

TALLER DE SISTEMAS CIBERFÍSICOS

CURSO 2020

SDN para Enrutamiento en el Datacenter

Autor:

Agustina PARNIZARI

Supervisores:

Eduardo Grampín

Alberto Castro

Leonardo Alberro

13 de marzo de 2021

Índice

1. Introducción	2
2. Marco Teórico	3
2.1. Redes Definidas por Software	3
2.2. OpenFlow	5
2.2.1. Open vSwitch	11
2.3. Ruteo basado en origen	12
2.3.1. Segment Routing	12
2.4. ONOS	16
3. Entorno de trabajo	19
3.1. SPRING-OPEN	19
3.1.1. Utilización del ambiente	19
3.2. Kathará	21
3.2.1. Generación de Topologías Spine-Leaf	22
3.2.2. Controlador Onos	23
3.2.3. Dispositivos OpenvSwitch	24
3.2.4. Utilización del ambiente	25
4. Pruebas Realizadas	28
4.1. Spring-Open Mininet	28
4.1.1. Conectividad	30
4.1.2. Falla de Enlaces	40
4.1.3. Prueba de Trafico TCP	42
4.1.4. Ingeniería de Tráfico	44
4.2. Kathará	50
4.2.1. Conectividad	51
4.2.2. Falla de Enlaces	55
4.2.3. Prueba de Trafico TCP	57
5. Conclusión	60

1. Introducción

El tráfico en los Datacenter ha cambiado en los últimos tiempos. Mientras que antes predominaba un tráfico Norte-Sur correspondiente con arquitecturas Cliente-Servidor donde la mayor carga de trabajo se presentaba en las redes WAN, hoy en día la mayor carga de trabajo se da en un sentido Este-Oeste dentro del Datacenter.

Esto se debe en gran parte al auge de servicios en la nube y de los proveedores de contenido. Los servidores interactúan entre si procesando cantidades de datos masivas, realizando respaldos y replicas o almacenando bases de datos distribuidas. Particularmente los sistemas ciberfísicos y la industria 4.0 utilizan servicios en la nube para el procesamiento de los datos que generan, por lo que se ven afectados ante los problemas que se presenten en esta infraestructura.

Por los motivos presentados, las topologías de Datacenter tradicionales y los protocolos de ruteo como OSPF, IS-IS y BGP ya no se ajustan a las necesidades actuales, como son la escala, resiliencia de la red ante fallas y distribución del tráfico.

Debido a ello han resurgido topologías de tipo Clos como Fat-Tree y Leaf-Spine. Este tipo de topologías tiene una forma de árbol en la cual cada rack del Datacenter contiene un ToR (Top of Rack) o Leaf, que se interconecta con los otros Leaf mediante switches de agregación (o Spines), estableciendo distintos caminos. Dichas topologías son robustas a fallas y permiten distribuir la carga mediante algoritmos equitativos como Equal Cost Multi Path (ECMP).

A su vez han surgido protocolos que explotan las cualidades de estas topologías. Se tienen soluciones distribuidas como BGP para el Datacenter, Open-Fabric o RIFT las cuales ya están siendo utilizadas o se encuentran en proceso de estandarización. Por otro lado se tienen soluciones centralizadas basadas en redes definidas por software (SDN) como Segment Routing, la cual se explorará en este trabajo.

El objetivo de este proyecto es entender el funcionamiento de una solución centralizada para el ruteo en el Datacenter, conocer herramientas utilizadas en este paradigma y crear un entorno de trabajo que permita probar su funcionamiento.

2. Marco Teórico

2.1. Redes Definidas por Software

Las Redes Definidas por Software o SDN (Software Defined Networks) plantean un paradigma en el cual se separa físicamente el plano de control del plano de forwarding o datos. El plano de control además tiene la capacidad de controlar varios dispositivos.

Su arquitectura se divide en tres capas, como se puede ver en la figura 1. En la capa de Aplicación corren las aplicaciones y servicios de la red, que van a utilizar los recursos disponibles. Las aplicaciones se comunican con la capa de control mediante APIs REST, extrayendo información de los recursos del controlador SDN. La capa de Control, que contiene el controlador SDN, se comunica directamente con la infraestructura mediante interfaces como OpenFlow (una de las tecnologías abiertas más utilizadas) para controlar a los dispositivos de red (switches, routers, etc) que se encuentran en la capa inferior o de Infraestructura.

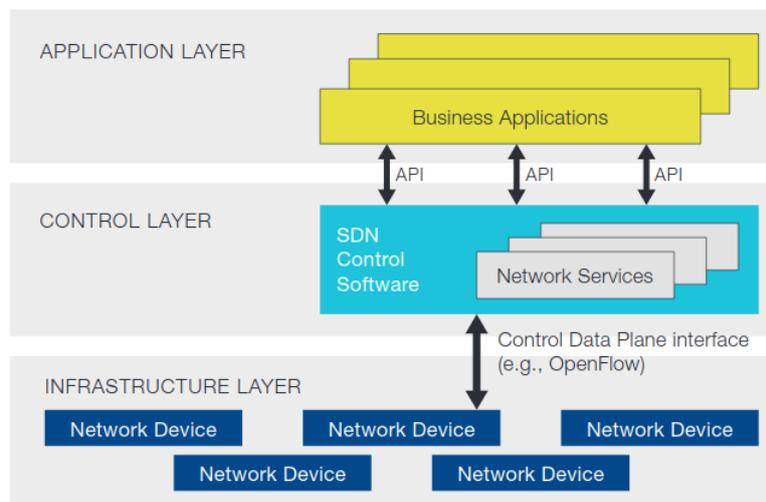


Figura 1: Arquitectura de SDN [1]

SDN presenta ciertos beneficios en comparación con el paradigma de red distribuido tradicional. Entre estos beneficios se pueden mencionar los siguientes:

- Directamente programable: al estar desacopladas las funciones de control de

las funciones de forwarding se puede configurar y programar el control de la red con herramientas de automatización.

- **Ágil:** Los administradores pueden realizar cambios sobre el tráfico de forma dinámica y aplicándolos a toda la red según las necesidades de ese momento.
- **Gestión centralizada:** La lógica de la red se centraliza en controladores que tienen una visión completa de la infraestructura. Esto facilita la gestión de los administradores que tienen acceso a una única herramienta para aplicar cambios en la red.
- **Configuración programada:** Los administradores puede gestionar, asegurar y optimizar los recursos de red mediante programas atomizados, que pueden desarrollar ellos mismos dado que no dependen de software propietario.
- **Estándares abiertos:** SDN simplifica el diseño y la operativa de red, dado que las instrucciones las ejecutan los controladores, generalmente con protocolos abiertos, en lugar de dispositivos y protocolos cerrados.

Por otro lado la adopción de SDN tiene algunos desafíos por enfrentar, dentro de los cuales se encuentran:

- **Seguridad:** La gestión centralizada presenta un único punto de falla. Si este es atacado toda la red se ve comprometida.
- **Cuello de botella:** Cuando se utiliza una única instancia de controlador SDN, puede convertirse en un cuello de botella ante una gran cantidad de tráfico.
- **Universalización de APIs:** No existe un estándar universalmente aceptado para la comunicación entre las aplicaciones y los controladores. Existen distintas APIs para distintos controladores SDN lo cual hace más difícil el desarrollo de nuevas aplicaciones que se comuniquen con distintos controladores.

Esta arquitectura presenta casos de uso interesantes, desde virtualización de redes donde se pueden encontrar herramientas comerciales como NSX de VMware o abiertas como Neutron de Openstack, su uso en Datacenters, redes WAN como la WAN privada de Google, hasta su uso en redes celulares 5G. Se puede consultar

en mas detalle sobre sus posibles casos de uso en [2].

Uno de los casos de uso predominantes es su aplicación en Datacenters donde se busca utilizar switches 'white-box' o abiertos. Estos switches al contrario de los switches 'black-box' tradicionalmente atados a las especificaciones de sus proveedores, permiten un control mediante estándares abiertos y programables como OpenFlow.

Los switches generalmente se encuentran en una topología leaf-spine como se puede ver en la figura 2. Mientras que entre los servidores de un mismo rack (Leaf) la comunicación es a nivel de capa de enlace, la interconexión entre los distintos racks es de capa de red. El ruteo no es realizado por los switches sino por un controlador centralizado el cual indica a los switches las reglas de forwarding que deben seguir.

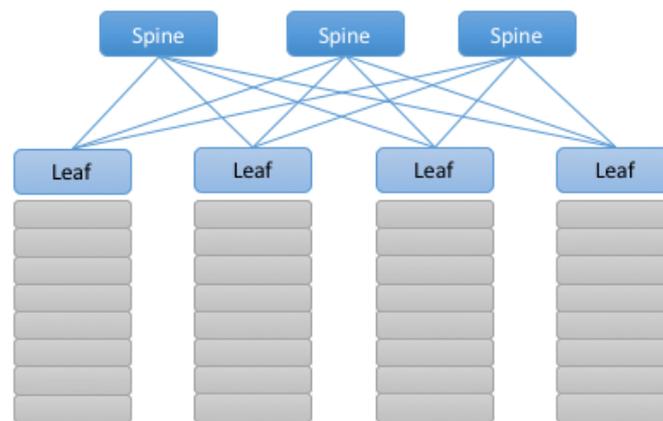


Figura 2: Ejemplo de leaf-spine Switching Fabric [2]

Se va a estudiar en detalle este caso de uso aplicando metodologías de ruteo basadas en origen para topologías leaf-spine.

2.2. OpenFlow

Openflow es un conjunto de protocolos y API utilizados en SDN para separar el plano de control del plano de forwarding. Los principales componentes de su arquitectura se muestran en la figura 3. Estos son el controlador Openflow, que genera las instrucciones del plano de forwarding, el protocolo de comunicación entre

el controlador y los dispositivos, y el switch como unidad cuya función es procesar paquetes.

El controlador OpenFlow genera las instrucciones para el plano de forwarding a partir de las decisiones de ruteo de las aplicaciones. Las instrucciones están contenidas en estructuras de datos en forma de tabla. Luego el protocolo envía a los switches dichas tablas y cada switch procesa los paquetes entrantes según las instrucciones obtenidas.

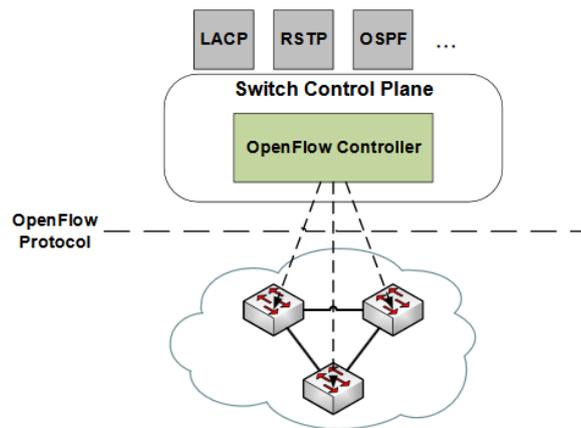


Figura 3: Arquitectura de OpenFlow [3]

Estructuras de datos

Las tablas de flujos y grupos son estructuras de datos que definen el comportamiento de los paquetes que ingresen y egresen de los switch en el plano de datos. Las mismas son generadas por el controlador SDN en el plano de control y se construyen mediante las API.

Las tablas de flujos analizan la información de los paquetes que ingresan al switch y ejecuta la acción que le corresponda. Se realiza un matcheo basado en el modelo TCP/IP, donde cada campo puede contener información de la interfaz física, dirección MAC, VLAN, dirección IP o puerto de transporte. Una vez se encuentra para el paquete coincidencia con un flujo se ejecuta una instrucción con un conjunto de acciones. La imagen 4 muestra la estructura de una tabla de flujos.

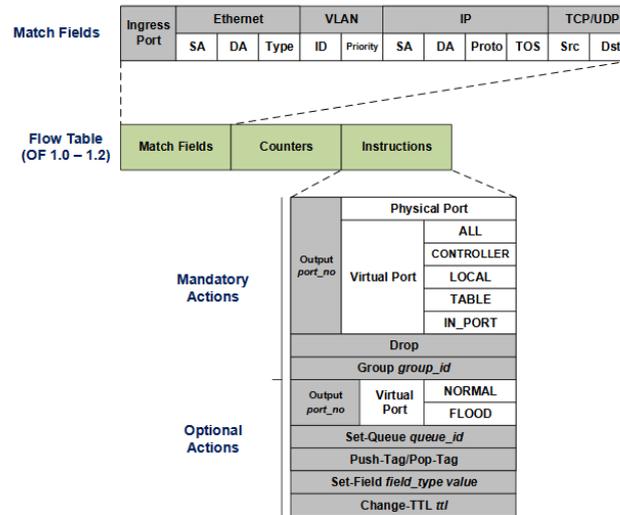


Figura 4: Tabla de Flujo OpenFlow [3]

El campo *Match* contiene la información de matcheo para comparar con el paquete y el campo *Counters* contiene contadores que permiten al sistema generar estadísticas (cantidad de paquetes, bytes, etc). En cuanto al campo *Instructions*, se puede observar que algunas acciones son obligatorias mientras que otras son opcionales. Se puede consultar en mayor detalle sobre la estructura en [3], pero resulta de interés comprender la acción *Output* que utiliza distintos tipos de puertos:

- El puerto físico es una interfaz de hardware del switch y a su vez una interfaz Ethernet.
- Un puerto lógico es una interfaz virtual de un Switch y una interfaz Ethernet.
- Un puerto reservado es una interfaz lógica que se asocia a una operación de forwarding específica y puede estar asociada a mas de una interfaz lógica o física. Los distintos puertos reservados son:
 - ALL: Envía paquetes a todas las interfaces.
 - CONTROLLER: Envía paquetes al controlador.
 - LOCAL: Envía paquetes a la interfaz de loopback.
 - TABLE: Envía paquetes a la proxima tabla de flujo.

- IN_PORT: Envía paquetes a la interfaz de entrada.
- NORMAL: Procesa los paquetes como un switch tradicional.
- FLOOD: Envía paquetes por todas interfaces salvo la de ingreso. Normalmente se utiliza para ARP y LLDP.

Por otro lado se tienen las tablas de grupos, que permiten enviar un paquete por diferentes puertos debido a diferentes propósitos. Si bien la acción FLOOD permite enviar un mensaje por varios puertos, esta limitada a protocolos como LLDP. Con las tablas de grupos sin embargo se pueden agrupar puertos para diferentes acciones, por ejemplo para utilizar ECMP.

En la imagen 5 se presenta la estructura de las tablas de grupo. Las entradas de las tablas contienen un identificador del grupo, un tipo de grupo, contadores para estadísticas al igual que en las tablas de flujo, y un conjunto de acciones o *buckets*. Los *buckets* contienen las instrucciones a seguir, que pueden incluir enviar los paquetes por uno o multiples puertos, o enviar el paquete a otro grupo para que lo procese, permitiendo encadenar acciones lo cual veremos más adelante que es necesario en Segment Routing.

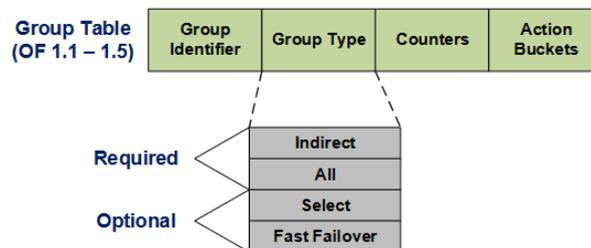


Figura 5: Tabla de Grupo OpenFlow [3]

Los grupos son de cuatro tipos distintos, donde dos de ellos son requeridos y siempre deben estar presentes mientras que el resto son opcionales. Los distintos tipos son:

- *Indirect*: Ejecuta un *bucket* del grupo asociado a un puerto de salida, permitiendo emular el comportamiento de capa de red, como apuntar al próximo destino IP.

- *All*: Se ejecutan todos los *buckets* del grupo y el paquete se envía por cada interfaz relacionada con un *bucket*, permitiendo emular el comportamiento de multicast o broadcast.
- *Select*: Ejecuta un *bucket* en el grupo. El mismo va cambiando y es seleccionado por un algoritmo de Round Robin o una función de Hash, permitiendo el balanceo de carga. Con este tipo de grupo se puede emular el comportamiento de ECMP por lo cual va a ser de especial interés.
- *Fast Failover*: Ejecuta el primer *bucket* cuya interfaz asociada está operativa. Si la interfaz falla se ejecuta el próximo *bucket* activo. Con este tipo de grupo se puede emular fast failover para tener un backup de links sin consultar al controlador SDN ante fallas.

Protocolo de red

El protocolo de red o canal Openflow representa una interconexión lógica entre el controlador Openflow y los switches. El controlador configura y gestiona los switches y recibe eventos de los mismos para conocer el estado de la red.

Los mensajes pueden ser iniciados por el controlador, a lo cual se le llama '*controller-to-switch*', con lo cual monitorean y configuran el comportamiento de los switch. También pueden ser inicializados por los switches para enviar información sobre el estado de la red al controlador, a lo cual se le llama mensajes asíncronos o pueden ser simétricos, es decir iniciados por ambas partes, lo cual generalmente se usa para establecer la conexión o enviar mensajes *keep alive*.

Es de interés para entender como funciona el protocolo dar una breve explicación de algunos mensajes relevantes. Por más detalle de los mismos se puede consultar [4].

- OFPT_HELLO: Este mensaje se utiliza por el controlador y los switches para establecer la conexión. Se identifica y establece que versión de OpenFlow se va a utilizar.
- OFPT_FEATURES_REQUEST / REPLY: El controlador envía la consulta a un switch para obtener sus *capabilities*. Este mensaje solo contiene el header

de OpenFlow. El switch responde indicando parámetros de interés para el controlador.

- `OFPT_MULTIPART_REPLY`: Utilizado para realizar consultas que puedan traer una cantidad de datos mayor a la soportada por el cuerpo de OpenFlow (limitada a 64KB). Generalmente se utiliza para obtener estadísticas y detalles de los switches. Dentro del mensaje se indica el tipo de datos que se quiere obtener. Por ejemplo `OFPMPP_PORT_DESC` obtiene el detalle de cada puerto del switch.
- `OFPT_GET_CONFIG_REQUEST / REPLY`: El controlador puede consultar la configuración de los switches. La consulta es vacía mientras que la respuesta contiene distintos parámetros de configuración.
- `OFPT_BARRIER_REQUEST / REPLY`: El controlador puede consultar si el switch termino de procesar ciertas operaciones. Antes de contestar el switch debe terminar de procesar todos los mensajes pendientes.
- `OFPT_FLOW_MOD`: El controlador modifica las tablas de flujos del switch con este mensaje.
- `OFPT_GROUP_MOD`: El controlador modifica las tablas de grupos del switch con este mensaje.
- `OFPT_PACKET_IN / OUT`: Cuando se envía tráfico del plano de datos al controlador lo recibe mediante un mensaje `PACKET_IN`. Cuando es el controlador el que desea enviar un mensaje envía un `PACKET_OUT`. Generalmente los mensajes ARP se envían al controlador que resuelve la consulta.

Switch OpenFlow

El switch OpenFlow es la unidad lógica que procesa paquetes a partir de las tablas de flujos y grupos vistos anteriormente. El switch se conecta al controlador mediante el canal Openflow, donde se manejan los mensajes del protocolo. En la figura 6 se puede ver su estructura.

