27-05-2021

White Paper:
GIX Implementation Based on White Box

Grant Agreement No.:	856726
Work Package	WP6
Task Item:	Task 1
Dissemination Level:	PU (Public)
Lead Partner:	RENATER
Document ID:	GN4-3-21-29376cd
Authors:	Xavier Jeannin (RENATER), Maxime Wisslé (RENATER), Tim Chown (Jisc), Ivana Golub (PSNC)

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 856726 (GN4-3).

Abstract

This document describes the implementation of a Global Internet Exchange point (GIX) based on white box hardware and a commercial network operating system, OcNOS. The objectives of the project were to renew the GIX with the same set of features, reduce the Total Cost of Ownership, increase independence from traditional network vendors, and provide flexibility with regard to future Layer 2 core architecture options, which were successfully met.

Table of Contents

Table of Figures

Table of Tables

Executive Summary

When renewing its Global Internet Exchange point (GIX), the SFINX, in 2020, the French National Research and Education Network, RENATER, implemented the new version based on white box hardware. The white boxes replaced the existing Brocade MLX switches providing the connection service to users at Layer 2 and the forwarding capacity required between SFINX users. The objectives of this work were to evaluate the white box solution in production, increase independence from traditional vendors, provide new alternative solutions for future deployments and reduce the Total Cost of Ownership (TCO).

The SFINX architecture is simple and based on a Layer 2 VLAN setup.

The technical implementation of the GIX use case was successful. The configuration is based on white boxes manufactured by Dell and a Network Operating System (NOS) called OcNOS from IP Infusion. At the operational level, it was possible to have the same level of maintenance both for hardware and software. The RENATER Network Operations Centre was able to integrate OcNOS easily.

The TCO was reduced. However, given the market's responsiveness to competition, the cost comparison will be repeated at intervals.

Since the new SFINX deployment described in this document was finished, the Router for Academia, Research and Education (RARE) Operating System [RARE] has been developed to the point where it is ready for production use. This new, open source NOS provides all the required features for the GIX use case and now seems sufficiently mature. It has therefore become an option that would be considered for the next generation of this GIX white box solution.

1 Introduction

The French National Research and Education Network (NREN), RENATER, implemented a Global Internet Exchange point (GIX), the Service for French Internet Exchange (SFINX), in 1995. The original version was based on traditional vendor hardware, including the proprietary Network Operating System (NOS) provided with the hardware. When renewing their GIX in 2020, RENATER decided to implement a white box solution instead. This white box solution had to provide all the existing SFINX features, with the additional requirement that it should not exceed the Total Cost of Ownership (TCO) of the previous infrastructure.

The SFINX is implemented on two sites located in two RENATER points of presence (PoPs) in Paris, interconnected by a 3 x 10G link aggregation (LAG). Seventeen clients (Internet Service Providers (ISPs)) are connected in the Paris1 PoP and 57 clients are connected in the Paris2 PoP. Future plans might include replacing the Layer 2 GIX core with a Layer 2 core delivered by an Ethernet Virtual Private Network (EVPN) over Internet Protocol (IP) core architecture.

In this document, Section 2 describes the project objectives and requirements for the new version of the SFINX. Section 3 outlines the implementation. The results of this implementation in production, in terms of Total Cost of Ownership, are presented in Section 4, with a wider summary of results, benefits and issues provided in Section 5.

2 Objectives and Requirements

This section describes the project objectives and requirements for the new version of the SFINX. To provide context, it begins by outlining the architecture of the original SFINX.

2.1 Original Architecture

Figure 2.1 presents the original SFINX architecture, which is based on Brocade MLX switches. The first switch (*sfinx1-swi-201*) is located at Interxion (IX1) (Paris1) and the second switch (*sfinx2-swi-201*) is located at Telehouse 2 (TH2) (Paris2). The switch *sfinx1-swo*-201 (IX1) uses 17 interfaces and the switch *sfinx2-swi-201* (TH2) has 57 active interfaces. An aggregation of 3 x 10G links connects the two SFINX sites.

