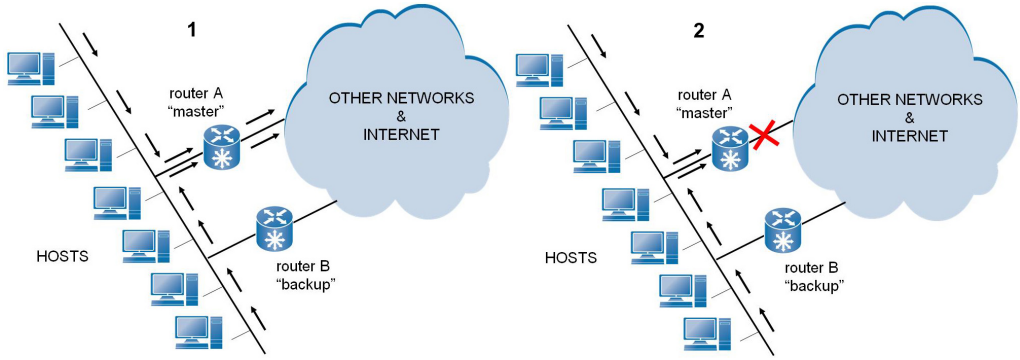

Figure 16: Example of downlink traffic disruption without VRRP object tracking

In the configuration represented by Figure 16, where there is no object tracking active, downlink may be recovered by a protection system in the external network, i.e. a routing protocol. But the uplink traffic, represented by arrows in the figure, is not reacting to the failure, resulting in disruption of communication.

VRRP object tracking offers a complete protection for both uplink and downlink, thus without any traffic disruption. The example shown in Figure 17 explains what would happen in the network if VRRP object tracking is active in case of node failure. Comparing with Figure 16, a different switchover is done if interface tracking is properly configured: no traffic disruption appears in the uplink.

At the beginning of Figure 17, Router A is the responsible of uplink traffic (1). When tracked interface goes down (2), the VRRP router reacts decreasing the priority of Router A. Thus, allowing the Router B to become master of the virtual router (3). As a consequence of this transition, Router B is now the responsible of forwarding the uplink traffic generated from the hosts (4).

Another object that may be tracked is a BFD session of a neighbor. On that case, neighbor tracking through BFD session permits a very fast switchover in case of Master failure, around 20 milliseconds. If the tracked neighbor goes down, BFD session is interrupted and the related VRRP priority is decremented. Immediately after that, new VRRP master is elected. Indeed, the addition of tracking of BFD sessions, combined with the fast preemption explained on 6.2.3, produce a recovery time totally independent from the advertisement interval. As BFD can handle very high packet rates, BFD neigh-

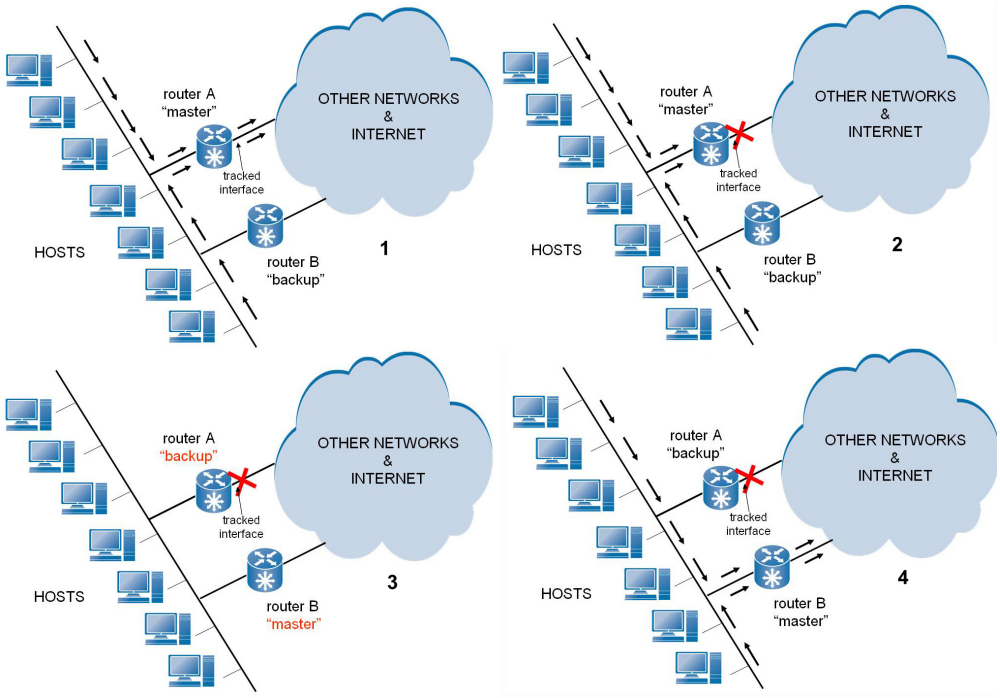

Figure 17: Example of uplink traffic rerouted when object tracking activated

bor tracking allows failover times of dozens of milliseconds to this enhanced VRRP.

6.2.2 Accept ping request

This feature, which can be enabled or disabled, is focused to any master router of a Virtual IP that is not the owner of that virtual address. If the option is switched on, the master of a non-owner Virtual IP takes the responsibility of answering all the ping requests directed to the Virtual IP address. Following the standard protocol (RFC 3768), only if the master is also the owner of the Virtual IP the router would answer a ping request. In other words, if an IP address is accessible because of VRRP, but the master router responsible of this address is not its owner, no reply 'echo ping request' is sent related to the reachable address. Therefore, to increase the utility of ICMP ping request combined with VRRP, this option allows the use of the ping application to check the reachability of any IP address participating on a VRRP network.

6.2.3 Fast preemption

That feature permits, under certain conditions, going through the failover process much faster than the regular standard. If the option is active, certain timers are not fired and the switchover process is done as fast as system capability allows, assuming some risks however. The most critical risk is the prob-

ability of protocol misbehaving due to backup routers doing simultaneously the transition to master state. An IETF draft propose a similar extension of VRRP combined with BFD in order to reduce transition time. [5]

6.2.4 Preemption delay

This option permits to add additional waiting time when a preemption case happens. Assuming that preemption normally occurs when a failing node is available again, the main goal of this waiting time is to allow the full recovery of the recently returned device. With this additional time, the stability of the node is assured, having enough to reestablish all its connectivity. Additional application of the waiting timer may be the creation of hierarchies when several backup nodes are waiting for preempting the current master. If fast preemption option is active, the organized configuration of preemption delays is the best way to avoid simultaneous transitions to master state, as there would not be any other active timer to decide which router is the correct master.

6.2.5 Faults, alarms and log of events

Faults management and logging events are among Tellabs 8600 routing subsystem features. Such subsystem defines a fault as an event notification sent by the software module where the event happened. Those notifications are helpful for management purposes, pretty similar to SNMP (Simple Network Management Protocol) [29] traps. There are many different fault states and alarms for unexpected and critical situations. VRRP implementation takes advantage of them creating a series of specific VRRP events that might occur during the execution of VRRP. Some alarms for VRRP critical situations have been created as well.