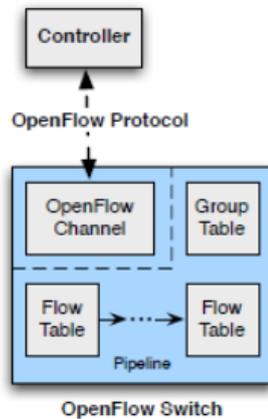


Figura 6: Switch OpenFlow [3]

Los paquetes son procesados a través de varias tablas de flujo. Estas tablas se procesan secuencialmente, a lo que se le llama *Pipeline* de OpenFlow. Los paquetes pueden ser enviados directo a una interfaz de salida o al controlador desde cualquiera de las tablas, o pueden ser descartados en este proceso.

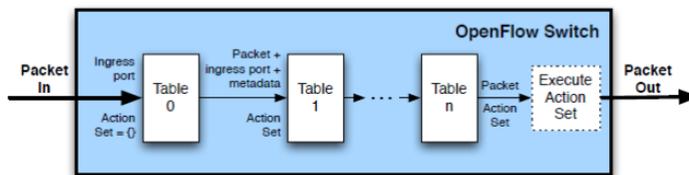


Figura 7: Pipeline de OpenFlow [3]

2.2.1. Open vSwitch

Open vSwitch es un switch de código abierto diseñado para utilizarse en producción como switch virtual. Entre sus funcionalidades tiene soporte de OpenFlow por lo que se puede utilizar en la arquitectura SDN.

La herramienta tiene soporte para sistemas operativos Linux a partir del kernel versión 3.10, por lo cual un sistema Linux con Open VSwitch instalado y las interfaces de red necesarias, conectado a un controlador SDN, actúa como switch

Openflow. Esto será de utilidad al momento de realizar pruebas en un entorno virtual. Se puede profundizar sobre Open vSwitch en [7].

Open vSwitch soporta las distintas versiones de OpenFlow y puede utilizarlas al mismo tiempo en distintos bridges lo cual permite una gran flexibilidad en su uso. Particularmente se utilizará OpenFlow version 1.3 dado que es compatible con los requerimientos de Segment Routing. Open vSwitch tiene soporte de esta versión de OpenFlow así como sus versiones anteriores.

2.3. Ruteo basado en origen

El ruteo basado en origen consiste en que el dispositivo de red de origen defina la ruta total o parcial de un paquete en la red, en lugar de que sea procesado por cada nodo intermedio en la red. El camino se agrega al cabezal del paquete y los nodos intermedios realizan forwarding a partir de esta información, en lugar de calcular el próximo salto en cada paso.

2.3.1. Segment Routing

Segment Routing es un método de ruteo basado en origen en el cual se tiene un conjunto de instrucciones o segmentos que guían a los paquetes a través de los dispositivos de red. Este concepto fue propuesto por Cisco y su estándar es mantenido por el grupo de la IETF llamado Source Packet Routing in Networking (SPRING).

Esta tecnología permite utilizar ingeniería de tráfico de una forma sencilla y escalable mediante un soporte de túneles y políticas, lo cual da una gran flexibilidad a los administradores de red.

Puede aplicarse a MPLS sobre IPv4 sin realizar cambios en su arquitectura, donde los segmentos son etiquetas MPLS y la lista de instrucciones es el Stack MPLS. Se puede aplicar también a IPv6 donde los segmentos son direcciones IPv6 y se debe agregar a los paquetes un cabezal de ruteo, conteniendo la lista de direcciones. En este trabajo se estudiara su aplicación a MPLS, razón por la cual se mencionarán algunos conceptos básicos de dicha metodología.

MPLS

MPLS (Multiprotocol Label Switching) es un mecanismo de conmutación de etiquetas que opera entre la capa de red y la capa de enlace. Se diseñó para unificar los servicios de transporte de datos para las redes basadas en circuitos y paquetes, reemplazando a tecnologías como ATM.

La conmutación de paquetes es el mecanismo de forwarding que utiliza, en el cual los paquetes contienen una etiqueta que indica a los nodos como procesar y reenviar los datos. La información de etiquetas se mantiene en una LIB (Label Information Base) en cada nodo y es distribuida mediante el protocolo LDP (Label-Distributed Protocol).

En la figura 8 se puede observar como opera MPLS. Todos los nodos realizan conmutación de etiquetas, lo cual se identifica como nodos LSR (Label Switch Router). Mientras que algunos nodos solo realizan forwarding de etiquetas MPLS, algunos nodos identificados como Edge Routers realizan la conversión entre IP y MPLS.

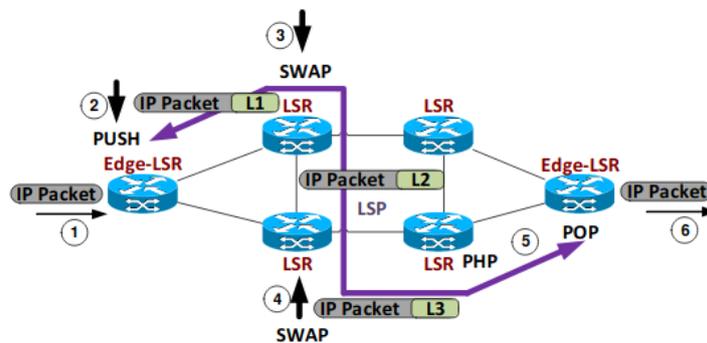


Figura 8: Modelo MPLS [3]

Las acciones que pueden tomar los nodos LSR en la conmutación de etiquetas son:

- PUSH: Se asigna una etiqueta a los paquetes IP que ingresan a la red MPLS.
- SWAP: Dirige el paquete al próximo nodo según la etiqueta MPLS.

- POP: Se quita la etiqueta MPLS del paquete IP. Generalmente esta acción la realiza el penúltimo nodo de la ruta y no el nodo de borde, lo que se conoce como PHP (Penultimate Hop Popping).

Modelo de Segment Routing

Se introducirá la arquitectura y funcionamiento de Segment Routing aplicado a MPLS. Por más detalles sobre dicha aplicación o su implementación en IPv6 se puede consultar en [8] y [3].

En la figura 9 se puede observar el comportamiento de Segment Routing utilizando MPLS. A partir del nodo de borde se genera una lista ordenada de segmentos identificados como SID (Segment ID), la cual atraviesa la red mientras se le realizan las siguientes operaciones:

- PUSH: Ejecutada por un nodo SR para introducir un segmento en la lista. Coincide con la acción PUSH MPLS.
- CONTINUE: Reenvía el paquete sin realizar cambios en la lista de segmentos, correspondiéndose con la acción SWAP MPLS.
- NEXT: Quita el segmento activo y pasa al próximo, correspondiendo con la acción POP MPLS.

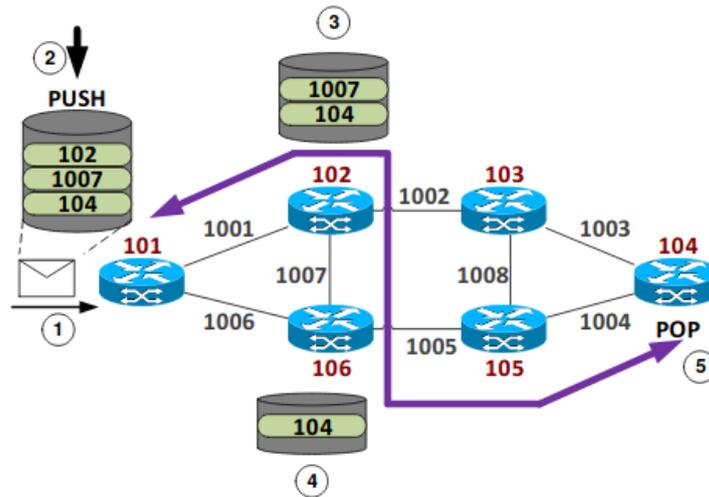


Figura 9: Segment Routing [3]

Como se menciono anteriormente, los segmentos se identifican por su SID, los cuales pueden ser de distintos tipos:

- Node-Sid: Identificador único de un nodo SR.
- Adjacency-Sid: Identifica un link de egreso de un nodo.
- Service-Sid: Identifica un servicio particular que provee un nodo. Puede ser por ejemplo un servicio de inspección de paquetes.
- Anycast-Sid: Identifica un grupo de nodos, lo cual permite funcionalidades como Multicast o ECMP.

Hasta ahora se vio el modelo de plano de datos de Segment Routing, es de interés conocer el funcionamiento del plano de control. Se tienen tres posibilidades para la implementación del mismo:

- Configuración estática: Uso de túneles configurados manualmente. Esto se utiliza sobre todo para encontrar problemas de forma temporal.
- Algoritmo distribuido: Los nodos calculan el camino más corto al destino utilizando otros algoritmos como OSPF e IS-IS y luego generan el stack de SID.

- SDN: Un controlador SDN realiza los cálculos teniendo una visión global de la red y envía el camino computado a los nodos.

Tanto en el caso distribuido como el caso centralizado, el algoritmo de cálculo de rutas por defecto es SPF para ECMP, permitiendo en el caso centralizado agregar políticas que sobre escriben su comportamiento, permitiendo una mayor flexibilidad en la aplicación de técnicas de ingeniería de tráfico.

En el presente trabajo se estudiará una implementación particular de Segment Routing sobre SDN, utilizando el controlador ONOS.

2.4. ONOS

ONOS (Open Networking Operating System) es un controlador SDN que se define como un sistema operativo dado que provee APIs, asignación de recursos, servicios y software orientado a usuarios, como un CLI y una GUI.

Se presentará el caso de uso de Ruteo en el Datacenter con sus aplicaciones y relación con el controlador ONOS. Se puede profundizar sobre la herramienta en [9].

Segment Routing en ONOS

Se presenta la arquitectura de Segment Routing implementada como aplicación de ONOS, la cual provee el plano de control para aplicar Segment Routing sobre SDN. La figura 10 presenta su estructura cuyos componentes son:

- Segment Routing Manager: Computa SPF de ECMP y genera las reglas de forwarding para los switches.
- ARP Handler: Se encarga de las consultas ARP. Contiene las tablas ARP de los routers que gestiona.
- ICMP Handler: Maneja las consultas ICMP dirigidas a los routers
- Generic IP Handler: Maneja los paquetes IP recibidos. Si el destino de los paquetes se encuentra dentro de las subredes de los routers, envía el paquete al host correspondiente. En caso contrario envía una consulta ARP al ARP Handler.

- Segment Routing Policy: Crea las políticas y asigna las reglas a las tablas de ACL de los routers. Si están asociadas a túneles, crea también los túneles.

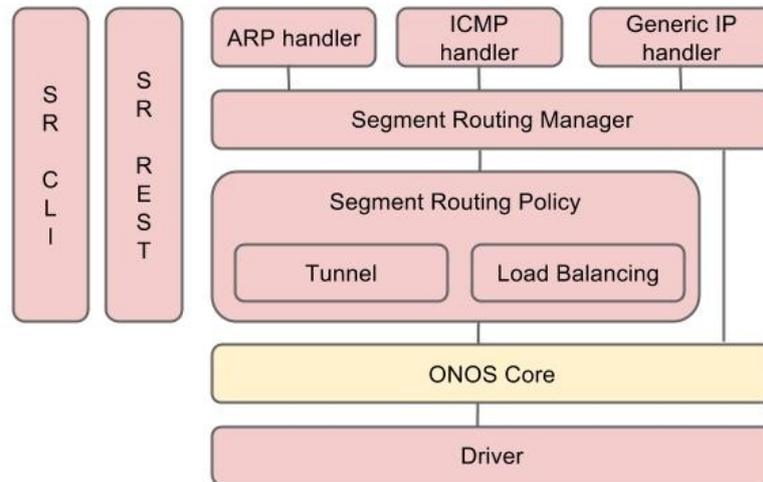


Figura 10: Arquitectura de Segment Routing en ONOS [10]

Un elemento importante de Segment Routing en Onos es el servicio de configuración, el cual se puede ver en la imagen 11. El mismo maneja la configuración necesaria para Segment Routing de todos los elementos de red. Se va a ver al momento de generar entornos de prueba que se necesita configurar los switches para que Onos aplique Segment Routing sobre ellos.

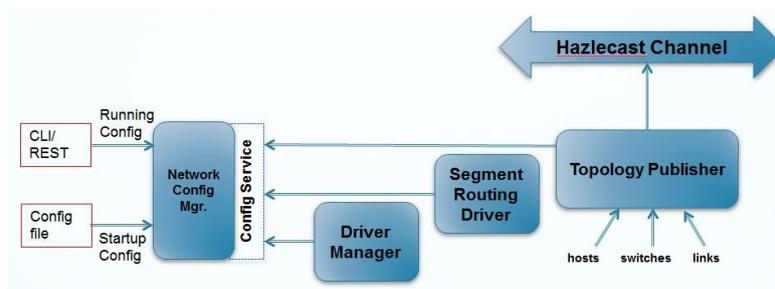


Figura 11: Servicio de configuración [10]

Trellis

Trellis es una suite de aplicaciones SDN en el Datacenter que implementa una topología leaf-spine sobre *white switches*. Las aplicaciones corren sobre un controlador ONOS.

Trellis interconecta el Datacenter a otras redes, incluyendo Internet, mediante BGP. A su vez conecta el Datacenter a redes de acceso. Como se puede ver en la imagen 12, no se trata de una topología de Datacenter convencional, sino que Trellis se encuentra en el borde de la red interconectando los distintos componentes.

Este despliegue ya esta siendo utilizado por proveedores de servicio Tier 1 y es un caso de uso interesante, dado que utiliza puramente SDN tanto dentro del Datacenter como por fuera.

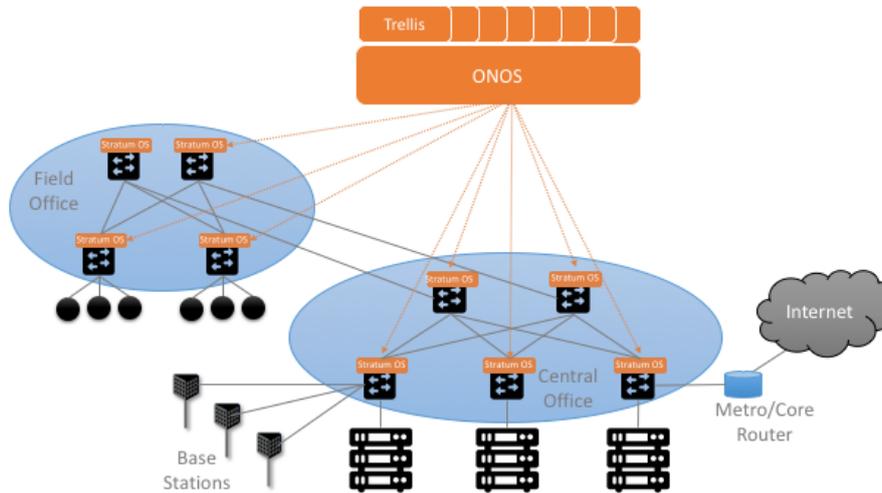


Figura 12: Despliegue de Trellis

3. Entorno de trabajo

Se utilizaron dos entornos de trabajo distintos para la implantación y prueba de funcionamiento de Segment Routing en Onos. En primer lugar se utilizó el entorno del proyecto SPRING-Open [11] propuesto por la Open Network Foundation (ONF [13]) el cual permitió comprender la aplicación y realizar pruebas que se detallarán en la sección 4. En segundo lugar se emuló un entorno de producción utilizando el framework Kathará [17] basado en contenedores.

3.1. SPRING-OPEN

El proyecto SPRING-OPEN presenta un ambiente que implementa Segment Routing utilizando una versión del controlador Onos interna de los desarrolladores, la cual no fue lanzada como una versión oficial. No es recomendable utilizar este entorno para un ambiente de producción pero es de gran utilidad para prototipos y pruebas funcionales. El entorno utiliza un cliente implementado en Python y está enfocado en el uso de la aplicación Segment Routing.

En cuanto a los dispositivos de red el proyecto permite tanto una implementación por Hardware utilizando switches Dell 4810 como switches virtuales utilizando Mininet [14] y OpenVSwitch CPqD con la versión de OpenFlow 1.3. Dado el alcance del proyecto y disponibilidad de recursos se utilizó el ambiente emulado.

El entorno se encuentra disponible incluido en una máquina virtual de Virtual Box con sistema operativo Ubuntu 14.04 incluyendo los elementos de software mencionados anteriormente instalados. Se le asignaron a la VM 2 CPU y 2GB de memoria RAM. Por simplicidad se utilizó dicha máquina virtual pero se pueden consultar las instrucciones de instalación de cada componente en [12].

3.1.1. Utilización del ambiente

El controlador debe ser configurado antes de inicializar con un archivo de configuración de la topología que se desee utilizar. El proyecto incluye varias configuraciones de prueba, pero se puede crear una nueva siguiendo la sintaxis propuesta en [15]. La misma se debe ubicar en el directorio `/spring-open/conf/`. Luego se debe editar el archivo `onos.properties` ubicado en dicho directorio cambiando el parámetro que se puede ver a continuación.

```
# Specify a network configuration file to be used
by the NetworkConfigManager
net.onrc.onos.core.configmanager.NetworkConfigManager.networkConfigFile
= conf/sr-3node.conf
```

Una vez configurado Onos se debe iniciar el controlador ejecutando el comando `./onos.sh start` dentro del directorio `spring-open`. Se puede verificar el estado del programa ejecutando `/spring-open/onos.sh status`, una ejecución correcta debería mostrar una salida similar a la que se puede ver en la figura 13.

```
mininet@mininet-vm:~$ ~/spring-open/onos.sh status
[ZooKeeper]
JMX enabled by default
Using config: /home/mininet/spring-open/conf/zoo.cfg
Mode: standalone

[RAMCloud coordinator]
0 RAMCloud coordinator running

[RAMCloud server]
0 RAMCloud server running

[ONOS core]
1 instance of onos running
```

Figura 13: Estado de Controlador Onos correcto

Una vez iniciado el servicio se puede iniciar la topología en Mininet. El proyecto ofrece varios ejemplos, sin embargo en este caso se utilizó un script personalizado que genera topologías del tipo Spine-Leaf. Los scripts se ubican en el directorio `/mininet/custom` y deben tener permiso de ejecución. En la figura 14 se puede ver un ejemplo de ejecución para una topología Spine-Leaf con dos Spine, dos Leaf y un servidor por Leaf. En la sección 4 se utilizarán otros ejemplos.

```

mininet@mininet-vm:~/mininet/custom$ sudo ./sr_2x2_spine_leaf-script.py
*** Creating network
*** Adding controller
*** Adding hosts:
h1 h2
*** Adding switches:
s1 s2 s3 s4
*** Adding links:
(5ms delay) (5ms delay) (h1, s1) (5ms delay) (5ms delay) (h2, s2) (5ms delay) (5ms delay) (s1, s3) (5ms
elay) (5ms delay) (s1, s4) (5ms delay) (5ms delay) (s2, s3) (5ms delay) (5ms delay) (s2, s4)
*** Configuring hosts
h1 h2
*** Starting controller
*** Starting 4 switches
s1 s2 s3 s4
*** Starting CLI:
mininet> []

```

Figura 14: Ejecucion de script Mininet

Una vez iniciada la topología se puede acceder al cliente Python de la forma en que se muestra en la figura 15. Dicho cliente permite ver los dispositivos configurados y la información de ruteo, así como permite configurar túneles y políticas. La especificación de comandos disponibles se encuentra en [16].

```

mininet@mininet-vm:~$ cd spring-open-cli/
mininet@mininet-vm:~/spring-open-cli$ source ./workspace/ve/bin/activate
(ve)mininet@mininet-vm:~/spring-open-cli$ sudo make start-sdncon
if ! lsof -iTCP:8000 -sTCP:LISTEN >/dev/null; then \
    ( \
        cd /home/mininet/spring-open-cli/sdncon; \
        [ -d /home/mininet/spring-open-cli/workspace/ve/cassandra/data/sdncon ] || python manage.py syncdb --noinput; \
        /home/mininet/spring-open-cli/build/start-and-wait-for-port.sh -p 8000 -l /home/mininet/spring-open-cli/workspa
ce/ve/log/sdncon.log python manage.py runserver 0.0.0.0:8000; \
    ); \
fi
+++ sdncon running
(ve)mininet@mininet-vm:~/spring-open-cli$ cd cli/
(ve)mininet@mininet-vm:~/spring-open-cli/cli$ ./cli.py
version200
default controller: 127.0.0.1:8000, SDN OS 1.0 - custom version
mininet-vm> sh router
# Router DPID Router Name Router IP Router Mac Edge Router Node SID
|-----|-----|-----|-----|-----|-----|
1 00:00:00:00:00:00:01 Leaf 1 172.16.101.1/32 00:00:00:00:00:01 true 101
2 00:00:00:00:00:00:02 Leaf 2 172.16.102.1/32 00:00:00:00:00:02 true 102
3 00:00:00:00:00:00:03 Spine 1 172.16.103.1/32 00:00:00:00:00:03 false 103
4 00:00:00:00:00:00:04 Spine 2 172.16.104.1/32 00:00:00:00:00:04 false 104
mininet-vm> []

```

Figura 15: Conexión a cliente de Onos

Una vez realizados los pasos anteriores el ambiente esta pronto para ser utilizado. Se profundizara sobre dicho uso en la sección 4 con ejemplos concretos.