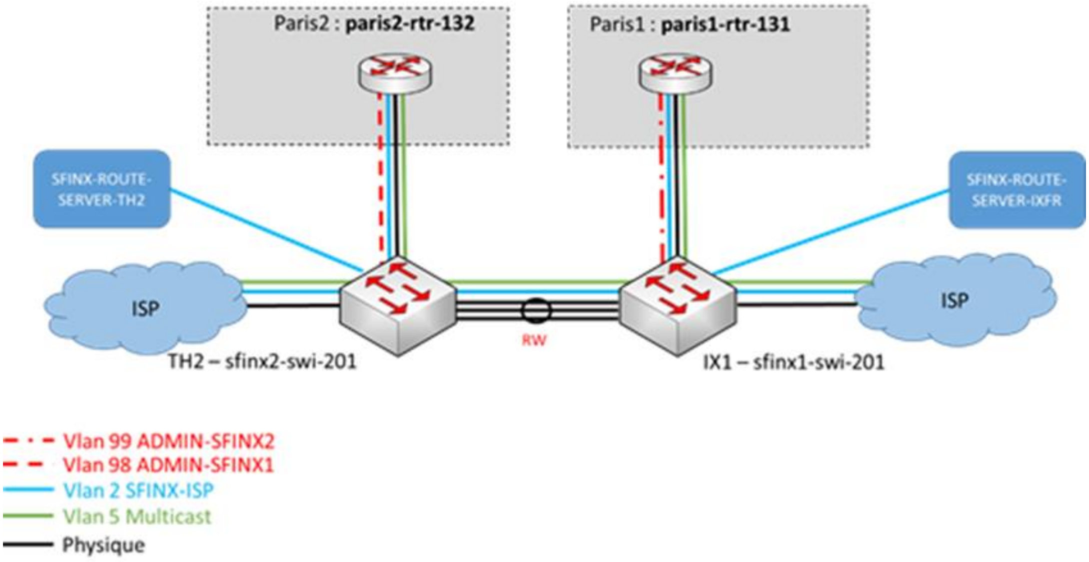

Figure 2.1: SFINX original architecture

The SFINX is based on a set of Virtual Local Area Networks (VLAN 2 for unicast traffic and VLAN 5 for multicast traffic) shared by the connected users / Internet Service Providers (ISPs). The users can exchange their routes thanks to direct peering between them or thanks to two route servers implemented with BIRD software [BIRD]. The RENATER Network Operations Centre (NOC) used "administrative VLANs" to configure the switches.

2.2 Objectives

The primary objective of the project was to renew the SFINX while maintaining all the existing features. In terms of budget, the Total Cost of Ownership (TCO) could not exceed the TCO of the previous infrastructure. In addition, RENATER aimed to increase its independence from traditional vendors, and provide new alternative solutions for future deployments.

2.3 New GIX Requirements

This section describes the requirements for the new SFINX implementation in terms of NOC operation and management, security, routing and switching, and performance and reliability.

2.3.1 Operation and Management

The operation and management requirements are:

- **Out-of-Band (OOB) access**. The RENATER NOC should be able to connect to the machine via a console and/or management port to be able to take control of and reconfigure the machine from scratch in the event of a disaster.
- **Management access**. Management access is used for the daily management operations (changing a configuration, checking something, connecting a new client, etc.). In order to be able to manage the SFINX's switches, the NOC should be able to connect to them via a standard, in-band management interface. The connection should be made via specific administrative VLANs (1 per switch) using the SSH protocol. Telnet access should be limited to the console/management port (Virtual Routing and Forwarding (VRF) management).
- **Authentication and authorisation.** The authentication and authorisation should be done via the same system used by other RENATER backbone equipment, TACACS.
- **Logging**. Logging is done through the in-band management interface and the log has to be sent to four log servers.
- **Monitoring**. Simple Network Management Protocol (SNMP) monitoring should be performed via the in-band management interface. The SNMP client has to be able to send its report to four SNMP servers.
- **Automation**. The switch has to be configurable remotely through a programmable interface.

2.3.2 Security

The security implementation is mainly based on Access Control Lists (ACLs) available on the Network Operating System (NOS):

- **IP layer**. Only the NOC can access the in-band management interface. Telnet can be used only on the management interface (Eth0). The Network Time Protocol (NTP) server, log server and SNMP server can access the white box switch only via the in-band management interface.

- **Media Access Control (MAC) layer**. The following protections at the MAC level are required: MAC ACL, Bridge Protocol Data Unit (BPDU) protection, broadcast storm protection.

2.3.3 Routing and Switching

The routing and switching requirements are:

- **VLAN**. The SFINX uses VLAN (802.1Q).
- **Spanning Tree Protocol (STP)**. STP is disabled. Rapid Spanning Tree Protocol (RSTP) (802.1w) per VLAN should be used. Compatibility with the RSTP per VLAN implementation on the current switch (Brocade) would help the migration phase.
- **Core**. The NOS must have the possibility, in a second project step, to change the core of the GIX architecture by replacing the Layer 2 core with an Ethernet Virtual Private Network (EVPN) over IP core.

2.3.4 Performance and Reliability

The requirements with regard to performance and reliability are:

- **Switching/forwarding**. The solution should be able to forward 30 Gbps of traffic.
- **Bandwidth**. The solution should be able to connect 65 clients at 1 Gbps or 10 Gbps, and should enable some clients to connect at 100 Mbps for testing ("free-trial clients").
- **Reliability**. The SFINX is maintained by a maintenance contract that ensures a mean time to repair (MTTR) of 4 hours for hardware failure, 24 hours a day, 7 days a week. In the event of software problems, the vendor provides a ticket service through which a case could be opened. The reliability must be the same as for the previous SFINX implementation.