Through a CLI command showing the VRRP related events is possible to check if the current state of the node implies any problem to the rest of the network or the node itself. It also permits to show the history of transitions of a node, making easier to check if there is any misbehavior or network misconfiguration. It is also possible to save into a log file all the events that might be interesting to review later on. With the CLI command logging all the events, the review and debugging of what happened internally in the node is trivial. Taking into account all the information contained in the log file, it is often possible to follow and explain the behavior of the node from the VRRP point of view.

6.3 CLI commands

As part of VRRP implementation, new CLI commands are added to the Tellabs 8600 software. They allow the configuration of different parameters of the

[5] IETF draft published on October 16th 2010 [8]

protocol and its testing once is configured. When deleting some configuration, the same command line must be used with the word no at the beginning.

6.3.1 Configuration commands

All configuration commands are accessible from interface mode, no VRRP configuration can be done if the actual interface is not specified before. Multiple VRRP sessions are permitted within an interface, but no Virtual Router ID may be repeated: it must be unique per interface.

`ip vrrp <1-255> A.B.C.D [priority <1-254>]`

This command establishes a VRRP session in the interface. VRID and Virtual IP have to be configured as there is not default value, and must be unique within the interface as well. Priority field is optional: if nothing indicated, default value is 100 for any interface not owning the virtual IP address.

`ip vrrp <1-255> preempt [wait <1-600000>] [fast]`

This command enables or disables the preemption. By default, preemption is active. RFC 3768, requires a waiting time of three advertisement intervals if preemption is based on advertisement that has non-zero priority. But with zero priority advertisement, preemption is done after a delay equal to the shortest skew time. If fast preemption is specified, preemption occurs immediately. Then, no time is waited if zero priority advertisement is received by an interface with fast preemption active. Officially, fast preemption is not recommended if the network has more than two VRRP nodes. Additional waiting time is optionally specified. Such option is highly recommended in case of several routers using fast preemption, as it makes possible to add some hierarchy among the candidates to master.

`ip vrrp <1-255> advertisement-interval <1-10>`

Such command allows the setting of advertisement interval of VRRP session in seconds. By default, VRRP advertisements are sent every second.

`ip vrrp <1-255> accept-data`

The command serves to enable or disable the control data acceptance. This option allows the interface to reply a ping request to the Virtual IP if the VRRP session is on Master state and the interface is not the owner of the Virtual IP. By default, this option is disabled.

6.3.2 Show commands

CLI commands showing VRRP information are briefly described below. Additionally there is an example with the information as it appears on the command line.

show ip vrrp interface

It shows a table with all VRRP sessions running in the node, even on interfaces that are shutdown. As follows, an example of such table showing some basic information of VRRP sessions in the current interface.

```
Router1(cfg-if[ge11/0/0])#show ip vrrp interface
 Interface   VrID   VIP           Pri   State    Master         Own IP         Own   Pre
 ge9/0/0      15    10.147.15.1   100   Init                    N/A            No    Yes
 ge11/0/0      5    10.147.5.1    200   Init                    N/A            No    Yes
 ge11/0/0     25    10.147.25.1   255   Master   self           10.147.25.1    Yes   Yes
 ge11/0/0    100    10.147.100.1  250   Backup   10.147.100.45  10.147.100.50  No    Yes
```

show ip vrrp interface IFNAME

Adding the interface name to the previous command, the command shows detailed information of all VRRP instances running in the specified interface.

```
Router1(cfg-if[ge9/0/0])#show ip vrrp interface ge11/0/0
Interface ge11/0/0, VrID 5
    State is Init
    Virtual IP address is 10.147.5.1, not owner, not accept-data
    Real IP address is N/A
    Virtual MAC address is 0000.5e00.0105
    Configured priority 200, current priority 200
    Track object dampening initial 0 ms, multiplier 1000ms, maximum 1000ms
    Advertisement interval is 1 seconds
    Preemption is enabled, delay 0 ms, fast mode off
    Master down timer is 3.219 seconds, not started
Interface ge11/0/0, VrID 25
    State is Master
    Virtual IP address is 10.147.25.1, not owner, not accept-data
    Real IP address is 10.147.25.2
    Virtual MAC address is 0000.5e00.0119
    Configured priority 195, current priority 195
    Track object dampening initial 0 ms, multiplier 1000ms, maximum 1000ms
    Advertisement interval is 1 seconds
    Preemption is enabled, delay 0 ms, fast mode off
    Master down timer is 3.238 seconds, not started
Interface ge11/0/0, VrID 100
    State is Backup
    Virtual IP address is 10.147.100.1, not owner, not accept-data
    Real IP address is 10.147.100.50
    Virtual MAC address is 0000.5e00.0164
    Configured priority 250, current priority 250
    Track object TRACK1 is up, decrement 10
    Track object dampening initial 0 ms, multiplier 1000 ms, maximum 1000ms
    Advertisement interval is 1 seconds
    Preemption is enabled, delay 0 ms, fast mode off
    Master router 10.147.100.45
    Priority 251
```

```
    Advertisement interval is 1 seconds
    Master down timer is 3.063 seconds, remaining 2.542 seconds
```

The following shows an useful example of this command. If a tracked object goes down, the related priority is modified automatically. In the example, the previous VRRP session changes its priority due to a tracked object, called TRACK1, which is gone down.

```
Interface ge11/0/0, VrID 100
    State is Backup
    Virtual IP address is 10.147.100.1, not owner, not accept-data
    Real IP address is 10.147.100.50
    Virtual MAC address is 0000.5e00.0164
    Configured priority 250, current priority 240
    Track object TRACK1 is down, decrement 10
    Track object dampening initial 0 ms, multiplier 1000ms, maximum 1000ms
```

It is possible to check current state and information about a tracked object with the command `show track TRACKNAME`

```
Router1(cfg-if[ge9/0/0])#show track TRACK1
Track TRACK1
    IP interface ge9/0/0
    State down
    5 changes, last change 00d00h02m ago
    Up delay 0ms, down delay 0ms
    Tracked by:
    VRRP: ge11/0/0 VrID 100 (decrement 10)
```

As follows, an example of master preemption deactivated. Router1 keeps the master state, even though there is another VRRP session in the Router2 with higher priority belonging to the same VRRP group (VRID=100, Virtual IP=10.147.100.1). Because preemption is disabled in the VRRP session of Router2, it does not take the master role despite of having the highest priority of the VRRP group.

```
Router2(cfg-if[ge3/1])#show ip vrrp int ge3/1
Interface ge3/1, VrID 100
    State is Master
    Virtual IP address is 10.147.100.1, not owner, not accept-data
    Real IP address is 10.147.100.2
    Virtual MAC address is 0000.5e00.0164
    Configured priority 240, current priority 230
    Track object TRACK2 is down, decrement 10
    Track object dampening initial 0 ms, multiplier 1000ms, maximum 1000ms
    Advertisement interval is 1 seconds
    Preemption is enabled, delay 0 ms, fast mode off
    Master down timer is 3.102 seconds, not started
Router1(cfg-if[ge11/0/0])#show ip vrrp int ge11/0/0
Interface ge11/0/0, VrID 100
```

```
    State is Backup
    Virtual IP address is 10.147.100.1, not owner, not accept-data
    Real IP address is 10.147.100.3
    Virtual MAC address is 0000.5e00.0164
    Configured priority 245, current priority 245
    Track object TRACK1 is up, decrement 10
    Track object dampening initial 0 ms, multiplier 1000ms, maximum 1000ms
    Advertisement interval is 1 seconds
    Preemption is disabled, delay 0 ms, fast mode off
    Master router 10.147.100.2
    Priority 230
    Advertisement interval is 1 seconds
    Master down timer is 3.043 seconds, remaining 2.901 seconds
```