3.2. Kathará

Como se menciona anteriormente el proyecto Sping-Open no esta pensado para utilizarse en producción. Por dicho motivo se busco generar un escenario en el cual

se emulara un ambiente de producción para probar la aplicación Segment Routing de Onos y como se comporta en 'el mundo real'.

Se dispone además de un conjunto de pruebas para algoritmos distribuidos en topologías Fat Tree realizadas en Kathará (BGP, Open Fabric y RIFT), por lo cual implementando la solución en este entorno se podrían realizar comparaciones de performance, mensajería y otros parametros de interés. Por esta razón se decidió implantar Segment Routing sobre SDN en esta herramienta.

El framework Kathará permite emular escenarios de red utilizando Docker. El modelo se compone de tres conceptos base: El dispositivo, el dominio de colisión y el escenario de red.

- Un dispositivo es un componente virtual que actúa como un dispositivo de red (un router, servidor o switch por ejemplo). Tiene una o más interfaces de red, CPU, memoria, disco virtual y corre un sistema operativo. Se implementan con contenedores Docker.
- Un dominio de colisión es una red virtual de capa dos actuando como conexión física entre los dispositivos.
- Un escenario de red es un conjunto de dispositivos conectados por dominios de colisión. Se representa mediante un directorio que contiene la configuración de la topología y de los distintos dispositivos.

Este modelo y el hecho de utilizar contenedores Docker hace que el ambiente pueda pasar a un entorno de producción de una forma directa, salvando la configuración de red que se deba realizar. La arquitectura de Kathara así como su comparación con otros emuladores se puede consultar en [21].

El entorno se ejecutó sobre un host Ubuntu 20.04 64 bits con 8 GB de memoria RAM y CPU Intel Core i5. Por detalles sobre su instalación y dependencias se puede consultar [19]. El framework se encuentra disponible para sistemas operativos Linux basados en Debian, Windows y MacOSX por el momento.

3.2.1. Generación de Topologías Spine-Leaf

Para realizar pruebas y permitir mayor flexibilidad en el ambiente se desarrolló una herramienta que genera una topología de tipo Spine-Leaf y permite

ingresar la cantidad de Spines, Leafs y Servidores por Leaf mediante línea de comando o un archivo de configuración *config.json*. Al ejecutar la herramienta se generan los archivos de configuración necesarios para Kathará y además se genera un archivo *netconf.json* que contiene la configuración que luego utiliza Onos para configurar a los Switches con Segment Routing.

Se utilizó como base un generador de topologías Fat-Tree para los protocolos RIFT, BGP y OpenFabric. Dicho generador está desarrollado en Python por el grupo de desarrolladores de Kathará.

Para el presente trabajo específicamente se desarrollaron las clases *Spine_Leaf* (definida en *Kathara/model/SR_Onos.py*) la cual genera y conecta los objetos de la topología, *Controller* (definida en */Kathara/model/node_types/Controller.py*) para definir el objeto Controlador y *SrConfigurator* (definida en */Kathara/protocol/sr/SrConfigurator.py*) que configura los archivos *lab.conf* y *startup* para el entorno.

Se modificaron luego aquellos parámetros necesarios para el ambiente y se creó la función *sr_generate_netconf_onos* dentro de *utils.py* para generar el archivo de configuración *netconf.json* de Onos.

3.2.2. Controlador Onos

El controlador Onos se define dentro del ambiente como un dispositivo de la red el cual utiliza la imagen *onos/kathara* y está conectado a los Spine y Leaf mediante dominios de colisión en lo que denominé red de gestión, la cual no interfiere en el plano de datos, utilizando la subred *192.168.0.0/16* para este propósito.

Cada una de sus interfaces está conectada a un dispositivo que ejecuta OpenvSwitch. Además de tener una interfaz a la cual se puede acceder desde el host que ejecuta Kathará.

La imagen de Docker utilizada se basa en la imagen modificada en [22]. Esta imagen se crea a partir de la versión de onos 1.15 y modifica las aplicaciones para iniciarlas por defecto, así como permite conexión entrante para los puertos necesarios.

En el archivo *Dockerfile* dentro del directorio *Onos* en [18] se encuentra la configuración de dicha imagen. Se habilitaron en este caso las siguientes aplicaciones de Onos:

- Openflow: Necesaria para control de los dispositivos.
- Net Config Host Provider: Permite configurar los Host a partir de un archivo de configuración.
- Route Service: Necesario para que Onos calcule las rutas.
- Segment Routing: Habilita el uso de la aplicación en los switches configurados para este propósito.

Se agrego además tcpdump para analizar los paquetes Openflow entre el controlador y los nodos.

Se encontró que la versión de Onos utilizada no permite Ingeniería de tráfico en Segment Routing al no contener los comandos en la CLI necesarios para crear túneles y políticas. Por lo cual no fue posible realizar pruebas de dichas funcionalidades.

3.2.3. Dispositivos OpenvSwitch

Para implementar los Spine y Leaf como switches OpenVSwitch se utilizo la imagen base de Kathará y se le agrego OpenVswitch instalando por *apt* los paquetes *openvswitch-common* y *openvswitch-switch*. La imagen docker se definió como *kathara/ovs* y se puede encontrar en *Openvswitch/Dockerfile* en el repositorio [18].

Es necesario iniciar el servicio al iniciar la topología, por lo cual en los archivos de configuración de los nodos OpenvSwitch se ejecuta el comando:

```
/usr/share/openvswitch/scripts/ovsctl start
```

Luego cada nodo ovs se debe configurar para comunicarse con el controlador y se deben configurar sus puertos. Para ello se configuró en cada archivo startup (el cual contiene la configuración de inicio de los dispositivos en Kathará) de los nodos Spine y Leaf con los siguientes comandos:

- Creación de bridge, se especifica dirección MAC de forma que coincida con la dirección del archivo *netconfig.json* de Onos. Por convención a su vez cada bridge lleva el nombre del nodo.

```
ovs-vsctl add-br nombre_nodo --set bridge nombre_nodo
other-config:hwaddr="mac_addr"
```

- Si bien no es una configuración de ovs, se configura una ip al bridge para que sea la IP de loopback en Segment Routing. Esta IP coincide con la del archivo *netconfig.json*, luego se levanta la interfaz.

```
ip addr ip_nodo dev nombre_nodo
ip link set nombre_nodo up
```

- Se configura id del datapath, con lo cual Onos va a identificar el Switch. Si no se especifica genera un ID aleatorio, pero es necesario en este caso generarlo para que coincida con el archivo *netconf.json*.

```
ovs-vsctl set bridge nombre_nodo other-config:datapath-id=id_nodo
```

- Indicar la direccion IP y puerto del controlador. En este caso se indica la IP del host que contiene Onos en la interfaz de control y el puerto 6653.

```
ovs-vsctl set-controller nombre_nodo tcp:ip_onos:6653
```

- Configuración de puertos:

```
ovs-vsctl add-port nombre_nodo interfaz
```

En el caso de los Leaf se debe agregar una VLAN des-etiquetada en los puertos conectados a servidores:

```
ovs-vsctl add-port nombre_nodo interfaz tag=1
vlan_mode=native-untagged
```

Una vez ejecutadas las configuraciones al inicio, los switches pueden utilizar el protocolo Openflow para comunicarse con el controlador. También le enviaran información utilizando el protocolo LLDP (Link Layer Discovery Protocol).

3.2.4. Utilización del ambiente

Para generar la topología primero se debe editar el archivo *conf.json* que se encuentra en el directorio */Kathara* del proyecto [18]. En el ejemplo siguiente *k_leaf* representa la cantidad de Spines, *k_top* la cantidad de Leafs (dado que dicha configuración parte del generador de Fat-Tree), *redundancy_factor* representa la cantidad de controladores Onos (por el momento solo se puede utilizar uno),

servers_for_rack la cantidad de servidores conectados a cada Leaf y *protocol* el protocolo utilizado. En este caso *sr* representa que se va a utilizar la aplicación Segment Routing de Onos.

```
{
  "k_leaf": 2,
  "k_top": 2,
  "redundancy_factor": 1,
  "servers_for_rack": 2,
  "protocol": "sr"
}
```

Luego se debe ejecutar el generador con el comando `./main.py` en dicho directorio. Esto generara un nuevo directorio cuyo nombre varia según la topología. En este ejemplo genera un directorio llamado *spine.leaf_2_2_1.sr*. Este directorio contiene dentro un archivo *topology_info.json* que contiene la configuración del laboratorio (igual al archivo *config.json*), el archivo de configuración de Onos mencionado anteriormente *netconf.json* y un directorio *lab* donde se encuentra propiamente la configuración del laboratorio.

Dentro de *lab* se encuentra el archivo *lab.conf* que contiene la configuración del escenario de red. Se definen parámetros como que imagen de Docker van a utilizar los dispositivos, la conexión entre los mismos y si algún dispositivo es *bridged*, es decir que tiene una interfaz conectada al bridge de Docker, el cual se puede comunicar con el host.

Particularmente en este ambiente el controlador utiliza la imagen *onos/kathara*, los Leaf y los Spine utilizan la imagen *kathara/ovs*. Los servidores no especifican la imagen por lo cual por defecto usan la imagen *kathara/base*. Además el controlador tiene una interfaz bridge para poder acceder a la interfaz grafica y el cliente desde el host.

En el directorio *lab* también se encuentran los archivos *startup* de cada nodo del escenario, los cuales contienen comandos que los dispositivos ejecutan al iniciar el laboratorio. Entre ellos la configuración de interfaces de red, inicio de servicios y configuración de los mismos como en el caso de OpenVSwitch ya especificado.

Para iniciar el laboratorio se debe ejecutar el comando `sudo kathara lstart -privileged`. En ese momento se genera el directorio *shared* el cual esta presente en

todos los dispositivos y permite compartir archivos entre estos y el host.

Una vez iniciado el laboratorio se puede ingresar a la terminal de cada nodo con el comando *kathara connect nombre_nodo*. Por ejemplo para ver el log de Onos se puede acceder ejecutando *kathara connect controller* y luego dentro del controlador ejecutar *tail -f apache-karaf-3.0.8/data/log/karaf.log*.

Una vez iniciado Onos se puede acceder mediante su interfaz gráfica. Se debe identificar la ip por defecto para el bridge docker en el host, por mas información consultar [20]. La interfaz por defecto en linux tiene la ip *172.17.0.2* y la interfaz web de Onos en este caso es *http://172.17.0.2:8181/onos/ui*. Se puede acceder también al cli de Onos por SSH al puerto 8101. En ambos casos las credenciales de acceso son usuario *onos* y clave *rocks*.

Para configurar la topología en Onos se debe enviar el archivo *netconf.json* a la API REST de Onos mediante el metodo POST. En este caso se utilizó el comando *curl* para dicho propósito. Esto se debe realizar luego de iniciada la API, para comprobar si ya inicio se puede ver el log de Onos, pero alcanza con comprobar que se puede acceder a la interfaz web. Se ejecuta el comando:

```
curl --user --onos:rocks -X POST -H 'Content-Type: application/json'
http://172.17.0.2:8181/onos/v1/network/configuration/ -d @netconf.json
```

Una vez configurado Segment Routing se realizaron las pruebas detalladas en la sección 4. El ambiente dispone de herramientas como Ping, treceeroute, ifconfig y tcpdump para analizar la red. Además los servidores tienen el servicio apache inicializado esperando conexiones en el puerto 80.

Para finalizar y liberar los recursos se debe ejecutar *kathara lclean*. En caso de que el entorno no finalice correctamente se debe ejecutar *kathara wipe* para destruir los objetos.

4. Pruebas Realizadas

Se presenta en esta sección el resultado de las pruebas realizadas en ambos ambientes.

Se utilizaron como ejemplo dos topologías del tipo Spine Leaf de distintos tamaños. La primera consiste en dos Spine, dos Leaf y dos servidores por Leaf, lo que llamaré como 2x2x2. La segunda topología permite simular un entorno más cercano al que se utilizaría en producción, dentro de los recursos que se pueden disponer en este caso. Comprende 4 Spines, 4 Leafs y 1 servidor por cada Leaf a lo que llamaré 4x4x1.

Se tienen diferencias en el diseño de las soluciones dado las posibilidades que cada ambiente permite, como configuración de direcciones IP y MAC, el control que se tiene sobre los dispositivos, los comandos disponibles y las herramientas a ejecutar.

4.1. Spring-Open Mininet

Se plantean para el entorno de Mininet dos topologías. En primer lugar la topología 2x2x2 que se puede ver en la figura 16 definida en los archivos *sr-spine-leaf-2x2x2-controller.conf* para configuración de Onos, *sr-spine-leaf-2x2x2-topo.py* para creación de la topología y *sr-2x2x2-spine-leaf-script.py* para ejecución del script.

En la imagen 16 se pueden ver los nombres de cada nodo junto con su identificador de segmento (Node SID) en el caso de los Leaf y Spine. A su vez se tiene una subred de prefijo 24 para cada servidor, dado que el ambiente no permite tener más de un puerto en el mismo dominio de broadcast (si bien el prefijo podría ser mas específico se decidió este valor para hacer mas sencilla la numeración).

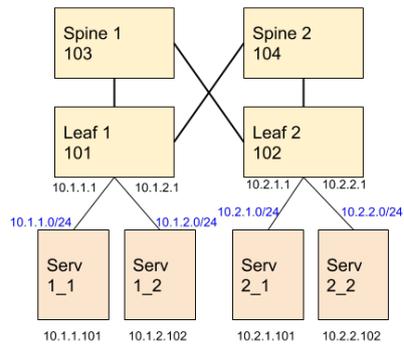


Figura 16: Spine-Leaf 2x2x2

En la imagen 17 por otro lado se tiene una topología 4x4x1. La misma se define en *sr-spine-leaf-4x4-controller.conf* para Onos, *sr-spine-leaf-4x4-topo.py* para la creación de la topología y *sr-4x4-spine-leaf-script.py* para el script de Mininet.

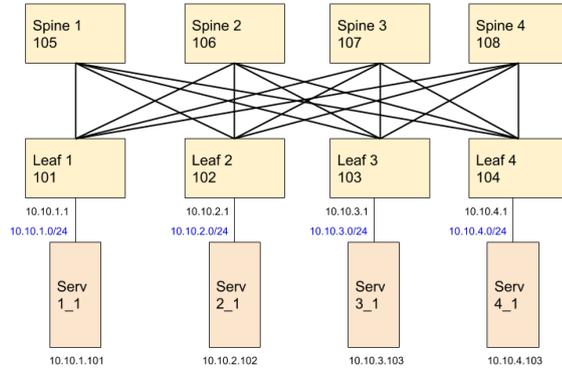


Figura 17: Spine-Leaf 4x4x1

4.1.1. Conectividad

Spine Leaf 2x2x2

En primer lugar se inicio la topología de la figura 16 y se probo la conectividad entre todos los host con el comando *pingall* de Mininet. En este ambiente los host se reconocen al intentar comunicarse con un router, por lo que esta acción ping es necesario para descubrirlos. Luego se ejecuto un ping entre el servidor *h_1.1* y el servidor *h_2.2* como se puede ver en la figura 18.

```

mininet@mininet-vm:~/mininet/custom$ sudo ./sr_2x2x2_spine_leaf-script.py
*** Creating network
*** Adding controller
*** Adding hosts:
h_1_1 h_1_2 h_2_1 h_2_2
*** Adding switches:
s1 s2 s3 s4
*** Adding links:
(5ms delay) (5ms delay) (h_1_1, s1) (5ms delay) (5ms delay) (h_1_2, s1) (5ms delay) (5ms del
5ms delay) (s1, s4) (5ms delay) (5ms delay) (s2, s3) (5ms delay) (5ms delay) (s2, s4)
*** Configuring hosts
h_1_1 h_1_2 h_2_1 h_2_2
*** Starting controller
*** Starting 4 switches
s1 s2 s3 s4
*** Starting CLI:
mininet> pingall
*** Ping: testing ping reachability
h_1_1 -> h_1_2 h_2_1 h_2_2
h_1_2 -> h_1_1 h_2_1 h_2_2
h_2_1 -> h_1_1 h_1_2 h_2_2
h_2_2 -> h_1_1 h_1_2 h_2_1
*** Results: 0% dropped (12/12 received)
mininet> h_1_1 ping h_2_2
PING 10.2.2.102 (10.2.2.102) 56(84) bytes of data.
64 bytes from 10.2.2.102: icmp_seq=1 ttl=62 time=53.3 ms
64 bytes from 10.2.2.102: icmp_seq=2 ttl=62 time=44.9 ms
64 bytes from 10.2.2.102: icmp_seq=3 ttl=62 time=44.9 ms
64 bytes from 10.2.2.102: icmp_seq=4 ttl=62 time=45.1 ms
64 bytes from 10.2.2.102: icmp_seq=5 ttl=62 time=46.1 ms