3 Implementation

This section outlines the Network Operating System (NOS), hardware and features implementation of the new GIX.

3.1 NOS

RENATER has chosen OcNOS as the Network Operating System for the SFINX. OcNOS is delivered by IP Infusion [IPINFUSION]. It is already deployed in the London Internet Exchange (LINX) and also has relatively rich Multi-Protocol Label Switching (MPLS) features. In coming to this decision, RENATER tested the different features of this NOS and set up a testbed of the new SFINX. In addition, IP Infusion offers software maintenance that is very competitive.

Since the migration was completed, the Router for Academia, Research and Education (RARE) Operating System [RARE] has gained in maturity to the point where it is ready for production. This new, open source NOS provides almost all the required features. It has therefore become an option that would be considered for the next generation of this GIX white box solution.

3.2 Hardware

For the hardware, the choice was led by the ability to support several Network Operating Systems and the capacity to deliver very efficient hardware maintenance. The switch Dell EMC S4048-ON (see Figure 3.1) from Dell supports the open source Open Network Install Environment (ONIE). ONIE is an interface that allows different NOSs to be installed on a single piece of equipment. This switch is equipped with a Broadcom Trident2 chipset, which at the time of implementation was one of the very few chipsets that supported a good number of NOSs: OcNOS, Cumulus Linux OS, Big Switch Networks Switch Light OS, Dell EMC Networking OS9, Dell EMC SmartFabric OS10 and Pluribus OS. This may be important if the NOS needs to be changed for some reason at a later date. In terms of interfaces, there are 48 x 10 Gbps ports with 6 x 40 GbE uplink ports (or 72 x 10 GbE ports in breakout mode) and up to 720 Gbps forwarding performance. The pairing of OcNOS and Trident2 is able to provide sufficient capacity in terms of MAC addresses, Address Resolution Protocol (ARP) table entries, IPv4 routes, IPv6 neighbour discovery table, and IPv6 routes.

As the SFINX has some trial clients connected at 100 Mbps and it was not possible to make the Dell EMC S4048-ON ports function at this speed, two Cisco 2960s were added to the setup to aggregate all the 100 Mbps connections.

Figure 3.1: Front view of Dell EMC S4048-ON

Dell was able to provide maintenance and replacement in two hours on the two SFINX sites.

3.3 Feature Implementation

The following feature implementation descriptions show how the new SFINX meets the requirements outlined in Section 2.3.

Out of Band (OOB) Access

RENATER deployed an OOB network on which all devices are connected through their serial port. The aim of this OOB is to take control of the machine when the in-band management does not work anymore and, if necessary, to reinstall the system via a defined process. The white box (WB) serial port was connected to the RENATER OOB network.

In case the NOS needs to be reinstalled on the WB, ONIE defines the behaviour of the WB when it boots. An IP address configuration can be done via Dynamic Host Configuration Protocol (DHCP) on the management port (1 Gbps) or manually. The NOS, OcNOS in this instance, can be obtained via the configured network from a predefined server. Another option that was used in this instance was to load the switch from a USB stick that contained the OcNOS operating system and the necessary configuration. This process was documented in detail for the RENATER NOC.

Management Access

The switches are not included in the RENATER backbone. SSH access cannot be restricted easily in OcNOS: while OcNOS can be configured to restrict SSH use to the management port, this is not practical, as the RENATER NOC uses SSH to manage RENATER's devices not on the management port but using an SSH connection inside a Virtual Routing and Forwarding (VRF), called admin VRF. This management based on the admin VRF is called in-band management as it uses a standard port. Taking into consideration this limitation, the NOC access management will be achieved through SSH access on a dedicated interface. A VLAN and an IP address will be set up on this interface, called the in-band management interface. Access to all other IP addresses via SSH must be forbidden by use of ACLs. Specific ACLs will be applied in order to limit SSH access of the in-band management interface to the RENATER NOC. Telnet will be limited to the management port (VRF management).

Authentication and Authorisation

One of the prerequisites was that authentication should be done using the same TACACS system RENATER uses for other services, and that the authorisation levels should be organised in the same way. Even though OcNOS supports TACACS, it does not provide the same level and granularity of authorisation and user rights. For example, all users can have the same rights but it was not possible to configure read-only rights for some users and full-access rights for others. Therefore, a compromise has been made: the same TACACS system is used, but with sub-optimal authorisation levels.

Logging

Logging was implemented without any problem through the in-band management interface. All the logs of the WB are sent to a log server to be used for potential debugging.