6.3.3 Track object commands

The following example show how to create and to specify the target of the tracked object as an IP route.

```
Router1(cfg-track[ROUTE1])target ip route 10.147.105.1/24
Router1(cfg-track[ROUTE1]) show track ROUTE1
Track ROUTE1
    IP route 10.147.105.1/24
    State down
    0 changes, last change 00d00h00m ago
    Up delay 0ms, down delay 0ms
    Tracked by:
```

To assign a VRRP session and a decrement value to a tracked object. Multiple tracked objects can be related with the same VRRP session. The value of its priority decreases with every tracked object on down state. In the example case, two tracked object are down. Therefore, the priority of the VRRP session decreases the sum of the two decrement values:25. Such decrease comes from tracked object ROUTE1 (priority decrement of 15) and TRACK1 (priority decrement of 10).

```
Router1(cfg-if[ge11/0/0])ip vrrp 100 track ROUTE1 decrement 15
Router1(cfg-if[ge11/0/0])ip vrrp 100 track TRACK1 decrement 10
Router1(cfg-if[ge11/0/0])show ip vrrp interface ge11/0/0
Interface ge11/0/0, VrID 100
    State is Backup
    Virtual IP address is 10.147.100.1, not owner, not accept-data
    Real IP address is 10.147.100.3
    Virtual MAC address is 0000.5e00.0164
    Configured priority 245, current priority 220
    Track object ROUTE1 is down, decrement 15
```

```
Track object TRACK1 is down, decrement 10
Track object dampening initial 0 ms, multiplier 1000ms, maximum 1000ms
Advertisement interval is 1 seconds
Preemption is enabled, delay 0 ms, fast mode off
Master router 10.147.100.2
Priority 240
Advertisement interval is 1 seconds
Master down timer is 3.141 seconds, remaining 2.862 seconds
```

It is possible to create a fault when a tracked object goes down. To configure such option, the following CLI command should be used when configuring the tracked object.

```
Router1(cfg-track[NEIGHBOR1])#emit-fault
```

ip vrrp <1-255> tracking delay init <0-180000> mul <0-180000> max <0-180000>
This command is used to configure delays when track changes on a VRRP session. Initial delay, multiplier and maximum delay must be indicated. As quick reaction to become Master is desirable when track goes down, it is not generally recommended the use of tracking delays. Too long values on such delays would cause simultaneous mastership transitions on several routers, thus wrong but transient behavior in the VRRP network.

ip vrrp <1-255> track TRACKNAME neighbor A.B.C.D
Such command is necessary to set neighbor monitoring of a VRRP session. The related VRRP session ignores messages from its neighbor unless required track has been up for some time. Therefore, VRRP messages are discarded when track is down, but also when track has been recently up. Such behavior prevents problems during boot-up from multiple nodes thinking they should be masters, quite frequent case if fast preemption option is used. If neighbor track to current master goes down, VRRP session transitions to Master state immediately. To avoid multiple masters on situations where primary router interface flaps, suitable configuration of preemption delays or disabling fast preemption is recommended.

It should be considered that only a single track per neighbor is permitted.

6.4 Summary

The actual implementation of VRRP on Tellabs 8600 systems includes some additional features that improve the standard behavior of the protocol. Among these, the ability of doing a faster preemption, the use of ICMP ping application to check the status of the virtual IP, and the possibility of adding tracking objects to VRRP sessions. To be able to revise VRRP configurations, many CLI commands showing different VRRP parameters are provided. There is also the possibility of checking the behavior of the protocol with a list of VRRP events is logged and internal alarms may be fired.

7 Test and results

Once the VRRP is already implemented on Tellabs 8600 switches, it is time to test in real hardware how it really behaves. In this chapter, different tests and their execution are described. These tests are designed to check the protocol stability and the probability of misbehavior. The last test is able to measure the actual improvement when enhaced features are active.

7.1 Test scenario

To test the VRRP feature, the following network scenario detailed in Figure 18 is proposed.

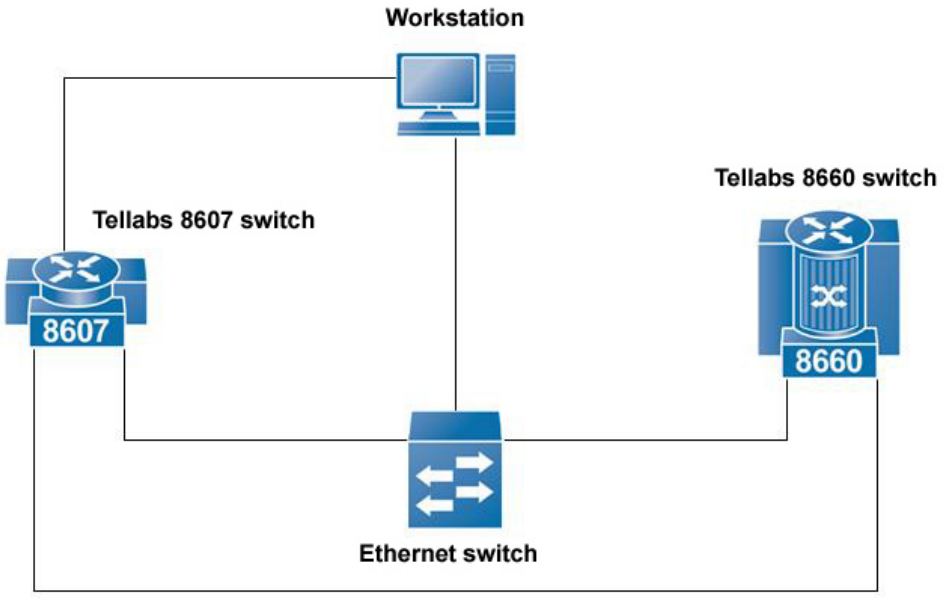

Figure 18: Scheme of network used during testing

For VRRP testing purposes, a network composed by two Tellabs nodes is built: Tellabs 8607 and Tellabs 8660 with the VRRP feature up and running. Both switches are properly equipped with several interface modules, including electrical Fast Ethernet RJ-45 ports and slots for SFP (Small form-factor pluggable transceiver) modules. In the case of the 8660 node, a redundant control card is installed in the box. Cabling includes UTP (Unshielded twisted pair) category 6 for electrical links and single-mode optical fiber for optical links. Physical transmission rate is up to 100 Mbps through electrical links and 1 Gbps through optical fiber. Tellabs SFP transceivers used in this network use a laser with 1310 nm of wavelength.

A packet analyzer running in the Ethernet switch is used in the network. The popular software called Wireshark is sniffing all the traffic coming

through one of the VRRP active links. The sniffing system allows flexibility and the controlled link may be switched to another one if necessary. Thus, it is simple to check and measure the performance of VRRP in the test network.