```

Figura 18: Salida de Mininet

En la figura 19 se muestra la información de los switch configurados como routers que ejecutan Segment Routing. Se puede ver el Datapath ID asignado en el archivo de configuración de Onos, el alias, las direcciones IP y MAC, la etiqueta MPLS asignada o NodeSid y el puerto con el cual reciben mensajes del controlador por la interfaz de gestión. En este ejemplo el Leaf 1 recibe los mensajes Openflow en el puerto 47151 de la interfaz de loopback de la maquina virtual.

```

mininet-vm> sh switch
# Switch DPID          Alias  Connected Since          Connected At  Type  Controller
-----
1 00:00:00:00:00:00:01 Leaf 1  Thu Jul 30 15:02:17 PDT 2020 127.0.0.1:47151 packet mininet-vm
2 00:00:00:00:00:00:02 Leaf 2  Thu Jul 30 15:02:17 PDT 2020 127.0.0.1:47154 packet mininet-vm
3 00:00:00:00:00:00:03 Spine 1  Thu Jul 30 15:02:17 PDT 2020 127.0.0.1:47153 packet mininet-vm
4 00:00:00:00:00:00:04 Spine 2  Thu Jul 30 15:02:17 PDT 2020 127.0.0.1:47152 packet mininet-vm
mininet-vm> sh router
# Router DPID          Router Name Router IP      Router Mac      Edge Router Node SID
-----
1 00:00:00:00:00:00:01 Leaf 1         172.16.101.1/32 00:00:00:00:00:01 true          101
2 00:00:00:00:00:00:02 Leaf 2         172.16.102.1/32 00:00:00:00:00:02 true          102
3 00:00:00:00:00:00:03 Spine 1        172.16.103.1/32 00:00:00:00:00:03 false         103
4 00:00:00:00:00:00:04 Spine 2        172.16.104.1/32 00:00:00:00:00:04 false         104

```

Figura 19: Informacion Switches Onos CLI

Para comprobar que el funcionamiento de Segment Routing es correcto se siguió el camino de los paquetes ICMP entre los dos servidores antes mencionados, así como las decisiones en cada router. En la figura 20 se marca la decisión que toma el Leaf 1 sobre los paquetes que tienen como ip de destino la subred $10.2.2.0/24$, a la cual pertenece el host $h_2.2$. Se indica que la instrucción a seguir se encuentra en


```

mininet-vms sh sw 00:00:00:00:00:00:04 table mpls
# Bytes Packets Dur(s) Priority MPLS Label MPLS BOS MPLS TC Instructions
-----
1 1020 10 3903 65535 101 true * {goto: {tableid: acl}, write: {pop_mpls: 0x800, dec_mw_ttl, copy_ttl_in, group: 2}}
2 0 0 3903 65535 101 false * {goto: {tableid: acl}, write: {pop_mpls: 0x8847, copy_ttl_in, group: 2, dec_mpls_ttl}}
3 1020 10 3903 65535 102 true * {goto: {tableid: acl}, write: {pop_mpls: 0x800, dec_mw_ttl, copy_ttl_in, group: 1}}
4 0 0 3903 65535 102 false * {goto: {tableid: acl}, write: {pop_mpls: 0x8847, copy_ttl_in, group: 1, dec_mpls_ttl}}
5 0 0 3903 65535 103 true * {goto: {tableid: acl}, write: {group: 3, dec_mpls_ttl}}
6 0 0 3903 65535 103 false * {goto: {tableid: acl}, write: {group: 3, dec_mpls_ttl}}
7 0 0 3903 65535 104001 true * {goto: {tableid: acl}, write: {pop_mpls: 0x800, dst_mac: 00:00:00:00:00:01, dec_mw_ttl, src_mac: 00:00:00:00:00:04, output: 1, copy_ttl_in}}
8 0 0 3903 65535 104001 false * {goto: {tableid: acl}, write: {dec_mpls_ttl, pop_mpls: 0x8847, dst_mac: 00:00:00:00:00:01, src_mac: 00:00:00:00:00:04, output: 1, copy_ttl_in}}
9 0 0 3903 65535 104002 true * {goto: {tableid: acl}, write: {pop_mpls: 0x800, dst_mac: 00:00:00:00:00:02, dec_mw_ttl, src_mac: 00:00:00:00:00:04, output: 2, copy_ttl_in}}
10 0 0 3903 65535 104002 false * {goto: {tableid: acl}, write: {dec_mpls_ttl, pop_mpls: 0x8847, dst_mac: 00:00:00:00:00:02, src_mac: 00:00:00:00:00:04, output: 2, copy_ttl_in}}
11 0 0 3904 0 * * * {goto: {tableid: acl}, write: {output: controller}}
mininet-vms sh sw 00:00:00:00:00:00:04 group 1
# Group Type Group Id Pkts Bytes Bucket Pkts Bucket Bytes Set Src Mac Set Dst Mac Push Mpls Set Bos COPY TTL Dec Mpls TTL Outport Group
-----
1 SELECT 1 10 980 10 980 00:00:00:00:00:04 00:00:00:00:00:02 2
mininet-vms

```

Figura 23: Tabla MPLS de Spine 2

Una vez el paquete ICMP llega al Leaf 2, el mismo decide según su tabla IP. Dado que el servidor *h_2_2* está directamente conectado lo envía por el puerto correspondiente. Luego el Leaf 2 responde con un paquete ICMP Echo Reply cuyo flujo es análogo al visto. Se pueden ver por más información las tablas MPLS e IP de cada router en el directorio *Pruebas/Conectividad/Mininet/2x2x2* de [18].

```

mininet-vms sh sw 00:00:00:00:00:00:02 table ip
# Bytes Packets Dur(s) Cookie Priority Dst IP Instructions
-----
1 0 0 3948 0 65535 172.16.101.1/32 {goto: {tableid: acl}, write: {group: 2}}
2 0 0 3948 0 65535 172.16.103.1/32 {goto: {tableid: acl}, write: {dec_mw_ttl, group: 7}}
3 0 0 3948 0 65535 172.16.104.1/32 {goto: {tableid: acl}, write: {dec_mw_ttl, group: 4}}
4 490 5 3944 0 65535 10.2.1.101/32 {goto: {tableid: acl}, write: {dst_mac: 00:00:00:00:02:24, src_mac: 00:00:00:00:00:02, output: 1}}
5 606 7 3773 0 65535 10.2.2.102/32 {goto: {tableid: acl}, write: {dst_mac: 00:00:00:00:22:24, src_mac: 00:00:00:00:00:02, output: 2}}
6 1666 17 3948 0 49104 10.1.1.0/24 {goto: {tableid: acl}, write: {group: 2}}
7 392 4 3948 0 49104 10.1.2.0/24 {goto: {tableid: acl}, write: {group: 2}}
8 196 2 3949 0 0 * {goto: {tableid: acl}, write: {output: controller}}

```

Figura 24: Tabla IP de Leaf 2

Se inspeccionaron los paquetes con *tcpdump* en las interfaces que conectan al Leaf 1 con los dos Spine para detectar si se utiliza ECMP y para verificar los paquetes MPLS. Se puede ver en las figuras 25 y 26 se transmiten paquetes y el tráfico desde el servidor *h_1_1* al servidor *h_2_2* contiene la etiqueta MPLS 102 del Leaf 2. Se comprobó por lo tanto que el funcionamiento en esta topología es el esperado para el algoritmo Segment Routing.

```

+-----+-----+-----+-----+
| 5 24.546655 | 10.1.1.101 | 10.2.2.102 | ICMP |
+-----+-----+-----+-----+
| 6 24.582805 | 10.2.2.102 | 10.1.1.101 | ICMP |
+-----+-----+-----+-----+
| 7 26.554291 | 10.1.1.101 | 10.2.2.102 | ICMP |
+-----+-----+-----+-----+
| 8 26.582293 | 10.2.2.102 | 10.1.1.101 | ICMP |
+-----+-----+-----+-----+

+ Frame 5: 102 bytes on wire (816 bits), 102 bytes captured (816 bits)
+ Ethernet II, Src: 00:00:00_00:00:01 (00:00:00:00:00:01), Dst: 00:00:00_00:00:03 (00:00:00:00:00:03)
+ MultiProtocol Label Switching Header, Label: 102, Exp: 0, S: 1, TTL: 63
0000 0000 0000 0110 0110 ..... = MPLS Label: 102
..... = MPLS Experimental Bits: 0
.....1 ..... = MPLS Bottom Of Label Stack: 1
..... 0011 1111 = MPLS TTL: 63
+ Internet Protocol Version 4, Src: 10.1.1.101, Dst: 10.2.2.102
+ Internet Control Message Protocol

```

Figura 25: Paquete MPLS de Leaf 1 hacia Spine 1

```

+-----+-----+-----+-----+
| 3 10.503762 | 10.1.1.101 | 10.2.2.102 | ICMP |
+-----+-----+-----+-----+
| 4 10.531782 | 10.2.2.102 | 10.1.1.101 | ICMP |
+-----+-----+-----+-----+
| 5 12.511416 | 10.1.1.101 | 10.2.2.102 | ICMP |
+-----+-----+-----+-----+
| 6 12.539676 | 10.2.2.102 | 10.1.1.101 | ICMP |
+-----+-----+-----+-----+
| 7 14.516409 | 10.1.1.101 | 10.2.2.102 | ICMP |
+-----+-----+-----+-----+
| 8 14.545293 | 10.2.2.102 | 10.1.1.101 | ICMP |
+-----+-----+-----+-----+

+ Frame 3: 102 bytes on wire (816 bits), 102 bytes captured (816 bits)
+ Ethernet II, Src: 00:00:00_00:00:01 (00:00:00:00:00:01), Dst: 00:00:00_00:00:04 (00:00:00:00:00:04)
+ MultiProtocol Label Switching Header, Label: 102, Exp: 0, S: 1, TTL: 63
0000 0000 0000 0110 0110 ..... = MPLS Label: 102
..... = MPLS Experimental Bits: 0
.....1 ..... = MPLS Bottom Of Label Stack: 1
..... 0011 1111 = MPLS TTL: 63
+ Internet Protocol Version 4, Src: 10.1.1.101, Dst: 10.2.2.102
+ Internet Control Message Protocol

```

Figura 26: Paquete MPLS de Leaf 1 hacia Spine 2

Spine Leaf 4x4x1

Se probó por otro lado la topología de la figura 17. En este caso se ejecuto el comando *pingall* y se siguieron los flujos en las tablas de los respectivos nodos. En particular se inspeccionaron las interfaces del Leaf 1 hacia cada Spine y se comprobó que los paquetes estaban etiquetados correctamente. Las capturas y las tablas de cada nodo se pueden observar con más detalle en el directorio *Pruebas/-Conectividad/Mininet/4x4x1*. En la figura 27 se puede ver la tabla ip del Leaf 1 y sus respectivos grupos. En cada caso se envía a los cuatro Spine los paquetes con la etiqueta correspondiente utilizando ECMP.

```

mininet-vm> sh switch 00:00:00:00:00:00:01 table ip
# Bytes Packets Dur(s) Cookie Priority Dst IP Instructions
-----
1 0 0 423 0 65535 172.16.106.1/32 {goto: {tableid: acl}, write: {dec_nw_ttl, group: 48}}
2 0 0 423 0 65535 172.16.105.1/32 {goto: {tableid: acl}, write: {dec_nw_ttl, group: 38}}
3 0 0 423 0 65535 172.16.107.1/32 {goto: {tableid: acl}, write: {dec_nw_ttl, group: 14}}
4 0 0 423 0 65535 172.16.108.1/32 {goto: {tableid: acl}, write: {dec_nw_ttl, group: 22}}
5 0 0 423 0 65535 172.16.102.1/32 {goto: {tableid: acl}, write: {group: 29}}
6 0 0 423 0 65535 172.16.104.1/32 {goto: {tableid: acl}, write: {group: 72}}
7 1968 20 391 0 65535 10.10.1.101/32 {goto: {tableid: acl}, write: {dst_mac: 00:00:00:00:01:01, src_mac: 00:00:00:00:00:01, output: 1}}
8 0 0 423 0 65535 172.16.103.1/32 {goto: {tableid: acl}, write: {group: 12}}
9 1078 11 423 0 49104 10.10.2.0/24 {goto: {tableid: acl}, write: {group: 29}}
10 882 9 423 0 49104 10.10.4.0/24 {goto: {tableid: acl}, write: {group: 72}}
11 980 10 423 0 49104 10.10.3.0/24 {goto: {tableid: acl}, write: {group: 12}}
12 98 1 432 0 0 * {goto: {tableid: acl}, write: {output: controller}}

mininet-vm> sh switch 00:00:00:00:00:00:01 group 29
# Group Type Group Id Pkts Bytes Bucket Pkts Bucket Bytes Set Src Mac Set Dst Mac Push Mpls Set Bos COPY TTL Dec Mpls TTL Outport Group
-----
1 SELECT 29 12 1176 3 294 00:00:00:00:01 00:00:00:00:05 102 true True True 2
2 SELECT 12 1176 3 294 00:00:00:00:01 00:00:00:00:06 102 true True True 3
3 SELECT 12 1176 3 294 00:00:00:00:01 00:00:00:00:07 102 true True True 4
4 SELECT 12 1176 3 294 00:00:00:00:01 00:00:00:00:08 102 true True True 5

mininet-vm> sh switch 00:00:00:00:00:00:01 group 12
# Group Type Group Id Pkts Bytes Bucket Pkts Bucket Bytes Set Src Mac Set Dst Mac Push Mpls Set Bos COPY TTL Dec Mpls TTL Outport Group
-----
1 SELECT 12 10 980 3 294 00:00:00:00:01 00:00:00:00:05 103 true True True 2
2 SELECT 10 980 3 294 00:00:00:00:01 00:00:00:00:06 103 true True True 3
3 SELECT 10 980 2 196 00:00:00:00:01 00:00:00:00:07 103 true True True 4
4 SELECT 10 980 2 196 00:00:00:00:01 00:00:00:00:08 103 true True True 5

mininet-vm> sh switch 00:00:00:00:00:00:01 group 72
# Group Type Group Id Pkts Bytes Bucket Pkts Bucket Bytes Set Src Mac Set Dst Mac Push Mpls Set Bos COPY TTL Dec Mpls TTL Outport Group
-----
1 SELECT 72 9 882 3 294 00:00:00:00:01 00:00:00:00:05 104 true True True 2
2 SELECT 9 882 2 196 00:00:00:00:01 00:00:00:00:06 104 true True True 3
3 SELECT 9 882 2 196 00:00:00:00:01 00:00:00:00:07 104 true True True 4
4 SELECT 9 882 2 196 00:00:00:00:01 00:00:00:00:08 104 true True True 5
mininet-vm>

```

Figura 27: Tablas y Grupos de Leaf 1

Plano de Control: OpenFlow

Se inspeccionó para la topología de la figura 16 con *tcpdump* en la interfaz *localhost* la comunicación a nivel de plano de control de los nodos, siguiendo el flujo de los paquetes Openflow. Según se puede ver en la imagen 19 el Leaf 1 recibe los mensajes del controlador en el puerto 47151.

En las figuras 28 y 29 se puede ver el intercambio de mensajes *OFPT_HELLO* con los cuales se establece la conexión, especificando la versión de OpenFlow a utilizar. Como se puede observar por los puertos de destino y origen del paquete TCP la conexión es iniciada por el Switch.

```

#EEO 47151-6633 [ACK] Se
221 15.433046 127.0.0.1 127.0.0.1 OpenFlow 74 Type: OFPT_HELLO
#EEO 47151-6633 [ACK] Se
222 15.433117 127.0.0.1 127.0.0.1 TCP 66 6633 -- 47151 [ACK] Se
#EEO 47151-6633 [ACK] Se
223 15.436614 127.0.0.1 127.0.0.1 OpenFlow 82 Type: OFPT_HELLO
#EEO 47151-6633 [ACK] Se
224 15.436969 127.0.0.1 127.0.0.1 TCP 66 47151 -- 6633 [ACK] Se
#EEO 47151-6633 [ACK] Se
225 15.437452 127.0.0.1 127.0.0.1 OpenFlow 74 Type: OFPT_FEATURES_R

#Frame 221: 74 bytes on wire (592 bits), 74 bytes captured (592 bits) on interface eth0
#Ethernet II, Src: 00:00:00:00:00:00 (00:00:00:00:00:00), Dst: 00:00:00:00:00:00 (00:00:00:00:00:00)
#Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
#Transmission Control Protocol, Src Port: 47151, Dst Port: 6633, Seq: 1, Ack: 1, Len: 8
#OpenFlow 1.3
  Version: 1.3 (0x04)
  Type: OFPT_HELLO (0)
  Length: 8

```

Figura 28: OpenFlow Hello de Leaf 1 hacia Onos

```

222 15.438117 127.0.0.1 127.0.0.1 TCP 66 6633 - 47151 [ACK] Seq=1 Ack=9 Win=4
223 15.436014 127.0.0.1 127.0.0.1 OpenFlow 82 Type: OFPT_HELLO
224 15.436069 127.0.0.1 127.0.0.1 TCP 66 47151 - 6633 [ACK] Seq=9 Ack=17 Win=
225 15.437452 127.0.0.1 127.0.0.1 OpenFlow 74 Type: OFPT_FEATURES_REQUEST

Frame 223: 82 bytes on wire (656 bits), 82 bytes captured (656 bits)
Ethernet II, Src: 00:00:00:00:00:00 (00:00:00:00:00:00), Dst: 00:00:00:00:00:00 (00:00:00:00:00:00)
Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
Transmission Control Protocol, Src Port: 6633, Dst Port: 47151, Seq: 1, Ack: 9, Len: 16
OpenFlow 1.3
Version: 1.3 (0x04)
Type: OFPT_HELLO (0)
Length: 16
Transaction ID: 4294967295
Element

```

Figura 29: OpenFlow Hello de Onos hacia Leaf 1

Luego de establecerse la conexión, como se puede ver en la figura 30, el controlador envía al switch un mensaje *OFPT_FEATURES_REQUEST* el cual recibe como respuesta en la imagen 30 información del switch, como su Datapath, el número de tablas y las funcionalidades que permite el mismo.

```

223 15.436014 127.0.0.1 127.0.0.1 OpenFlow 82 Type: OFPT_HELLO
224 15.436069 127.0.0.1 127.0.0.1 TCP 66 47151 - 6633 [ACK] Seq=9 Ack=17
225 15.437452 127.0.0.1 127.0.0.1 OpenFlow 74 Type: OFPT_FEATURES_REQUEST

Frame 225: 74 bytes on wire (592 bits), 74 bytes captured (592 bits)
Ethernet II, Src: 00:00:00:00:00:00 (00:00:00:00:00:00), Dst: 00:00:00:00:00:00 (00:00:00:00:00:00)
Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
Transmission Control Protocol, Src Port: 6633, Dst Port: 47151, Seq: 17, Ack: 9, Len: 8
OpenFlow 1.3
Version: 1.3 (0x04)
Type: OFPT_FEATURES_REQUEST (5)
Length: 8
Transaction ID: 4294967294

```

Figura 30: OpenFlow Features Request de Onos hacia Leaf 1

```

227 15.437775 127.0.0.1 127.0.0.1 OpenFlow 98 Type: OFPT_FEATURES_REPLY
228 15.437775 127.0.0.1 127.0.0.1 OpenFlow 82 Type: OFPT_MULTIPART_REQUEST

Frame 227: 98 bytes on wire (784 bits), 98 bytes captured (784 bits)
Ethernet II, Src: 00:00:00:00:00:00 (00:00:00:00:00:00), Dst: 00:00:00:00:00:00 (00:00:00:00:00:00)
Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
Transmission Control Protocol, Src Port: 47151, Dst Port: 6633, Seq: 9, Ack: 25, Len: 32
OpenFlow 1.3
Version: 1.3 (0x04)
Type: OFPT_FEATURES_REPLY (6)
Length: 32
Transaction ID: 4294967294
datapath_id: 0x0000000000000001
n_buffers: 256
n_tables: 64
auxiliary_id: 0
Pad: 0
capabilities: 0x0000004f
.....1 = OFPC_FLOW_STATS: True
.....1.1 = OFPC_TABLE_STATS: True
.....1.1.1 = OFPC_PORT_STATS: True
.....1.1.1.1 = OFPC_GROUP_STATS: True
.....1.0 = OFPC_IP_REASM: False
.....1.1.1 = OFPC_QUEUE_STATS: True
.....0 = OFPC_PORT_BLOCKED: False
Reserved: 0x00000000