Monitoring

SNMP monitoring was easy to implement and it is allowed only on the in-band management interface. The WB is polled by four SNMP collectors gathering the throughput of all the switch's interfaces.

Automation

The automation support was tested successfully using Ansible. OcNOS provides an Ansible module that communicates with the white box thanks to SSH. It is not currently used by the RENATER NOC to configure the SFINX.

Security

The Access Control List (ACL) support available on OcNOS allows implementation of the security required by the SFINX:

- **IP layer**. Only the NOC can access the in-band management interface. Telnet can be used only on the management interface (Eth0). The NTP server, log server and SNMP server can access the WB switch only through the in-band management interface.
- **MAC layer**.
  - MAC ACLs are configured to control which hosts are connected to the port and can access the SFINX Layer 2 network. The MAC address filter is implemented on each connected port to control the device connected to the port.
  - Rapid Spanning Tree Protocol (RSTP) protection prevents the users from injecting BPDUs in the SFINX (bpdu-guard).
  - Broadcast storm protection was implemented.

On the SFINX, the security is also based on the fact no one except authorised people can physically access the machine and the ports of the machine.

Routing and Switching

- A set of seven VLANs were implemented on the two switches to manage the SFINX.

  During the migration phase, in order to ensure a smooth migration, it was planned that the Rapid Spanning Tree Protocol (RSTP) (802.1w) per VLAN would be implemented on the two white boxes. However, the two RSTP (802.1w) per VLAN implementations were identified as not compatible. Because the RENATER NOC did not want to change the configuration of the old machine (assessing this as being too risky), RSTP was not configured on the white boxes and (to remove the requirement for spanning tree) it was decided not to create any physical loop when the new white boxes were introduced into the SFINX architecture.

  During the tests, an issue was discovered that slowed down the machine. The NOS supplier identified that the origin of the problem was due to the BIOS of the Dell machine. The NOS supplier was able to provide a patch at the NOS level that corrected the problem without upgrading the machine BIOS.

- OcNOS allows the possibility to change the core of the GIX architecture by replacing the Layer 2 core with an EVPN over IP core.

3.4 Migration Plan

The migration process was carefully planned, with additional verifications introduced after each step, in order to ensure a smooth and stable transition. Figures 3.3, 3.4 and 3.5 below show the migration steps involved.

Figure 3.2 shows the old architecture, prior to the migration.

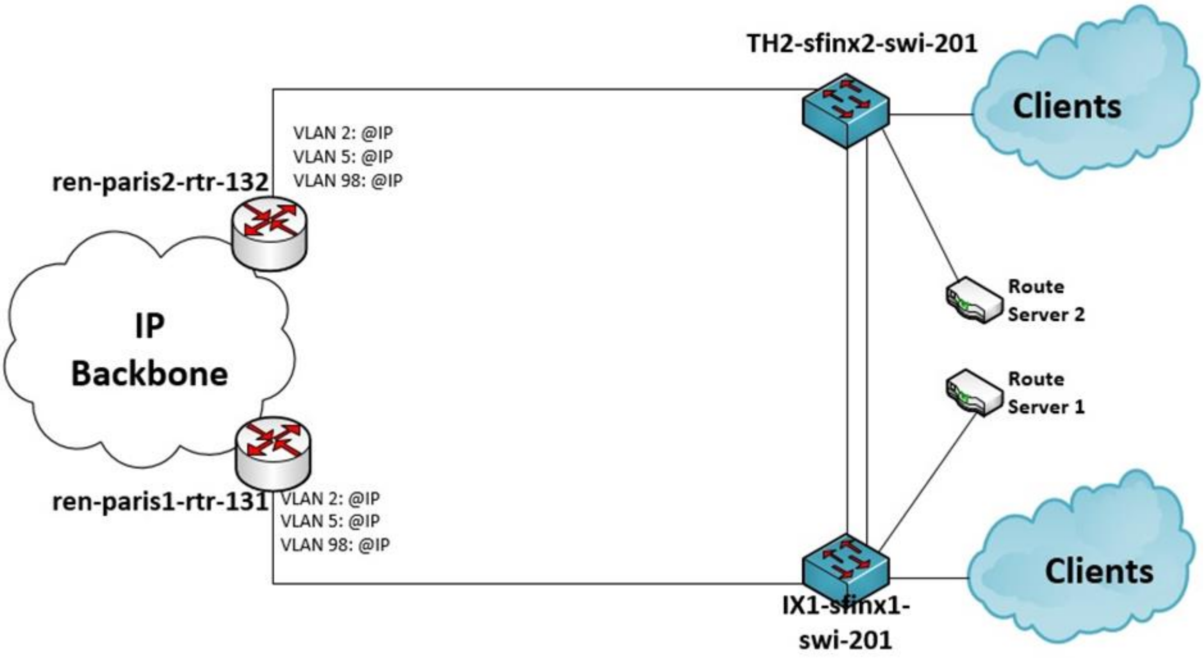

Figure 3.2: Architecture before the migration

Figure 3.3 shows the first of the migration steps as the white boxes were introduced but before the clients and the root servers were migrated. The switches are marked in the diagrams as White-box IX and White-box TH2.