It is important to remark how difficult is trying to measure the perfomance of the isolated VRRP feature. This is because its complex dependencies on other subsystems and impossibility of disabling other new features in the software.

7.2 Test A

The test A measures the VRRP advertisement interval when the network is in steady state. No state transitions, no changes on topology, no failures are provoked. In core operator networks, where the service is assured the 99.999% of time, VRRP will work on that steady state during the 99.999% of time. Thus, it is important to test the correct behavior of VRRP feature when there are no changes in the network.

7.2.1 From the software point of view

VRRP is implemented as a medium priority task inside the routing software, as any other routing protocol present in the software. Thus, neither critical nor important task nor protocol would be affected by VRRP. Consequently, high priority tasks of the software may delay the theoretical behavior of VRRP, thus being considered as non real time task. That might represent, in some cases, few milliseconds of delay when sending VRRP advertisements. Similarly, the amount of memory used by the VRRP feature is very low compared to other parts of the software. For example, the routing table of every interface may contain several thousands of entries, even hundreds of thousands. On the other hand, there are normally few VRRP sessions configured on a single interface.

7.2.2 From the network point of view

The bandwidth consumption of VRRP protocol is almost negligible. The fixed size of advertisement allows an easy calculation of bandwitdh requirements of the protocol. At IP layer, the VRRP advertisement is 40 bytes. The first half is corresponding to the IP header and the second half means the VRRP data. The maximum frequency that the RFC 3768 specifies the sending the advertisements is 1 per second. Thus, a bandwitdh of 320 bps is needed per VRRP session. As the theoretical maximum of sessions per VRRP router is 255, the maximum bandwitdh required for VRRP advertisements is 255 x 320 bps = 81.6 kbps. However, VRRP is implemented to be independent in every interface. Therefore, 255 is the permitted number of VRRP sessions in every interface. To manage this amount of VRRP sessions is rather complicated.

For this reason, the maximum number of VRRP sessions per interface will be limited to a lower value: for example, 8.

Anyway, the most common links used today on aggregation and access networks are Fast Ethernet and Gigabit Ethernet, up to 10 Gigabit Ethernet in some cases. Such links can absorve the VRRP traffic without any trouble. Following are indicated the percentage of bandwidth consumption for different cases.

7.2.3 Percentage of use in different types of link

Fast Ethernet (FE):
Theoretical maximum bandwidth consumption: 81.6 kbps/IP capacity of 10/100 FE, resulting in 0,8%
Standard consumption, with 8 VRRP sessions: 2.56 kbps/IP capacity of 10/100 FE, resulting in 0,03%

Gigabit Ethernet (GE):
Theoretical maximum bandwidth consumption: 81.6 kbps/IP capacity of GE, resulting in 0,08%
Standard consumption, with 8 VRRP sessions: 2.56 kbps/IP capacity of GE, resulting in 0,003%

10 Gigabit Ethernet:
Theoretical maximum bandwidth consumption: 81.6 kbps/IP capacity of 10GE, resulting in 0,008%
Standard consumption, with 8 VRRP sessions: 2.56 kbps/IP capacity of 10GE, resulting in 0,0003%

7.2.4 Procedure

All VRRP parameters are configured by default. No modification in the network is done during the measurements. The VRRP traffic is captured during several minutes. No unexpected behavior is detected: no events are logged, no faults are created, no mastership transition happens.

7.2.5 Results

VRRP advertisements are sent from the Master in regular intervals with a very small variation. The variability of the advertisement interval is a delay around 4 ms in almost all the cases. This delay is a consequence of VRRP protocol being part of a non-real time task of Tellabs 8600 routing subsystem. For example, with a defined interval of 1 second, the measured interval is, in most cases, between 1,001 and 1,004 seconds. Then, the example shows a tolerance of 0.4% in the advertisement interval. Exceptionally, few advertisement intervals are out of range, up to 15 milliseconds longer than expected. Such a long

delay when sending VRRP advertisements is very unlikely to occur, however. In addition to be uncommon, it would only affect the correct behavior of the protocol in very few cases: when a VRRP backup node is already missing two packets and the third one is sent by the VRRP master with such a long delay. In that case, the protocol behavior is dependent of the backup node priorities and the delay itself. Basically, if the delay of third advertisement is longer than a skew timer, a mastership transition will occur when the shortest skew timer among the backup router expires.

Another important consideration that needs to be done is the probability of link failure. Depending on the importance of the network, many other protection systems may be implemented. In that case, it is very unlikely that a failure achieves the IP layer, thus affecting the VRRP protocol. For example, the fast, expensive and highly deployed protection mechanism Multiplex Section Protection, described in the ITU-T (Telecommunication Standardization Sector) recommendation G.783, is too fast to be achieved at IP layer. Such mechanism permits a recovery on SONET (Synchronous optical networking) systems in less than 50 ms, but the prohibitive cost of this protection keeps it in the core network. Considering the case of two advertisements lost and the third one being delayed, it is easy to figure out how improbable to happen is a failing state that will take two advertisement intervals to recover. Normally, an important network failure will provoke the loose of several consecutive VRRP advertisements at least, not just two.

Following there is a calculation of probability for incorrect VRRP transition due to long delay. Even though this incorrect transition is very unlikely to happen, probability is not zero and may be specified. A remarkable assumption to be done is that the probability distribution of priorities is assumed to be the uniform distribution. In other words, all priorities have identical probabilities to appear. In spite of being a bit unrealistic, it makes the calculations easier and it does not assume any priority policy, parameter that is totally dependent of the network operator.

7.2.6 Probability of incorrect transition

The shortest skew timer ocurrs when the Backup router has a priority of 254. Taking into consideration the formula published in the RFC 3768, the theoretical value of skew timer is 7.81 ms. However, when using a real device, the mastership transition delay measured with backup priority of 254 is rather long: average of 17 ms.

The theoretical value of skew timer for a backup router with priority 250 is 23.44 ms. On a real device, the mastership transition delay measured with backup priority of 250 is 28 ms. In that case, theoretical and measured value are much closer.

In order to calculate the probability of incorrect transition:
P of using backup with priority = 254 is **1/254**
Measured P of advertisement delay longer than 17 ms = **1/25**
Measured P of advertisement delay longer than 28 ms = **1/200**
P of two consecutive VRRP advertisements lost but receiving the third = **P***

P of transition due to advertisement delay when backup priority = 254:

$$P_{254} = \text{P* x 1/25 x 1/254} = \textbf{1/6350 x P*}$$

P of transition due to advertisement delay when backup priority = 250:

$$P_{250} = \text{P* x 1/200 x 1/254} = \textbf{1/50800 x P*}$$

The total probability of having an incorrect transition is the sum of all one-priority probabilities

$$P_{total} = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + ... + P_{252} + P_{253} + P_{254}$$

7.2.7 Analysis of results

The measured percentage of advertisement intervals with longer delay than 17 ms is 4%. It is important to consider a possible value of **P** = 0.0001 or similar and the fact that P254 is much higher than P250. Then, when calculating the total probability of misbehavior, it is clearly evident that priorities lower than 250 affects much less the total priority. A priority policy using only priorities lower than 250 would drastically reduce the probability of incorrect transition. Indeed, the probability of wrong mastership transition when a priority is lower than 250 can be considered almost negligible, even in highly available operator networks

As conclusion, in a reliable network as mobile operator networks, this short and unfrequent delay on VRRP advertisements will not affect at all the correct behavior of network traffic.