```

Figura 31: OpenFlow Features Reply de Leaf 1 a Onos

Después el controlador solicita al Leaf 1 que envíe la descripción de sus puertos, con un mensaje *OFPT_MULTIPART_REQUEST* del tipo *OFMP_PORT_DESC* (figura 32). El Leaf retorna la información de sus puertos como se puede ver en la imagen 33, en el cual se indican el número de cada puerto, su dirección MAC y configuración entre otros datos.

```

228 15.437775 127.0.0.1 127.0.0.1 OpenFlow 82 Type: OFPT_MULTIPART_REQUEST, OFPMP_PORT_DESC
229 15.442782 127.0.0.1 127.0.0.1 OpenFlow 492 Type: OFPT_MULTIPART_REPLY, OFPMP_PORT_DESC
230 15.444144 127.0.0.1 127.0.0.1 OpenFlow 94 Type: OFPT_GET_CONFIG_REQUEST
231 15.444780 127.0.0.1 127.0.0.1 OpenFlow 74 Type: OFPT_BARRIER_REPLY
232 15.444961 127.0.0.1 127.0.0.1 OpenFlow 78 Type: OFPT_GET_CONFIG_REPLY
233 15.445231 127.0.0.1 127.0.0.1 TCP 66 6633 - 47151 [ACK] Seq=69 Ack=397 Win=45856 Len=0
234 15.445625 127.0.0.1 127.0.0.1 OpenFlow 82 Type: OFPT_MULTIPART_REQUEST, OFPMP_DESC

Frame 228: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: 00:00:00:00:00:00 (00:00:00:00:00:00), Dst: 00:00:00:00:00:00 (00:00:00:00:00:00)
Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
Transmission Control Protocol, Src Port: 6633, Dst Port: 47151, Seq: 25, Ack: 41, Len: 16
OpenFlow 1.3
Version: 1.3 (0x04)
Type: OFPT_MULTIPART_REQUEST (18)
Length: 16
Transaction ID: 4294967293
Type: OFPMP_PORT_DESC (13)
Flags: 0x0000
Pad: 00000000

```

Figura 32: OpenFlow Multipart Request de Onos a Leaf 1

```

229 15.442782 127.0.0.1 127.0.0.1 OpenFlow 492 Type: OFPT_MULTIPART_REPLY, OFPMP_PORT_DESC
230 15.444144 127.0.0.1 127.0.0.1 OpenFlow 94 Type: OFPT_GET_CONFIG_REQUEST
231 15.444780 127.0.0.1 127.0.0.1 OpenFlow 74 Type: OFPT_BARRIER_REPLY
232 15.444961 127.0.0.1 127.0.0.1 OpenFlow 78 Type: OFPT_GET_CONFIG_REPLY
233 15.445231 127.0.0.1 127.0.0.1 TCP 66 6633 - 47151 [ACK] Seq=69 Ack=397 Win=45856 Len=0
234 15.445625 127.0.0.1 127.0.0.1 OpenFlow 82 Type: OFPT_MULTIPART_REQUEST, OFPMP_DESC

Frame 229: 492 bytes on wire (3216 bits), 492 bytes captured (3216 bits) on interface 0
Ethernet II, Src: 00:00:00:00:00:00 (00:00:00:00:00:00), Dst: 00:00:00:00:00:00 (00:00:00:00:00:00)
Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
Transmission Control Protocol, Src Port: 47151, Dst Port: 6633, Seq: 41, Ack: 41, Len: 336
OpenFlow 1.3
Version: 1.3 (0x04)
Type: OFPT_MULTIPART_REPLY (19)
Length: 336
Transaction ID: 4294967293
Type: OFPMP_PORT_DESC (13)
Flags: 0x0000
Pad: 00000000
Port
Port no: 1
Pad: 00000000
Mac addr: d0:e0:35:e4:e5:e0 (d0:e0:35:e4:e5:e0)
Pad: 0000
Name: s1-eth1
Config: 0x00000000
.....0 = OFPPC_PORT_DOWN: False
.....0 = OFPPC_NO_RECV: False
.....0 = OFPPC_NO_FWD: False
.....0 = OFPPC_NO_PACKET_IN: False
State: 0x00000004
Current: 0x00000000

```

Figura 33: OpenFlow Multipart Reply Port Description de Leaf 1 a Onos

En la figura 34 se puede observar que el controlador envía un mensaje *OFPT_FLOW_MOD* con el cual modifica los flujos del Leaf, instalando los flujos generados por Segment Routing. Luego en la figura 35 se modifican los grupos del Leaf con el mensaje *OFPT_GROUP_MOD*, con las acciones a tomar para cada grupo.

No.	Time	Source	Destination	Protocol	Length	Info
899	19.810067	127.0.0.1	127.0.0.1	TCP	66	47151 → 6633 [ACK] Seq=4332 Ack=10720
900	19.823606	127.0.0.1	127.0.0.1	OpenFlow	286	Type: OFPT_PACKET_IN
901	19.845467	127.0.0.1	127.0.0.1	OpenFlow	282	Type: OFPT_FLOW_MOD
902	19.848169	127.0.0.1	127.0.0.1	OpenFlow	294	Type: OFPT_PACKET_OUT
903	19.848279	127.0.0.1	127.0.0.1	TCP	66	47151 → 6633 [ACK] Seq=4552 Ack=10720
904	19.858980	127.0.0.1	127.0.0.1	OpenFlow	230	Type: OFPT_PACKET_IN
905	19.861283	127.0.0.1	127.0.0.1	OpenFlow	282	Type: OFPT_FLOW_MOD
906	19.864885	127.0.0.1	127.0.0.1	OpenFlow	148	Type: OFPT_PACKET_OUT

```

Frame 904: 230 bytes on wire (1840 bits), 230 bytes captured (1840 bits) on interface 0
Ethernet II, Src: 00:00:00:00:00:00 (00:00:00:00:00:00), Dst: 00:00:00:00:00:00 (00:00:00:00:00:00)
Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
Transmission Control Protocol, Src Port: 47151, Dst Port: 6633, Seq: 4552, Ack: 10584, Len: 164
OpenFlow 1.3
  Version: 1.3 (0x04)
  Type: OFPT_PACKET_IN (10)
  Length: 164
  Transaction ID: 0
  Buffer ID: OFP_NO_BUFFER (4294967295)
  Total length: 42
  Reason: OFPR_NO_MATCH (0)
  Table ID: 1
  Cookie: 0x0000000000000000
  Match
    Type: OFPMT_OXM (1)
    Length: 92
    OXM field
    OXM field
    OXM field
    OXM field
    OXM field
    Class: OFPXMC_OPENFLOW_BASIC (0x0000)
    0011 000. = Field: OFPXMT_OFB_ARP_SHA (24)
    .... .0 = Has mask: False
    Length: 6
    Value: 00:00:00:00:11:02 (00:00:00:00:11:02)
    OXM field
    OXM field
    OXM field
    OXM field
    Class: OFPXMC_OPENFLOW_BASIC (0x0000)
    0010 111. = Field: OFPXMT_OFB_ARP_TPA (23)
    .... .0 = Has mask: False
    Length: 4
    Value: 10.1.2.1
  Pad: 00000000
  
```

Figura 36: OpenFlow Packet IN ARP

No.	Time	Source	Destination	Protocol	Length	Info
899	19.810067	127.0.0.1	127.0.0.1	TCP	66	47151 → 6633 [ACK] Seq=4332
900	19.823606	127.0.0.1	127.0.0.1	OpenFlow	286	Type: OFPT_PACKET_IN
901	19.845467	127.0.0.1	127.0.0.1	OpenFlow	282	Type: OFPT_FLOW_MOD
902	19.848169	127.0.0.1	127.0.0.1	OpenFlow	294	Type: OFPT_PACKET_OUT
903	19.848279	127.0.0.1	127.0.0.1	TCP	66	47151 → 6633 [ACK] Seq=4552
904	19.858980	127.0.0.1	127.0.0.1	OpenFlow	230	Type: OFPT_PACKET_IN
905	19.861283	127.0.0.1	127.0.0.1	OpenFlow	282	Type: OFPT_FLOW_MOD
906	19.864885	127.0.0.1	127.0.0.1	OpenFlow	148	Type: OFPT_PACKET_OUT

```

Frame 906: 148 bytes on wire (1184 bits), 148 bytes captured (1184 bits) on interface 0
Ethernet II, Src: 00:00:00:00:00:00 (00:00:00:00:00:00), Dst: 00:00:00:00:00:00 (00:00:00:00:00:00)
Internet Protocol Version 4, Src: 127.0.0.1, Dst: 127.0.0.1
Transmission Control Protocol, Src Port: 6633, Dst Port: 47151, Seq: 10720, Ack: 4716, Len: 82
OpenFlow 1.3
  Version: 1.3 (0x04)
  Type: OFPT_PACKET_OUT (13)
  Length: 82
  Transaction ID: 1948
  Buffer ID: OFP_NO_BUFFER (4294967295)
  In port: OFPP_ANY (4294967295)
  Actions length: 16
  Pad: 000000000000
  Action
    Type: OFPAT_OUTPUT (0)
    Length: 16
    Port: 2
    Max length: 32767
    Pad: 000000000000
  Data
    Ethernet II, Src: 00:00:00:00:00:01 (00:00:00:00:00:01), Dst: 00:00:00:00:11:02 (00:00:00:00:11:02)
    Address Resolution Protocol (reply)
      Hardware type: Ethernet (1)
      Protocol type: IPv4 (0x0800)
      Hardware size: 6
      Protocol size: 4
      Opcode: reply (2)
      Sender MAC address: 00:00:00:00:00:01 (00:00:00:00:00:01)
      Sender IP address: 10.1.2.1
      Target MAC address: 00:00:00:00:11:02 (00:00:00:00:11:02)
      Target IP address: 10.1.2.102
  
```

Figura 37: OpenFlow Packet Out ARP

4.1.2. Falla de Enlaces

Se probó que ocurre ante la caída de un link entre un Spine y un Leaf. La aplicación Segment Routing debería actualizar los flujos de los nodos para que no se pierda la conectividad siempre que se siga teniendo un camino posible entre dos servidores.

Spine Leaf 2x2x2

En el primer caso se bajó la interfaz que conecta al Spine 1 con el Leaf 2, como se puede ver en la figura 38, y se comprobó el cambio en el Leaf 1, el cual en teoría debería modificar su flujo para que se envíen los paquetes al Spine 2.

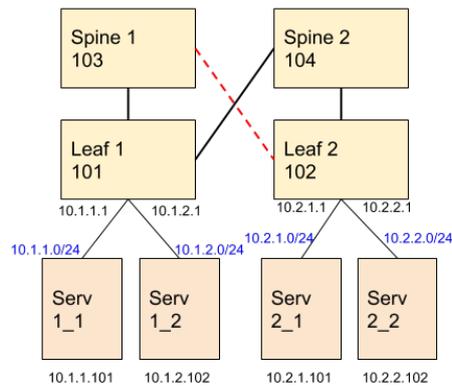


Figura 38: Falla de Link

En la imagen 39 se puede observar marcado en verde el estado de los flujos del Leaf 1 antes de bajar la interfaz y en azul luego de la falla. Se puede ver como en la tabla IP cambia el grupo 2, el cual envía el paquete MPLS con etiqueta 102 a ambos Spine, por el grupo 5 que solo envía el paquete MPLS al Spine 2. No se perdieron paquetes durante ni luego del cambio.

1 0.000000	96:9b:14:e9:1a:ca	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:03	PC/0001_120
2 0.000000	8a:0f:a7:da:0f:19	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120
3 15.022378	96:9b:14:e9:1a:ca	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:03	PC/0001_120
4 15.024976	8a:0f:a7:da:0f:19	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120
5 17.585800	10.1.1.101	10.2.2.102	TCP	70	38472 → 80 [ACK] Seq=1 Min=58 Len=0 TSval=17727541 TSecr=17727535	
6 17.631780	10.2.2.102	10.1.1.101	TCP	80	[TCP ACKed unseen segment] 80 → 38472 [ACK] Seq=1 Ack=109 Win=57 Len=0 TSval=17727540 TSecr=17727541	
7 17.631780	10.1.1.101	10.2.2.102	TCP	70	[TCP ACKed unseen segment] [TCP Previous segment not captured] 38472 → 80 [ACK] Seq=109 Ack=18 Min=58 Len=0	
8 30.039922	96:9b:14:e9:1a:ca	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:03	PC/0001_120
9 30.040406	8a:0f:a7:da:0f:19	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120
10 45.075488	96:9b:14:e9:1a:ca	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:03	PC/0001_120
11 45.075273	8a:0f:a7:da:0f:19	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120
12 45.020900	10.1.1.101	10.2.2.102	TCP	82	[TCP ACKed unseen segment] 38472 → 80 [FIN, ACK] Seq=109 Ack=18 Min=63 Len=0 TSval=17734546 TSecr=17727541	
13 45.020900	10.2.2.102	10.1.1.101	TCP	80	[TCP ACKed unseen segment] 80 → 38472 [ACK] Seq=109 Ack=18 Min=63 Len=0 TSval=17734546 TSecr=17727541	
14 60.004904	8a:0f:a7:da:0f:19	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120

Figura 42: Trafico TCP entre Leaf 1 y Spine 1

1 0.000000	d2:d9:83:72:f4:e6	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:04	PC/0001_120
2 0.000000	ae:d2:be:32:2f:17	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0004_120
3 15.022717	d2:d9:83:72:f4:e6	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:04	PC/0001_120
4 15.024039	ae:d2:be:32:2f:17	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0004_120
5 17.541806	10.1.1.101	10.2.2.102	TCP	70	38472 → 80 [SYN] Seq=0 Min=29200 Len=0 MSS=1460 SACK_PERM=1 TSval=17727530 TSecr=0 WS=512	
6 17.567903	10.2.2.102	10.1.1.101	TCP	74	80 → 38472 [SYN, ACK] Seq=0 Ack=1 Min=28960 Len=0 MSS=1460 SACK_PERM=1 TSval=17727535 TSecr=17727	
7 17.568511	10.1.1.101	10.2.2.102	HTTP	178	GET / HTTP/1.1	
8 17.614898	10.2.2.102	10.1.1.101	TCP	83	80 → 38472 [PSH, ACK] Seq=1 Ack=109 Win=29154 Len=17 TSval=17727547 TSecr=17727541 [TCP segment of a reassembled PDU]	
9 17.614898	10.1.1.101	10.2.2.102	TCP	518	[TCP ACKed unseen segment] [TCP Previous segment not captured] 38472 → 80 [ACK] Seq=109 Ack=110 Min=22250 Len=0 TSval=17727547 TSecr=17727541	
10 17.631159	10.1.1.101	10.2.2.102	TCP	82	38472 → 80 [ACK] Seq=109 Ack=18 Min=32250 Len=0 TSval=17727552 TSecr=17727547 SLE=1466 SRE=1917	
11 30.040175	d2:d9:83:72:f4:e6	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:04	PC/0001_120
12 30.042414	ae:d2:be:32:2f:17	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0004_120
13 45.075539	d2:d9:83:72:f4:e6	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:04	PC/0001_120
14 45.075810	ae:d2:be:32:2f:17	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0004_120
15 45.045392	10.2.2.102	10.1.1.101	TCP	80	[TCP ACKed unseen segment] 80 → 38472 [ACK] Seq=1017 Ack=110 Min=29184 Len=0 TSval=17734554 TSecr=17727541	
16 60.002956	d2:d9:83:72:f4:e6	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:04	PC/0001_120
17 60.005192	ae:d2:be:32:2f:17	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0004_120

Figura 43: Trafico TCP entre Leaf 1 y Spine 2

Spine Leaf 4x4x1

En este caso se puede ver en las figuras 44 a 47 que el tráfico se distribuye entre los cuatro Spines y se observan también pedidos de retransmisión, además de tener paquetes TCP del mismo flujo yendo por distintos caminos.

18 60.144270	2a:a4:ee:25:05:06	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0002_120
19 75.159884	96:78:10:05:f9:47	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:05	PC/0001_120
20 75.163945	2a:a4:ee:25:05:06	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0002_120
13 75.579722	10.2.2.102	10.1.1.101	TCP	74	54722 → 80 [SYN] Seq=0 Min=20000 Len=0 MSS=1460 SACK_PERM=1 TSval=18499186 TSecr=0 WS=512	
14 75.596145	10.1.1.101	10.2.2.102	TCP	78	80 → 54722 [SYN, ACK] Seq=0 Ack=1 Min=28960 Len=0 MSS=1460 SACK_PERM=1 TSval=18499192 TSecr=18499186 WS=512	
15 75.642448	10.1.1.101	10.2.2.102	TCP	520	[TCP ACKed unseen segment] [TCP Previous segment not captured] 80 → 54722 [FIN, PSH, ACK] Seq=1406 Ack=110 Min=29184 Len=0 TSval=18499197 TSecr=18499192	
16 75.678680	10.2.2.102	10.1.1.101	TCP	70	[TCP ACKed unseen segment] [TCP Previous segment not captured] 54722 → 80 [ACK] Seq=110 Ack=10 Min=32250 Len=0 TSval=18499197 TSecr=18499192	
18 90.181785	2a:a4:ee:25:05:06	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0002_120

Figura 44: Trafico TCP entre Leaf 1 y Spine 1

8 45.124721	1a:f9:18:c5:50:fa	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:06	PC/0001_120
9 60.140806	1a:f9:18:c5:50:fa	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:06	PC/0001_120
10 60.140344	5a:a1:d9:ab:00:77	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120
11 75.163364	1a:f9:18:c5:50:fa	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:06	PC/0001_120
12 75.165476	5a:a1:d9:ab:00:77	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120
13 75.625518	10.2.2.102	10.1.1.101	TCP	66	54722 → 80 [ACK] Seq=1 Ack=1 Min=58 Len=0 TSval=18499197 TSecr=18499192	
14 75.625518	10.1.1.101	10.2.2.102	TCP	68	[TCP ACKed unseen segment] [TCP Previous segment not captured] 80 → 54722 [ACK] Seq=1 Ack=1 Min=58 Len=0 TSval=18499203 TSecr=18499197	
15 80.134822	1a:f9:18:c5:50:fa	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:06	PC/0001_120
16 90.130609	5a:a1:d9:ab:00:77	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120
17 105.231938	1a:f9:18:c5:50:fa	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:06	PC/0001_120
18 105.242332	5a:a1:d9:ab:00:77	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120
19 120.239050	1a:f9:18:c5:50:fa	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:06	PC/0001_120
20 120.239264	5a:a1:d9:ab:00:77	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0003_120
21 123.698847	10.2.2.102	10.1.1.101	TCP	78	[TCP ACKed unseen segment] [TCP Previous segment not captured] 54722 → 80 [FIN, ACK] Seq=110 Ack=18 Min=63 Len=0 TSval=18499203 TSecr=18499197	
22 123.720298	10.1.1.101	10.2.2.102	TCP	70	[TCP ACKed unseen segment] [TCP Previous segment not captured] 80 → 54722 [ACK] Seq=1917 Ack=111 Win=57 Len=0 TSval=18499203 TSecr=18499197	