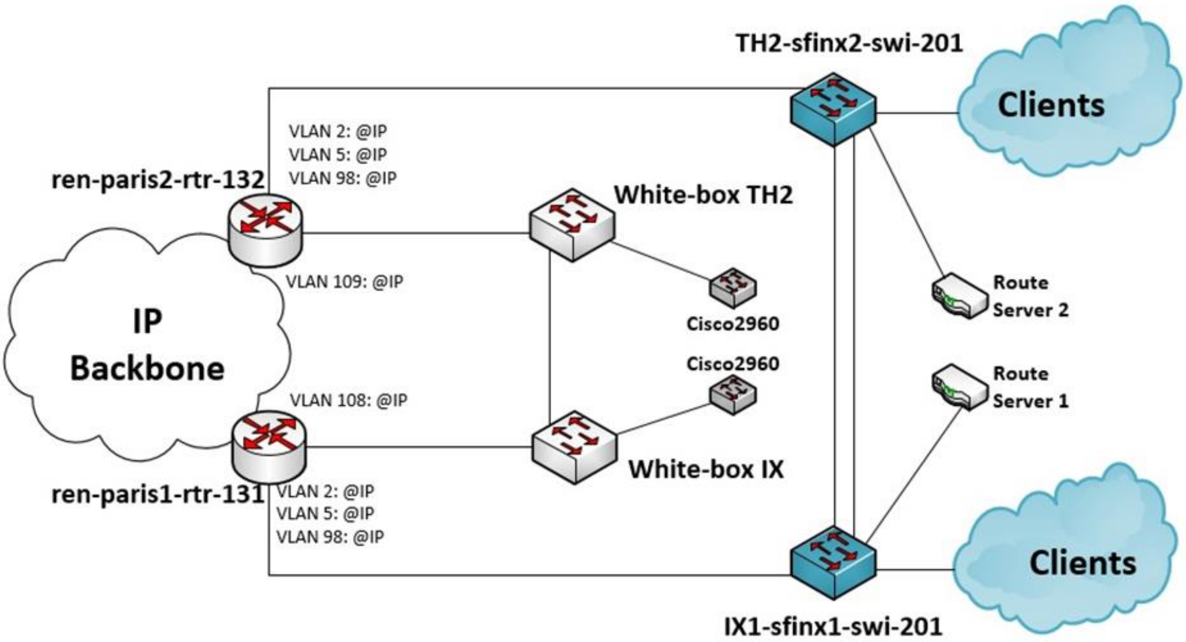

Figure 3.3: White boxes connected to the RENATER backbone

Figure 3.4 shows how the white boxes were connected to the production infrastructure. As it was impossible to implement RSTP, all topological loops were avoided.