7.3 Test B

The test B measures the duration of a mastership transition in case that the current master releases its privilege sending a VRRP advertisement with a special value for priority field: 0. Theoretical value for this duration is exactly the skew timer. As the skew timer depends on priority, the duration of such transition is directly related to the value of backup priority. Consequently, this test measures the transition time when the master is willing to change its status. The delay between the advertisement with zero priority, sent by the old master, and the first VRRP advertisement sent by the new master. And it compares the transition time with the priority values used by the backup routers that become masters.

7.3.1 Procedure

Priority values have been chosen to represent typical values that will be probably used on real networks, even though that will depend on network operator policies. Initially with steps of 5 and then using 10, the values used for testing are: 254, 250, 245, 240, 230, 220, 210, 200, 190, 180, 170, 160, 150 and finally 100, the default backup priority. For every priority, 9 mastership transitions are provoked, resulting on 9 measured transition times per priority. The value showed in Table 2 is the median of the samples. The median is chosen instead of using the average to avoid the influence of some non coherent values. These particular values are totally out of range, with some cases where the measured delay is twice the rest of the samples. These unexpected samples are probably motivated by parts of hardware which are still prototypes. Moreover, part of the software is currently in development and needs additional debugging the completely functional. At this point, it is important to remember that VRRP feature is a non real time task inside Tellabs software. Thus, unexpected delays are likely to happen in a regular basis when sending VRRP advertisements. In this test, about 5% of the samples have been considered out of range, tipically about 100 milliseconds above the median.

7.3.2 Results

Table 2: Measurements of Master transition due to priority 0 in different backup priority cases

Backup priority	**Measured transition (ms)**	**Theoretical transition (ms)**
254	16	7.8125
250	28	23.4375
245	47	42.96875
240	70	62.5
230	106	101.5625
220	145	140.625
210	185	179.6875
200	225	218.75
190	264	257.8125
180	302	296.875
170	340	335.9375
160	380	375
150	420	414.0625
100	616	609.375

In Table 2, the results of measuring skew timers on real devices. The table compares measured values and theoretical values for different backup priorities.

7.3.3 Analysis of results

Table 2 contains a comparison between theoretical values of the skew timer and the measured ones on different priority cases. These results show that almost all mastership transitions are done matching the theoretical time, thus may be used without any trouble. There are few exceptions however. When the backup priority becomes very close to owner priority, the difference between theoretical value and measured value rapidly increases. This anomaly is only remarkable when the backup priority is above 250, therefore only few values are affected (from 251 to 254). The usage of such priority values could drive to unexpected protocol behavior in some cases. This is because the very short difference between two transitions inside the interval. For example, when two backup routers with priorities 254 and 253 are waiting for becoming master. As known because of measurements, both transitions have too similar duration. If the fluctuation of advertisement intervals is taken into account, it is perfectly possible that the backup router with lower priority (253) will become master.

Then, in order to minimize the risk of misbehavior, a good policy when configuring priorities on backup routers is avoiding the use of values between 251 and 254. Anyway, the probability of problems caused by these results is low. Additionally, it would be a good practise to avoid differences smaller than 2 between priorities inside the same VRRP group.

7.4 Test C

The analysis of the duration of traffic disruption with different VRRP configurations is the objective of the test C. The parameter being measured is the delay between the master failure and the gratuitous ARP sent by the new master. As previously said, when a new node becomes master, a gratuitous ARP message containing the real MAC address and the Virtual IP is sent from the new master. The purpose of the this message is to force switches to learn the real MAC address of the new master. Immediately after listenning this message, all the nodes start considering the new master node as the device responsible for the Virtual IP and Virtual MAC.

7.4.1 Procedure

A priority decrease of Master is done to provoke a change on the mastership. Two different ways of priority decreasing are used in this test. Some mastership transitions are generated by direct priority decrement through CLI

command. Other VRRP transitions are done by shutting down a tracked interface, which provokes a priority decrement of the VRRP session tracking such interface. Both procedures are used 50% of the times, 20 times in total. It is also important to take into account that 50% of the cases have the fast preemption option switched on, the rest being without any fast option activated. In all measurements, waiting timer is set to zero. Thus, this test analyzes 4 types of cases: direct decrement with fast option, direct decrement without fast option, tracked object decrement with fast option and track decrement without fast option.

In some extra additional cases, the mastership change is provoked by increasing the backup priority. When this priority becomes higher than the master priority, a mastership transition happens. These additional cases show that the mastership transition is immediate when fast option is switched on, but it is not possible to analyze accurately the delay because there is no packet sent when the priority change actually happens. The only packets sent to the network are the VRRP advertisement from the new master and the gratuitous ARP announcing the MAC address of the new master. As additional information, it may be interesting to measure the delay between both packets. Finally, the measured delay between these two packets is 4 milliseconds, being this value very stable during all the test cases.

7.4.2 Behavior of VRRP protocol with the FAST option active

The option of fast preemption permits the mastership transition without waiting the expiration of master dead timer. Even though it is not recommended to use it under certain circumstances, this option allows failover times with an average of few hundreds of milliseconds. In case of receiving a VRRP advertisement with lower priority than its own priority, the node switches to master state immediately, without waiting the expiration of any timer. Therefore, the new master starts sending VRRP advertisements with new priority just after detecting the priority change. On a similar way, if the priority of the backup node becomes higher than the priority of master, the transition is done without any waiting time. The new master starts its duties immediately.

The previous examples assume that there is not waiting preemption timer configured. If configured, the only difference on the explained cases would be the delay of the mastership transition: it would not be immediate as in the previous examples. Instead, it would have a delay corresponding to the value of preemption waiting timer.

7.4.3 Behavior of VRRP protocol with the FAST option inactive

On that case, the behavior of the implemented protocol matches the standarized VRRP v2 (RFC 3768). The backup router will always wait until the master dead timer expires. In other words, the failover time will be around three times the advertisement interval. The only exception would be when

a zero priority advertisement is sent from the master router. Such a case implies a failover time equivalent to the skew timer.

7.4.4 Transition time

When priority is modified in the current master, possibly through CLI command or because of tracked object, a new advertisement is immediately sent by this node with the new priority value. If fast option is active and priority is higher on a backup router, this router switches to master state without delay. It sends, therefore, a new VRRP advertisement just after receiving the last priority value from the already old master. The delay between the VRRP advertisement announcing the modification of master priority and the first VRRP advertisement from the new master is measured and tagged in Figure 19 as **transition time**.