Figura 45: Trafico TCP entre Leaf 1 y Spine 2

5 30.047526	e0:f2:2b:a3:49:0d	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0004_120
6 30.050404	3e:09:d3:50:e9:28	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:07	PC/0001_120
7 45.065712	3e:09:d3:50:e9:28	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:07	PC/0001_120
8 45.065905	e0:f2:2b:a3:49:0d	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0004_120
9 45.526047	10.2.2.102	10.1.1.101	HTTP	375	GET / HTTP/1.1	
10 45.543638	10.1.1.101	10.2.2.102	TCP	87	80 → 54722 [PSH, ACK] Seq=1 Ack=110 Min=57 Len=17 TSval=18499203 TSecr=18499197 [TCP segment of a reassembled PDU]	
11 60.082369	3e:09:d3:50:e9:28	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:07	PC/0001_120
12 60.090975	e0:f2:2b:a3:49:0d	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0004_120
13 75.115899	3e:09:d3:50:e9:28	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:07	PC/0001_120
14 75.145484	e0:f2:2b:a3:49:0d	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0004_120

Figura 46: Trafico TCP entre Leaf 1 y Spine 3

2 0.004513	6a:80:66:27:7e:b0	LLDP_Multicast	LLDP	67 MA/00:00:00:00:00:01	PC/0005 120
3 15.019764	8e:7d:46:ed:72:eb	LLDP_Multicast	LLDP	67 MA/00:00:00:00:00:00	PC/0001 120
4 15.023438	6a:80:66:27:7e:b0	LLDP_Multicast	LLDP	67 MA/00:00:00:00:00:01	PC/0005 120
5 15.530100	10.10.2.102	10.10.1.101	TCP	66 54722 -- 89 [ACK] Seq=1 Ack=1 Win=58 Len=0	TSval=18490209 TSecr=18490203
6 30.048554	8e:7d:46:ed:72:eb	LLDP_Multicast	LLDP	67 MA/00:00:00:00:00:00	PC/0001 120
7 30.056704	6a:80:66:27:7e:b0	LLDP_Multicast	LLDP	67 MA/00:00:00:00:00:01	PC/0005 120
8 45.073245	8e:7d:46:ed:72:eb	LLDP_Multicast	LLDP	67 MA/00:00:00:00:00:00	PC/0001 120
9 45.100409	6a:80:66:27:7e:b0	LLDP_Multicast	LLDP	67 MA/00:00:00:00:00:01	PC/0005 120
10 60.097635	8e:7d:46:ed:72:eb	LLDP_Multicast	LLDP	67 MA/00:00:00:00:00:00	PC/0001 120
11 60.111077	6a:80:66:27:7e:b0	LLDP_Multicast	LLDP	67 MA/00:00:00:00:00:01	PC/0005 120

Figura 47: Trafico TCP entre Leaf 1 y Spine 4

Este problema es común en ECMP y en los algoritmos de distribución de carga en general. Es importante por lo tanto tener en cuenta este tipo de problemas cuando se utilizan protocolos sensibles a la conexión.

4.1.4. Ingeniería de Tráfico

Se realizaron pruebas introduciendo túneles y políticas aplicadas a los mismos para cambiar el comportamiento normal de ECMP con Shortest Path First por un comportamiento específico deseado.

Para ello se utilizo el cliente de Onos con la opción de configuración habilitada, se crearon los túneles necesarios y luego se le aplicaron políticas. Se comprobó el comportamiento del protocolo ejecutando un ping entre servidores a los cuales aplicaba alguna política.

La configuración de un túnel, por ejemplo el túnel *azul* de la figura 49 se realiza de la siguiente forma:

```
mininet-vm> enable
mininet-vm# configure
mininet-vm(config)# tunnel azul
mininet-vm(config-tunnel)# node 101
mininet-vm(config-tunnel)# node 103
mininet-vm(config-tunnel)# node 102
mininet-vm(config-tunnel)# exit
```

Para crear una política se configura una subred de origen, una subred de destino, una prioridad y el túnel al cual se asigna. Por ejemplo para la política *azul* de la figura 49 se ejecutaron los siguientes comandos.

```
mininet-vm(config)# policy azul policy-type tunnel-flow
mininet-vm(config-policy)# flow-entry ip 10.1.1.0/24 10.2.1.0/24
mininet-vm(config-policy)# tunnel azul
```

```
mininet-vm(config-policy)# priority 1000
mininet-vm(config-policy)# exit
```

Spine Leaf 2x2x2

Se crearon para esta topología cuatro túneles, para tener dos túneles en cada sentido y obligar a que todo el tráfico entre los servidores *h_1.1* y *h_2.1* vaya por el camino marcado en azul de la figura 48, mientras que el tráfico entre los servidores *h_1.2* y *h_2.2* vaya por el camino marcado en verde. Todo el tráfico no definido en las políticas debería utilizar el algoritmo SPF.

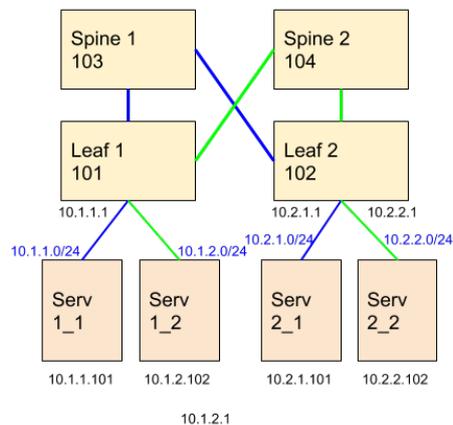


Figura 48: Ingeniería de tráfico en Spine Leaf 2x2x2

En la figura 49 se pueden observar los túneles y políticas creados. Los túneles *azul* y *azul-reverso* marcan el camino azul de la figura 48 y se puede notar que el recorrido de etiquetas MPLS es el mismo pero en distintos sentidos. Lo mismo ocurre para los túneles *verde* y *verde-reverso* que marcan el camino en verde.

```

mininet-vm(config)# sh tunnel
# Id Policies Tunnel Path [Head-->Tail] Label Stack [Outer-->Inner]
-----
1 azul azul [101, 103, 102] [[102]]
2 azul-rev azul-reverso [102, 103, 101] [[101]]
3 verde verde [101, 104, 102] [[102]]
4 verde-rev verde-reverso [102, 104, 101] [[101]]
mininet-vm(config)# sh policy
# Policy Id Policy Type Priority Dst Mac Src Mac Ether Type Dst IP IP Protocol Src IP Dst TcpPort Src TcpPort Tunnel Used
-----
1 azul TUNNEL_FLOW 1000 * * 0x2048 10.2.1.0/24 * 10.1.1.0/24 * * azul
2 azul-reverso TUNNEL_FLOW 1000 * * 0x2048 10.1.1.0/24 * 10.2.1.0/24 * * azul-rev
3 verde TUNNEL_FLOW 1000 * * 0x2048 10.2.2.0/24 * 10.1.2.0/24 * * verde
4 verde-reverso TUNNEL_FLOW 1000 * * 0x2048 10.1.2.0/24 * 10.2.2.0/24 * * verde-rev

```

Figura 49: Túneles y Políticas

Las figuras 50 y 51 muestran las tablas de flujo de ACL de los Leaf, junto con los grupos seleccionados en cada caso, marcando por su color el tunnel correspondiente. Se puede ver como se aplica a cada Leaf la política correspondiente según el sentido de la conexión. Se ejecutaron ping entre los host *h_1_1* y *h_2_1* y entre los host *h_1_2* y *h_2_2* en los cuales se comprobó que el tráfico iba por el camino correcto y no se estaba utilizando SPF para estos casos.

```

mininet-vm(config)# sh sw 00:00:00:00:00:00:01 table acl
# Bytes Packets Dur(s) Cookie Priority In Port Src MAC Dst MAC EthType Src IP Dst IP Protocol Src Port Dst Port Instructions
-----
1 30968 316 670 0 1000 * * * ip(0x800) 10.1.1.0/8 10.2.1.0/24 * * * {clear actions, write: {group: 11}}
2 5880 60 208 0 1000 * * * ip(0x800) 10.1.2.0/8 10.2.2.0/24 * * * {clear actions, write: {group: 12}}
3 92318 935 1472 0 0 * * *
mininet-vm(config)# sh sw 00:00:00:00:00:00:01 group
# Group Type Group Id Pkts Bytes Bucket Pkts Bucket Bytes Set Src Mac Set Dst Mac Push Mpls Set Bos COPY TTL Dec Mpls TTL Outport Group
-----
1 SELECT 1 0 0 0 0 00:00:00:00:01 00:00:00:00:04 103 true True True 4
2 SELECT 2 0 0 0 0 00:00:00:00:01 00:00:00:00:04
3 SELECT 3 172 16856 172 16856 00:00:00:00:01 00:00:00:00:03
4 SELECT 4 0 0 0 0 00:00:00:00:01 00:00:00:00:04 102 true True True 4
5 SELECT 5 212 20776 106 10388 00:00:00:00:01 00:00:00:00:03 102 true True True 3
6 SELECT 6 212 20776 106 10388 00:00:00:00:01 00:00:00:00:04 102 true True True 4
7 SELECT 7 0 0 0 0 00:00:00:00:01 00:00:00:00:03 104 true True True 3
8 SELECT 8 0 0 0 0 00:00:00:00:01 00:00:00:00:03 102 true True True 3
9 SELECT 9 0 0 0 0 00:00:00:00:01 00:00:00:00:03
10 SELECT 10 0 0 0 0 00:00:00:00:01 00:00:00:00:04
11 SELECT 11 316 30968 316 30968 00:00:00:00:01 00:00:00:00:03 102 true True True 3
12 SELECT 12 60 5880 60 5880 00:00:00:00:01 00:00:00:00:04 102 true True True 4

```

Figura 50: Tabla ACL de Leaf 1 y grupos correspondientes

```

mininet-vm(config)# sh sw 00:00:00:00:00:00:02 table acl
# Bytes Packets Dur(s) Cookie Priority In Port Src MAC Dst MAC EthType Src IP Dst IP Protocol Src Port Dst Port Instructions
-----
1 10388 106 475 0 1000 * * * ip(0x800) 10.2.1.0/8 10.1.1.0/24 * * * {clear actions, write: {group: 10}}
2 5880 60 145 0 1000 * * * ip(0x800) 10.2.2.0/8 10.1.2.0/24 * * * {clear actions, write: {group: 11}}
3 124958 1275 1488 0 0 * * *
mininet-vm(config)# sh sw 00:00:00:00:00:00:02 group
# Group Type Group Id Pkts Bytes Bucket Pkts Bucket Bytes Set Src Mac Set Dst Mac Push Mpls Set Bos COPY TTL Dec Mpls TTL Outport Group
-----
1 SELECT 1 0 0 0 0 00:00:00:00:02 00:00:00:00:04 103 true True True 4
2 SELECT 2 0 0 0 0 00:00:00:00:02 00:00:00:00:04
3 SELECT 3 0 0 0 0 00:00:00:00:02 00:00:00:00:03
4 SELECT 4 381 37338 191 18718 00:00:00:00:02 00:00:00:00:03 101 true True True 4
5 SELECT 5 381 37338 190 18620 00:00:00:00:02 00:00:00:00:04 101 true True True 4
6 SELECT 6 0 0 0 0 00:00:00:00:02 00:00:00:00:03 104 true True True 3
7 SELECT 7 0 0 0 0 00:00:00:00:02 00:00:00:00:03 101 true True True 3
8 SELECT 8 0 0 0 0 00:00:00:00:02 00:00:00:00:04 101 true True True 4
9 SELECT 9 0 0 0 0 00:00:00:00:02 00:00:00:00:03
10 SELECT 10 106 10388 106 10388 00:00:00:00:02 00:00:00:00:03 101 true True True 3
11 SELECT 11 60 5880 60 5880 00:00:00:00:02 00:00:00:00:04 101 true True True 4

```

Figura 51: Tabla ACL de Leaf 2 y grupos correspondientes

Luego de ver el comportamiento de Segment Routing aplicando dichas políticas, se vio que podría ser una solución para el problema con el tráfico TCP y ECMP

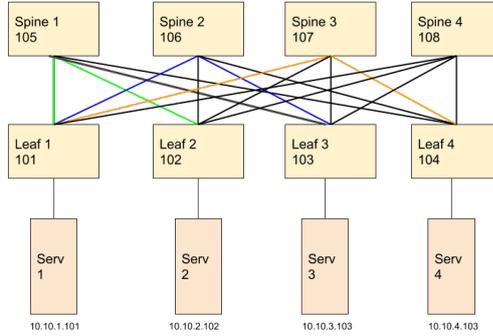


Figura 54: Ingeniería de tráfico en Spine Leaf 4x4x1

```
mininet-vm(config)# sh tunnel
# Id      Policies      Tunnel Path [Head-->Tail] Label Stack [Outer-->Inner]
-----
1 azul    azul          [101, 106, 103]      [[103]]
2 azul-rev azul-reverso  [103, 106, 101]      [[101]]
3 naranja naranja     [101, 107, 104]      [[104]]
4 naranja-rev naranja-rev [104, 107, 101]      [[101]]
5 verde  verde        [101, 105, 102]      [[102]]
6 verde-rev verde-rev    [102, 105, 101]      [[101]]
mininet-vm(config)# sh policy
# Policy Id Policy Type Priority Dst Mac Src Mac Ether Type Dst IP IP Protocol Src IP Dst TcpPort Src TcpPort Tunnel Used
-----
1 azul      TUNNEL_FLOW 1000 * * 0x2048 10.10.3.0/24 * 10.10.1.0/24 * * azul
2 azul-reverso TUNNEL_FLOW 1000 * * 0x2048 10.10.1.0/24 * 10.10.3.0/24 * * azul-rev
3 naranja  TUNNEL_FLOW 1000 * * 0x2048 10.10.4.0/24 * 10.10.1.0/24 * * naranja
4 naranja-rev TUNNEL_FLOW 1000 * * 0x2048 10.10.1.0/24 * 10.10.4.0/24 * * naranja-rev
5 verde    TUNNEL_FLOW 1000 * * 0x2048 10.10.2.0/24 * 10.10.1.0/24 * * verde
6 verde-rev TUNNEL_FLOW 1000 * * 0x2048 10.10.1.0/24 * 10.10.2.0/24 * * verde-rev
```

Figura 55: Túneles y Políticas

Se puede observar en la imagen 56 la tabla de flujos ACL del Leaf 1 junto con los grupos correspondientes a cada túnel. Se ejecuto ping desde el host *h_1.1* al resto de los host y se puede identificar la cantidad de paquetes ICMP que se enviaron por cada grupo. Las tablas de las figuras 57, 58 y 59 muestran la política inversa y el grupo utilizado en cada caso.

1 0.000000	96:70:10:05:fd:47	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:05	PC/0001	120
2 0.018202	2a:a4:ee:25:06:06	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0002	120
3 15.052510	96:70:10:05:fd:47	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:05	PC/0001	120
4 15.060296	2a:a4:ee:25:06:06	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0002	120
5 23.213322	10.10.2.102	10.10.1.101	TCP	74	55934 → 80 [SYN, Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM=1 TSval=18873040 TSecr=0 WS=512		
6 23.238901	10.10.1.101	10.10.2.102	TCP	78	80 → 55934 [SYN, ACK] Seq=0 Ack=1 Win=28960 Len=0 MSS=1460 SACK_PERM=1 TSval=18873047 TSecr=18873049 WS=512		
7 23.271701	10.10.2.102	10.10.1.101	TCP	66	55934 → 80 [ACK] Seq=1 Ack=1 Win=29696 Len=0 TSval=18873053 TSecr=18873047		
8 23.272020	10.10.2.102	10.10.1.101	HTTP	176	GET / HTTP/1.1		
9 23.288709	10.10.1.101	10.10.2.102	HTTP	70	80 → 55934 [ACK] Seq=1 Ack=110 Win=29184 Len=0 TSval=18873061 TSecr=18873053		
10 23.295207	10.10.1.101	10.10.2.102	TCP	87	80 → 55934 [RST, ACK] Seq=1 Ack=110 Win=29184 Len=0 TSval=18873062 TSecr=18873053 [TCP segment of a reassembled PDU]		
11 23.295498	10.10.1.101	10.10.2.102	HTTP	520	[TCP Previous segment not captured] Continuation		
12 23.323561	10.10.2.102	10.10.1.101	TCP	66	55934 → 80 [ACK] Seq=110 Ack=18 Win=29096 Len=0 TSval=18873068 TSecr=18873062		
13 23.323828	10.10.2.102	10.10.1.101	TCP	78	[TCP Window Update] 55934 → 80 [ACK] Seq=110 Ack=18 Win=32256 Len=0 TSval=18873068 TSecr=18873062 SLE=1466 SRE=1917		
14 30.063838	96:70:10:05:fd:47	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:05	PC/0001	120
15 30.104134	2a:a4:ee:25:06:06	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0002	120
16 45.123338	96:70:10:05:fd:47	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:05	PC/0001	120
17 45.139571	2a:a4:ee:25:06:06	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0002	120
18 60.143602	96:70:10:05:fd:47	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:05	PC/0001	120
18 60.173092	2a:a4:ee:25:06:06	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0002	120
20 71.204293	10.10.2.102	10.10.1.101	TCP	78	55934 → 80 [FIN, ACK] Seq=110 Ack=18 Win=32256 Len=0 TSval=18885038 TSecr=18873062 SLE=1466 SRE=1917		
21 71.224131	10.10.1.101	10.10.2.102	TCP	70	80 → 55934 [ACK] Seq=1917 Ack=111 Win=29184 Len=0 TSval=18885044 TSecr=18885038		
22 75.176476	96:70:10:05:fd:47	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:05	PC/0001	120
23 75.191452	2a:a4:ee:25:06:06	LLDP_Multicast	LLDP	67	MA/00:00:00:00:00:01	PC/0002	120

Figura 60: Política para tráfico HTTP entre Leaf 1 y Leaf 2

Sobre la utilización de políticas y túneles un problema que se observo en general y se explica en [8] es que si un link que es parte de un túnel y tiene aplicada una política falla, el protocolo comienza a tomar caminos sub-óptimos o directamente se pierde la conectividad. Se probó generar políticas secundarias con menor prioridad que tomaran otro camino, pero el resultado fue el mismo.

4.2. Kathará

Se realizaron las mismas pruebas para el ambiente Kathará con algunos cambios. Las imágenes 61 y 62 muestran las topologías utilizadas, donde se puede observar que en primer lugar los servidores conectados a un Leaf pueden conectarse en un mismo dominio de broadcast, lo cual no era posible en el ambiente Mininet. El controlador se conecta a cada switch mediante un link formando la red de gestión y a su vez se conecta al host de Kathará para que sea posible acceder a la interfaz gráfica.

Para las pruebas asumo que el ambiente esta configurado para utilizar la aplicación Segment Routing.