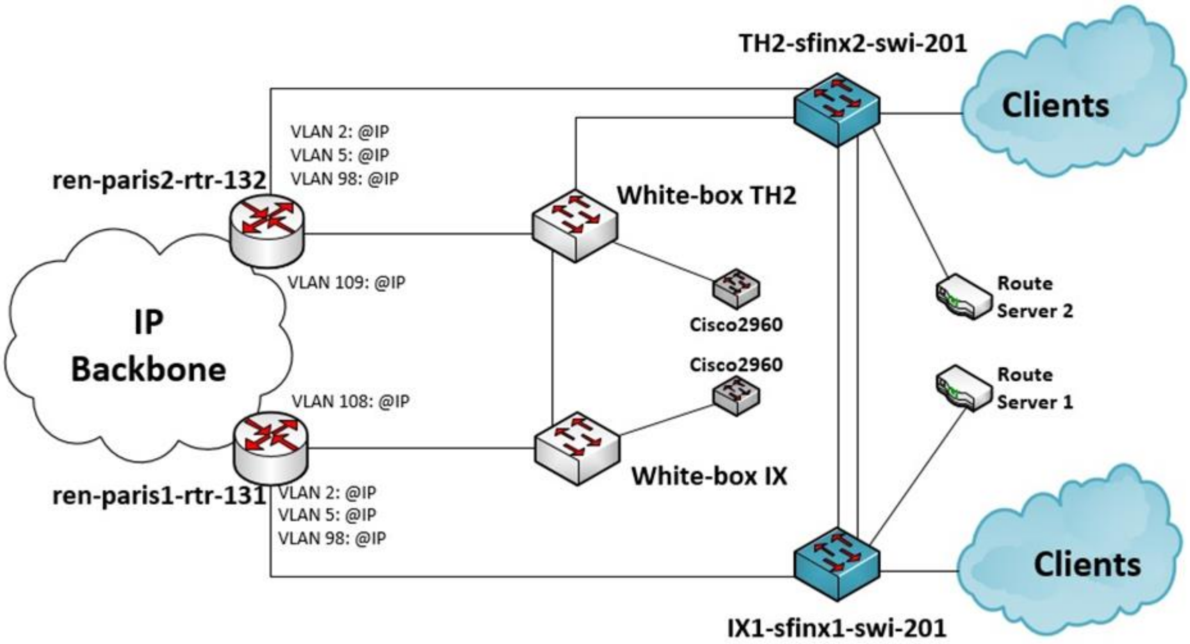

Figure 3.4: Introduction of the white boxes into the SFINX

Figure 3.5 shows the last step of the migration. One can see that the old machines did not deliver service except for clients at Telehouse 2 who were using the free trial at 100 Mbps (free-trial clients). Due to the specific local context, their new connections will require new wiring.

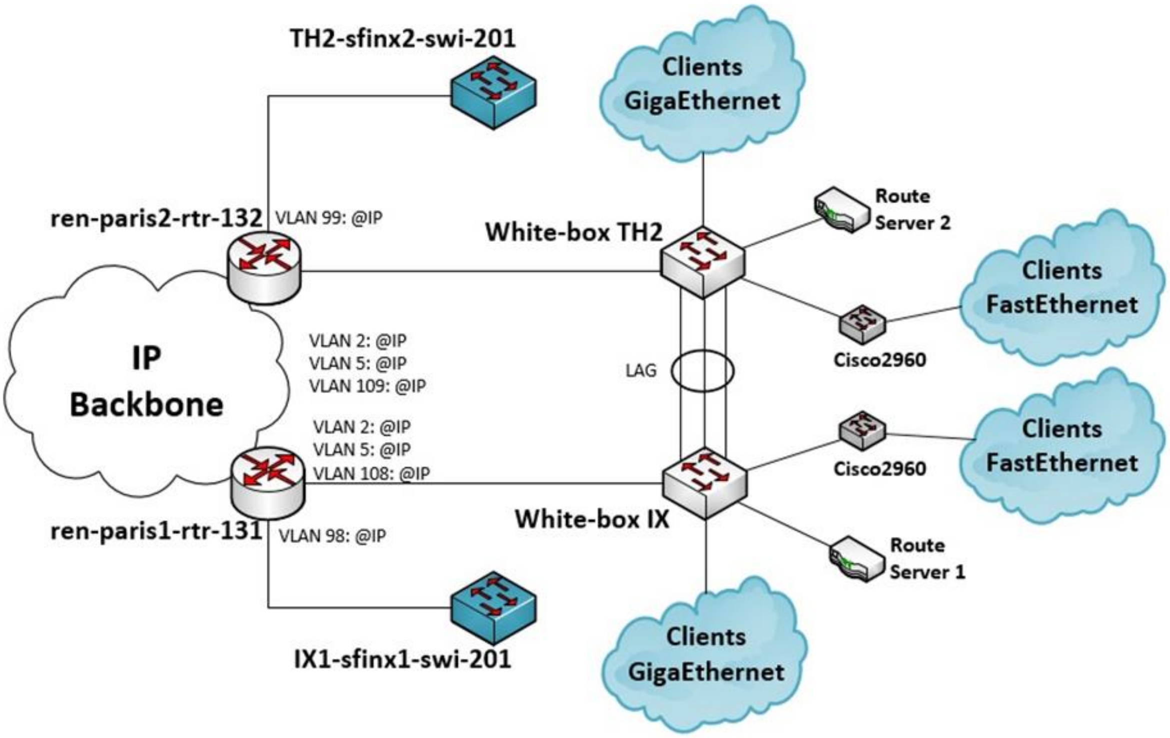

Figure 3.5: Last step of the migration

All the clients on Telehouse 2, except the free-trial clients, were migrated.

At the time of writing, the new SFINX has been running for several months, without any problems. In addition, no performance issue has been observed, although the current traffic level does not put pressure on the hardware.

The statistics of one of the clients connected at 20 Gbps are presented in Figure 3.6.