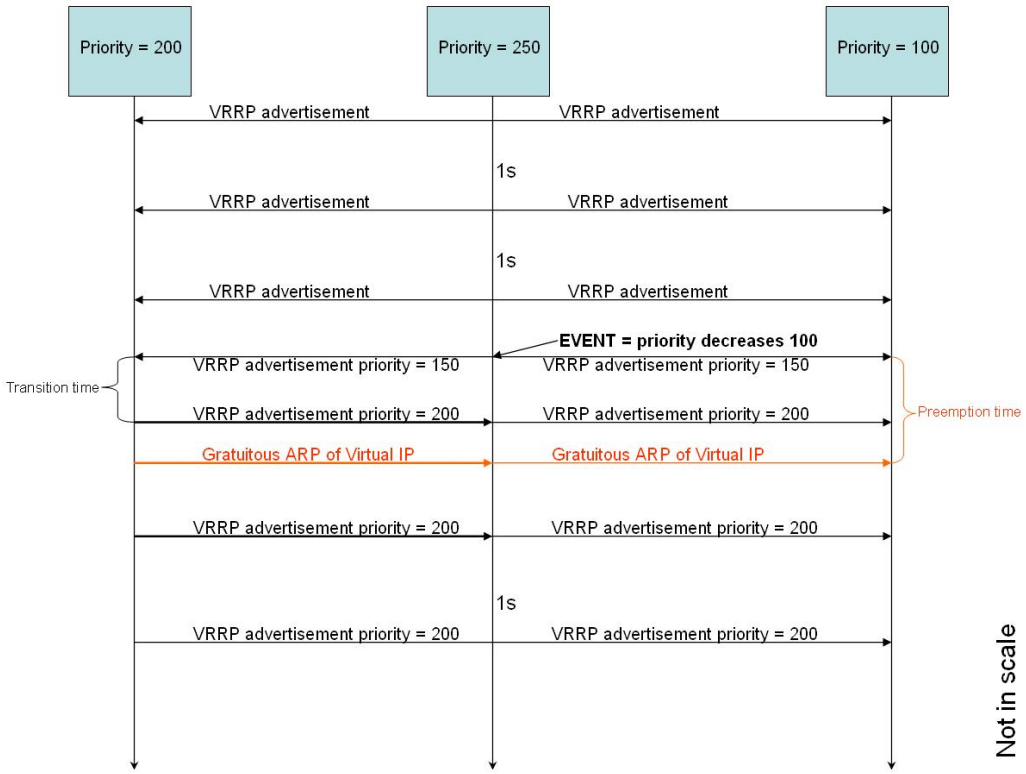

Figure 19: Communication flow between participants on VRRP group with Fast preemption

7.4.5 Preemption time

The switchover of the traffic will happen when all the network hosts are aware about the change of the mastership. This only will happen when the ARP entry containing the new MAC address as the responsible for the Virtual IP

is installed. For that reason, the delay between the VRRP advertisement announcing the modification of master priority and the gratuitous ARP sent by the new master is measured and considered as the traffic disruption time. In Figure 19 is tagged as **preemption time**.

7.4.6 Results

When Fast option is active, the average results are as follows:

– 2-3 milliseconds of measured transition time.

– 11-14 milliseconds of measured preemption time.

When Fast option is not active and the preemption waiting timer is zero, the behavior of the VRRP feature matches the behavior described in the RFC 3678. Therefore, the average of transition times is two and a half advertisement intervals, which is equivalent to 2,5 seconds at least. In Figure 20 is possible to check the communication flow of the standard VRRP, with rather long failover time. That period is about two hundred times longer than the switchover when fast preemption is active.

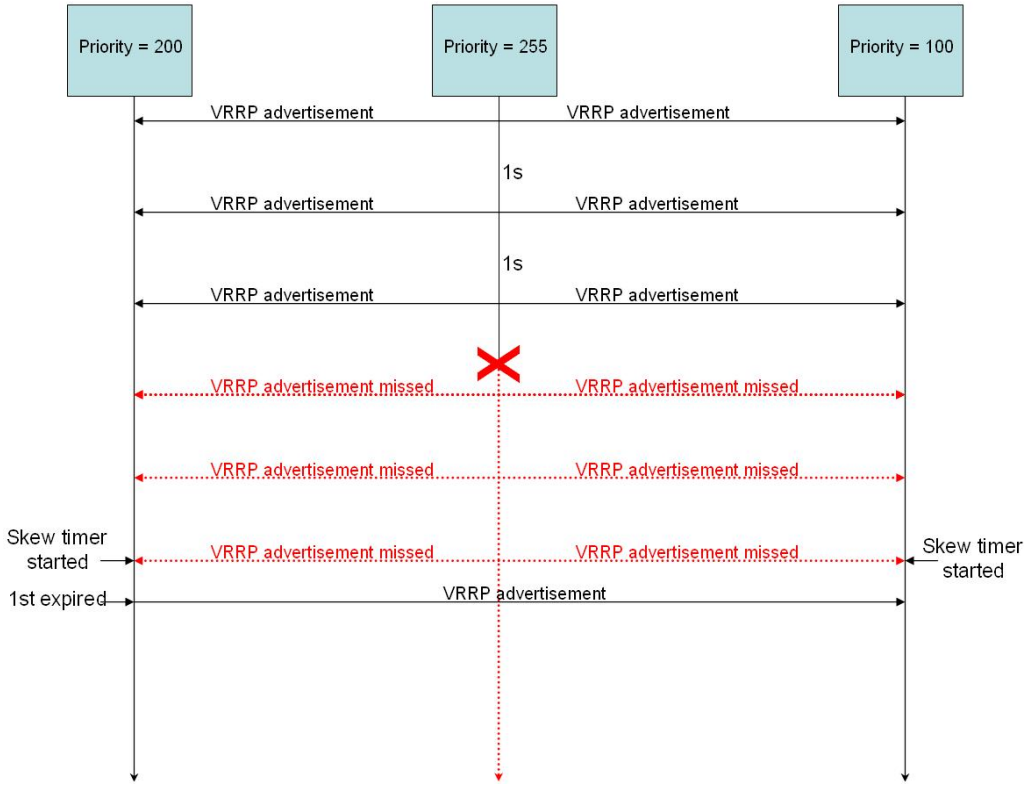

Figure 20: Communication flow between participants on VRRP group without Fast preemption

7.4.7 Analysis of results

The addition of the fast preemption option permits a traffic switching much faster than standard VRRP. Then, this enhanced VRRP is advisable to be implemented on any operator network with high availability requirements. Additional configuration needs to be taken if several backup devices in the same Virtual Router are configured with fast option active. Even though is rather uncommon on commercially profitable networks, such situation may happen and needs to be treated more specifically. Indeed, some hierarchy must be implemented among all the devices participating in the same virtual router through different preemption delays. At some point, a global policy related to preemption delays needs to be defined by the operator.

7.5 Test results

Globally, VRRP running in the test network make possible a fast and proper failover, as the network absorbes any component failure without impact to overall service. The additional features implemented improve the velocity of the failover process in a such a way that the performance of this VRRP is comparable to other standards without the downside of additional complexity of configuration. Moreover, it keeps full compatibility with the standarised VRRP as it accomplishes all its requirements.

Briefly, standard VRRP is not comparable to a routing protocol with very short convergence time. However, this enhanced VRRP is an optimized protection for failing routers at networks edge, where static routes are used and no dynamic routing is available.

8 Conclusions

8.1 Strengths of VRRP feature

The implementation of VRRP feature on Tellabs 8600 software takes into consideration the Radio Acces Network where this kind of node is found. Then, considering the aggregation network scope, the enhanced VRRP supplies some additional services and compatibilities that are potentially useful to any mobile network operator.