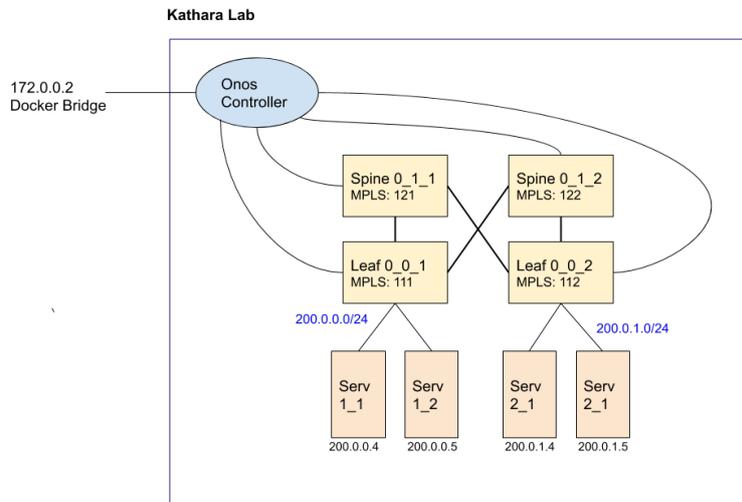


Figura 61: Spine Leaf 2x2x2 con Controlador

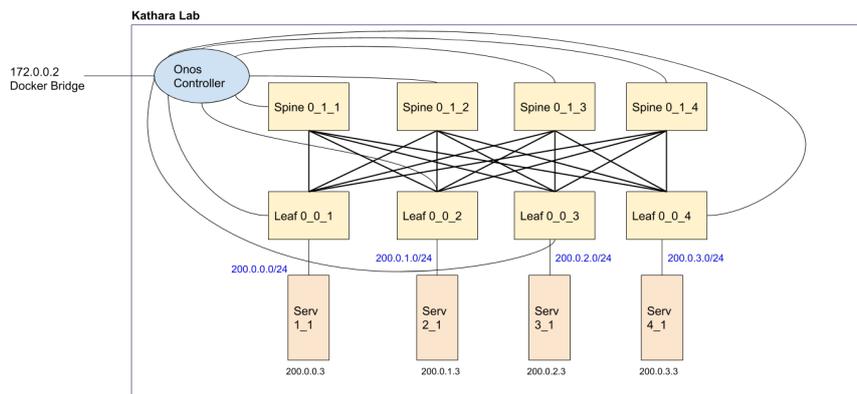


Figura 62: Spine Leaf 4x4x1 con Controlador

4.2.1. Conectividad

Spine Leaf 2x2x2

Se ejecutó un ping desde el host *serv_1.1* al host *serv_2.1*. En la figura 63 se puede ver un resumen de los flujos y grupos seleccionados para el *leaf_0.0.1*.

Primero se muestra la tabla de flujos con la entrada correspondiente a la subred *200.0.1.0/24*, donde se encuentra el servidor *serv_2.1*. El flujo indica que se debe tomar la acción del grupo *0x70000007*. Dicho grupo es del tipo *Select* y presenta dos caminos para el paquete, uno tomando la acción especificada en el grupo *0x92000004* y el otro por el grupo *0x92000006*. Se deberían tomar ambos caminos al tratarse de un *Select*, sin embargo se ve tráfico solo en uno de ellos, lo cual se discutirá más adelante. Siguiendo el grupo con trafijo ICMP, se puede ver que la acción que se toma para dicho grupo es la acción PUSH MPLS, agregando la etiqueta 112, NodeSid del *leaf_0.0.2*. Por último se muestra la MAC de destino que corresponde a la MAC del *spine_0.1.1* y la interfaz de salida que es la interfaz que conecta al Leaf con dicho Spine.

STATE	PACKETS	DURATION	FLOW PRIORITY	TABLE NAME	SELECTOR	TREATMENT
Added	36	274	48010	30	ETH_TYPE:ipv4, IPV4_DST:200.0.1.0/24	def[GROUP:0x70000007], transition:TABLE:60, cleared:false
GROUP ID	APP ID	STATE	TYPE	PACKETS		
0x70000007	org.onosproject.segmentrouting	ADDED	Select	84	Bytes: 0 Packets: 0 Actions: [GROUP:0x92000004] Bytes: 8232 Packets: 84 Actions: [GROUP:0x92000006]	
0x92000006	org.onosproject.segmentrouting	ADDED	Indirect	124	Bytes: 12152 Packets: 124 Actions: [VLAN_POP, MPLS_PUSH:mpls_unicast, MPLS_LABEL:112, GROUP:0x90000005, VLAN_PUSH:vlan, VLAN_ID:4095]	
0x90000005	org.onosproject.segmentrouting	ADDED	Indirect	163	Bytes: 15974 Packets: 163 Actions: [ETH_DST:02:00:00:79:20:17, ETH_SRC:02:00:00:FF:A9:68, VLAN_ID:4094, GROUP:0xfe0002]	
0xfe0002	org.onosproject.segmentrouting	ADDED	Indirect	193	Bytes: 18914 Packets: 193 Actions: [VLAN_POP, OUTPUT:2]	

Figura 63: Definición de Flujos y Grupos desde Leaf 1 a subred 200.0.1.0/24

Una vez el trafico llega al *spine_0.1.1*, este realiza un POP de la etiqueta al ser el penúltimo salto y envía el paquete ICMP al *leaf_0.0.2*. Se pueden ver en detalle los flujos y grupos de los spine en el archivo *Pruebas/Conectividad/Kathara/2x2x2/2x2x2 Tablas.ods*.

Luego que el paquete llega ya desetiquetado al *leaf_0.0.2*, el flujo para la IP del servidor indica que se envíe a dicho servidor por la interfaz directamente conectada al mismo.

STATE	PACKETS	DURATION	FLOW PRIORITY	TABLE NAME	SELECTOR	TREATMENT
Added	237	479	64010	30	ETH_TYPE:ipv4, IPV4_DST:200.0.1.4/32	def[GROUP:0x20000008], transition:TABLE:60, cleared:false
GROUP ID	APP ID	STATE	TYPE	PACKETS		
0x20000008	org.onosproject.segmentrouting	ADDED	Indirect	301		
Bytes: 29498 Packets: 301 Actions: [ETH_DST:02:00:00:74:50:20] ETH_SRC:02:00:00:EE:75:24, VLAN_ID:1, GROUP:0x10003]						
0x10003	org.onosproject.segmentrouting	ADDED	Indirect	338		
Bytes: 33504 Packets: 338 Actions: [VLAN_POP:[OUTPUT:3]]						

Figura 64: Definición de Flujos y Grupos desde Leaf 2 a servidor 2_1

En las figuras 65 y 66 se muestra la inspección de las interfaces que conectan al *leaf_0_0_1* con el *spine_0_1_1* y el *spine_0_1_2* respectivamente. Se puede observar que los paquetes ICMP echo request estan encapsulados por el paquete MPLS dirigiéndose al primer spine, mientras que la respuesta llega des etiquetada por el segundo (dado que el *spine_0_1_2* ya realizó la acción POP MPLS de la etiqueta 111 del *leaf_0_0_1*).

34	24.244231	200.0.0.3	200.0.3.3	ICMP	102 Echo (ping) request...
35	24.789181	02:eb:9f:67:c9:42		LLDP_Multicast	130 MA/00:00:00:00:00:1...
36	24.799256	02:eb:9f:67:c9:42		Broadcast	0x8942 130 Ethernet II
37	25.102121	200.0.0.3	200.0.3.3	ICMP	102 Echo (ping) request...
38	26.116723	200.0.0.3	200.0.3.3	ICMP	102 Echo (ping) request...
39	26.600922	02:eb:9f:67:c9:42		LLDP_Multicast	LLDP 130 MA/00:00:00:00:00:0...
40	26.600951	02:eb:9f:67:c9:42		Broadcast	0x8942 130 Ethernet II
41	27.140752	200.0.0.3	200.0.3.3	ICMP	102 Echo (ping) request...
42	27.899748	02:eb:9f:67:c9:42		LLDP_Multicast	LLDP 130 MA/00:00:00:00:00:1...
<p>Frame 34: 102 bytes on wire (816 bits), 102 bytes captured (816 bits)</p> <p>Ethernet II, Src: 02:00:00:89:00:18 (02:00:00:89:00:18), Dst: 02:00:00:84:7d:bd (02:00:00:84:7d:bd)</p> <p>MultiProtocol Label Switching Header, Label: 114, Exp: 0, S: 1, TTL: 64</p> <p>0000 0000 0000 0111 0010 = MPLS Label: 114</p> <p>..... 000..... = MPLS Experimental Bits: 0</p> <p>..... 001..... = MPLS Bottom Of Label Stack: 1</p> <p>..... 0100 0000 = MPLS TTL: 64</p> <p>Internet Protocol Version 4, Src: 200.0.0.3, Dst: 200.0.3.3</p> <p>Internet Control Message Protocol</p>					
105	71.816841	200.0.3.3	200.0.0.3	ICMP	98 Echo (ping) reply ...
106	72.840838	200.0.3.3	200.0.0.3	ICMP	98 Echo (ping) reply ...
107	73.860844	200.0.3.3	200.0.0.3	ICMP	98 Echo (ping) reply ...
<p>Frame 98: 98 bytes on wire (784 bits), 98 bytes captured (784 bits)</p> <p>Ethernet II, Src: 02:00:00:3d:80:8f (02:00:00:3d:80:8f), Dst: 02:00:00:89:00:18 (02:00:00:89:00:18)</p> <p>Internet Protocol Version 4, Src: 200.0.3.3, Dst: 200.0.0.3</p> <p>Internet Control Message Protocol</p>					

En las capturas se detectó que el tráfico ICMP en cada interfaz es en un sentido y no se esta distribuyendo la carga paquete a paquete como ocurría con Mininet. Esto se debe a que en las nuevas versiones de OpenVSwitch el próximo destino para ECMP se elije por flujo, mitigando los problemas de TCP vistos anteriormente. Esto repercute en el balanceo de carga y se debe hacer un estudio más exhaustivo de dicho problema.

Spine Leaf 4x4x1

Se ejecutó un ping entre el host *serv_1.1* y el host *serv_4.1* de la topología presentada en la imagen 62. Se puede ver en la entrada de la tabla de flujos de la imagen 67 que para la subred *200.0.3.0/24* corresponde el grupo *0x70000038*, el cual es un *Select* con cuatro caminos definidos. Como se vio en la prueba anterior, solo uno de los grupos muestra tener tráfico, el grupo *0x9200002D*. Para dicho grupo se ejecuta la acción PUSH MPLS de la etiqueta 114, NodeSid del *leaf_0.0.4*. Luego se envía el paquete MPLS por el puerto 4 al *spine_0.1.4*.

STATE	PACKETS	DURATION	FLOW PRIORITY	TABLE NAME	SELECTOR	TREATMENT
Added	13	2,110	48010	30	ETH_TYPE:ipv4, IPV4_DST:200.0.3.0/24	def{GROUP:0x70000038}, transition:TABLE:60, cleared:false
GROUP ID	APP ID	STATE	TYPE	PACKETS		
0x70000038	org.onosproject.segmentrouting	ADDED	Select	13	Bytes: 0 Packets: 0 Actions: [GROUP:0x92000027] Bytes: 1274 Packets: 13 Actions: [GROUP:0x9200002d] Bytes: 0 Packets: 0 Actions: [GROUP:0x92000030] Bytes: 0 Packets: 0 Actions: [GROUP:0x92000035]	
0x9200002d	org.onosproject.segmentrouting	ADDED	Indirect	13	Bytes: 1274 Packets: 13 Actions: [VLAN_POP, MPLS_PUSH:mpls_unicast, MPLS_LABEL:114] GROUP:0x9000002a, VLAN_PUSH:wan, VLAN_ID:4095]	
0x9000002a	org.onosproject.segmentrouting	ADDED	Indirect	13	Bytes: 1274 Packets: 13 Actions: [ETH_DST:02:00:00:84:7D:BD] ETH_SRC:02:00:00:89:00:18, VLAN_ID:4094, GROUP:0xffe0004]	
0xffe0004	org.onosproject.segmentrouting	ADDED	Indirect	13	Bytes: 1274 Packets: 13 Actions: [VLAN_POP, OUTPUT:4]	

Figura 67: Definición de Flujos y Grupos desde Leaf 1 a subred 200.0.3.0/24

En la imagen 68 se pueden ver el flujo y grupos para la respuesta del servidor *serv_1.4*. Nuevamente se tienen cuatro caminos en el grupo *0x70000077* del tipo *Select* y se observa tráfico en uno de ellos. La acción de dicho grupo es etiquetar los paquetes con el NodeSid del *leaf_0.0.1* y enviarlo al *spine_0.1.2* conectado en el puerto 2.

STATE	PACKETS	DURATION	FLOW PRIORITY	TABLE NAME	SELECTOR	TREATMENT
Added	13	2,490	48010	30	ETH_TYPE:ipv4, IPV4_DST:200.0.0.0/24	def[GROUP:0x70000077], transition:TABLE:60, cleared:false
GROUP ID	APP ID	STATE	TYPE	PACKETS		
0x70000077	org.onosproject.segmentrouting	ADDED	Select	13	Bytes: 0 Packets: 0 Actions: [GROUP:0x92000064] Bytes: 0 Packets: 0 Actions: [GROUP:0x9200006c] Bytes: 1274 Packets: 13 Actions: [GROUP:0x92000071] Bytes: 0 Packets: 0 Actions: [GROUP:0x92000075]	
0x92000071	org.onosproject.segmentrouting	ADDED	Indirect	13	Bytes: 1274 Packets: 13 Actions: [VLAN_POP, MPLS_PUSH:mpls_unicast, MPLS_LABEL:111, GROUP:0x9000006e, VLAN_PUSH:vlan, VLAN_ID:4095]	
0x9000006e	org.onosproject.segmentrouting	ADDED	Indirect	13	Bytes: 1274 Packets: 13 Actions: [ETH_DST:02:00:00:3D:80:8F, ETH_SRC:02:00:00:2B:16:26, VLAN_ID:4094, GROUP:0xffe0002]	
0xffe0002	org.onosproject.segmentrouting	ADDED	Indirect	13	Bytes: 1274 Packets: 13 Actions: [VLAN_POP, OUTPUT:2]	

Figura 68: Definición de Flujos y Grupos desde Leaf 4 a subred 200.0.0.0/24

En el directorio *Pruebas/Conectividad/Kathara/4x4x1* se encuentra detallada la información de las tablas de flujos y grupos para cada Spine y Leaf. A su vez se encuentran las capturas de tráfico de las interfaces del *leaf_0_0_1* con cada Spine.

4.2.2. Falla de Enlaces

Para simular la desconexión de enlaces se utilizó el comando de linux *ip link set dev interface down* en las interfaces de los Spine, conectándose a la consola del nodo Kathará correspondiente.

Spine Leaf 2x2x2

Para validar el funcionamiento de Segment Routing ante fallas de link se ejecutó primero un ping entre el servidor *serv_1.1* y el servidor *serv_2.1*. Durante este tiempo se bajó la interfaz que conecta al *spine_0_1_1* con el *leaf_0_0_2*, como se muestra en la figura 69 , lo cual implicó ejecutar en el Spine el comando *ip link set dev eth2 down*.

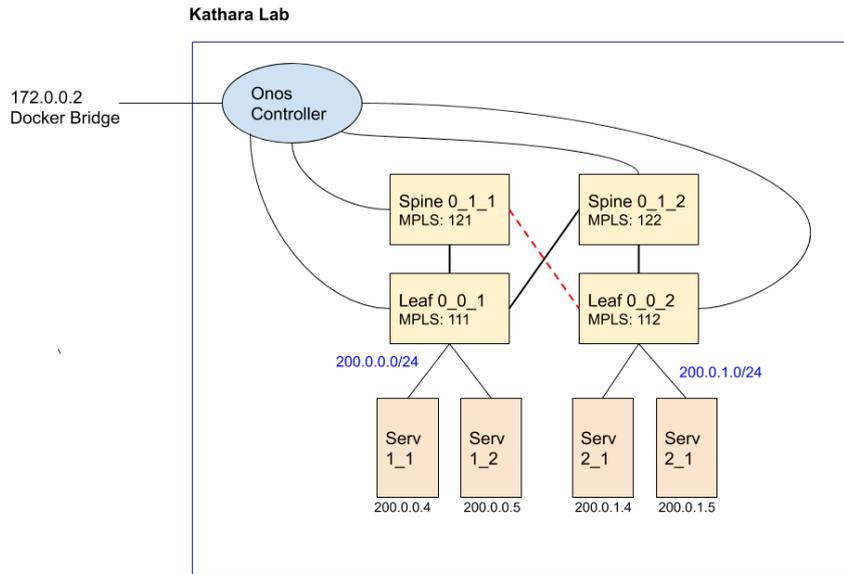


Figura 69: Falla de Link en Spine Leaf 2x2x2

Comparando con la figura 63 se puede ver que en la figura 70 se tiene solo un camino definido para el grupo `0x70000007` el cual envía el NodeSid etiquetado al `spine_0_1_2`. Durante la caída del link no se perdieron paquetes ICMP entre los hosts.

GROUP ID	APP ID	STATE	TYPE	PACKETS
0x70000007	org.onosproject.segmentrouting	ADDED	Select	58
Bytes: 5684 Packets: 58 Actions: [GROUP:0x92000006]				
0x92000006	org.onosproject.segmentrouting	ADDED	Indirect	561
Bytes: 54814 Packets: 561 Actions: [VLAN_POP, MPLS_PUSH:mpls_unicast, MPLS_LABEL:112, GROUP:0x90000005, VLAN_PUSH:vlan, VLAN_ID:4095]				
0x90000005	org.onosproject.segmentrouting	ADDED	Indirect	590
Bytes: 57656 Packets: 590 Actions: [ETH_DST:02:00:00:79:20:17, ETH_SRC:02:00:00:FF:A9:68, VLAN_ID:4094, GROUP:0xffe0002]				

Figura 70: Flujos y Grupos de Leaf 1 Luego de caída de Link

Spine Leaf 4x4x1

Se bajó en este caso la interfaz que conecta el `spine_0_1_1` con el `leaf_0_0_3` como se puede ver en la figura 71. Mientras se ejecutó un ping entre los servidores `serv_1_1` y `serv_1_3`, el cual no se vio afectado durante la caída del link. Se comprobó antes de bajar el link que el tráfico ICMP estaba pasando por el mismo.

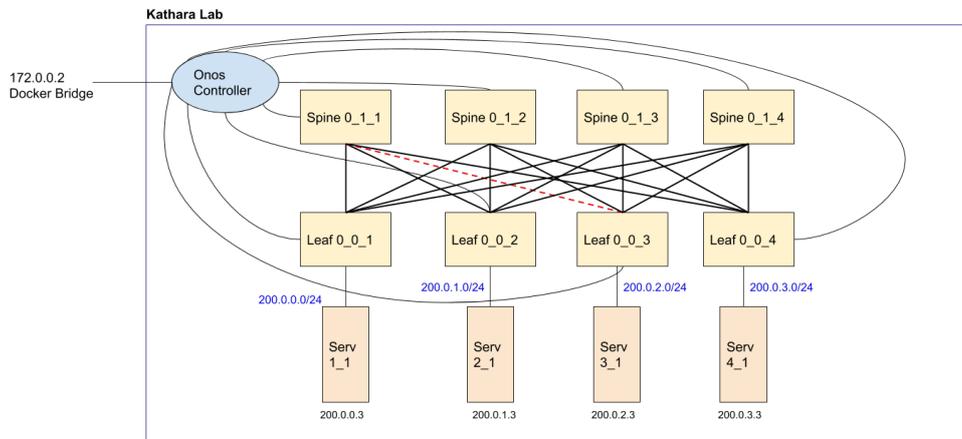


Figura 71: Falla de Link en Spine Leaf 4x4x1

Luego de bajar el link se puede observar en la imagen 72 que el grupo seleccionado tiene tres caminos en lugar de 4 por la cantidad de Spines. No se vio afectada la conectividad entre ambos host.