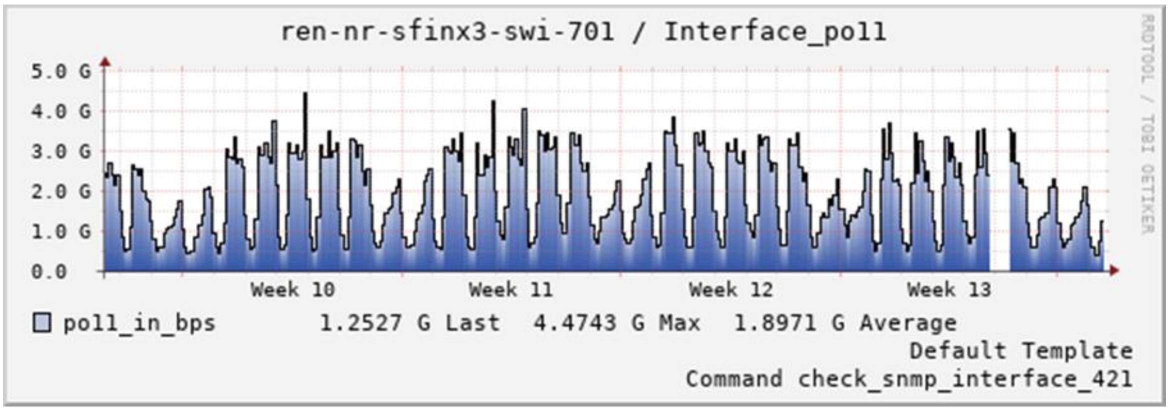

Figure 3.6: Statistics of a client connected at 20 Gbps, March 2021

The SFINX is connected to the RENATER backbone and the statistics from March 2021 are presented in Figure 3.7.

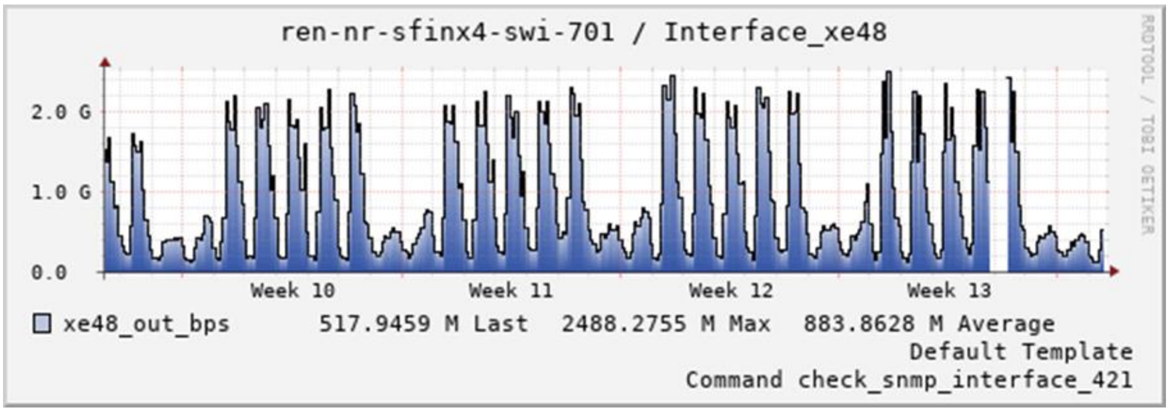

Figure 3.7: Statistics of incoming traffic towards RENATER from SFINX, March 2021

The new white-box-based SFINX solution has therefore been considered to be working successfully, satisfying both the main requirements and quality of service performance levels.

4 Total Cost of Ownership Assessment

The white paper “White Box Total Cost of Ownership” [TCO_WP] and related calculator spreadsheet [TCO_CALC] were used to identify the elements to be taken into consideration for the TCO computation. For commercial confidentiality reasons it is not possible for RENATER to disclose the specific prices and costs pertaining to the TCO computations of the new and previous GIX solutions. However, Tables 4.1 and 4.2 below present a relative comparison of the solutions’ CAPEX and OPEX costs elements.

Costs	Variable	Cost Comparison
White box equipment / hardware	Price per device	WB < traditional vendor
	Number of devices	idem
NOS (licence)	Price per device	Included in traditional vendor hardware cost, but still the cost of the WB hardware plus the NOS is less than the cost of the traditional vendor hardware
	Number of devices	idem
Optical transceivers (SFPs)	Price per piece	idem
	Number of SFPs	idem
Additional software licences	Price per device	Not applicable
	Number of devices	Not applicable

Table 4.1: Comparison of CAPEX costs of previous and new solution

Annual Costs	Variable	Cost Comparison
Electrical power and cooling consumption	Device consumption / year in kW/h	WB < traditional vendor
	Cost of kW/h	idem
Operational activities (payroll, personnel for monitoring, operations, applying updates)	Hourly rate	idem
	Number of machines maintained in production by 1 engineer	idem
Hosting rack unit	Monthly price per rack unit	idem

Annual Costs	Variable	Cost Comparison
	Number of rack units	2 units for WB and more than 12 for traditional vendor
NOS maintenance	No. of years to consider	WB < traditional vendor
Hardware maintenance	No. of years to consider	WB < traditional vendor
Staff training	Training cost per engineer	Training was not necessary as OcNOS was very close to Cisco IOS. It was an opportunity to withdraw another NOS exploitation.
	Number of engineers	idem

Table 4.2: Comparison of OPEX costs of previous and new solution

Overall, the white box solution was the less expensive of the two. The important point is that staff training was not a problem; only the installation process based on Open Network Install Environment (ONIE) needed to be documented. The integration of white boxes into the NOC tools was straightforward. The new NOS (OcNOS) management was easy for the RENATER NOC to adopt; it also gives the opportunity to stop using the proprietary NOS provided with the Brocade hardware and to increase independence from traditional network vendors.