- Integrated Layer 2 (Ethernet) and Layer 3 (IP) forwarding
- Works with global routing and Virtual Router Forwarding (VRF)
- Permits multiple nodes to use same virtual IP address
- Many encapsulation possibilities for VRRP protected traffic. VRRP feature supports lots of traffic types like Ethernet, IPv4, VLAN, Pseudowire, Link Agreggation Group.
- Redundant forwarding service for clients not running routing protocols
- BFD support for fast fault detection, permitting switching times around 20 milliseconds
- VRRP feature may work in Integrated Routing and Bridging (IRB) interfaces.

With the additional components of VRRP feature, like fast preemption or BFD tracking of neighbors, the recovery time is acceptable by operator requirements. Indeed, Tellabs implementation of VRRP should assure a failure recovery in few dozens of milliseconds, much faster than VRRP standard protocol, which needs few seconds to recover. As protection at the network edge should have a good trade-off between scalability, cost and disruption duration, the use of this enhanced VRRP in access and aggregation networks is highly recommended: it perfectly matches the requirements of mobile operators.

8.2 VRRP downsides

Although its adequate characteristics as First Hop Redundancy protocol, VRRP version 2 has some features which should be improved in the future. Indeed, the issues related to compatibility with IPv6 and recovery times shorter than one second are recently solved by VRRP version 3 (RFC 5798). However, many improvements are still waiting to be implemented in standard VRRP. For that reason, the IETF Working Group responsible for VRRP is still very active in order to enhance the protocol. During 2010, many documents related to VRRP have been published by IETF:

- VRRP version 3 (RFC 5798), which is a Proposed Standard for VRRP (March 2010). [6]
- Active Internet Draft with definitions of Managed Objects for VRRP version 3 (July 2010). [7]
- Active Internet Draft with extensions relating BFD and VRRP to achieve fast transitions (October 2010). [8]
- Active Internet Draft with extensions to use VRRP with graceful restart (December 2010). [9]

VRRP is still a weak protocol from security point of view, as no encryption or signature is used. Some documents discuss about the advantages and downsides of using passwords in VRRP routers. But lot of problems in case of password misconfiguration seem to dissuade from using any kind of password or private key. Anyway, the TTL value of VRRP advertisements ensures the protection against attacks from external networks.

8.3 VRRP future

New trends on redundancy on mobile network architectures seem to be the use of TDM or ATM tunnelled by redundant pseudowires on multi-chassis APS (Automatic Protection Switching) devices. But the usability and usefulness of VRRP is beyond doubt. Indeed, VRRP is more relevant with the most recent technologies: HSDPA and LTE are both using native IP traffic. Benefits from VRRP implementation are also used by 3G nodes not implementing HSDPA. They are commonly using ATM, which is tunnelled with ATM pseudowires. GSM base stations are also getting the benefits of VRRP. Although they are often using TDM, their flows are normally SAToP (Structure-Agnostic TDM over Packet) [15] tunnelled using the TDMoIP (TDM over IP) standard. In the end, the whole mobile operator network receives the benefit of using the virtual redundancy provided by this enhanced VRRP.

References

[1] O'Reilly, Kevin Dooley, Ian J. Brown, *Cisco IOS Cookbook*, 2006.

[2] Ayikudy Srikanth, Adnan Adam Onart, *VRRP: Increasing Reliability and Failover with the Virtual Router Redundancy Protocol* Pearson Education, 2002.

[3] Scott Andes, Daniel Castro, *Opportunities and innovations in the mobile broadband economy* September 2010
http://www.itif.org/publications/opportunities-and-innovations-mobile-broadband-economy

[4] Ville Hallivuori, *A concept for protecting stateful routing protocols* Master Thesis, Helsinki University of Technology, 2001

[5] Hinden, R., *Virtual Router Redundancy Protocol (VRRP)* Internet Engineering Task Force, RFC 3768, April 2004
http://www.ietf.org/rfc/rfc3768.txt

[6] Nadas, S., *Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6* Internet Engineering Task Force, RFC 5798, March 2010
http://www.ietf.org/rfc/rfc5798.txt

[7] Kalyan Tata, *Definitions of Managed Objects for VRRPv3* Internet Engineering Task Force, Internet-Draft, July 2010
http://datatracker.ietf.org/doc/draft-ietf-vrrp-unified-mib/

[8] H. Zhai, *Extensions to VRRP for Fast Transition of Failed Master* Internet Engineering Task Force, Internet-Draft, October 2010
http://tools.ietf.org/html/draft-zhai-vrrp-extension-ft-fm-00

[9] Q. Zhang, D. Zhang, *Extending the Virtual Router Redundancy Protocol for Graceful Restart* Internet Engineering Task Force, Internet-Draft, December 2010
http://datatracker.ietf.org/doc/draft-zhang-vrrp-gr/

[10] Deering S., *ICMP Router Discovery Messages (IRDP)* Internet Engineering Task Force, RFC 1256, September 1991
http://www.ietf.org/rfc/rfc1256.txt

[11] Thaler D., C. Hopps, *Multipath Issues in Unicast and Multicast (ECMP)* Internet Engineering Task Force, RFC 2991, November 2000
http://www.ietf.org/rfc/rfc2991.txt

[12] Li, T., Cole, B., Morton, P., D. Li, *Cisco Hot Standby Router Protocol (HSRP)* Internet Engineering Task Force, RFC 2281, March 1998
http://www.ietf.org/rfc/rfc2281.txt

[13] Smoot Carl-Mitchell, John S. Quarterman, *Using ARP to Implement Transparent Subnet Gateways* Internet Engineering Task Force, RFC 1027, October 1987
http://www.ietf.org/rfc/rfc1027.txt

[14] S. Bryant, P. Pate, *Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture* Internet Engineering Task Force, RFC 3985, March 2005
http://www.ietf.org/rfc/rfc3985.txt

[15] A. Vainshtein, YJ. Stein, *Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)* Internet Engineering Task Force, RFC 4553, June 2006
http://www.ietf.org/rfc/rfc4553.txt

[16] E. Rosen, A. Viswanathan, R. Callon, *Multiprotocol Label Switching Architecture (MPLS)* Internet Engineering Task Force, RFC 3031, January 2001
http://www.ietf.org/rfc/rfc3031.txt

[17] Postel, J. *Internet Protocol (IP)* Internet Engineering Task Force, RFC 791, September 1981
http://www.ietf.org/rfc/rfc791.txt

[18] Postel, J. *Internet Control Message Protocol (ICMP)* Internet Engineering Task Force, RFC 792, September 1981
http://www.ietf.org/rfc/rfc792.txt

[19] Plummer, D. *Address Resolution Protocol (ARP)* Internet Engineering Task Force, RFC 826, November 1982
http://www.ietf.org/rfc/rfc826.txt

[20] D. Katz, D. Ward, *Bidirectional Forwarding Detection (BFD)* Internet Engineering Task Force, RFC 5880, June 2010
http://www.ietf.org/rfc/rfc5880.txt

[21] Y. Rekhter, T. Li, S. Hares, *Border Gateway Protocol (BGP)* Internet Engineering Task Force, RFC 4271, January 2006
http://www.ietf.org/rfc/rfc4271.txt

[22] K. Nichols, S. Blake, F. Baker, D. Black, *Differentiated Services* Internet Engineering Task Force, RFC 2474, December 1998
http://www.ietf.org/rfc/rfc2474.txt