0x7000001d	org.onosproject.segmentrouting	ADDED	Select	81
Bytes: 7938 Packets: 81 Actions: [GROUP:0x92000014]				
Bytes: 0 Packets: 0 Actions: [GROUP:0x92000018]				
Bytes: 0 Packets: 0 Actions: [GROUP:0x9200001b]				

Figura 72: Flujos y Grupos de Leaf 1 Luego de caida de Link

4.2.3. Prueba de Trafico TCP

Dado que en el caso de Kathará se vio que el tráfico de OpenVswitch se distribuye por flujos, resulta de interés ejecutar pruebas con el protocolo TCP y comparar con el ambiente Mininet.

Los servidores Kathará inician con el servicio Apache levantado escuchando conexiones en el puerto 80. En este caso no se trata de un script de simulación como en Mininet, sino que se trata de servidores web sobre contenedores. Para probar se ejecutó *wget* desde los host y se capturo el tráfico en las interfaces de los Leaf.

Spine Leaf 2x2x2

En primer lugar se probó sobre la topología de la figura 61, realizando una consulta http desde el servidor *serv_1.2* al servidor *serv_2.1*, ejecutando el comando *wget 200.0.1.4*. En las imágenes 73 y 74 se puede observar en el sentido de los paquetes TCP que si bien la carga se distribuye no lo hace paquete a paquete, sino por sentido de la conexión. Además se resuelve correctamente la consulta http. Se puede ver en más detalle el tráfico en las capturas del directorio *Pruebas/Trafico TCP/Kathara/2x2x1*.

88	51.461420	200.0.0.5	200.0.1.4	TCP	70	48158	-	80	[SYN]	Seq=0 Win=64240 Len=0 MSS=1460 SACK_PERM=1 TSval=946736195 TSecr=0 WS=128
89	51.462344	200.0.0.5	200.0.1.4	TCP	70	48158	-	80	[ACK]	Seq=1 Ack=1 Win=64256 Len=0 TSval=946736197 TSecr=4223011862
90	51.462396	200.0.0.5	200.0.1.4	HTTP	204	GET	/	HTTP/2.1		
91	51.462810	200.0.0.5	200.0.1.4	TCP	70	48158	-	80	[ACK]	Seq=135 Ack=7241 Win=61312 Len=0 TSval=946736197 TSecr=4223011863
92	51.462864	200.0.0.5	200.0.1.4	TCP	70	48158	-	80	[ACK]	Seq=135 Ack=11013 Win=58752 Len=0 TSval=946736197 TSecr=4223011863
93	51.463738	200.0.0.5	200.0.1.4	TCP	70	48158	-	80	[FIN, ACK]	Seq=135 Ack=11013 Win=64128 Len=0 TSval=946736198 TSecr=4223011863
94	51.463926	200.0.0.5	200.0.1.4	TCP	70	48158	-	80	[ACK]	Seq=136 Ack=11014 Win=64128 Len=0 TSval=946736198 TSecr=4223011864

Figura 73: Trafico TCP entre Leaf 1 y Spine 1

66	46.062099	200.0.1.4	200.0.0.5	TCP	74	80	-	48158	[SYN, ACK]	Seq=0 Ack=1 Win=65160 Len=0 MSS=1460 SACK_PERM=1 TSval=4223011862 TSecr=946736195 WS=
67	46.062361	200.0.1.4	200.0.0.5	TCP	66	80	-	48158	[ACK]	Seq=1 Ack=135 Win=65152 Len=0 TSval=4223011863 TSecr=946736197
68	46.062675	200.0.1.4	200.0.0.5	TCP	7306	80	-	48158	[PSH, ACK]	Seq=1 Ack=135 Win=65152 Len=7240 TSval=4223011863 TSecr=946736197 [TCP segment of a r-
69	46.062740	200.0.1.4	200.0.0.5	HTTP	3838	HTTP/1.1	200	OK	(text/html)	
70	46.063778	200.0.1.4	200.0.0.5	TCP	66	80	-	48158	[FIN, ACK]	Seq=11013 Ack=136 Win=65152 Len=0 TSval=4223011864 TSecr=946736198

Figura 74: Trafico TCP entre Leaf 1 y Spine 2

Spine Leaf 4x4x1

Se probó luego para la topología de la figura 62 realizar consultas http del servidor *serv_1.1* a los servidores *serv_2.1*, *serv_3.1* y *serv_4.1*.

Se puede observar en la imagen 77 que las consultas a los servidores *serv_2.1* y *serv_3.1* se enviaron al *spine_0.1.2*. Las respuestas a ambas consultas llegaron por medio del *spine_0.1.4* como se registra en la figura 77. Por otro lado tanto la consulta al servidor *serv_1.4* como su respuesta enviaron el tráfico por el *spine_0.1.3* como se muestra en la imagen 76. En todos casos la conexión TCP se mostró estable sin pedidos de retransmisión de paquetes o pérdidas.

54	38.055474	200.0.0.3	200.0.1.3	TCP	70	39550	-	80	[ACK]	Seq=1 Ack=1 Win=64256 Len=0 TSval=3486390667 TSecr=481341669
55	38.055564	200.0.0.3	200.0.1.3	HTTP	204	GET	/	HTTP/1.1		
56	38.380935	200.0.0.3	200.0.1.3	TCP	70	39550	-	80	[ACK]	Seq=135 Ack=5793 Win=62080 Len=0 TSval=3486390601 TSecr=481342007
57	38.380917	200.0.0.3	200.0.1.3	TCP	70	39550	-	80	[ACK]	Seq=135 Ack=11013 Win=58752 Len=0 TSval=3486390601 TSecr=481342007
58	38.390031	200.0.0.3	200.0.1.3	TCP	70	39550	-	80	[FIN, ACK]	Seq=135 Ack=11013 Win=64128 Len=0 TSval=3486390602 TSecr=481342007
59	38.418080	200.0.0.3	200.0.1.3	TCP	70	39550	-	80	[ACK]	Seq=136 Ack=11014 Win=64128 Len=0 TSval=3486390630 TSecr=481342036
60	40.300150	02:eb:9f:67:c9:42	LLDP_Multicast	LLDP	130	PA/00:00:00:00:00:01	PC/32	120		
61	40.300156	02:eb:9f:67:c9:42	Broadcast	0x0942	130	Ethernet II				
62	40.700734	02:eb:9f:67:c9:42	LLDP_Multicast	LLDP	130	PA/00:00:00:00:00:12	PC/31	120		
63	40.700802	02:eb:9f:67:c9:42	Broadcast	0x0942	130	Ethernet II				
64	41.177726	200.0.0.3	200.0.2.3	TCP	70	58028	-	80	[SYN]	Seq=0 Win=64240 Len=0 MSS=1460 SACK_PERM=1 TSval=2225820530 TSecr=0 WS=128
65	41.178762	200.0.0.3	200.0.2.3	TCP	70	58028	-	80	[ACK]	Seq=1 Ack=1 Win=64256 Len=0 TSval=2225820532 TSecr=2653471022
66	41.178839	200.0.0.3	200.0.2.3	HTTP	204	GET	/	HTTP/2.1		
67	41.188554	200.0.0.3	200.0.2.3	TCP	70	58028	-	80	[ACK]	Seq=135 Ack=7241 Win=61312 Len=0 TSval=2225820541 TSecr=2653471032
68	41.188612	200.0.0.3	200.0.2.3	TCP	70	58028	-	80	[ACK]	Seq=135 Ack=11013 Win=58752 Len=0 TSval=2225820542 TSecr=2653471032
69	41.197337	200.0.0.3	200.0.2.3	TCP	70	58028	-	80	[FIN, ACK]	Seq=135 Ack=11013 Win=64128 Len=0 TSval=2225820550 TSecr=2653471032
70	41.197534	200.0.0.3	200.0.2.3	TCP	70	58028	-	80	[ACK]	Seq=136 Ack=11014 Win=64128 Len=0 TSval=2225820550 TSecr=2653471041
71	43.400500	02:eb:9f:67:c9:42	LLDP_Multicast	LLDP	130	PA/00:00:00:00:00:01	PC/32	120		
72	43.400510	02:eb:9f:67:c9:42	Broadcast	0x0942	130	Ethernet II				

Figura 75: Trafico TCP entre Leaf 1 y Spine 2

50	36.752210	200.0.0.3	200.0.3.3	TCP	70	39304 → 80	[SYN]	Seq=0	Win=64240	Len=0	MSS=1460	SACK_PERM=1	TSval=2809522537	TSecr=0	WS=128	
51	36.753003	200.0.0.3	200.0.0.3	TCP	74	80 → 39304	[SYN, ACK]	Seq=0	Ack=1	Win=65160	Len=0	MSS=1460	SACK_PERM=1	TSval=1003282625	TSecr=2809522537	WS=
52	36.753241	200.0.0.3	200.0.0.3	TCP	70	39304 → 80	[ACK]	Seq=1	Ack=1	Win=64256	Len=0	TSval=2809522538	TSecr=1003282625			
53	36.753365	200.0.0.3	200.0.0.3	HTTP	204	GET / HTTP/1.1										
54	36.753365	200.0.0.3	200.0.0.3	TCP	66	80 → 39304	[ACK]	Seq=1	Ack=135	Win=65152	Len=0	TSval=1003282626	TSecr=2809522538			
55	37.080353	200.0.0.3	200.0.0.3	TCP	7306	80 → 39304	[PSH, ACK]	Seq=1	Ack=135	Win=65152	Len=7240	TSval=1003282953	TSecr=2809522538	[TCP segment of a		
56	37.080398	200.0.0.3	200.0.0.3	TCP	70	39304 → 80	[ACK]	Seq=135	Ack=7241	Win=61312	Len=0	TSval=2809522866	TSecr=1003282953			
57	37.080445	200.0.0.3	200.0.0.3	HTTP	3838	HTTP/1.1 200 OK	(text/html)									
58	37.080467	200.0.0.3	200.0.0.3	TCP	70	39304 → 80	[ACK]	Seq=135	Ack=11013	Win=58752	Len=0	TSval=2809522866	TSecr=1003282953			
59	37.081392	200.0.0.3	200.0.0.3	TCP	70	39304 → 80	[FIN, ACK]	Seq=135	Ack=11013	Win=64128	Len=0	TSval=2809522867	TSecr=1003282953			
60	37.081599	200.0.0.3	200.0.0.3	TCP	66	80 → 39304	[FIN, ACK]	Seq=11013	Ack=136	Win=65152	Len=0	TSval=1003282954	TSecr=2809522867			
61	37.081643	200.0.0.3	200.0.0.3	TCP	70	39304 → 80	[ACK]	Seq=136	Ack=11014	Win=64128	Len=0	TSval=2809522867	TSecr=1003282954			

Figura 76: Trafico TCP entre Leaf 1 y Spine 3

11	7.054324	200.0.1.3	200.0.0.3	TCP	74	80 → 39550	[SYN, ACK]	Seq=0	Ack=1	Win=65160	Len=0	MSS=1460	SACK_PERM=1	TSval=401341669	TSecr=3486389662	WS=
12	7.054671	200.0.1.3	200.0.0.3	TCP	66	80 → 39550	[ACK]	Seq=1	Ack=135	Win=65152	Len=0	TSval=401341674	TSecr=3486389667			
13	7.387928	200.0.1.3	200.0.0.3	TCP	5858	80 → 39550	[PSH, ACK]	Seq=1	Ack=135	Win=65152	Len=5792	TSval=401342007	TSecr=3486389667	[TCP segment of a r		
14	7.388026	200.0.1.3	200.0.0.3	HTTP	5206	HTTP/1.1 200 OK	(text/html)									
15	7.417055	200.0.1.3	200.0.0.3	TCP	66	80 → 39550	[FIN, ACK]	Seq=11013	Ack=136	Win=65152	Len=0	TSval=401342036	TSecr=3486390002			
16	7.498930	02:eb:9f:67:c9:42	LLDP_Multicast	LLDP	130	NA/00:00:00:00:00:14	PC/31 120									
17	7.499019	02:eb:9f:67:c9:42	Broadcast	0x8942	130	Ethernet II										
18	9.299487	02:eb:9f:67:c9:42	LLDP_Multicast	LLDP	130	NA/00:00:00:00:00:01	PC/34 120									
19	9.299511	02:eb:9f:67:c9:42	Broadcast	0x8942	130	Ethernet II										
20	10.177735	200.0.2.3	200.0.0.3	TCP	74	80 → 58028	[SYN, ACK]	Seq=0	Ack=1	Win=65160	Len=0	MSS=1460	SACK_PERM=1	TSval=2653471022	TSecr=2225820530	WS=
21	10.177961	200.0.2.3	200.0.0.3	TCP	66	80 → 58028	[ACK]	Seq=1	Ack=135	Win=65152	Len=0	TSval=2653471023	TSecr=2225820532			
22	10.187548	200.0.2.3	200.0.0.3	TCP	7306	80 → 58028	[PSH, ACK]	Seq=1	Ack=135	Win=65152	Len=7240	TSval=2653471032	TSecr=2225820532	[TCP segment of a		
23	10.187630	200.0.2.3	200.0.0.3	HTTP	3838	HTTP/1.1 200 OK	(text/html)									
24	10.196526	200.0.2.3	200.0.0.3	TCP	66	80 → 58028	[FIN, ACK]	Seq=11013	Ack=136	Win=65152	Len=0	TSval=2653471041	TSecr=2225820550			

Figura 77: Trafico TCP entre Leaf 1 y Spine 4

Por lo visto en las pruebas del ambiente Kathará el trafico se distribuye por flujo tomando siempre un solo camino en cada sentido de la conexión, el cual puede ser el mismo entre dos nodos o distinto. El ambiente es robusto ante fallas de links y apto para utilizar el protocolo de transporte TCP. Sin embargo con las pruebas realizadas no se llega a apreciar un balanceo de carga. Una opción para comprobar el comportamiento de ECMP es realizar pruebas con grandes cargas de tráfico y analizar si se comienzan a tomar más caminos. Esta prueba no se realizó en el taller dado que requiere de mayor cantidad de recursos.

5. Conclusión

En el desarrollo de el presente trabajo se adquirieron conocimientos teóricos sobre SDN y su aplicación en Datacenters, además de técnicas como MPLS y Segment Routing donde el ruteo se basa en el origen. Por otro lado se profundizó en herramientas como OpenFlow, Onos y Open vSwitch, las cuales son ampliamente utilizadas hoy en día en ambientes productivos.

La solución Segment Routing en conjunto con SDN propone una solución centralizada para el ruteo en Datacenters donde se separa el plano de datos del plano de control. Esto permite una mayor flexibilidad al momento de aplicar ingeniería de tráfico para propósitos específicos realizando cambios en toda la red. A su vez permite una escala menos costosa dado que alcanza con tener switches que implementen OpenFlow para aplicar la solución. Al ser una herramienta de código abierto no se esta sujeto a los proveedores de la industria.

Por otro lado se comprendió el funcionamiento de entornos de prueba como Mininet y Kathará, encontrando la forma de trasladar el funcionamiento de un entorno simulado y utilizado para prototipos, a un entorno que emula un ambiente de producción mediante tecnologías ampliamente utilizadas como los contenedores. Dicho entorno puede pasar a un entorno de prueba en Hardware real de una forma directa

En la sección de pruebas se encontró que el entorno Mininet presenta ciertas desventajas en el tratamiento del tráfico TCP, dado que el algoritmo ECMP utiliza un algoritmo del tipo Round Robin. El entorno Kathará presento mejoras ante este aspecto ya que ECMP distribuye el tráfico por flujos, pero no se detecta un balanceo de la carga.

Trabajo Futuro

Se pueden detectar varias líneas de trabajo a continuar y profundizar a partir del trabajo realizado:

- No se logró aplicar políticas para ingeniería de tráfico sobre el ambiente Kathará. Se debe encontrar y probar una versión de ONOS compatible con

el entorno que lo permita.

- Analizar el funcionamiento de ECMP en Kathará en cuanto a la distribución de carga. Las pruebas realizadas no contenían una carga de tráfico considerable para detectar si la carga se distribuye equitativamente.
- Comparar esta solución centralizada con algoritmos distribuidos ya implementados en el ambiente Kathará (BGP, OpenFabric y RIFT) aplicado a topologías de Datacenter. Interesaría comparar el comportamiento y medir cantidad de mensajería del plano de control, consumo de recursos (Memoria, CPU) y throughput de red entre otros datos.
- Extender el ambiente Kathará para emular Trellis.

Referencias

- [1] Software-Defined Networking: The New Norm for Networks, Withe Paper, April 13, 2012, <https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf>
- [2] Software-Defined Networks: A Systems Approach, Acceso: 2020 <https://sdn.systemsapproach.org/uses.html>
- [3] Research on path establishment methods performance in SDN-based networks, David José Quiroz Martiña, 2015.
- [4] OpenFlow Switch Specification version 1.3.5 (Protocol version 0x04), March 26 2015, ONF TS-023 <https://www.opennetworking.org/wp-content/uploads/2014/10/openflow-switch-v1.3.5.pdf>
- [5] OpenFlow, Acceso: 2020 http://flowgrammable.org/sdn/openflow/#tab_switch
- [6] OpenFlow – Basic Concepts and Theory, Acceso: 2020 <https://overlaid.net/2017/02/15/openflow-basic-concepts-and-theory/>
- [7] Open vSwitch Documentation, Acceso 2020 <https://docs.openvswitch.org/en/latest/>
- [8] Segment Routing, Iipo Litmanen, Heksinski Metropolia Univiersity of Applied Sciences, April 2017
- [9] Open Network Operating System, Martin Pacheco, Sebastián Passaro, 2019.
- [10] SPRING Open Project Software Architecture, Acceso: 2020 <https://wiki.onosproject.org/display/ONOS/Software+Architecture>
- [11] ONF's SPRING-OPEN project, Acceso: 2020 <https://wiki.onosproject.org/display/ONOS/%5BArchived%5D+ONF%27s+SPRING-OPEN+project>

- [12] Installation Guide, Acceso: 2020
<https://wiki.onosproject.org/display/ONOS/Installation+Guide>
- [13] Open Network Foundation, Acceso: 2020
<https://www.opennetworking.org/>
- [14] Mininet, Acceso: 2020
<http://mininet.org/>
- [15] Configuring Onos, Acceso: 2020
<https://wiki.onosproject.org/pages/viewpage.action?pageId=2130918>
- [16] Using the CLI, Acceso: 2020
<https://wiki.onosproject.org/display/ONOS/Using+the+CLI>
- [17] Kathara, Acceso: 2020
<https://www.kathara.org/>
- [18] Repositorio Gitlab Taller. Acceso: 2020
<https://gitlab.fing.edu.uy/agustina.parnizari/ciberfisicos>
- [19] Kathara Linux Install, Acceso: 2020
<https://github.com/KatharaFramework/Kathara/wiki/Linux>
- [20] Use bridge networks, Acceso: 2020
<https://docs.docker.com/network/bridge/>
- [21] Kathara: A Lightweight Network Emulation System, Mariano Scazzariello, Lorenzo Ariemma, Tommaso Caiazzi, NOMS 2020 - 2020 IEEE IFIP Network Operations and Management Symposium
- [22] Taller de Onos en Docker, Sebastian Passaro, Modulo Taller, 2019
- [23] Analysis of an Equal-Cost Multi-Path Algorithm, C. Hopps,
<https://tools.ietf.org/html/rfc2992>