Even though the white box solution was less expensive when the project started, one has to keep in mind that network device vendors are able to quickly adapt their prices to those of new competitors. The situation could change considerably in accordance with the market context and thus the cost comparison has to be repeated at intervals once the project is deployed.

5 Conclusions

This section summarises the results and benefits achieved and the issues encountered during the transition from the old SFINX system to the new, white-box-based one. Several aspects were considered: technical, operational, financial (specifically, cost) and strategic.

The main objective was achieved: the SFINX was renewed with a similar set of features and for a significantly lower cost, at the same time providing RENATER with significantly more independence from its traditional network vendors.

At the technical level, one can notice that OcNOS does not provide the same set of features as traditional vendor equipment. It was also noticed that when the configuration is changed extensively, it could become unstable. Nevertheless, after the configuration was implemented, the NOS has run in production for several months without any problem. For now, the setup is relatively simple; only a Layer 2 switching solution was put in production. It was not possible to integrate the white box machines into the existing TACACS system, but a workaround was put in place.

At the operational level, there was a question about how quickly and easily the NOC engineers could become familiar with a new NOS. In fact, the NOS is very similar to the previously used Cisco IOS and it was very easy to learn to work with the new switch. Except for the TACACS limitation, the integration into the NOC tools was straightforward (OOB access, monitoring, SSH connection). The hardware maintenance was available from the supplier, as well as the NOS maintenance from the software supplier. If it is not obvious where to direct a required maintenance request, the RENATER NOC will first ask the software supplier who is the best party to identify the origin of the problem. As an example, a BIOS problem was correctly solved thanks to this process.

In terms of cost, the white box solution has lower costs for both CAPEX and OPEX. The employees did not have any difficulties mastering the new operating system introduced with the white box solution, as it resembles the existing NOS with which the engineers are already familiar. The maintenance fees for both hardware and software are relatively lower for the white box solution.

At the strategic level, by implementing the white box solution the RENATER management has given itself greater flexibility to address its deployment requirements even if, for now, the NOSs available on white boxes do not provide the appropriate features for the core of the RENATER backbone. However, after this GIX deployment project was finished, the RARE Operating System [RARE] has been developed to the point where it is sufficiently mature for production use. This new, open source NOS will be a good candidate for the next generation of the SFINX.

References

[BIRD] https://bird.network.cz/

[IPINFUSION] https://www.ipinfusion.com/

[RARE] https://wiki.geant.org/display/RARE/Home

[TCO_CALC] https://www.geant.org/Resources/Documents/TCO-Calculator.xlsx?web=1

[TCO_WP] “White Paper: White Box Total Cost of Ownership” https://www.geant.org/Resources/Documents/GN4-3_White-Paper_White-Box-TCO.pdf

Glossary

ACL	Access Control List
ARP	Address Resolution Protocol
BIOS	Basic Input/Output System
BIRD	BIRD Internet Routing Daemon
BPDU	Bridge Protocol Data Unit
CAPEX	Capital Expenditures
DHCP	Dynamic Host Configuration Protocol
EVPN	Ethernet Virtual Private Network
GIX	Global Internet Exchange point
IOS	Internetwork Operating System
IP	Internet Protocol
ISP	Internet ServiceProvider
L*n*	Layer *n*
LAG	Link Aggregation
LINX	London Internet Exchange
MAC	Media Access Control
MPLS	Multi-Protocol Label Switching
MTTR	Mean Time To Repair
NOC	NetworkOperations Centre
NOS	Network Operating System
NREN	National Research and Education Network
NTP	Network Time Protocol
OcNOS	Open Compute Network Operating System
ONIE	Open Network Install Environment
OOB	Out of Band
OPEX	Operating Expenses
PoP	Point of Presence
RARE	Router for Academia, Research and Education
RSTP	Rapid Spanning Tree Protocol
SFINX	Service for French Internet Exchange
SFP	Small Form-factor Pluggable
SNMP	Simple Network Management Protocol
SSH	Secure Shell
STP	Spanning Tree Protocol
TACACS	Terminal Access Controller Access Control System
TCO	Total Cost of Ownership
USB	Universal Serial Bus
VLAN	Virtual Local Area Network
VRF	Virtual Routing and Forwarding
WB	White Box