[23] H. Krawczyk, M. Bellare, R. Canetti, *Keyed-Hashing for Message Authentication* Internet Engineering Task Force, RFC 2104, February 1997
http://www.ietf.org/rfc/rfc2104.txt

[24] H. C. Kalt, *Internet Relay Chat (IRC)* Internet Engineering Task Force, RFC 2810, April 2000
http://www.ietf.org/rfc/rfc2810.txt

[25] L. Andersson, I. Minei, B. Thomas, *Label Distribution Protocol (LDP)* Internet Engineering Task Force, RFC 5036, October 2007
http://www.ietf.org/rfc/rfc5036.txt

[26] J. Moy, *Open Shortest Path First (OSPF)* Internet Engineering Task Force, RFC 2328, April 1998
http://www.ietf.org/rfc/rfc2328.txt

[27] G. Malkin, *Routing Information Protocol (RIP)* Internet Engineering Task Force, RFC 2453, November 1998
http://www.ietf.org/rfc/rfc2453.txt

[28] A. Farrel, A. Ayyangar, JP. Vasseur, *Resource Reservation Protocol-Traffic Engineering (RSVP-TE)* Internet Engineering Task Force, RFC 5151, February 2008
http://www.ietf.org/rfc/rfc5151.txt

[29] D. Harrington, R. Presuhn, B. Wijnen, *Simple Network Management Protocol (SNMP)* Internet Engineering Task Force, RFC 3411, December

2002
http://www.ietf.org/rfc/rfc3411.txt

[30] P. Pan, G. Swallow, A. Atlas, *Fast Reroute Extensions to RSVP-TE for LSP Tunnels (MPLS FRR)* Internet Engineering Task Force, RFC 4090, May 2005
http://www.ietf.org/rfc/rfc4090.txt

[31] ITU-T documentation *ITU-T Recommendation I.122: Framework for Frame Mode Bearer Services* ITU-T Recommendation, ITU-T I.122, March 1993
http://www.itu.int/itu-t/recommendations/index.aspx?ser=I.122

[32] ISO documentation *Intermediate System to Intermediate System* ISO standard, ISO/IEC 10589:2002, November 2002
http://standards.iso.org/ittf/PubliclyAvailableStandards/c030932_ISO_IEC_10589_2002(E).zip

[33] Cisco Documentation *Gateway Load Balancing Protocol Overview* Last review (18/10/2010)
http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6550/prod_presentation0900aecd801790a3_ps6600_Products_Presentation.html

[34] Cisco Documentation *Cisco GLBP Load Balancing Options* Last review (18/10/2010)
http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6554/ps6600/product_data_sheet0900aecd803a546c.html

[35] Cisco Documentation *Load Sharing with HSRP* Last review (18/10/2010)
http://www.cisco.com/en/US/tech/tk648/tk362/technologies_configuration_example09186a0080094e90.shtml

[36] Cisco Documentation *Hot Standby Router Protocol Features and Functionality* Last review (18/10/2010)
http://www.cisco.com/en/US/tech/tk648/tk362/technologies_tech_note09186a0080094a91.shtml

[37] Juniper Documentation *How VRRP works* Last review (18/10/2010)
http://www.juniper.net/techpubs/software/erx/junose53/

swconfig-routing-vol1/html/vrrp-config4.html

[38] Nortel Documentation *Configuring VRRP and using Tracking for Failover* Last review (18/10/2010)
http://support.nortel.com/go/main.jsp?cscat=DOCDETAIL&id=297129&poid=11902

[39] Cisco Documentation *What Is VRRP?* Last review (18/10/2010)
http://www.cisco.com/en/US/products/hw/vpndevc/ps2284/products_tech_note09186a0080094490.shtml

[40] Cisco Documentation *Configuring VRRP* Last review (18/10/2010)
http://www.cisco.com/en/US/docs/ios/ipapp/configuration/guide/ipapp_vrrp.html

[41] IBM Documentation *Virtual Router Redundancy Protocol on VM Guest LANs* Last review (18/10/2010)
http://www.redbooks.ibm.com/abstracts/redp3657.html?Open

[42] 3Com Documentation *Configuring the Virtual Router Redundancy Protocol* Last review (18/10/2010)
http://support.3com.com/infodeli/tools/bridrout/u_guides/html/nb111/family/features/vrrp.htm

[43] Li et al., *Standby Router Protocol* United States Patent US5473599, December 1995
http://www.freepatentsonline.com/5473599.pdf

[44] Protocol documentation *HSRP, Hot Standby Router Protocol* Last review (18/10/2010)
http://www.networksorcery.com/enp/protocol/hsrp.htm

[45] Guide to IP Layer Network Administration with Linux *Breaking a network in two with proxy ARP* Last review (18/10/2010)
http://linux-ip.net/html/adv-proxy-arp.html

[46] Symatech documentation *Proxy ARP* Last review (18/10/2010)
http://www.symatech.net/proxy-arp

[47] Cisco documentation *Proxy ARP* Last review (18/10/2010)
http://www.cisco.com/en/US/tech/tk648/tk361/technologies_tech_note09186a0080094adb.shtml

[48] OpenBSD documentation *The Common Address Redundancy Protocol (CARP)* Last review (18/10/2010)
http://www.openbsd.org/faq/faq6.html#CARP

[49] OpenBSD documentation *Manual page of Common Address Redundancy Protocol (CARP)* Last review (18/10/2010)
http://www.openbsd.org/cgi-bin/man.cgi?query=carp&sektion=4

[50] OpenBSD documentation *Firewall Redundancy with CARP and pfsync* Last review (18/10/2010)
http://www.openbsd.org/faq/pf/carp.html

[51] NetBSD documentation *Introduction to the Common Address Redundancy Protocol (CARP)* Last review (18/10/2010)
http://www.netbsd.org/docs/guide/en/chap-carp.html

[52] ITU-T recommendation *General description of asynchronous transfer mode (ATM)* Last review (9/12/2010)
http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-I.150-199902-I!!PDF-E&type=items

[53] ITU-T recommendation *Characteristics of synchronous digital hierarchy (SDH)* Last review (9/12/2010)
http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.783-200603-I!!PDF-E&type=items

[54] Tellabs documentation *Tellabs 8600 Managed Edge System* Last review (18/10/2010)
http://www.tellabs.com/products/8000/tlab8600sysoverview.pdf

[55] Tellabs documentation *Tellabs 8605 Access Switch Data sheet* Last review (18/10/2010)
http://www.tellabs.com/products/8000/tlab8605as.pdf

[56] Tellabs documentation *Tellabs 8607 Access Switch Data sheet* Last review (18/10/2010)
http://www.tellabs.com/products/8000/tlab8607as.pdf

[57] Tellabs documentation *Tellabs 8620 Access Switch Data sheet* Last review (18/10/2010)
http://www.tellabs.com/products/8000/tlab8620as.pdf

[58] Tellabs documentation *Tellabs 8630 Access Switch Data sheet* Last review (18/10/2010)
http://www.tellabs.com/products/8000/tlab8630as.pdf

[59] Tellabs documentation *Tellabs 8660 Edge Switch Data sheet* Last review (18/10/2010)
http://www.tellabs.com/products/8000/tlab8660es.pdf